Robust Detection of Degenerate Configurations whilst Estimating the Fundamental Matrix

P H S Torr, A Zisserman
Robotics Research Group
Department of Engineering Science
Oxford University,
Parks Road, Oxford,
OX1 3PJ, UK.
phst@robots.oxford.ac.uk
Fax ++44 1865 273908
Tel ++44 1865 273127

S J Maybank
Department of Computer Science
Reading University,
Reading,
RG6 6AY, UK.

Abstract

We present a new method for the detection of multiple solutions or degeneracy when estimating the *Fundamental Matrix*, with specific emphasis on robustness to data contamination (mismatches). The Fundamental Matrix encapsulates all the information on camera motion and internal parameters available from image feature correspondences between two views. It is often used as a first step in structure from motion algorithms. If the set of correspondences is degenerate, then this structure cannot be accurately recovered and many solutions explain the data equally well. It is essential that we are alerted to such eventualities. As current feature matchers are very prone to mismatching the degeneracy detection method must also be robust to outliers.

In this paper a definition of degeneracy is given and all two view non-degenerate and degenerate cases are catalogued in a logical way by introducing the language of varieties from algebraic geometry. It is then shown how each of the cases can be robustly determined from image correspondences via a scoring function we develop. These ideas define a methodology which allows the simultaneously detection of degeneracy and outliers. The method is called, PLUNDER-DL, and is a generalisation of the robust estimator RANSAC.

The method is evaluated on many differing pairs of real images. In particular it is demonstrated that proper modelling of degeneracy in the presence of outliers enables the detection of mismatches which would otherwise be missed. All processing including point matching, degeneracy detection and outlier detection is automatic.

Keywords: fundamental matrix, degeneracy, robust methods, motion analysis, epipolar geometry, model selection, MDL.

Contents

1 Introduction

Robotic vision has its basis in geometric modelling of the world, and many vision algorithms attempt to estimate these geometric models from perceived data. A largely unaddressed problem is the detection in the data of degeneracy with respect to the model estimated. A precise definition of degeneracy will be given later. For the moment it suffices to note that degeneracy arises when the fitted model contains too many parameters. It is important to detect degeneracy and select a more appropriate model with a fewer number of parameters. This paper presents a general robust statistically based estimator that detects degeneracy. Here 'robust' means that the method gives good results even when the data contains outliers, i.e. data that does not conform to the assumed probability distribution.

The methodology developed is applied to a common and important problem in robotic vision applications, namely that of estimating the *fundamental matrix* $\mathbf{F}$ [11, 18, 45]. The fundamental matrix encapsulates the epipolar geometry between two images of a scene, and may be computed directly from corresponding points in the two images. The *ideal* fundamental matrix is determined by the internal (camera) and external (motion) parameters alone. It does not depend on the depths of the scene points (structure). However the *estimated* fundamental matrix is calculated from measured image correspondences, not ideal ones, and its accuracy and sensitivity *do* vary with structure. It is important that the estimation algorithm can detect whether there are many fundamental matrices compatible with the image correspondences, or just one. If there are many this has obvious disadvantages from the point of view of reconstruction, and motion estimation, both of which can be accomplished once the fundamental matrix is available. The resulting structure or motion varies with the choice of fundamental matrix made from the many that fit the data.

Now the key definition of degeneracy is made. It is assumed that each image correspondence falls into one of two sets: a set of true correspondences perturbed by noise (usually Gaussian), and a set of outliers (mismatches) with a different noise distribution. The data are termed *degenerate* with respect to a *model* if the underlying set of noise free or true correspondences do not admit to a unique *solution* with respect to that model. As the noise free values of the data are unknown, the problem is to deduce from the observations (a) the set of inliers and (b) whether the underlying noise free inliers are indeed degenerate, with respect to the model chosen.

The problem of detecting degeneracy is considerably complicated by the presence of outliers. A simple two dimensional example will illustrate this. Figure 1 shows three cases of line-fitting to

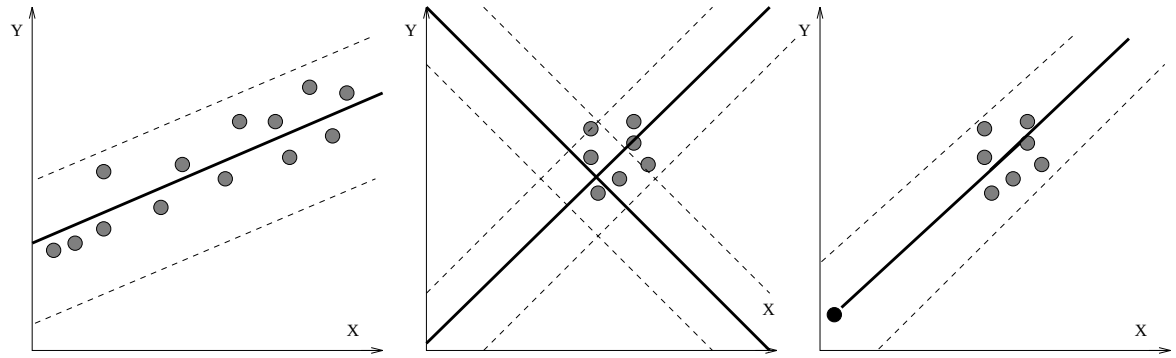


Figure 1: *Line fitting to noisy 2D data sets. In all cases the statistical thresholds for points to be consistent with the line are indicated by dashed lines. Lines that fit* **all** *the data are shown. (a) A non-degenerate data set, with no ambiguity in determining the best line fit. (b) A degenerate data set. (c) A single outlier renders a degenerate data set* apparently *non-degenerate.*

two dimensional data sets (i.e. the *model* is a line). A line is considered a good fit if the points all lie within the threshold indicated by the dotted lines. This threshold is typically related to the variance of the noise affecting the points. Figure 1 (a) shows a set which might be considered non-degenerate, and for which a line model is appropriate. Figure 1 (b) shows degenerate data, in that there are a wide range of acceptable fitted lines i.e. many *solutions*, all passing through the centroid of the data. A noise model is essential if this type of degeneracy is to be detected; in the absence of a noise model there is no criterion for deciding if a fit is good and hence no criterion for deciding if there are multiple fits. Note that if the noise in Figure 1 (a) is very high relative to the dispersion of the points then this might indeed be a degenerate set. The need for methods which can flag degeneracy in the presence of outliers is demonstrated by Figure 1 (c) where even one outlier can mask the degeneracy. It might be thought that a point rather than a line would better model the data in (b) or (c). Indeed, in Figure 1 (b) it can be seen all the points lie within a small circle leading to the inference that a point model is appropriate. Could the points in (c) be said to be consistent with a point model? An ingenuous approach to fitting is to select only the model consistent with the most points; but this approach, in the presence of noise and outliers, always leads to the selection of the non-degenerate model. Figures 1 (a), (b) and (c) show this to be case, because the line model has more consistent data than the point model. The need for a more sophisticated approach motivates the research presented in this paper.

It is clear then that the existence of degeneracy depends upon three elements: (1) the geometric model of the world, (2) the error criterion and (3) the statistical model. In the previous example,

the geometric model is a line, the error criterion is the sum of the squares of the perpendicular distances of the points to the line, and the statistical model is zero mean isotropic Gaussian noise. It is also clear that the detection of degeneracy and the detection of outliers are inextricably linked; this is now demonstrated with reference to the fundamental matrix $\mathbf{F}$.

