Performance Characterizaton of Fundamental Matrix Estimation Under Image Degradation

P H S Torr and A Zisserman

Robotics Research Group

Department of Engineering Science

Oxford University, Parks Road, Oxford, OX1 3PJ, UK.

Abstract

The fundamental matrix represents the epipolar geometry between two images. We describe an algorithm for simultaneously estimating the fundamental matrix and corresponding points automatically from the two images. The performance of this algorithm is then assessed as the images are degraded by JPEG lossy compression. A number of performance measures are proposed and evaluated over image pairs corresponding to different camera motions and scene types.

1 Introduction

In order for a vision module to be of general practical use, its performance must be understood over a wide variety of conditions. Förstner [9] clearly highlights that the future of Computer Vision algorithms is dependent on adequate performance characterisation. The lack of performance characterisation of vision algorithms is responsible for the hesitation of industry to use computer vision as one of its tools. Reasons for this situation are manifold: the lack of commonly accepted criteria for evaluation, the lack of a methodology for testing, the lack of translating the experience in testing of other engineering areas to Computer Vision, and possibly also the non-acceptance of empirical or theoretical comparisons of vision algorithms, including their replication, as original research.

This paper characterises the performance of an algorithm for estimating the fundamental matrix from image pairs. The fundamental matrix is the algebraic representation of the epipolar

geometry existing between two views, and it can be computed directly from image correspondences. We require the fundamental matrix as part of a sequence processing system for extracting 3D graphical models from uncalibrated image sequences [2]. These models are used for virtual reality applications. Neither the internal calibration of the camera nor the motion of the camera are known.

As image sequences are now often supplied JPEG or MPEG compressed it is important to quantify the degradation in performance, if any, that occurs as the degree of compression is varied. To this end we describe and apply a number of measures for assessing the degradation of the estimates of the fundamental matrix, and the 3D projective structure recovered from image correspondences, for *real* images as they are JPEG compressed.

These measures fall into two categories: The first category is when something is known of the ground truth. In this case the ground truth is the true fundamental matrix (which can be computed from the camera calibration and motion) and the true structure of the 3D scene. This category is assessed by comparing the fundamental matrix and 3D structure obtained by the estimation algorithm to the ground truth fundamental matrix and 3D structure. The second category involves images for which there is no *a priori* information about the ground truth. In this case a decision has to be made on what to use as a substitute for the ground truth.

The actual ground truth (category one) is usually not known (e.g. when viewing old video footage about which no metric information is available). For the class of scenes where the ground truth is unknown, we take the maximum likelihood estimate of the fundamental matrix and structure estimated from the uncompressed image pairs as the substitute ground truth, and then monitor how these alter for the compressed images. There has been little previous work on determining the difference of two fundamental matrices with the exception of the stability studies of [4, 19, 20, 29].

The paper is organised as follows: In Section 2 the algorithm for estimating the fundamental matrix is summarized. The input to the algorithm is an image pair, the output is the fundamental matrix, its covariance, and the 3D projective structure of the matched point features. The algorithm makes strong use of robust statistics in order to eliminate mismatches. Section 3 describes the four test pairs of images that are used. In Section 4 the analysis of variance is introduced. This analysis is used to decide whether a significant change in the variance of the error has occurred under image compression. In Section 5 measures of the first category are introduced. Here we have only assumed partial knowledge of the ground truth: namely knowledge of the object structure. In Section 6 measures of the second category are introduced. The results on the four test images are presented in Section 7, and discussed in Section 8.

2 Fundamental Matrix Estimation

In this section the fundamental matrix is introduced, and the algorithm for estimating it is then detailed. The input is the image pair, and the output is an estimate of the fundamental matrix, its covariance, and 3D projective structure. All processing is fully automatic, and is summarised in Figure 1.

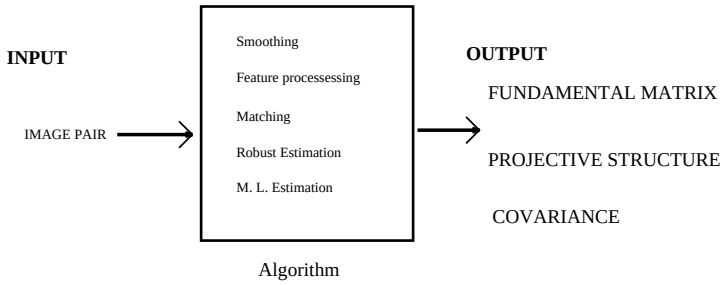


Figure 1: *A schematic of the input and output of the algorithm, the input is an image pair and the output is an estimated fundamental matrix and its covariance, point matches and hence projective structure.*

2.1 The Fundamental Matrix

Suppose two images are acquired of a 3D scene, with the two cameras (or a single camera moving) related by a rotation and non-zero translation. An image point $\underline{\mathbf{x}}_i$ in the first view (we adopt the convention of signifying noise free data by underlining it) corresponds to an image point in the second $\underline{\mathbf{x}}_i^{'}$, if they are images of the same 3D point. The points $\underline{\mathbf{x}}_i$ and $\underline{\mathbf{x}}_i^{'}$ are represented by homogeneous coordinates, $\underline{\mathbf{x}}_i = (\underline{x}_i, \underline{y}_i, 1)^\top$, and $\underline{\mathbf{x}}_i^{'} = (\underline{x}_i^{'}, \underline{y}_i^{'}, 1)^\top$. The points are related by

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

where $\mathbf{F}$ is a 3×3 matrix of rank 2. It is known as the fundamental matrix [6, 12] and can be computed from seven or more such correspondences.

2.2 Simultaneous Computation of F and matches

In many computer vision algorithms it is assumed that a least squares framework is sufficient to deal with data corrupted by noise. However, in many applications, visual data are not only noisy, but also contain outliers, data that are in gross disagreement with a postulated noise model. Outliers can so distort a fitting process that the fitted parameters become arbitrary. In the computation of the fundamental matrix, outliers typically arise from gross errors such as correspondence mismatches or the inclusion of movement inconsistent with the majority. The latter might be caused by features being on occluding contours, shadows or independently moving objects.

The algorithm employed here is robust to such outliers. It proceeds as follows: (a) Initial feature matching based on cross correlation and proximity; (b) Application of a robust estimator to get an initial estimate of $\mathbf{F}$; (c) Use of $\mathbf{F}$ to guide further matching. (d) Recomputation of $\mathbf{F}$ using all inlying matches to give an maximum likelihood estimate (MLE). These stages will now be explained in more detail. Note at all stages the computed $\mathbf{F}$ has rank 2.

Typically there are several hundred corners in a 512×512 image of which about two thirds will provide good matches. Using corners computed to sub-pixel accuracy, the typical distance of a point from its epipolar line is $\sim$0.2-0.4 pixels.

Interest points (corners) are detected in both images using the corner detector of Harris and Stephens [11]. This calculates an interest operator defined according to an auto-correlation

of Gaussian smoothed images. Corners are defined at the local minima of two dimensional quadratic patches fitted to the autocorrelation surface across the image. 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.

There is potentially a significant presence 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 — RANSAC has proved the most successful [8, 25, 30]: A putative fundamental matrix (up to three solutions) is computed from a random set of seven corner correspondences (the minimum number required to compute a fundamental matrix). The support for this fundamental matrix is determined by the number of correspondences in the initial match set within a threshold distance of their epipolar lines. This is repeated for many random sets, and the fundamental matrix with the largest support is accepted. The outcome is a set of corner correspondences consistent with the fundamental matrix, and a set of mismatches (outliers).

Further matches are then obtained, now using the fundamental matrix to determine the search region. The epipolar constraint provides a far more restrictive search region than that used for the initial matching. Consequently, a less severe threshold can be used on the matching attributes. In this case, matches are sought for unmatched corners searching only epipolar lines. For example, the epipolar line l' in the second image corresponding to the point x in the first image is given by $\mathrm{l}' = \mathbf{F}\mathrm{x}$. This generates a larger set of consistent matches.

Given a set of *measured* correspondences $\{\mathbf{x}_i \leftrightarrow \mathbf{x}_i'\}, i = 1, \ldots n$, it can be shown that the MLE for the fundamental matrix is that which minimises

$$D(\mathbf{F}, \{\mathbf{x}_i \leftrightarrow \mathbf{x}_i'\}) = \sum_{i=1}^{i=n} d_e\left(\hat{\mathbf{x}}_i, \mathbf{x}_i\right)^2 + d_e\left(\hat{\mathbf{x}}_i', \mathbf{x}_i'\right)^2 \quad \text{subject to } \hat{\mathbf{x}}_i'^{\top}\mathbf{F}\hat{\mathbf{x}}_i = 0 \ \text{ for all } i, \tag{2}$$

where $d_e()$ is the Euclidean distance in the image plane and $\{\hat{\mathbf{x}}_i \leftrightarrow \hat{\mathbf{x}}_i'\}$ are the ML estimates of the exactly corresponding points $\{\underline{\mathbf{x}}_i \leftrightarrow \underline{\mathbf{x}}_i'\}$ before corruption by measurement error. Once $\mathbf{F}$ is estimated, structure may be recovered as outlined in Hartley and Sturm [13] and Kanatani [16], and reviewed in Appendix B. These methods are also used to find the maximum likelihood estimates of the true (noise free) correspondences given the noisy observations i.e. $\{\hat{\mathbf{x}}_i \leftrightarrow \hat{\mathbf{x}}_i'\}$ given $\{\mathbf{x}_i \leftrightarrow \mathbf{x}_i'\}$. These estimates will be referred to (following [16]) as the optimally corrected points and will be useful for defining the ground truth, when no other information is given. The error for an individual correspondence will be denoted

$$d_i = d(\mathbf{F}, \mathbf{x}_i \leftrightarrow \mathbf{x}_i') = \sqrt{d_e\left(\hat{\mathbf{x}}_i, \mathbf{x}_i\right)^2 + d_e\left(\hat{\mathbf{x}}_i', \mathbf{x}_i'\right)^2} \tag{3}$$

with the subscript sometimes being dropped for convenience. So that $D = \sum_i d^2$.

Due to the frequent occurrence of outliers, in the guise of mismatches, the error actually minimized is

$$D_R = \sum_i \gamma\left(\frac{d}{\sigma}\right) \tag{4}$$

where $\gamma(d)$ is a robust Huber function [15]:

$$\gamma(e) = \begin{cases} e^2 & e < 1.96 \\ 1.96^2 & e \geq 1.96\,. \end{cases} \tag{5}$$

and σ is the (estimated) standard deviation of the inliers. The Huber cost function (5) allows the minimization to be conducted on all features whether they are outliers or inliers. It can be seen that this error function removes the effect of large errors on the minimization. The value 1.96 is chosen so that if an inlier follows a Gaussian distribution it is only rejected 5% of the time. Use of the Huber function limits the effects of outliers on the minimization. (If it is not used the results significantly degrade.) Typically, as the minimization progresses borderline outliers are redesignated inliers.

In Figure 7 the error function D_R is plotted under varying levels of compression to demonstrate the effect of diminished information on the convexity and mode structure of the cost function used in the MLE.

A closed form solution to this minimisation is generally intractable, so it is necessary to use non-linear gradient descent. This requires a parameterisation for the fundamental matrix. The particular parameterisation used is important since it defines the covariance matrix, which will be used in evaluating the performance. The parameterization is that of Luong *et al* [18], where $\mathbf{F}$ in expressed in terms of the non-homogeneous coordinates of the epipoles $\mathbf{e} = (e_1, e_2)$, $\mathbf{e}' = (e'_1, e'_2)$ and three of the four coefficients of the homography between the epipolar lines:

$$\mathbf{F} = \begin{bmatrix} b & a & -ae_2 - be_1 \\ -d & -c & ce_2 + de_1 \\ de'_2 - be'_1 & ce'_2 - ae'_1 & ae_2e'_1 + be_1e'_1 - ce_2e'_2 - de'_2e_1 \end{bmatrix} \tag{6}$$

where $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is the homography between the epipolar lines. As the homography has only three degrees of freedom, its matrix's determinant is constrained to be 1, giving $ad - bc = 1$, from which it can be seen that the fundamental matrix defined in Equation (6) has determinant zero.

The non-linear minimization is performed (using the method of Gill and Murray [10, 28]) on the seven independent parameters: $a, b, c, e_1, e_2, e'_1, e'_2$. A method for obtaining a first order approximation to d is explained in Appendix A.

3 Experimental Data

Four image pairs are evaluated. These differ in both the scene content and camera motion so that the measures could be evaluated under a number of conditions. The first pair has special structure, the second special motion, and the remaining two pairs are general. The images are shown in figure 2. Each pair is evaluated under nine levels of JPEG compression with quality $Q = 90, 80 \ldots 10$. JPEG is "lossy," meaning that the decompressed image has been degraded. A useful property of JPEG is that the degree of lossiness can be varied by adjusting a compression parameter Q. In fact, quality scales are not standardized across JPEG programs. The quality settings discussed in this article apply to the free IJG JPEG software. The size of

the calibration image under varying levels of compression is shown in Figure 5 (a). Note that the gain is dramatic for $Q = 70$ and then much more gradual after that. This graph is typical of the amount of compression obtained for varying Q on the images presented. In Figure 3 the uncompressed images are shown next to images compressed with $Q = 50$ and 10, a block like structure can be observed in images compressed with $Q = 10$ which is an artefect of the compression process.

Calibration Grid The images are of two orthogonal Tsai grids, with a translation between images of about 2cm and a rotation of about 2° around a vertical axis. The calibration object is some 80cm away. The images are 512×512 pixels. This is the image pair for which the ground truth of the 3D Euclidean structure is known.

The processing of these images differs from that of the other three pairs. In this case the corners are extracting by intersecting straight lines of a template, tuned to the calibration pattern, which is fitted to the image. There are two consequences of this method of extraction: first, the position of the corners is recovered more accurately, as will be seen when it comes to measureing $D(\mathbf{F}, \{\mathbf{x}_i \leftrightarrow \mathbf{x}_i'\})$ — although this anticipated accuracy will be affected by JPEG compression; Second, and more importantly for the performance characterisation, there are no mismatches (outliers) between the pairs since the correspondence is determined by the template for each image. For this image pair it is not necessary to utilise the robust part of the algorithm, and the accuracy of the non-linear MLE alone can be evaluated.

Road Pair This is an outdoor scene of a road. The camera is undergoing pure translation. The images were acquired by a camera mounted on a car translating in a straight line down the road. The image size is 256×256 pixels.

This is a special motion case because for a pure translation the fundamental matrix has only two degrees of freedom. The matrix can be parameterised by the epipole which has the same position in both images (i.e. $e_1 = e_1', e_2 = e_2'$).

Chapel Pair This is an outdoor scene of a chapel at Oxford. Camera motion is a lateral displacement of about 3-4cm (translation and rotation). The images were acquired by a hand-held camcorder. The images have been digitised at 760×550 pixels.

Basement Pair This is an interior scene. The camera is translating along the line of sight and rotating to the right about a vertical axis. The images were acquired by a camera mounted on an AGV. The image size is 512×512 pixels.

4 Analysis of Variance

In this paper it will prove desirable to test whether two estimated standard deviations σ_U and σ_C of two sets of observations, size n_c and n_u, are significantly different. For instance it might be that the standard deviation σ_U of the inliers is known between a given pair of reference images, and it is asked whether the standard deviation σ_C of the inliers for a compressed pair is

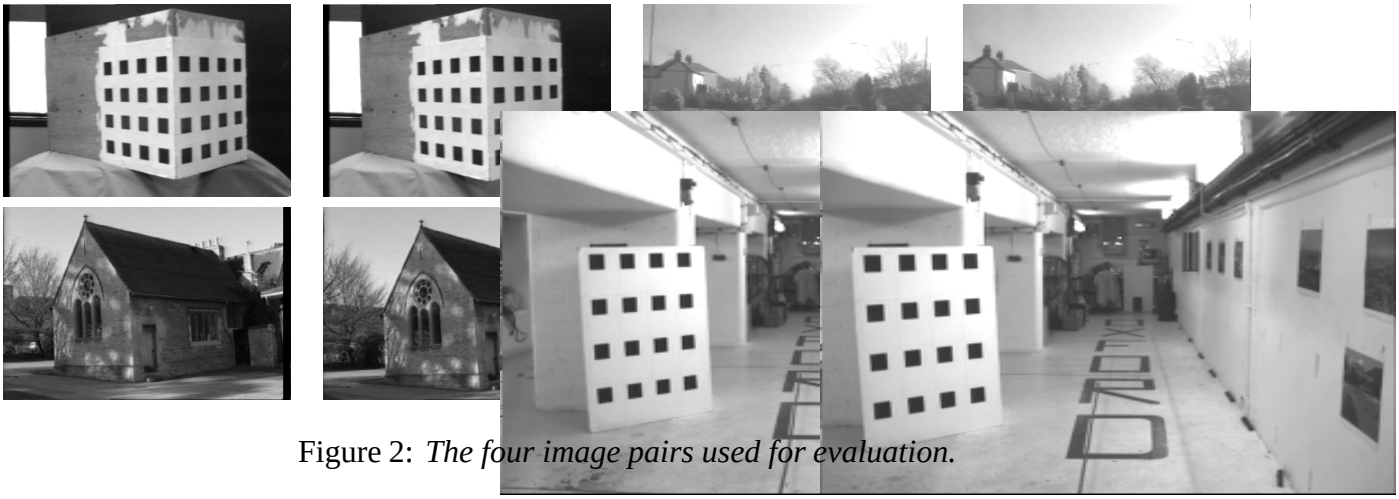


Figure 2: *The four image pairs used for evaluation.*

significantly greater. This can be framed as a hypothesis test. The Hypothesis tested is that the two standard deviations are equal

$$H_0 \quad : \quad \sigma_U = \sigma_C$$

against the alternative, that the two standard deviations are not equal

$$H_1 \quad : \quad \sigma_U \neq \sigma_C$$

under the assumptions that the two sets of observation are Gaussianly distributed. Under H_0, $\sigma_U = \sigma_C$ then

$$F = \frac{\sigma_C^2}{\sigma_U^2} \tag{7}$$

follows an F-distribution with n_c and n_u degrees of freedom [17]. A typical example of the threshold with $n_c = 200$ and $n_u = 200$ leads to the following tests, if

$$\frac{\sigma_C}{\sigma_U} \leq 1.26 \tag{8}$$

then $\sigma_U = \sigma_C$ at the $\alpha = 99\%$ level of confidence. If the value of σ_C fails this test then we can say with 99% confidence that the result has significantly deterioated, in that the two standard deviations are significantly different. By this mechanism it is possible to assess what level of JPEG compression still gives reasonable results.