There is no unique solution for $\mathbf{F}$ if the camera is rotating about its focal point without any translation. In this case the model that constrains the noise free correspondences is a plane-plane projectivity which is compatible with a two parameter family of fundamental matrices. If $\mathbf{F}$ is estimated from matches between such image pairs then mismatches and noise will conspire to make the data look non-degenerate. There will be too many inliers and too few outliers in the result, as the model selected is over parameterised. Conversely if one were to estimate a projectivity in the case where the camera is both translating and rotating then there would appear to be too few inliers and too many outliers. The two models for the correspondences, fundamental matrix or projectivity, are analogous to the line and point models discussed above. Should the noise free data conform to a projectivity then there will be no unique solution for the fundamental matrix (any set of point matches well fitted by a projectivity will be well fitted by a two dimensional family of fundamental matrices), just as there is no unique solution for a line if the noise free data conform to a point.

The remainder of this paper is as follows: The fundamental matrix is described in Section 2, and its degenerate cases in Section 2.1. In Section 3 a goodness of fit test is described, leading to a test for degeneracy for noisy but outlier free data. When considering multiple models for the data it is important to ensure that the error terms for different models are comparable; thus Section 3.1 examines the different error terms previously adopted in the literature [11, 18, 23, 34, 45, 52, 53], and derives the maximum likelihood error. The goodness of fit for a model is defined in Section 4.1. The problem of detecting degeneracy when estimating $\mathbf{F}$ is discussed in Section 5. In Section 6 previous work on robust estimators is reviewed. The most robust estimators to date, such as RANSAC [13], use random sampling. A deficiency of this class of estimators is their disregard for the issues relating to degeneracy. In their seminal paper Fischler and Bolles split interpretation of data into two distinct activities: "First there is the problem of finding the best match between the data and one of the available models (the classification problem); Second, there is the problem of computing the best values for the free parameters of the selected model (the parameter estimation problem). In practice, these two problems are not independent—a solution to the parameter estimation problem is often required to solve the classification problem." They provide a solution to the second but not the first of these two problems: the classification problem. Up until now little successful work

has appeared concerning this problem, which is surprising given the large effect that degeneracies may have in a working system. Section 7 sets out a random sampling estimator, with acronym PLUNDER-DL, that detects degeneracy even in the presence of outliers. PLUNDER-DL extends the RANSAC algorithm to the model selection problem using an information theoretic scoring function. It follows the PLUNDER methodology set out in [48, 49] in which a robust estimate is made of each model and a method devised to compare the models. Here in PLUNDER-DL a description length method is used to score each model. The efficacy of the PLUNDER-DL algorithm is demonstrated on two view model selection, and results are presented in Section 9.

2 Degeneracy and the Fundamental matrix

Suppose that a camera takes two noise free images of a 3D object, and undergoes a rotation and non-zero translation between the two images. The set of image points $\{\underline{\mathbf{x}}_i\}, i = 1, \ldots n$, in the first image (we adopt the convention of signifying noise free data by underlining it) is transformed to the set $\{\underline{\mathbf{x}}_i'\}$, in the second, where $\underline{\mathbf{x}}_i$ and $\underline{\mathbf{x}}_i'$ have 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 two sets of image points are related by

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

where $\mathbf{F}$ is a 3×3 matrix of rank 2. It is known as the fundamental matrix [11, 18].

Recall that data are degenerate if the true data (the noise free correspondences) admit multiple solutions for the fundamental matrix. It can be seen that an understanding is required of cases that lead to the true data being degenerate. The next subsection analyses all such cases.

2.1 Geometric Degeneracy

A perfect (i.e. noise free and no outliers) set of correspondences: $\{\underline{\mathbf{x}}_i\} \leftrightarrow \{\underline{\mathbf{x}}_i'\}, i = 1, \ldots n$, is geometrically degenerate with respect to $\mathbf{F}$ if it fails to define a unique epipolar transform, or equivalently if there exist linearly independent rank two matrices, $\mathbf{F}_j$, $j = 1, 2$, such that

$$\underline{\mathbf{x}}_i'^\top \mathbf{F}_1 \underline{\mathbf{x}}_i = \underline{\mathbf{x}}_i'^\top \mathbf{F}_2 \underline{\mathbf{x}}_i = 0 \qquad (1 \leq i \leq n) \ . \tag{2}$$

Equation (1) can be solved for $\mathbf{F}$ using the 8-point algorithm of [32], reformulated as a standard least squares problem. The equation is written in the form

$$\mathbf{Zf} = \mathbf{0} \tag{3}$$

where $\mathbf{f} = (f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8, f_9)^\top$ contains the entries of $\mathbf{F}$,

$$\mathbf{F} = \begin{bmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \\ f_7 & f_8 & f_9 \end{bmatrix}, \tag{4}$$

and $\mathbf{Z}$ is the $n \times 9$ measurement matrix

$$\mathbf{Z} = \begin{bmatrix} x'_1x_1 & x'_1y_1 & x'_1 & y'_1x_1 & y'_1y_1 & y'_1 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ x'_nx_n & x'_ny_n & x'_n & y'_nx_n & y'_ny_n & y'_n & x_n & y_n & 1 \end{bmatrix}. \tag{5}$$

Let N be the right null space of $\mathbf{Z}$ (which is also the null space of the image correspondence moment matrix $\mathbf{M} = \mathbf{Z}^\top\mathbf{Z}$). The geometric degeneracies are catalogued according to the dimension of N:

$\underline{\dim(N) = 1}$: There is no degeneracy: this is the case if the image correspondences arise from eight or more 3D points in general position.

$\underline{\dim(N) = 2}$: This is the case if only seven 3D points are observed in general position, or if both the optical centres and any number of 3D points lie on a quadric surface referred to as the *critical surface* [35]. Suppose N is spanned by the vectors $\mathbf{u}_1$ and $\mathbf{u}_2$, corresponding to the matrices $\mathbf{U}_1$ and $\mathbf{U}_2$ respectively. The matrices $\mathbf{U}_1$ and $\mathbf{U}_2$ need not have rank 2. The fundamental matrices which solve (3) lie in the one-parameter family of matrices: $\alpha\mathbf{U}_1 + (1-\alpha)\mathbf{U}_2$. Imposing the additional constraint that the fundamental matrix is rank two,

$$\det[\alpha\mathbf{U}_1 + (1-\alpha)\mathbf{U}_2] = 0 \tag{6}$$

yields a cubic equation in α, which has either one or three real solutions (the complex solutions are discarded) [20]. If there is only one real solution then the data are non-degenerate.

This method of computing one or three fundamental matrices for the minimum number of points (seven) is used in the robust algorithm of Section 6.

The critical surface is in general a quadric surface. It contains at least one real line, namely the line through the first optical centre in the direction of a (spurious) translation. It follows that the critical surface is a hyperboloid of one sheet or a degenerate form of such a surface. Degenerate hyperboloids include plane pairs, cones, cylinders and hyperbolic paraboloids. If the object points are on a critical surface then the image correspondences are given by a quadratic transformation [35]

$$\mathbf{x}' = (\mathbf{F}_1\mathbf{x}) \times (\mathbf{F}_2\mathbf{x}) \quad , \tag{7}$$

where $\mathbf{F}_1$ and $\mathbf{F}_2$ are two of the fundamental matrices compatible with the data.