5 Calibration grid

Figure 4 (a)(b) shows the two uncompressed images of the calibration grid, and (c) shows the matches, obtained from straight line template fitting, superimposed on the second view. There are 128 matches, and only these point correspondences are used in the estimation algorithm. Figure 4 (d) (e) shows the structure of the uncompressed reference object recovered in a Quasi-Euclidean [3] frame. Figure 4 (f) (g) show the reconstruction for a compressed image, with

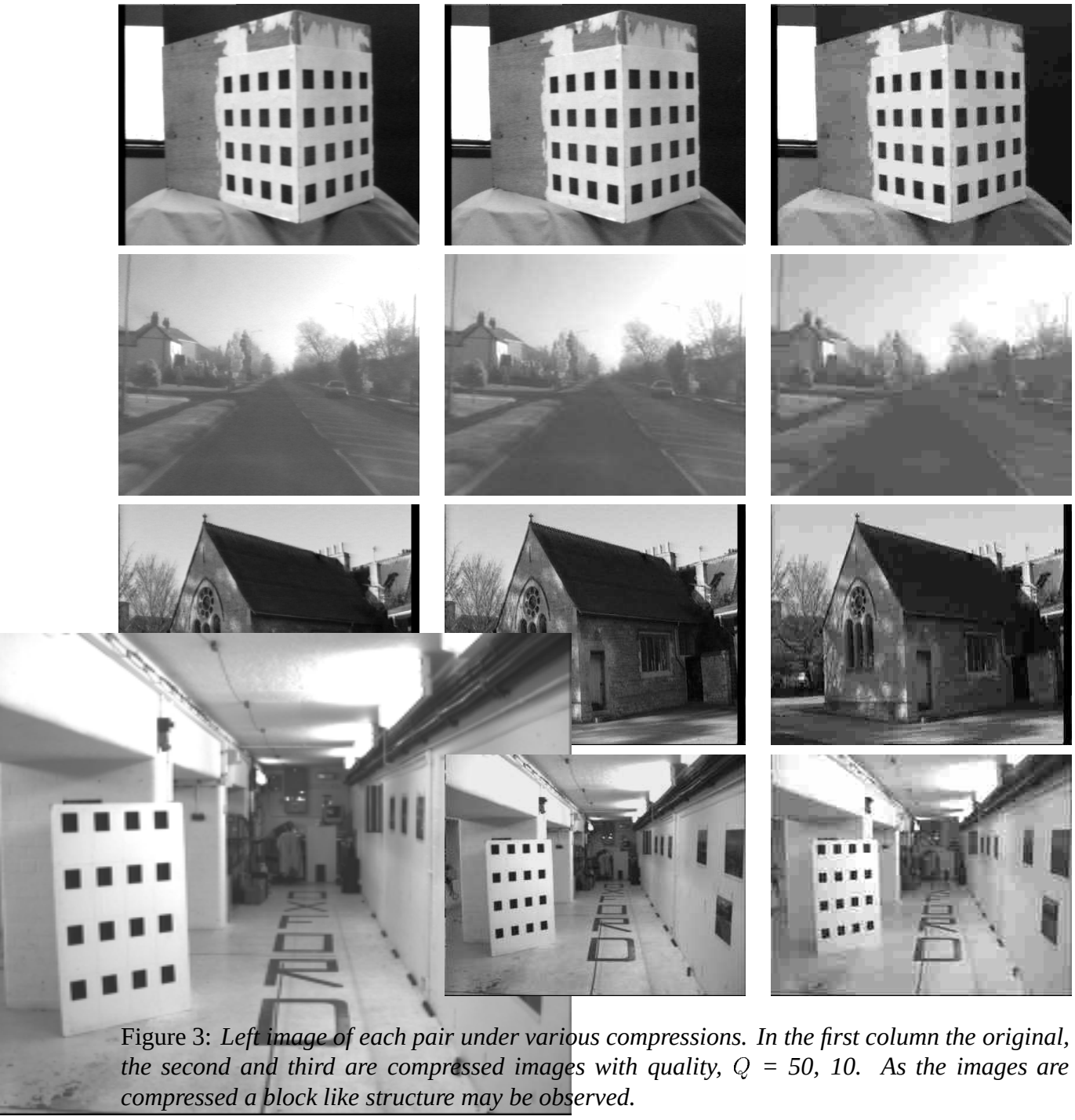


Figure 3: *Left image of each pair under various compressions. In the first column the original, the second and third are compressed images with quality, Q = 50, 10. As the images are compressed a block like structure may be observed.*

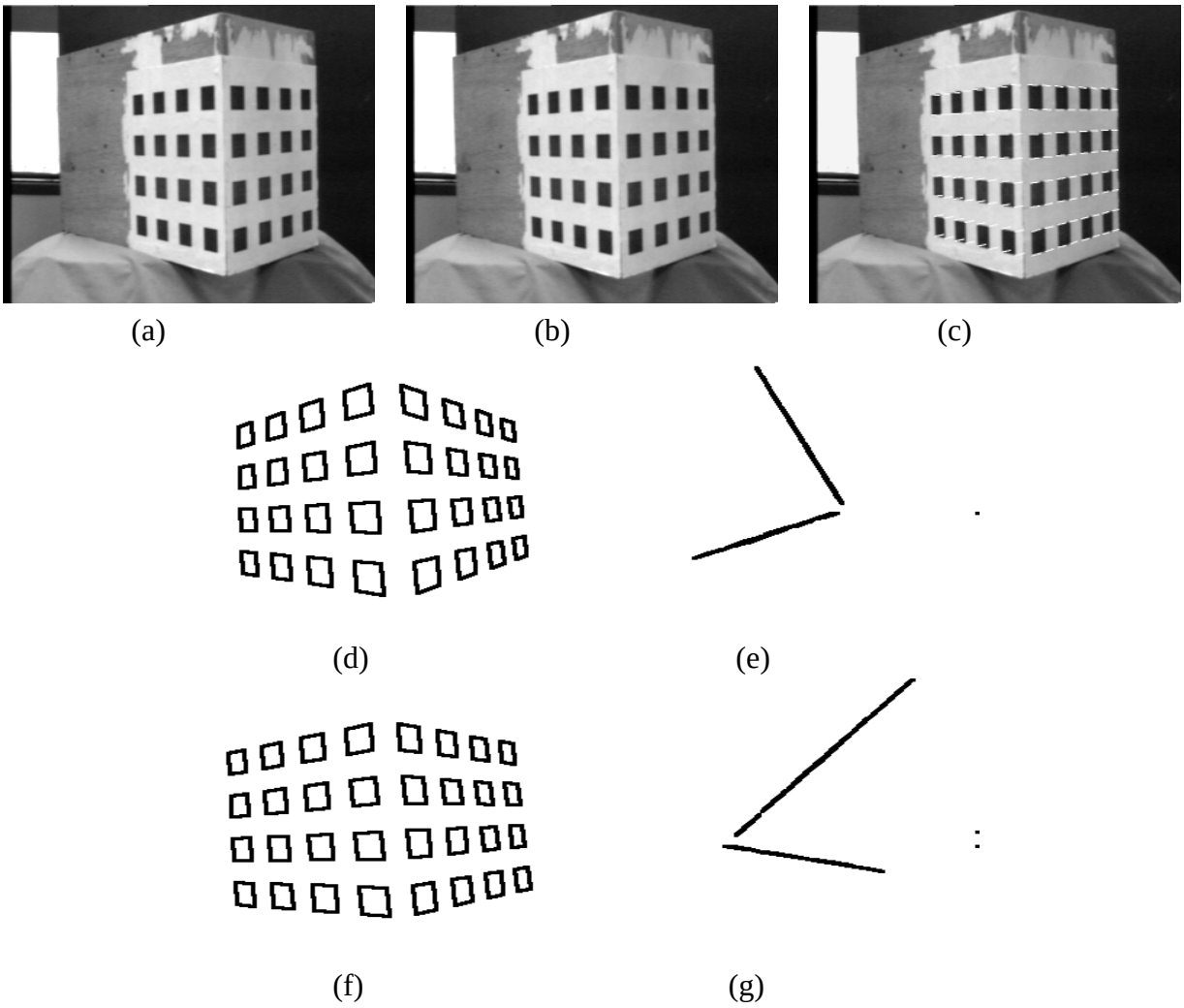
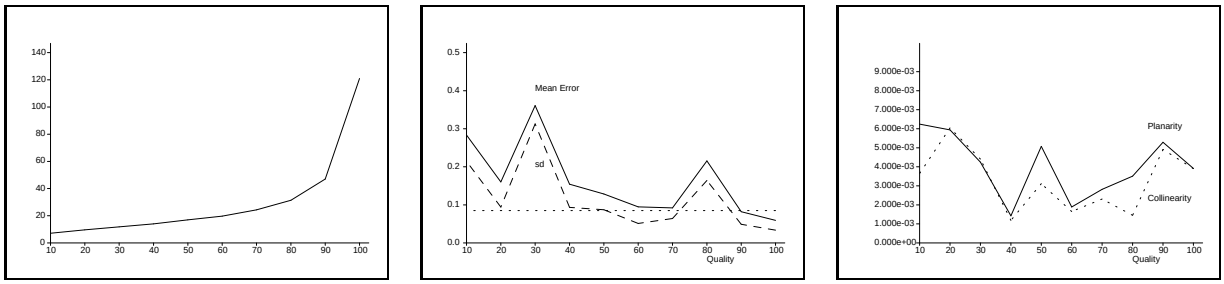


Figure 4: *(a) (b) Two uncompressed views of a calibration grid. (c) Lines joining matched corners superimposed on the second view. (d), (e) The recomputed structure for the grids (e) shows an overhead view, it can be seen that the reconstructed points all lie very close to a plane. Note that only the points at the vertex of each grid square are used in the reconstruction. In the figure they are joined to aid visualisation. (f), (g) The reconstructed grids for the compressed version of the image with* $Q = 10$*. Even though the compression is large the reconstruction still looks good, and the points are still planar.*



(a) (b) (c)

Figure 5: *The x-axis gives the quality, Q. For each graph the y-axis gives (a) The size of the first image after compression (b) The mean and standard deviations of the distance between the computed and actual grid structure, in centimeters. Those standard deviations below the dotted line satisfy an F test (c) The coplanarity and collinearity tests described in the text.*

$Q = 10$, at a first inspection the structure still looks very good. Next several performance measures are discussed.

Accurate 3D Euclidean point positions are known for the calibration grid. This known structure provides a reference frame in which measures on the recovered structure can be applied. First the recovered the 128 points of the projective structure are transformed to their corresponding points in the Euclidean frame. The projective transformation to achieve this is determined by least squares minimising the distance between the transformed and Euclidean point positions.

The distance between the known and transformed structure provides the first error measure. Figure 5 (b) gives the means and standard deviations of this distance error for varying levels of compression $Q = 90, 80 \ldots 10$. There is a very close correlation between the standard deviation and the mean in these examples. The F-test given in Section 4 was performed by comparing the standard deviations of the 128 matches in the compressed images to that of the uncompressed, the threshold thus obtained is drawn on the Figure. The compressed images with $Q = 10, 20, 30, 40, 60, 80$ fail the F test, which indicates that there is a significant degradation in the result for these compressed images. But note that the standard deviations for the images with $Q = 20, 40$ are very much borderline cases.

Further measures, such as collinearity and coplanarity, are made after the transformation to the Euclidean coordinate frame as this ensures that such measurements for each level of compression are made in a single reference coordinate frame. The collinearity measure $L = (\sigma_j^2 + \sigma_k^2)^{1/2}/\sigma_i$ and the coplanarity measure $P = \sigma_k/(\sigma_i^2 + \sigma_j^2)^{1/2}$ of a set of points are obtained by using SVD to obtain the principal axes i, j, k together with the variance σ_i, σ_j, σ_k of point positions along each principal axis. A straight line is thus expected to have $L = 0$ and a plane to have $P = 0$. Figure 5 (c) shows these two measures. It can be seen that the compression does not seem to detrimentally affect these values, both being closely correlated.

The general conclusion that can be drawn for these results is that the image can be compressed even with $Q = 10$ and the results not significantly degraded. The experiments were repeated on the calibration grid for different motions and the results were found to be similar. Of course the calibration object images are exceptional in that (a) the high contrast affords very clearly defined features to extract, (b) the method of intersecting lines produces very accurately located features. This has the advantage that we can clearly define a mapping between the calibrated

object and points detected in the image, but the disadvantage that such clearly defined features are atypical of more complex scenes. Thus the next section addresses the situation where this mapping is not available.

6 Selection of a Performance Measure

It remains now to define performance measures to assess the results of the estimator on uncalibrated images with varying levels of compression. This is a key task as the vision module encapsulating the estimation of $\mathbf{F}$ may be used for several different tasks (estimating structure has already been described, others include matching [2, 30], segmentation [25], and calibration [7]). Thus a set of general, task independent, error measures must be developed. Within this section several measures are derived and their relative merits discussed. The general procedure will be as follows. The ground truth is the (unknown) structure and motion of the scene. These cannot be directly estimated, but maximum likelihood estimates of the projective structure and fundamental matrix can be made from the images with most information (the uncompressed images), and these can be taken as the ground truth. The question then becomes the following: How can we compare two projective structures or two fundamental matrices meaningfully, one estimated for the compressed the other for the uncompressed data? There is no distance metric defined in projective space so it does not make sense to make the comparison there. Furthermore there is no guarantee that the same features will be detected when the image is compressed, as there was in the case of the calibration grid images. However, it does make sense to compare distance on the image plane, as both the images for the compressed and uncompressed data will be the same size. The maximum likelihood estimates of the structure projected into the image planes are referred to as the optimally corrected correspondences [16], being the maximum likelihood estimate of the true correspondence $\{\underline{\mathbf{x}}_i \leftrightarrow \underline{\mathbf{x}}_i'\}$. They have the advantage that they can be computed directly (without reference to the structure) and it is shown how to do this efficiently in Appendix B. Thus in subsection 6.2 deviance measures are developed between the optimally corrected correspondences and the computed fundamental matrices. In subsection 6.3 the more traditional approach of covariances matrices is discussed.

6.1 Standard Deviation of Inliers Measure—IM

The most obvious, and perhaps simplest measure, is the standard deviation of $d(\mathbf{F}, \mathbf{x}_i \leftrightarrow \mathbf{x}_i')$ for the inliers. This standard deviation is referred to as measure **IM**. The problem is that this is not known *a priori* and must be estimated from the data which is a mixture of two probability distributions, one for the inliers and one for the outliers. The output of the robust estimator provides an initial guess at $\mathbf{F}$, the inliers and the standard deviation of the inliers. The problem with this initial estimate of the standard deviation is that, although it is useful for weeding out gross outliers it is an inconsistent estimate of the true standard deviation (meaning that as the number of observations increases it does not tend to the true value). To understand why this is so the robust estimate of the standard deviation is now explained.

Given a set of Gaussian observations $d_i, i = 1 \ldots n$. It is known that $\text{med}_i |d_i| / \Phi^{-1}(0.75)$ is an asymptotically consistent robust estimator of $\underline{\sigma}$ when the inlying errors d_i are distributed as

$N(0, \underline{\sigma}^2)$, where Φ is the cumulative distribution function for the Gaussian probability density function[1]. It has been shown [22] empirically that the correction factor of $\left(1 + \frac{5}{d_f}\right)$ improves the estimate of the standard deviation, where $d_f = 7$ is the number of degrees of freedom (parameters in $\mathbf{F}$). On noting that $1/\Phi^{-1}(0.75) = 1.4826$ the estimate of $\underline{\sigma}$ is

$$\hat{\sigma} = 1.4826 \left(1 + \frac{5}{7}\right) \text{med}_i |d_i|. \tag{9}$$

This robust estimate of the standard deviation is highly useful when nothing is known about the noise distribution, as it allows the automated setting of thresholds to determine whether features are inlying. There is a complication though, which renders the first estimate of the standard deviation of the inliers inconsistent. The estimate of σ from (9) assumes the errors are all from the same Gaussian distribution. As some outliers, with a higher variance, will be included in this set the estimate of σ is typically too high. To compensate for this, once the final result is obtained (after non-linear minimization) the estimate of the standard deviation is improved by the EM algorithm [5], which was developed for mixture models such as this. It is assumed that the inlier and outlier distribution are Gaussian but with different parameters (Modelling the probability density function of outliers is a difficult and unresolved problem. As there is little *a priori* information about them. It might be more appropriate to model their probability density function as diffuse or uniform. However this may be approximated by a Gaussian with a suitably large standard deviation). The EM algorithm is guaranteed to increase the likelihood of the estimated standard deviation given the data, and is briefly described in Appendix C. Synthetic tests have shown that it greatly increases the accuracy of the estimate of the underlying standard deviation of the inliers. For the original and all compressed images the standard deviation of the inliers for the estimated $\mathbf{F}$ are calculated and then $\mathbf{IM} = \sigma$. Successive measures **IM** for the compressed images may be compared to the original by a suitable F test.

6.2 Deviation Measures—DM1, DM2

Whereas measure **IM** examines the deviation of the data from the fit to each set of correspondences, it might be more appropriate to examine the deviation of the fit for the compressed data from the original fit. One way to do this is by using covariance matrices, described below. Another way is described in this section.

For the uncompressed data, $\mathbf{F}_U$ and the maximum likelihood estimates $\{\hat{\mathbf{x}}_U\}$ of the true (noise free) correspondences $\{\underline{\mathbf{x}}_U\}$ are estimated. These correspondences are viewed as the ground truth set of correspondences. Note that for these correspondences there is zero error, i.e. $D(\mathbf{F}_U, \{\hat{\mathbf{x}}_U\}) = 0$.