$\underline{\dim(N) = 3}$: In this case corresponding points are related by a projectivity (which may also be an affinity) between the image planes,

$$\mathbf{x}' = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \mathbf{x} \ , \tag{8}$$

A proof is provided in Appendix A which corrects the erroneous proof given in [12]. This case can arise from an observation of a special type of structure (a plane) or from the camera undergoing a special type of motion (a rotation about its optic centre).

$\underline{\dim(N) = 4}$: In this case the 3D points are not in general position, for example they may lie on a line.

2.2 Varieties and Null Spaces for F

Now a different way of describing the degeneracies will be examined. It is based on the joint image or measurement space [51]. This will lead to a clear geometrical criteria for deciding whether a set of point matches are degenerate. Each pair of corresponding points $\underline{\mathbf{x}}$, $\underline{\mathbf{x}}'$ defines a single point in a measurement space $\mathcal{R}^4$, formed by concatenating the inhomogeneous coordinates of $\mathbf{x}$, $\mathbf{x}'$. Typically the set of all physically realisable measurements is compact. In this work most of the images are 512×512 pixels, thus a point (x, y, x', y') in $\mathcal{R}^4$ arising from a pair of matching image points has

$0 \leq x, y, x', y' \leq 512$. It might be considered somewhat eldritch to join the coordinates of points from the two images, but this makes sense because degeneracy is a property of pairs of matching points not of the points in a single image. The image correspondences $\{\underline{\mathbf{x}}_i\} \leftrightarrow \{\underline{\mathbf{x}}_i'\}, i = 1, \ldots n$, induced by a rigid motion define an algebraic variety V in $\mathcal{R}^4$. The properties of V depend on the dimension of the null space N defined in Section2.1. For example if $\dim(N) = 1$ then V is a quadric hypersurface defined by

$$\begin{pmatrix} x & y & x' & y' & 1 \end{pmatrix} \begin{bmatrix} 0 & 0 & f_1 & f_4 & f_7 \\ 0 & 0 & f_2 & f_5 & f_8 \\ f_1 & f_2 & 0 & 0 & f_3 \\ f_4 & f_5 & 0 & 0 & f_6 \\ f_7 & f_8 & f_3 & f_6 & 2f_9 \end{bmatrix} \begin{pmatrix} x \\ y \\ x' \\ y' \\ 1 \end{pmatrix} = 0 \quad . \tag{9}$$

which is obtained by rewriting (1). The variety V is quadric as (9) is degree two in x, y, x', y', and is dimension three. Each point of V yields a unique point in the world that projects down to the associated image points. The (Euclidean) coordinates (X, Y, Z) of world points provide a three dimensional coordinate system on V.

If $\dim(N) = 2$ and the 3D points are on a critical surface, then as noted in Section 2.1, there are three solutions $\mathbf{F}_1$, $\mathbf{F}_2$ and $\mathbf{F}_3$ for $\mathbf{F}$ although two may be complex conjugates. Then V is the intersection of the three quadric hypersurfaces for $\mathbf{F}_1$, $\mathbf{F}_2$ and $\mathbf{F}_3$. The variety V is two dimensional due to the fact that all the 3D points (X, Y, Z) lie on a 2D surface. The coordinates of world points on the critical surface provide a two dimensional coordinate system for V.

If $\dim(N) = 3$ and the correspondences are governed by a projective transformation as in (8) then V is the intersection of two quadric hypersurfaces: A quadratic polynomial is obtained when the scale factor is eliminated from the homogeneous equations (8) defining a projectivity. Two such quadratic surfaces are obtained in Appendix C one for x' (41) and one for y' (42). The resulting intersection is a two dimensional variety. The coordinates of the object plane provide a two dimensional coordinate system for V. The case of a purely rotating camera is included by making the plane at infinity the object plane.

If $\dim(N) = 3$ and the correspondences are governed by an affine transformation then V is the intersection of two hyperplanes i.e. it is a linear subspace of dimension two. This follows from (14) which yields one linear constraint between x', x, y and another between y', x, y, each of which defines a hyperplane in $\mathcal{R}^4$.

The variety V has dimension one only if the object points lie on a curve.

The models are summarised in Table 1. The table gives for each model: the degree of V, the number of degrees of freedom in the model, and the minimal set of correspondences needed in order to establish V. Then the equation for V is given. A natural hierarchy of models is defined by considering the dimension of V. It can be seen that the fundamental matrix, translation fundamental matrix and affine fundamental matrix (defined in the next section) all yield varieties with dimension three. The quadratic, projective and affine image transformations yield varieties with dimension two, the first and second of degree four and the third linear. For perfect data defining a variety of dimension three there is a unique solution for the fundamental matrix. If the data only define a variety of dimension two then there are multiple solutions. In other words, when estimating $\mathbf{F}$, non-degenerate models are associated with varieties of dimension three, and degenerate models are associated with varieties of dimension two.

The definition of the affine camera model, and its degeneracy the image-image affinity are made in the next subsection. The affine camera is useful when perspective effects are swamped by noise, hence before it is classified as non-degenerate.

2.3 Affine Camera

The affine camera arises when the field of view is small and all the distances to the viewed objects are large relative to their relief. This situation has long been modelled by parallel projection (for a detailed account of camera models see [37]), which sets the optic centre at infinity so that all the rays to the camera are parallel. From two images the data generally admits a one parameter family of scene reconstructions - the bas-relief ambiguity. The ambiguity is a characteristic of parallel projection and not simply a result of having too few image correspondences. In the estimation process this corresponds to all the points in $\mathcal{R}^4$ lying near to a dimension 3 hyperplane.

The affine camera [37, 41] generalises the orthographic and scaled orthographic models. It is a composition of the following: a 3D affine transformation of the world coordinate system, parallel projection onto the image plane, and a 2D affine transformation of the image. It has a fundamental matrix $\mathbf{F}_A$ of the form

$$\underline{\mathbf{x}}_i'^{\top}\mathbf{F}_A\underline{\mathbf{x}}_i = \underline{\mathbf{x}}_i'^{\top}\begin{bmatrix} 0 & 0 & a_1 \\ 0 & 0 & a_2 \\ a_3 & a_4 & a_5 \end{bmatrix}\underline{\mathbf{x}}_i = a_1\underline{x}' + a_2\underline{y}' + a_3\underline{x} + a_4\underline{y} + a_5 = 0 \quad . \tag{16}$$

1. **Varieties V of dimension 3.**

   (a) *Fundamental matrix*, degree 2, DOF 7, correspondences 7, for 1 or 3 solutions, 8 for unique solution.

$$\mathbf{x}'^{\top}\mathbf{F}\mathbf{x} = 0 \quad \text{where} \quad \mathbf{F} = \begin{bmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \\ f_7 & f_8 & f_9 \end{bmatrix} \tag{10}$$

   (b) *Affine fundamental matrix*, degree 1, DOF 4, correspondences 4.

$$\mathbf{x}'^{\top}\mathbf{F}_A\mathbf{x} = 0 \quad \text{where} \quad \mathbf{F}_A = \begin{bmatrix} 0 & 0 & a_1 \\ 0 & 0 & a_2 \\ a_3 & a_4 & a_5 \end{bmatrix}. \tag{11}$$

   (c) *Translation fundamental matrix*, degree 2, DOF 2, correspondences 2.

$$\mathbf{x}'^{\top}\mathbf{F}_T\mathbf{x} = 0 \quad \text{where} \quad \mathbf{F}_T = \begin{bmatrix} 0 & g_3 & -g_2 \\ -g_3 & 0 & g_1 \\ g_2 & -g_1 & 0 \end{bmatrix}. \tag{12}$$

2. **Varieties V of dimension 2.**

   (a) *Quadratic image transformation*, $\mathbf{Q}$, degree 4, DOF 14, correspondences 7;
   $\mathbf{x}' = \mathbf{F}_1\mathbf{x} \times \mathbf{F}_2\mathbf{x}$, where $\mathbf{F}_1$, $\mathbf{F}_2$ are 3×3 fundamental matrices.