Given an estimate of $\mathbf{F}_C$ for a compressed image $D(\mathbf{F}_C, \{\hat{\mathbf{x}}_U\})$ provides a measure of the deviation of $\mathbf{F}_C$ from $\mathbf{F}_U$ in the region of correspondences. If the estimated $\mathbf{F}_C$ for the compressed data is close to $\mathbf{F}_U$ then the standard deviation of $d(\mathbf{F}_C, \hat{\mathbf{x}}_U)$ will not be large. The standard deviation of this distance is measure **DM1**. Note that $D(\mathbf{F}_C, \{\hat{\mathbf{x}}_U\}) = 0$ and $\mathbf{DM1} = 0$ if and only if $\mathbf{F}_C = \lambda \mathbf{F}_U$.

[1] $N(0, \underline{\sigma}^2)$ signifies a Gaussian or normal distribution mean 0 and variance $\underline{\sigma}^2$.

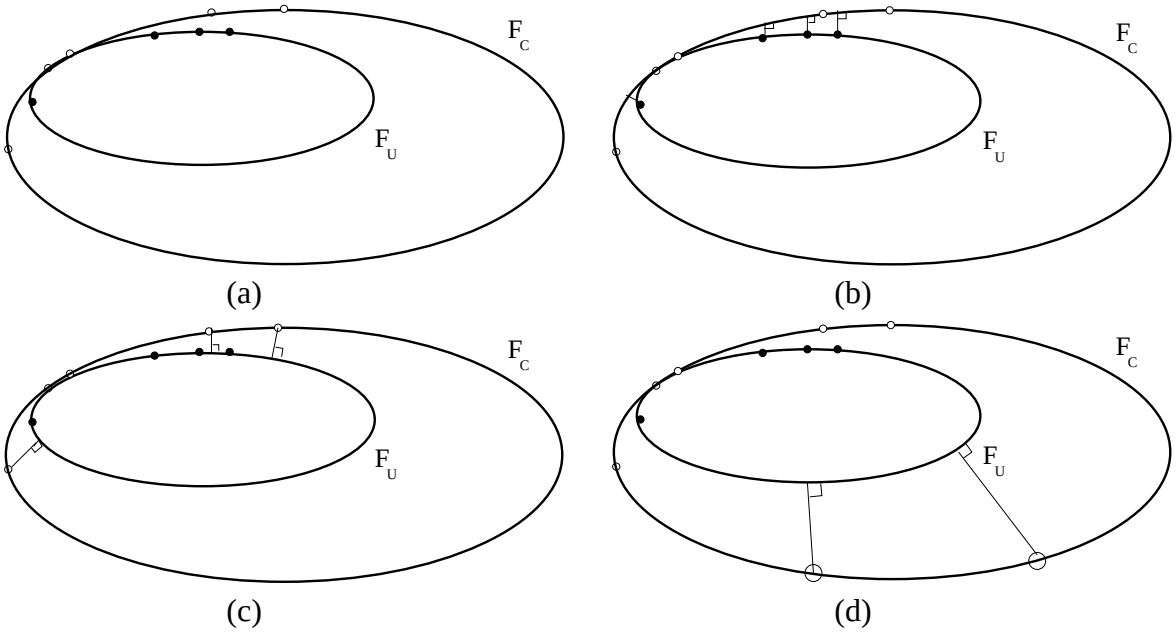


Figure 6: *A lower dimensional depiction of the deviation measures. (a) Two conics F_U, F_C, along with several points observed on the conics. (b) (c) ,* **DM1**, **DM2**. *The deviation of one conic from another can be measured as the standard deviation of the orthogonal distances of the observation on the first conic to the second.* **DM1** *being the standard deviation of the distance of points on F_U to the conic F_C.* **DM2** *being the standard deviations of the points on F_C to the conic F_U. (d) Alternatively new random points may be generated on the conic and the standard deviation of these measures monitored.*

Another approach is to store the original estimate $\mathbf{F}_U$ for the uncompressed data; then $D(\mathbf{F}_U, \{\mathbf{x}_C\})$ provides a different measure of the difference between $\mathbf{F}_U$ and $\mathbf{F}_C$. The standard deviation of $d(\mathbf{F}_U, \mathbf{x}_C)$ is measure **DM2**.

A two dimensional analogue of these measures is shown in Figure 6 where an ellipses represents in two-space the fundamental matrix (which is a quadric 4-space (x, y, x', y')). As can be seen, both of these measures only give the deviation of the two fundamental matrices from each other within the region of the observations. Zhang [29] proposes choosing points at random on the curve and monitoring their variance, such an approach bears some similarity to the approach mooted by proponents of the AIC criterion for model evaluation [1], which measures the expected value of new random observations. This approach is shown in Figure 6 (d). Here two curves are shown that fit well within the region of the observations but diverge elsewhere. In this case random points that are taken far from the data on the first curve will show a high deviation from the second curve. The problem with this approach is that the measure is determined largely by the scheme selected for randomly generating new observations. This method is not employed here.

6.3 Covariance Matrix Measure—CM

The two previous approaches have been based on analysis of variance of the errors defined in various ways. In this section a method involving covariance matrices is established. Let $\mathbf{f}$ be the 7 parameters of the epipolar parameterization described in above written as a vector:

$$\mathbf{f} = \begin{pmatrix} a & b & c & e_1 & e_2 & e'_1 & e'_2 \end{pmatrix}^\top \quad . \tag{10}$$

If the true covariance matrix Σ_F and true set of parameters $\underline{\mathbf{f}}$ are known, a fit $\mathbf{f}$ could be assessed by the measure:

$$\mathbf{CM} = (\underline{\mathbf{f}} - \mathbf{f})^\top \underline{\Sigma}_F^{-1} (\underline{\mathbf{f}} - \mathbf{f}) \tag{11}$$

which is a χ^2 variable with 7 degrees of freedom [16]. The true covariance and true set of parameters are unknown, but both are provided by the non-linear minimization [10, 28].

The covariance of the parameters may be calculated from the inverse Hessian of D_R at the solution. The Hessian matrix of D_R is the matrix of second partial derivatives with respect to our solution, and its (i, j)th element is given by $\mathbf{G}(i, j) = \partial^2 D_R / (\partial \mathbf{f}_i \partial \mathbf{f}_j)$. In many least squares applications the is well approximated by $\mathbf{G} = 2\mathbf{J}^\top \mathbf{J}$, where $\mathbf{J}$ is matrix of first partial derivatives of the error function and its (i, j)th component is $\partial d_i / \partial \mathbf{f}_j$, where d_i is the error of the ith correspondence. This is due to the special structure of the Hessian for non-linear least squares problems [28] as

$$\mathbf{G} = 2\left(\mathbf{J}^\top \mathbf{J} + \sum_{i=1}^{n} d_i \mathbf{H}_{d_i}\right) \quad , \tag{12}$$

where $\mathbf{H}_{d_i}$ is the Hessian of d_i. In the neighbourhood of the solution (with outliers removed) $|d_i|$ is often small or zero (for an exact fit). In such cases $\mathbf{J}^\top \mathbf{J}$ is an adequate approximation to the Hessian $\mathbf{G}$ (obviating the need to calculate the second derivatives), which is a first order approximation to the Fisher information matrix. The estimated covariance matrix Σ is given by $\Sigma = \sigma^2 \left(\mathbf{J}^\top \mathbf{J}\right)^{-1}$ where σ^2 is the estimated variance of the residuals d. This can be estimated

by numerical methods on the data set, excluding the outliers, and is provided as a by product of the numerical minimization. If Σ is singular or near singular then there exist multiple solutions for $\mathbf{F}$, as pointed out by Horn [14].

The measure is obtained as follows. First the covariance matrix and solution are estimated for the uncompressed data and used to define the ground truth: $\underline{\Sigma}_F$, $\underline{\mathbf{f}}$. Then, for each estimate of $\mathbf{f}$ at a given level of compression the expression (11) gives the **covariance measure—CM**. If the error function is approximately quadratic in the vicinity of the solution, **CM** may be approximated as a χ^2 with seven degrees of freedom. If **CM** ≥ 12.02 then we can be 90% certain that the solution provided by the compressed data is very different from that provided by the uncompressed.

Another measure that is examined is the condition number of the covariance matrix of a given fit, being the ratio of the largest to smallest eigenvalue of that covariance matrix. This measure is referred to as **CN**.

7 Results

Within this section the measures are tabulated for the derived performance measures on the test images.

Calibration Grid In Section 5 measures were derived explicitly for calibration grids. The uncalibrated measures **DM1**, **DM2**, **CM**, are now applied to the fundamental matrix estimated for the calibration grid shown in Section 5. Previously the known structure was used to help in estimating the fundamental matrix, but in this section the standard algorithm as described in Section 2 is used to estimate $\mathbf{F}$. The results are shown in Table 1. Apart from the compressed images with $Q = 60, 70$ where the algorithm converged incorrectly the results are generally good across the spectrum of compressions. Again good results are obtained due to the high contrast nature of the image, and the high degree of accuracy achieved by interesting lines to identify features.

The algorithm for estimating the fundamental matrix proved very unstable on all these images. This is because the matches all lie on two planes. Previous analysis by Luong and Faugeras [19] has shown that such a case produces high instability, and needs to be catered for by special purpose homography fitting algorithms as described in [19, 26].

Road Pair For a camera undergoing pure translation corresponding points move along epipolar lines, so that lines joining the matched points and the epipole are collinear. The epipolar lines for each point are shown in Figure 7 (c). There are 74 matches shown in Figure 7 (d), the inliers in (e), and the outliers in (f) generated by the robust algorithm.