   (b) *Projective image transformation*, degree 4, DOF 8, correspondences 4.

$$\mathbf{x}' = \mathbf{H}\mathbf{x} \quad \text{where} \quad \mathbf{H} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix}. \tag{13}$$

   (c) *Affine image transformation*, degree 1, DOF 6, correspondences 3.

$$\mathbf{x}' = \mathbf{K}\mathbf{x} \quad \text{where} \quad \mathbf{K} = \begin{bmatrix} k_1 & k_2 & k_3 \\ k_4 & k_5 & k_6 \\ 0 & 0 & 1 \end{bmatrix}. \tag{14}$$

   (d) *Image translation*, degree 1, DOF 1, correspondence 1.

$$\mathbf{x}' = \mathbf{L}\mathbf{x} \quad \text{where} \quad \mathbf{L} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix}. \tag{15}$$

   (e) *No Motion*, degree 1, DOF 0, correspondences 0.

Table 1: *Models that are fitted to sets of correspondences. The models (top to bottom) are given in order of decreasing generality. For example the affine fundamental matrix and the translational fundamental matrix are both special cases of the fundamental matrix.*

This defines a hyperplane in the measurement space $\mathcal{R}^4$. The problem of estimating $\mathbf{F}_A$ is linear, and requires $n \geq 4$ point correspondences. When appropriate, use of $\mathbf{F}_A$ circumvents the computation of the quadratic (in image coordinates) coefficients of $\mathbf{F}$ which are inherently ill-conditioned.

The image-image affinity is a degenerate case of the affine camera. Recall that the true data are degenerate if there are multiple solutions for $\mathbf{F}$. If the observations conform to an affine camera then degeneracy occurs if the point correspondences are given by an image-image affinity (17)

$$\mathbf{x}' = \begin{bmatrix} k_1 & k_2 & k_3 \\ k_4 & k_5 & k_6 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x}. \tag{17}$$

Conversely if the data conform to an image-image affinity then the data are degenerate, either the camera is rotating or the $3D$ points lie on a distant plane, under orthography. If the data are consistent with (16) but not with (17) then the data are non-degenerate but there is a large uncertainty in determining the coefficients of the quadratic terms in $\mathbf{F}$. This means that although the true $\mathbf{F}$ might be arbitrary (i.e. the camera can undergo any motion) a linear solution for $\mathbf{F}$ (with $f_1 = 0$, $f_2 = 0$, $f_4 = 0$, $f_5 = 0$) is acceptable in that it fits the data. Thus the data need not be degenerate by our definition for it to satisfy the affine camera. In fact, in this case it is only classified as degenerate if it satisfies the image-image affinity model.

Does the fact that a matrix $\mathbf{F}_A$ fits the data necessarily mean that the data arise from an affine camera? The answer is no. An $\mathbf{F}$ matrix has affine form (16) if and only if the translation is parallel to the image plane and there is no rotation out of the image plane. This means that although the algorithm can indicate that $\mathbf{F}_A$ fits the data, it cannot be said without *a priori* knowledge whether the camera has this special form and is observing a scene with large perspective effects, or whether it is observing a distant object with small perspective effects.

3 Fitting to Noisy Data

In the last section the causes of degeneracy given perfect data were categorised, but in any application the models must be estimated from data contaminated by noise. Within this section various error measures are discussed so that in the next section a test for goodness of fit and hence degeneracy can be developed. Estimation of the fundamental matrix necessarily involves the minimization of some error criterion. In Sections 3.1 and 3.2 several criteria are discussed and their relative

merits assessed. In Section 4.1 it is shown how inferences can be made about whether noisy, but outlier free, data are degenerate.

3.1 Error Term for F

Four commonly used error terms (e.g. see [34]) for evaluating an estimate of $\mathbf{F}$ given noisy data, are now examined. Given n noisy image point correspondences between the two images $\{\mathbf{x}_i\} \leftrightarrow \{\mathbf{x}_i'\}$ these error terms for a given $\mathbf{F}$ are:

1. The sum of squares of *algebraic distances*: $R = \sum_i r_i^2$, where $r_i = \mathbf{x}_i'^\top \mathbf{F} \mathbf{x}_i$, and the coefficients of $\mathbf{F}$ are suitably normalized.

2. The sum of squares of orthogonal distances of correspondences (x, y, x', y') to the variety V defined by $\mathbf{F}$ in $\mathcal{R}^4$, $D_V = \sum_i d_V(\mathbf{x}_i, \mathbf{x}_i', V)^2 \equiv \sum_i d_{V_i}^2$.

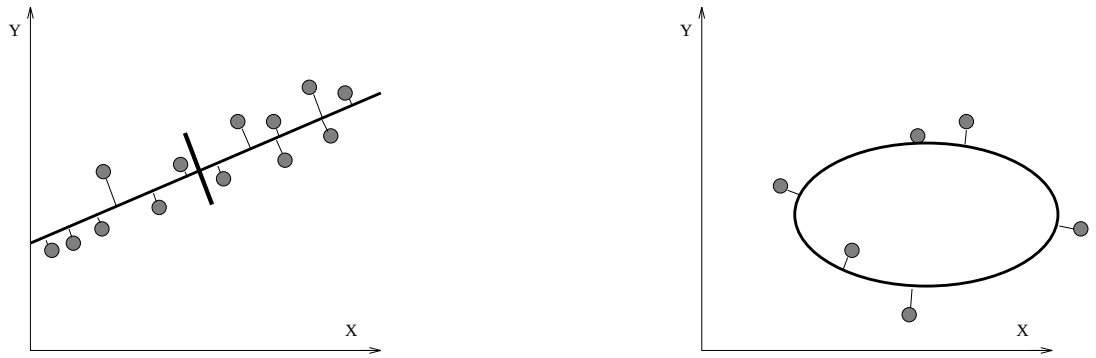
3. The sum of squares of *Sampson's distances* [40]: $D_S = \sum_i (r_i/|\nabla r_i|)^2 = \sum_i d_{S_i}^2$.

4. The sum of squares of *Luong's distances* [34]: $D_L = \sum_i (d_{L_i}^2 + d_{L_i}'^2)$, where d_{L_i} d_{L_i}' are the perpendicular distances of the ith point to its estimated epipolar line in each image.

Each error term is now explained in more detail.

3.1.1 Algebraic Distance and Normalization

In this subsection minimization of the sum of squares of the algebraic distance is discussed. The minimization cannot be done without some normalization of the coefficients, in order to avoid the trivial solution $\mathbf{f} = \mathbf{0}$ (where $\mathbf{0}$ is a vector with all elements zero). With $n \geq 8$ image correspondences, the value of $\mathbf{F}$ which minimizes R can be found by linear (in the parameters—not the data) least squares [32]. Given the measurement matrix (5) the solution for the fundamental matrix is obtained from the eigenvector corresponding to the least eigenvalue of the moment matrix $\mathbf{M} = \mathbf{Z}^\top \mathbf{Z}$. This minimizes $R = \sum r_i^2$ with normalization $\mathbf{f}^\top \mathbf{f} = \sum_i f_i^2 = 1$. Ideally the normalization of the coefficients should be invariant to rigid transformations of the coordinate system (x, y, x', y'). This is because the choice of centre and orientation of axes in each image is arbitrary. Unfortunately there is no normalisation which renders the algebraic distance invariant to the choice of coordinate system. In the past the linear solution for the epipolar geometry minimizing the algebraic distance



(a) (b)

Figure 2: *Both (a) (b) show perpendicular distances from data points to the fit. Assuming Gaussian noise, minimizing D_V the sum of squares of these distances gives the maximum likelihood estimate. In (a) the eigenvector of the moment matrix,* $\mathbf{M}$*, corresponding to the minimum eigenvalue minimizes D_V. Here the normalized algebraic distance r_i equals the perpendicular distance of the point to the line. In (b) the algebraic distance does not equal the perpendicular distance of a point to the conic, and consequently the eigenvector of the moment matrix does not minimize D_V.*

has been considered too ill conditioned to provide useful results. Furthermore Luong [34] claims that such linear methods produce a biased solution, and that the error function D_L is preferable.

This received wisdom has been challenged recently in [19] and by our own results too, in that suitable preconditioning of the data radically improves the results, obtained from the algebraic distance. The preconditioning involves transforming and scaling the image points such that the coordinates are of a similar magnitude. Details may be found in [5].

3.1.2 Geometric Distances

The measure R does not possess any physical significance (because the r_i do not represent measurable geometric distances). In order to make this clear, consider the two dimensional example shown in Figure 2. In the case of line fitting, the minimization of the algebraic distance is equivalent to minimizing D_V the sum of squares of perpendicular distances to the line [26], providing that the data are measured from a coordinate system with origin at the datas' centroid. For conic and quadric fitting this equivalence no longer holds. It can be shown [26] that given Gaussian noise the best fitting (maximum likelihood) curve or surface is such that the sum of squares of perpendicular distances of points to the is a minimum. This corresponds to minimizing the reprojection error: if

$(\hat{x}, \hat{y}, \hat{x}', \hat{y}')$ is a point on the variety V (i.e. it exactly satisfies the epipolar constraint), then

$$d_V = \min((x - \hat{x})^2 + (x' - \hat{x}')^2 + (y - \hat{y})^2 + (y' - \hat{y}')^2) \tag{18}$$

This error term is invariant to rigid transformations of the coordinate system, in $\mathcal{R}^4$, which is a desirable property, as described in the previous section. Unfortunately, the joins of the points to the variety in $\mathcal{R}^4$ are not parallel (unless the variety is a hyperplane), they may not be unique (e.g. for a point lying at the centre of a hypersphere), and a closed form solution for the distance of each point to the variety is generally unobtainable. In fact Hartley and Sturm [21] prove that the distance is given by the root of a degree six polynomial, for $\mathbf{F}$. The iterative algorithm for solving the polynomial is accurate, but relatively costly in time.

Next two less costly approximations to this distance D_v are examined. The first was proposed by Sampson: A first order approximation to the orthogonal distance leads to the error term $d_s = (r/|\nabla r|)^2$. Weng *et al* [53] extended this to the fundamental matrix as follows, the modulus of the gradient, $|\nabla r|$, is easily computed:

$$\begin{aligned} |\nabla r| &= (r_x^2 + r_y^2 + r_{x'}^2 + r_{y'}^2)^{\frac{1}{2}} \\ r_x = f_1 x' + f_4 y' + f_7 \qquad & r_y = f_2 x' + f_5 y' + f_8 \\ r_{x'} = f_1 x + f_2 y + f_3 \qquad & r_{y'} = f_4 x + f_5 y + f_6. \end{aligned}$$

The terms r_x, r_y, $r_{x'}$, $r_{y'}$ are the partial derivatives of r. Note that d_s is undefined at the point in $\mathcal{R}^4$ given by the two epipoles, as here $|\nabla r|$ is zero. This can lead to instabilities in practice and is corrected by excluding any match within one pixel of the epipole.

The second approximation is the error measure D_L defined in Luong [34]. With perfect data $\{\underline{\mathbf{x}}_i\} \leftrightarrow \{\underline{\mathbf{x}}_i'\}, i = 1, \ldots n$, each point $\underline{\mathbf{x}}$ defines an epipolar line $\mathbf{F}\underline{\mathbf{x}}$ upon which its corresponding point $\underline{\mathbf{x}}'$ should lie. Given noisy data, $\mathbf{x}'$ in general will not lie on $\mathbf{F}\mathbf{x}$. Luong proposes minimizing the sum of squares of the distances of the line $\mathbf{F}\mathbf{x}$ to $\mathbf{x}'$, and the line $\mathbf{F}^\top\mathbf{x}'$ to $\mathbf{x}$. We call these quantities the epipolar distances, d_{L_i}, d'_{L_i}. These distances are invariant to rigid transformations of the image coordinates.

For a particular datum note the similarity of D_L

$$D_L = \sum r^2 \left(\frac{1}{r_x^2 + r_y^2} + \frac{1}{r_{x'}^2 + r_{y'}^2} \right) . \tag{19}$$

to D_S

$$D_S = \sum r^2 \left(\frac{1}{r_x^2 + r_y^2 + r_{x'}^2 + r_{y'}^2} \right) \quad . \tag{20}$$

The distance D_V will be adopted from hereon as the error criterion, as it has a geometric significance, and leads to the maximum likelihood solution. However, in practise it is often computationally convenient to use D_S to approximate D_V. After an initial solution is obtained by RANSAC (described later), the solution is improved using an iterative search on the fundamental matrix, parameterized with seven parameters (4 for the epipoles and 3 for the epipolar homography) as described in [34].

3.2 Error Term for Dimension Two Varieties

Similar arguments, as given in the last section for $\mathbf{F}$, follow for the dimension two varieties. The optimal error term, namely the sum of squares of distances of data points in $\mathcal{R}^4$ to the variety, could be computed exactly by solving a polynomial, but, as for the fundamental matrix, a first order approximation gives an acceptable result with less computation. The derivation of the first order approximation is given in Appendix C. Having described the maximum likelihood error, a test for the goodness of fit of the parameters is described in the next section.

4 Detecting Degeneracy in Outlier Free Data

The cases of exact degeneracy have been enumerated in Section 2.1. In this section the effects of noise on the estimation of degeneracy are examined. Degeneracy cannot now be exactly determined—i.e. with the added noise it becomes impossible to decide with 100% certainty whether the 3D points approximate one of the configurations outlined in Section 2.1. Within this section traditional χ^2 thresholds for the acceptability of each model are given and their failing revealed.