For pure translation, the fundamental matrix has a special skew symmetric form

$$\mathbf{F}_T = \begin{bmatrix} 0 & g_3 & -g_2 \\ -g_3 & 0 & g_1 \\ g_2 & -g_1 & 0 \end{bmatrix}, \tag{13}$$

where $\mathbf{e} = (g_1, g_2, g_3)^\top$ is the epipole (having two degrees of freedom). To evaluate the change in the error surface as the image is compressed, the robust error function (5) is evaluated with the epipole varying over the image. In Figures 7 (f) (g) (h) the robust Huber function given in (5) is shown as a contour plot, superimposed on the images. It can be seen that the error surface becomes increasingly shallow as the compression increases and develops several local minima. The uncertainty increases along the y axis as the compression is increased. This is can be understood as follows. The epipole can be thought of as determined by the intersection of the lines $\mathbf{x} \times \mathbf{x}'$. It can be seen in Figures 7 (d) that most of these lines are near horizontal, because the disposition of correspondences just happens to be such that the matches are on the left and right sides of the image. Thus the x coordinate of the epipole has a much lower standard deviation than the y coordinate. This effect is exacerbated as the accuracy of the points decreases. In other examples this "starring" effect has been noticed to a lesser or greater extent, in that if a sector of the image is sparser in matches than the rest, the epipole will be less certainly determined in the direction of that quadrant, as shown in Figure 8.

The performance measures are shown in Table 2, all provide acceptable results with $Q = 100 \ldots 30$. For $Q < 30$ the algorithm failed due to insufficient matches being detected. The covariance matrices for this example are much better conditioned than for the fundamental matrices estimated with more parameters (a total of seven in the general case) of the following two examples, highlighting the advantages of explicitly using a reduced model. For the uncompressed data the covariance matrix is

$$\Sigma = \begin{bmatrix} 54.6 & -32.1 \\ -32.1 & 280.0 \end{bmatrix}. \tag{14}$$

it can be seen that the y coordinate of the epipole is much less certain than the x coordinate, this is borne out by the valley like nature of the error surface. Note that the covariance matrix is the inverse of the Hessian of the error surface.

Chapel Pair In Figure 9 the matches, inliers and outliers are shown on compressed versions of the original image, using JPEG compression with quality set at 100, 70 and 20. At quality 20 the matching process performs poorly.

Graphs of the measures DM1, DM2 and IM are shown in Figure 10. It can be seen that all have a fairly similar shape. The standard deviation of the inliers for the uncompressed image is $\sigma_U = 0.28$. Examining the measures DM1 and DM2 and comparing them by an F-test to σ_U the deviations of both these measures is not too severe until a compression of 40, at this point the solution degrades rapidly. Measures DM1 and DM2 give similar results. For quality 90–50 there were a similar number of inliers. The condition number of the covariance matrix did not provide any significant information. The measures are tabulated in Table 3.

The conclusion is that this image may be compressed with $Q = 50$, which results in an 84% storage reduction, without significantly effecting the error measure.

Basement Pair Table 4 gives numerical values of the measures. The standard deviation of the inliers for the uncompressed data is $\sigma_U = 0.1606$ which indicates that compression 10–70 provide reasonable results for measures DM1 and DM2. The condition numbers CN of the covariance matrices revealed little correlation with the compression. The measure CM again is

Calibration grid pair

Quality Q	Matches	Inliers	IM	DM1	DM2	CM	CN
100	128	128	0.020	0.020	0.020	0.0	16701
90	128	128	0.025	0.025	0.025	0.064	22106
80	128	128	0.084	0.073	0.072	0.064	1089255
70	128	128	0.166	0.172	0.150	477003	77222
60	128	128	0.079	0.083	0.082	300.72	190800
50	128	128	0.016	0.016	0.016	0.001	11498
40	128	128	0.012	0.022	0.023	0.228	16166
30	128	128	0.013	0.022	0.024	46.638	1115n
20	128	128	0.007	0.037	0.039	0.23	11209
10	128	128	0.019	0.067	0.068	0.4436	4565

Table 1: *The measures for the calibration grid images shown in Figure 5. Due to the simple nature of these images the results are not significantly degraded.*

Road pair

Quality Q	Matches	Inliers	IM	DM1	DM2	CM	CN
100	74	63	0.488	0.488	0.488	0.0	1.28
90	60	53	1.052	1.750	2.206	2.3	6.93
80	67	56	0.511	2.400	1.026	0.3	4.76
70	65	53	0.660	2.368	0.418	0.5	2.13
60	61	44	0.710	2.113	0.438	0.3	2.57
50	62	47	0.775	2.062	1.273	0.7	1.14
40	72	49	0.730	1.645	2.746	0.4	5.93
30	69	52	0.797	3.519	2.531	1.7	1.84

Table 2: *The measures for the road images shown in Figure 7. At $Q = 10, 20$ the algorithm failed to converge at all, producing fewer than 10 inliers for each image leading to very poor results.*

very low, due to great uncertainty in the covariance matrix, but indicates that compression 10–70 provide reasonable results, in agreement with measures DM1 and DM2. A compression with quality set to 70 leads to an 86% reduction in image storage, similar to the previous suggested compression.

8 Discussion

The main contribution of this paper is a methodology for determining the similarity of two different estimates of the fundamental matrix for a scene obtained either by different methods or from different sets of observations. Overall the error metrics DM1, DM2 and CM are highly correlated. The covariance matrix is only based on first order approximations yet proves useful as the error surfaces are generally well approximated by a quadratic surface at the solution.

It seems that, although compression will always degrade the result, a compression of up

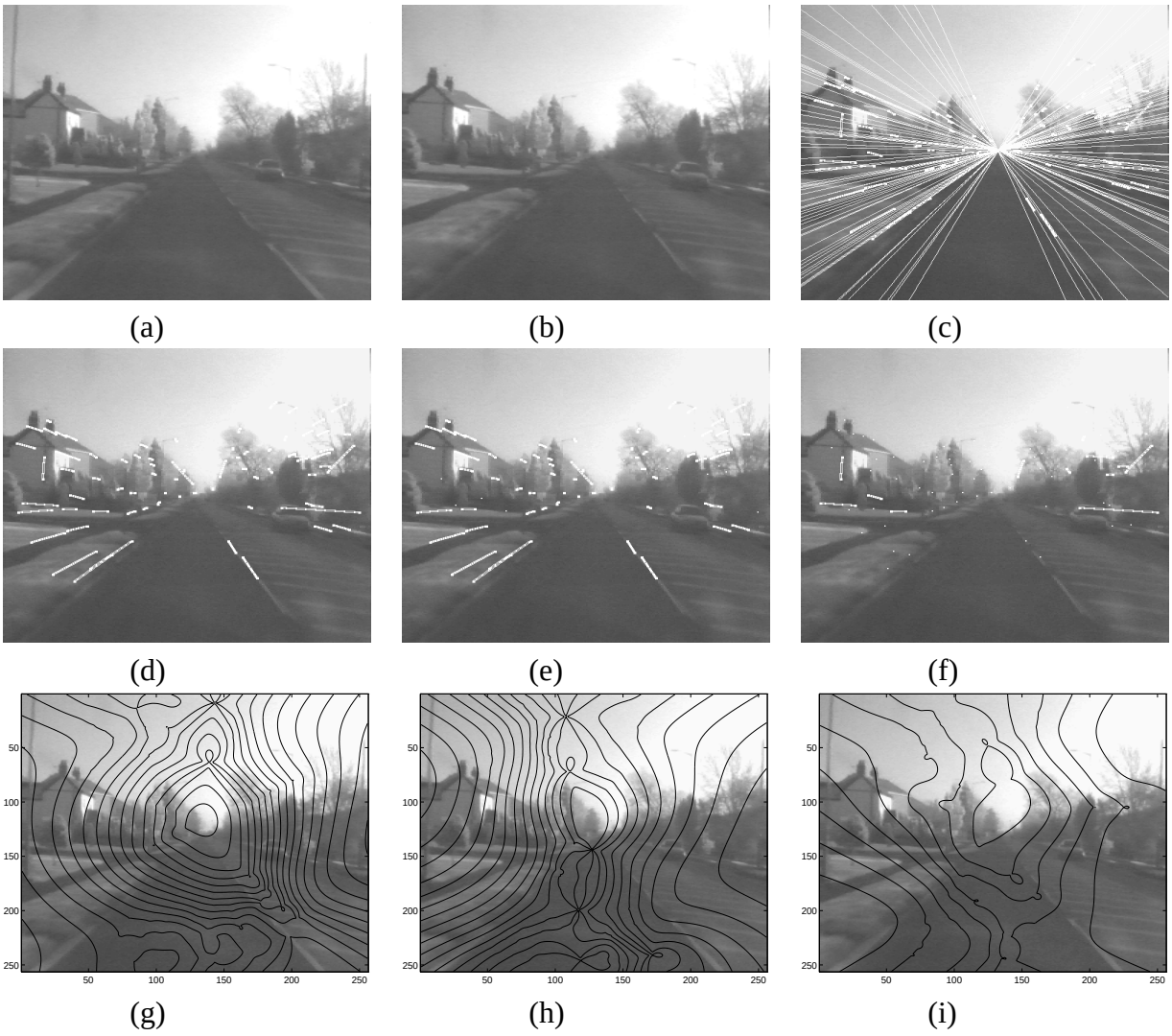


Figure 7: *(a) (b) two images taken from a car translating down a road, (c) the estimated epipolar geometry, (d) the matches, (e) inliers (f) outliers, (g) (h) (i) contour plots of the Huber error surface as the epipole varies across the image, for the uncompressed image, and* $Q = 80, 50$. *The uncertainty increases along the* y *axis as the compression is increased.*