In the absence of outliers, degeneracy could also be detected by examining the covariance of the estimated fundamental matrix. (The covariance approach is described in [9]). However this approach is only based on a first order approximations, whereas the χ^2 test is exact. Another approach using AIC was suggested by Kanatani [1, 24, 25] but this does not deal with outliers, whereas the approach we outline will. Extension to the AIC to cope with outliers is discussed in Torr [50].

4.1 Determining Goodness of Fit

Within this section the traditional χ^2 method of determining the goodness of fit is outlined, and its deficiencies revealed. Assuming that the image coordinates are perturbed by Gaussian noise with mean zero and known variance σ^2 (thus there are no outliers), then the distance of a perturbed point correspondence from the algebraic variety defined by the fundamental matrix also follows a Gaussian distribution. Hence the sum of squares of these distances follows a χ^2 distribution. The acceptability of a given $\mathbf{F}$ may be judged in terms of its errors. The true fundamental matrix is unknown and an estimate must perforce be used in its place. If the estimate is sufficiently close to the true fundamental matrix then the sum of squares of distances to the estimate d_V, divided by their standard deviation, has a χ^2 distribution with $d_f = n - 7$ degrees of freedom, where n is the number of image correspondences, and seven is the number of degrees of freedom of the fundamental matrix. The suitability of a given $\mathbf{F}$ may be tested under the assumption that a fit is good if it lies within the *critical region* of all solutions, defined as those with sum of squares of distances below $T_F = \chi^2_{d_f}(\alpha)$, where α is the required degree of confidence, i.e. the critical region in the space of $\mathbf{F}$ is given by

$$\{\mathbf{F} | D_V(\mathbf{F}) \leq T_F\} \tag{21}$$

typically $\alpha = 95\%$. We have given two other types of fundamental matrix that might occur in practice, the error $D_V(\mathbf{F}_T)$ has $d_f = n - 2$ degrees of freedom, $D_V(\mathbf{F}_A)$ has $d_f = n - 4$ degrees of freedom.

As an example Figure 3 (a) (b) shows two images from a sequence generated by Harris [16] of a toy truck on a turntable. Figure 3 (c) (d) shows two $\mathbf{F}$'s that fit the data equally well, in that the sums of squares of distances D_V to the appropriate variety satisfy a χ^2 test in each image. The sum of squares of distances being 3.5 for (c) and 4.1 for (d). The threshold for the χ^2 test is 5.2, so both are acceptable solutions (at a 95% level). The epipolar lines in (d) (but not (c)) are consistent with an affine camera (in that they are parallel).

From this example it can be seen that the affine camera, whilst not being a configuration that would arise for noise free data, is a highly useful model for data in which the perspective effects have been swamped by noise. In this case although the noise free matches are non-degenerate, the noisy matches would lead to highly unstable estimates of $\mathbf{F}$. By adopting $\mathbf{F}_A$ only the parts of $\mathbf{F}$ that can be estimated reliably are used to describe the data.

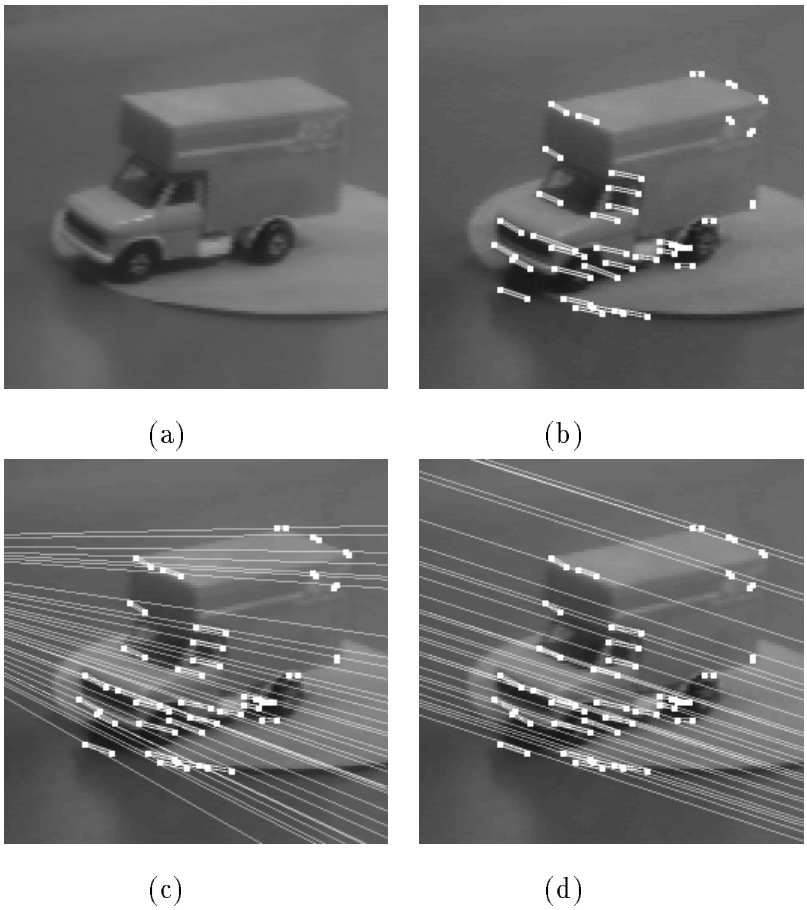


Figure 3: *Rotating toy truck sequence (a) first image (b) second image with correspondences from (a) to (b) superimposed, (c) (d) Two epipolar geometries that fit the correspondences shown in Figure 3 (b) equally well. It can be seen that the data are degenerate. (d) is an affine camera epipolar geometry with parallel epipolar lines.*

It might be that the data set are within the critical region of another model i.e. a plane. To determine this a projectivity is fitted to the data, and the resulting sum of squared error is compared to a threshold: T_H. If the error lies below the threshold then the data are deemed degenerate. Sometimes it might happen that a configuration is actually non-degenerate but lies so close to a plane (for example) that with the added noise it gives a sum of squared error lying below T_H. Here also it would be inferred that the data was degenerate because there is insufficient information to infer non-degeneracy.

For a given projectivity $\mathbf{H}$ the error measure has $d_f = 2n - 8$ degrees of freedom, $2n$ because the codimension of V is two [1], and reduced by 8 because a projectivity has 8 degrees of freedom. The critical region of solution for quadratic transformations is

$$\{\mathbf{H} | D_V(\mathbf{H}) \leq \chi^2_{(2(n-4),\alpha)} = T_Q\} \tag{22}$$

4.2 Model Selection Using χ^2

Given that the sum of squares of error is a χ^2 variable with a known number of degrees of freedom, the probability of each χ^2 and hence each model can be found. The model corresponding to the most likely χ^2 can be accepted. For example if a set of 100 correspondences has $D_V(\mathbf{F}) = 106.0$ and $D_V(\mathbf{H}) = 207.6$ then as $D_V(\mathbf{F})$ is a χ^2 with 100-7 = 93 degrees of freedom and $D_V(\mathbf{H})$ is a χ^2 with $2 \times 100 - 8 = 192$ degrees of freedom, if $\sigma = 1.0$

$$\Pr(D_V(\mathbf{F}) \geq 106.0) = \Pr(\chi^2_{93} \geq 106.0) = 0.165 \tag{23}$$

where $\Pr(D_V(\mathbf{F}) \geq 106.0)$ is the probability of the sum of squares of error being greater or equal to 106.0 given that the correct model is a fundamental matrix; which is evaluated by integrating the area to the right of 106.0 under the χ^2 curve with 93 degrees of freedom. Similarly

$$\Pr(D_V(\mathbf{H}) \geq 207.6) = \Pr(\chi^2_{192} \geq 207.6) = 0.215 \tag{24}$$

indicating that $\mathbf{H}$ is a more likely fit.

However it has been suggested [2, 8] that the χ^2 statistic is biased for model selection, generally over fitting the model. Furthermore this algorithm will fail if there are outliers as the sum of squares

[1] The codimension multiplies the number of degrees of freedom for the χ^2 test. This is because only the perturbation normal to V can be detected. The perturbation tangential to V is 'invisible' to the χ^2 test.

of errors is no longer a χ^2 distribution. Any robust method to try and partition the data set into inliers and outliers based on a threshold will misclassify some outliers as inliers and *vice versa*. As an example consider the situation where the fundamental matrix has 100 inliers with $D_V(\mathbf{F}) = 106.0$ as above, but the homography only has 20 inliers with $D_V(\mathbf{H}) = 5.5$, with $2 \times 20 - 8 = 32$ DOF; then

$$\Pr(D_V(\mathbf{H}) \geq 5.5) = \Pr(\chi^2_{32} \geq 5.5) = 0.66 \tag{25}$$

which indicates that the homography would be more likely, this is obviously not the case given that the homography has far fewer inliers, hence it is evident that a method that takes into account the inlier and outlier distribution must be formulated.

5 Detecting Degeneracy in the Presence of Outliers

Many previous approaches to the estimation of structure and motion rely on accurate or hand picked matches as input. In a fully automated system it is inevitable that point (corner) matchers produce some mismatches. This is often because the initial corner matching is done by cross correlation of the image intensities between the two images.

5.1 Initial Matching

Corner points are found using the Harris [17] corner detector. For each point at (x, y) in image 1 the match with highest cross correlation in image 2 is selected within a square search region centred on (x, y) of side $2b$ (in our case $b = 40$). This means that the set of all correspondences as input to the algorithm must have the form: $(x, y, x - b \leq x^{'} \leq x + b, y - b \leq y^{'} \leq y + b)$. For these matches we have found that Gaussian noise on each coordinate accurately models the noise process i.e. a true coordinate $\underline{x}$ is perturbed into a noisy coordinate x with isotropic Gaussian noise with the same standard deviation on each coordinate.

5.2 Outlier Model

In Section 2 degeneracy was defined for perfect data. Now the definition is extended to noisy data, that may include outliers. It is assumed that the data arise from two noise free sets, $\{\underline{I}\}$ a set of inliers, and $\{\underline{O}\}$ a set of outliers, plus a noise process. In this paper $\{\underline{I}\}$ and $\{\underline{O}\}$ are

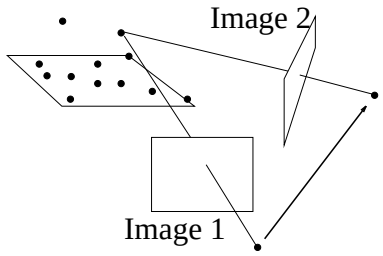


Figure 4: *A configuration consisting of many coplanar points, and two points off the plane. There are two reasonable interpretations: i) the configuration is 3D; ii) the configuration is planar, and two points have been mis-matched (outliers) so that they lie off the plane.*

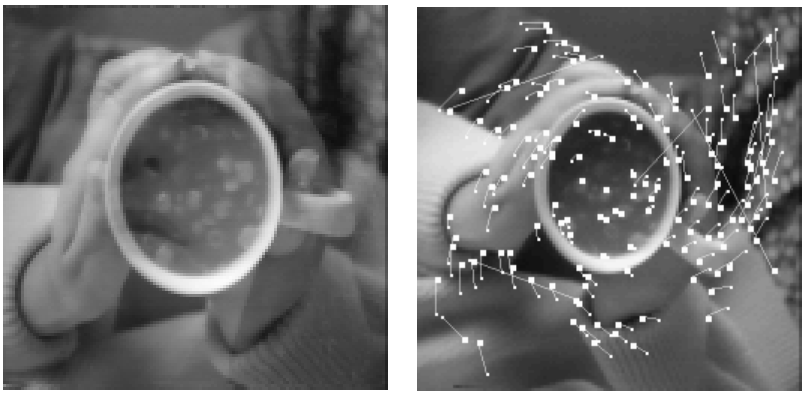
point correspondences between two images. The set $\{\underline{I}\}$ conforms to some geometric constraint, in this case the constraint arising from the assumption that the world in view is rigid. The set $\{\underline{O}\}$ contains outliers caused by the deficiencies in the corner matcher.

Let the observed values of $\{\underline{I}\}$ be $\{I\}$. It is assumed that each measured point is subject to isotropic Gaussian noise. The task of detecting degeneracy is equivalent to that of inferring the model underlying $\{I\}$. Naïve tests to detect degeneracy (for instance those based on non-robust covariance matrices) fail when there are outliers.

An example of a difficult configuration is shown in Figure 4. The majority of the 3D points lie on a plane and two lie off it. Without extra information an observer cannot deduce whether this is a non-degenerate configuration or a degenerate configuration (most of the correspondences appear to follow a projectivity) in the presence of outliers. Figure 5 shows a second example, of a cup viewed from a camera undergoing cyclotorsion about its optic axis, together with a translation along the optic axis. The predominant motion is degenerate, up to noise, however this degeneracy cannot be easily detected because of the presence of outliers. The best fitting projectivity has a sum of squares of errors outside of the critical region defined in (22). The best fitting fundamental matrix has a sum of squares of errors within the critical region (21). However with the outliers removed the sum of squares lies within the critical region (22), indicating a projectivity.

These examples show that the problem of detecting degeneracy has two intertwined elements which must be simultaneously resolved namely a) discovering the outliers; and b) selecting the correct model (i.e. degeneracy).

The next section describes some effective robust methods to estimate $\mathbf{F}$, the various other models can be estimated using similar methods. It is then shown how to use these estimates to choose the correct model.



(a) (b)

Figure 5: *(a) first image and (b) second image with matches. The camera motion is composed of a large rotation of the camera about the optic axis and translation along the line of sight.*

6 Robust Estimation of the Fundamental Matrix

In [45] a large review of robust estimators was made. It was found that random sampling methods give the best estimator in the face of outliers. In this section three types of random sampling estimator: MINPRAN, LMS and RANSAC are described and compared. In each case it is considered how each estimator might be used to detect degeneracy and the shortcomings of each are described.