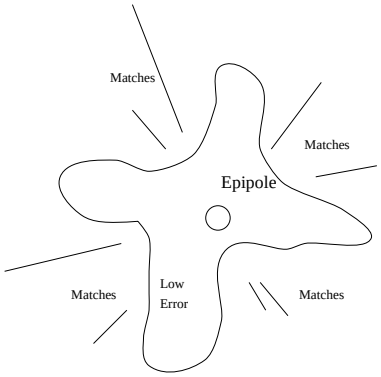


Figure 8: *Assuming that the motion is pure translation, if error function is plotted on the image as the epipole is varied, then, the error surface gives low values in the direction of any sectors sparse on matches. This is shown in the Figure. This effect does not show up in covariance matrices which are always elliptical in shape.*

Chapel pair

Quality Q	Matches	Inliers	IM	DM1	DM2	CM	CN
100	677	410	0.527	0.527	0.527	0.0	3865
90	652	335	0.493	0.150	0.225	1.2	3622
80	642	341	0.842	0.563	0.83	6.1	1734
70	624	374	0.794	0.610	0.944	3.4	1708
60	629	338	0.816	0.658	1.084	2.3	1412
50	608	339	1.027	0.295	0.436	7.2	2779
40	587	253	1.528	3.842	4.204	22.3	71
30	294	204	0.484	0.077	0.0717	5.3	22.6
20	518	326	3.161	0.791	1.915	16.2	11450
10	461	378	8.237	1.589	0.791	11.4	6118

Table 3: *The measures for the chapel images shown in Figure 9.*

Basement pair

Quality Q	Matches	Inliers	IM	DM1	DM2	CM	CN
100	323	243	0.1606	0.1606	0.1606	0.00	228.2
90	316	219	0.2285	0.0017	0.0022	1.36	6.9
80	288	186	0.3114	0.0099	0.0145	1.52	11.6
70	298	207	0.334	0.0089	0.0846	5.19	17.3
60	295	196	0.3135	0.0112	0.0244	7.07	8.1
50	306	214	0.4216	0.0203	0.0681	4.8	13.1
40	302	207	0.4843	0.0309	0.1382	11.52	16.3
30	294	204	0.4844	0.0769	0.0717	5.31	22.6
20	296	189	0.6526	0.2346	0.1324	39.33	57.8
10	158	108	0.6239	0.2812	0.1782	19.2	28.6

Table 4: *The measures for the basement images.*

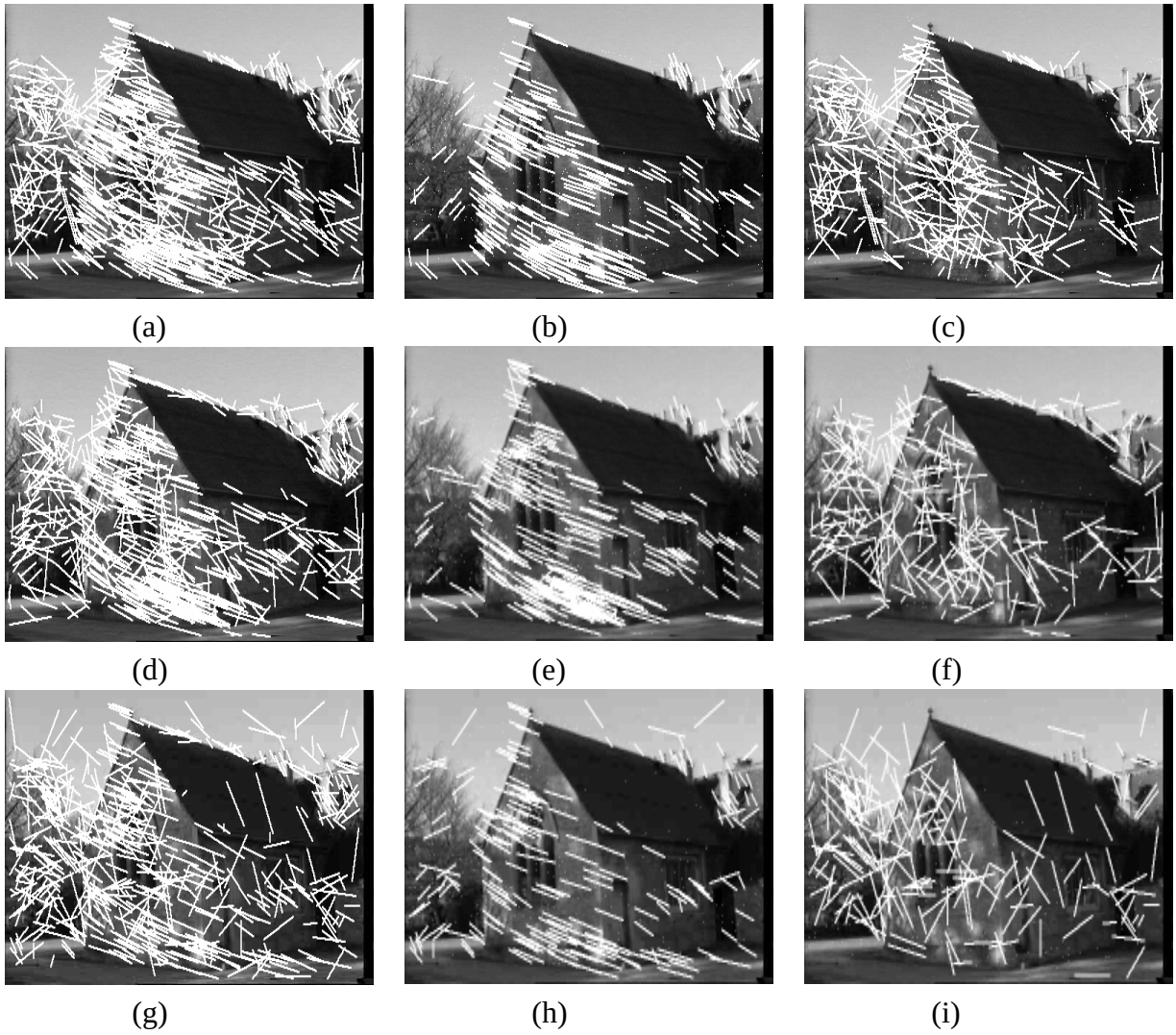


Figure 9: *Matches for the chapel pair under various compressions. The first column (a), (d), (g) show the detected matches, the second column shows the inliers, the third column the outliers. The first row has quality 100, the second 70, the third 20.*

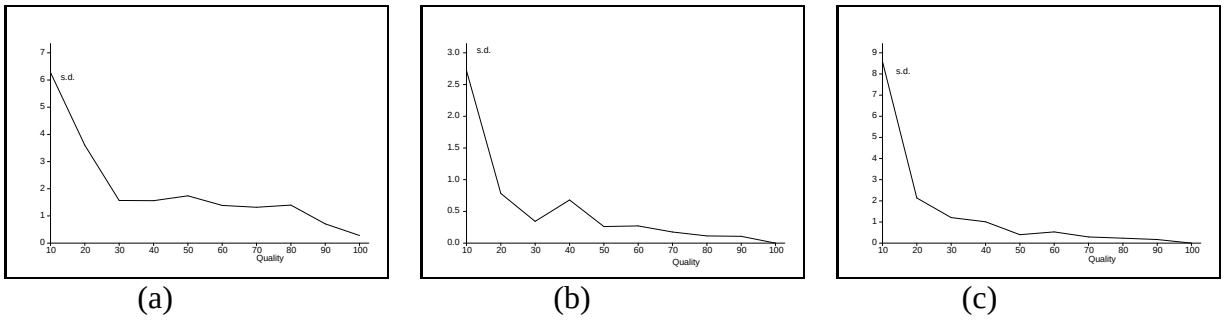


Figure 10: *Graphs of the measures on the chapel images. (a) IM, (b) DM1, (c) DM2. The y-axis gives the standard deviations, the x-axis gives the Q the quality.*

to 85%, by setting $Q = 30$ or 40, may be tolerated and useful results still obtained, but this is the limit of the possible performance. For $Q < 30$ a tiling effect begins to dominate the image texture, causing the appearance of spurious features. Examination of Figure 5 reveals that the largest percentage of compression occurs from $Q = 100$ to $Q = 70$ being a 3:1 reduction. Given that further compression beyond $Q = 70$ results in a small decrease in size for worsening results, it is suggested that $Q = 70$ is the most an image should be compressed without risk of substantial degradation. Of course what can be termed significant is highly task specific, for some processes—such as obstacle avoidance—course estimates may be sufficient, for others—such as parts inspection—all the information must be used. Thus the results presented here should be taken merely as guidelines. The same tests run on images compressed by image subsampling/averaging on a 3:1 ratio produced extremely poor results in comparison, which suggests that it always better to use a good image compression mechanism rather than subsampling, if the objective is merely to reduce the storage space required for images. Future areas of research might examine other types of compression in more detail, such as smoothed subsampling, and MPEG compression. It might also be worth investigating ways of generating synthetic images where the ground truth can be known exactly.

Acknowledgements We would like to thank Paul Beardsley and Andrew Fitzgibbon for software and for helping to produce some of the images in this paper. The road images were supplied by Pilkington. Financial support for this project was provided by ACTS project VANGUARD.

A Error Criterion

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. 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 $\mathcal{R}^4$.

Given a set of correspondences the (unbiased) maximum likelihood 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 $\mathcal{R}^4$ [13, 16, 17, 21, 23, 24, 27]. A closed form solution for this distance cannot be found, but a first order approximation can be made (it is in fact the solution of a sixth order polynomial as explained in the Appendix B). This was first proposed by Sampson: A first order approximation to the orthogonal distance leads to the error term $d_s = r/|\nabla r|$, where r is the algebraic residual. Note that d_s is undefined at the point in $\mathcal{R}^4$ given by the two epipoles in each image, as here $|\nabla r|$ is zero. Tests run on real and synthetic data show agreement typically of three or four significant figures between the first order approximation d and the exact geometric distance. The error d_s is typically 0.01 to 0.1 pixels depending on the type of scene. The approximation is increasingly good the nearer the points in $\mathcal{R}^4$ are to the variety V and the smaller the curvature of V. The sign of the error term d is significant, and it indicates what side

of the variety the point (in $\mathcal{R}^4$) lies on. As d_s is a signed variable it can be treated as a proper Gaussian variable.

The distance $D_S = \sum_i r^2 \left(\frac{1}{r_x^2 + r_y^2 + r_{x'}^2 + r_{y'}^2} \right)$ is adopted as the error criterion, as it has a geometric significance, providing a first order approximation to the maximum likelihood solution.

B Optimal Correction

Within this appendix it is explained how the optimally corrected image points are obtained, being the maximum likelihood estimate of the true correspondences given $\mathbf{F}$ and a set of observed correspondences. This is used when the ground truth is unknown and maximum likelihood estimates of it must be obtained from the uncompressed data.

Given the fundamental matrix and the image correspondences, the 3D projective structure can be recovered. Hartley and Sturm [13] provide an optimal method for computing structure. Given a measured correspondence $\mathbf{x} \leftrightarrow \mathbf{x}'$ and $\mathbf{F}$ they seek points $\hat{\mathbf{x}}$, $\hat{\mathbf{x}}'$ that minimize the function $d_h = d(\mathbf{x}, \hat{\mathbf{x}}) + d(\mathbf{x}', \hat{\mathbf{x}}')$ where d() represents the Euclidean distance on the image plane, subject to the epipolar constraint $\hat{\mathbf{x}}'^{\top} \mathbf{F} \hat{\mathbf{x}} = 0$. Assuming a Gaussian distribution $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}'$ are the most likely values of the true correspondences. Once $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}'$ are found, the point $\mathbf{X}$ may be found by triangulation, since the corresponding backprojection rays will meet precisely in space. Hartley and Sturm show how $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}'$ may be found as the solution of a degree 6 polynomial.

In fact this distance d_h is exactly the same as the orthogonal distance of the correspondence examined in the previous section. This suggests the first order approximation to $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}'$ is given by

$$\begin{pmatrix} \hat{x} \\ \hat{y} \\ \hat{x'} \\ \hat{y'} \end{pmatrix} \approx \begin{pmatrix} x \\ y \\ x' \\ y' \end{pmatrix} - r \begin{pmatrix} r_x \\ r_y \\ r_{x'} \\ r_{y'} \end{pmatrix} . \tag{15}$$

This turns out to be equivalent to the optimally corrected correspondence of Kanatani [16]. Comparisons of the computation of $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}'$ by Hartley and Sturm's method and that of Kanatani have again shown agreement to 3 or 4 significant figures, as mentioned, but Kanatani's method is several orders of magnitude faster, being a linear method, and hence preferable for rapid evaluation purposes.

C EM Algorithm

When data might arise from more than one probability distribution then it is appropriate to model it as a mixture model. There are several approaches to estimating the parameters of a mixture model. The EM algorithm is used here as it is relatively easy to implement and speedy to converge providing the estimate of $\mathbf{F}$ is sufficiently good, and it has been found that the robust estimators described above do provide a good estimate of the solution.

Assuming that we have estimated accurately the set of parameters, each datum has an associated error d. A mixture model is assumed for the set of errors, with two categories ω_j, $j = 0$ for inliers, $j = 1$ for outliers. The probability of an element being in a given category is $\Pr(\omega_j)$, with $\Pr(\omega_0) + \Pr(\omega_1) = 1$. Both are assumed Gaussian with mean $\mu_j = 0$ and standard deviation σ_j. (It is assumed that both errors have zero mean for the maximum likelihood estimate. Recall that we can assign a sign to the error depending on what side of the variety it lies in $\mathcal{R}^4$. The zero mean assumption has been borne out in practise for the inliers, for the outliers the estimated distribution tends to be extremely long tailed (almost uniform) and the mean cannot be determined.) Given the classification of points between inliers and outliers by the robust estimator the initial estimate of the Gaussian models may be provided by

$$\sigma_0^2 = \sum_{\text{inliers}} \frac{d^2}{n_i - 7} \quad \sigma_1^2 = \textstyle\sum_{\text{outliers}} \frac{d^2}{n_o} \quad \Pr(\omega_0) = \frac{n_i}{n} \text{ and } \Pr(\omega_1) = \frac{n_o}{n}$$

where n_i and n_o are the numbers of inliers and outliers provided by the robust estimator respectively. The EM algorithm proceeds in two iterative steps:

1. Choose an initial estimate of the parameters, $\Pr(\omega_j)$, σ_j.
2. Obtain maximum likelihood estimate of σ_j holding $\Pr(\omega_j)$ constant, $\hat{\sigma}_i^2 = \frac{\sum_{k=1}^{n} \Pr(\omega_i|d_{Si},\sigma_i) d_{Si}^2}{\sum_{k=1}^{n} \Pr(\omega_i|d_{Si},\sigma_i)}$.
3. Obtain expected values of $\Pr(\omega_j)$ holding σ_j constant, $\Pr(\hat{\omega}_i) = \frac{1}{n} \Pr(\omega_i|d_{Si},\sigma_i)$.
4. Repeat until convergence.

Where $\Pr(\omega_i|d_{Sk},\sigma_i) = \frac{p(d_{Sk}|\omega_i,\sigma_i)\Pr(\omega_i)}{\sum_{i=1} sp(d_{Sk}|\omega_i,\sigma_i)\Pr(\omega_i)}$, and each datum follows an appropriate Gaussian: $p(d_{Sk}|\omega_i,\sigma_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(\frac{-d_{Sk}^2}{2\sigma_i^2}\right)$. If there are fewer than 5% inliers the EM algorithm is not used, as it has been found that the robust estimation of standard deviation is accurate. Generally the EM algorithm has a dramatic effect on the standard deviation sometimes halving the estimated standard deviation.

References

[1] H. Akaike. A new look at the statistical model identification. *IEEE Trans. on Automatic Control*, Vol. AC-19(6):716–723, 1974.

[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.

[3] P. A. Beardsley, A. Zisserman, and D. W. Murray. Sequential update of projective and affine structure from motion. Technical Report OUEL 2012/94, Dept of Eng Science, University of Oxford, 1994.

[4] G. Csurka, C. Zeller, Z. Zhang, and Faugeras O. Characterizing the uncertainty of the fundamental matrix. Rapport de Recherche 2560, INRIA, 1996.

[5] 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.

[6] 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.

[7] O.D. Faugeras, Q.T. Luong, and S.J. Maybank. Camera self-calibration: theory and experiments. In G. Sandini, editor, *Proc. 2nd European Conference on Computer Vision , LNCS 588, Santa Margherita Ligure*, pages 321–334. Springer-Verlag, 1992.

[8] 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.

[9] W. Förstner. 10 pros and cons against perfomance characterization of vision algorithms. In *Workshop on Performance Characterization of Vision Algorithms, Cambridge*, pages 1–22, 1996.

[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, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. *Conference on Computer Vision and Pattern Recognition*, 1992.

[13] R. I. Hartley and P. Sturm. Triangulation. In *American Image Understanding Workshop*, pages 957–966, 1994.

[14] B.K.P. Horn. Relative orientation. *International Journal of Computer Vision*, 4:59–78, 1990.

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

[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] Q. T. Luong and O. D. Faugeras. Determining the fundamental matrix with planes: Instability and new algorithms. *CVPR*, 4:489–494, 1993.

[20] Q. T. Luong and O. D. Faugeras. A stability analysis for the fundamental matrix. In J. O. Eckland, editor, *Proc. 3rd European Conference on Computer Vision, LNCS 800/801, Stockholm*, pages 577–586. Springer–Verlag, 1994.

[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] 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.

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

[26] P. H. S. Torr, A. Zisserman, and S. Maybank. Robust detection of degeneracy. In E. Grimson, editor, *ICCV95*, pages 1037–1044. Springer–Verlag, 1995.

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

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

[29] Z Zhang. A review of the fundamental matrix. Technical Report to be Published, 1996.

[30] 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.