6.1 Using RANSAC to Estimate F

A highly robust first example of a fitting algorithm is the random sample consensus paradigm (RANSAC) [13] for parameter estimation. Given that a large proportion of the data may consist of outliers, the approach is the opposite to conventional smoothing techniques. Rather than using as much data as possible to obtain an initial solution and then attempting to identify outliers, a minimal subset of the data is used to estimate the parameters (e.g. two points for a line, seven correspondences for a fundamental matrix). Sufficient of the minimal subsets are chosen to ensure that there is a 95% chance that at least one of the subsets contains no outliers. The best set of parameters is the one that maximizes the number of points whose error is below a threshold. In order to determine whether or not a data point is consistent with a given set of parameters, the error of the point is compared to a user set threshold, typically 1.96σ, where σ is the estimated standard deviation of the error in locating the point. If the error is Gaussian, then this is a 95% critical region. The error estimate used is Sampson's approximation described in Section 3.1. Once the outliers are removed, the remaining points are combined and a non-linear estimator is used to

1. Repeat for m samples, where m is determined by Table 3.
   (a) Select a random sample of 7 correspondences and compute an estimate $\mathbf{F}$ of the fundamental matrix.
   (b) Calculate the distance d in $\mathcal{R}^4$ from each data point to the variety described by $\mathbf{F}$.
   (c) In the case of the RANSAC estimator calculate the number of inliers consistent with $\mathbf{F}$, determined by Equation (26).
2. Select the estimate of $\mathbf{F}$ with the biggest consistent data set. In the case of ties select the solution which has the lowest standard deviation of inlying errors.
3. Re-estimate $\mathbf{F}$ using all the data identified as consistent. (A different more computationally expensive estimator may be used at this point e.g. iterative minimization.)

Table 2: *A brief summary of RANSAC. Other random sampling algorithms are similar.*

find a solution.

The use of RANSAC to estimate the fundamental matrix was first reported in Torr [44], and later [10] and is summarised in Table 2. To estimate the fundamental matrix a set of corresponding points (corners) are obtained as described in Section 5.1. Samples of seven point correspondences are drawn repeatedly at random and used to form the data matrix $\mathbf{Z}$ as in (5). An examination of the null space of $\mathbf{M} = \mathbf{Z}^\top \mathbf{Z}$ yields one or three solutions as shown in Section 2.1. The number of consistent correspondences is found for each solution; consistent points are deemed inliers, and the remaining points are deemed outliers. Techniques to find outliers are all founded upon some knowledge of the standard deviation σ of the image noise. Generally, given σ, outliers are found using the test:

$$\text{correspondence} = \begin{cases} \text{non outlier} & d_i^2 \leq t^2 \\ \text{outlier} & \text{otherwise,} \end{cases} \tag{26}$$

where $t^2 = 3.84\sigma^2$ is a user defined threshold. The value 3.84 corresponds to the 95% confidence level for a χ^2 with one degree of freedom (as the codimension is 1). This means that an inlier will only be incorrectly rejected (a Type II error) 5% of the time. The solution that gives the most

inliers is then used as the starting point for a non-linear minimization to improve the fit of $\mathbf{F}$ on the set of inliers.

Intuitively, it might be thought that the selection of 7 points and the solution of a cubic leads to a very ill-conditioned estimate of $\mathbf{F}$, and that the utilization of more than 7 points might improve matters. This approach is followed in [10], where 8 points are used in each sample and $\mathbf{F}$ estimated by linear methods. We have compared 7 point and 8 point sampling. The results are evaluated by calculating the error of the *estimated* $\mathbf{F}$ with respect to the *noise free* points (the ground truth is known). The variance of the d_V of a *perfect* point, over these 100 trials, for 7 point sampling was one quarter of that for 8 point sampling. This indicates the 7 point method is more accurate. Another advantage of the 7 point sampling is that the constraint $\det(\mathbf{F}) = 0$ holds. When outliers are introduced the benefits of the 7 point method increase because a significantly greater number of trials are needed for 8 point samples in order to obtain at least one good sample, with a high probability. In the light of these results the minimum data set of 7 is chosen for each sampling.

We now calculate how many samples of 7 points we require. Fischler and Bolles [13] and Rousseeuw [39] propose slightly different means of calculation, but both give broadly similar numbers. In our work we use the method of calculation given in [39]. Ideally we should consider every possible sample, but this is usually computationally infeasible, and so we choose m, the number of samples, sufficiently high to give a probability, Υ, in excess of 95% that at least one sample is free from outliers. The expression for Υ is

$$\Upsilon = 1 - (1 - (1 - \epsilon)^{p_f})^m, \tag{27}$$

where ϵ is the fraction of data consisting of outliers, and p_f is the number of points in each sample. Table 3 gives examples of the number m of samples required to ensure $\Upsilon \geq 0.95$ for given p_f and ϵ. It can be seen that the robust algorithm may require less samples than there are outliers, as the number of samples is linked to the proportion rather than number of outliers. Consequently the computational cost of the sampling can be acceptable even when the number of outliers is large. It can also be seen that the smaller the data set needed to instantiate a model, the fewer samples are required for a given Υ. If the fraction of data consisting of outliers is unknown, as is usual in point matching, an educated worst case estimate of the fraction must be made in order to determine the number of samples to be taken. This estimate can be updated as larger consistent sets are found, e.g. if our worst guess is 50%, $\epsilon = 0.5$, and we discover a set with 80% inliers then the updated estimate is $\epsilon = 0.2$. The beauty of RANSAC is that it conducts a highly guided search of the

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

Table 3: *The number m of samples required to ensure $\Upsilon \geq 0.95$ for given p_f and ϵ, where Υ is the probability that all the data points selected in at least one sample are inliers.*

solution space leading to very rapid convergence on a robust solution.

6.2 Other Random Sampling Algorithms

Surprisingly, RANSAC originated in the field of computer vision, where perhaps a greater understanding of the practical aspects of robust estimation exists, and it was not until a few years later that a similar highly robust estimator was developed in the field of statistics, namely Rousseeuw's least median of squares (LMS) estimator [39]. RANSAC uses the number of inliers as a cost function to be maximised. The method of random sampling can be applied with a different cost function, as exemplified by LMS and MINPRAN (minimize probability of randomness), both of which are now described. To avoid confusion the notation is as follows: LMS stands for least median of squares, and LS stands for least mean squares.

LMS is similar to the RANSAC algorithm as given in Table 2 except that the solution giving the lowest median error is selected as the estimate at step 2. The advantage of LMS is that it requires no setting of thresholds or *a priori* knowledge of the variance of the error. Indeed the variance σ may be robustly estimated as follows: It is known that $\text{med}_i|d_i|/\Phi^{-1}(0.75)$ is an asymptotically consistent estimator of σ when the d_i are distributed as $N(0, \sigma^2)$, where Φ is the cumulative distribution function for the Gaussian probability density function[2]. It has been shown [39] empirically that when $n \approx 2p_f$ the correction factor of $\left(1 + \frac{5}{d_f}\right)$ improves the estimate of the standard deviation, where d_f is the number of degrees of freedom. On noting that $1/\Phi^{-1}(0.75) = 1.4826$ the estimate

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

1. Repeat for m samples, where m is determined by Table 3.
   (a) Select a random sample of 7 data points and compute an estimate $\mathbf{F}$ of the fundamental matrix.
   (b) Calculate the distance d in $\mathcal{R}^4$ from each data point to the variety described by $\mathbf{F}$.
   (c) Sort the absolute values of the errors $|d|$ into a n length vector: $d(1), d(2), \cdots d(n)$.
   (d) For $i = 1$ to n
      - If $d(i) < d_{\min}(i)$ then $d(i) = d_{\min}(i)$
2. For $i = 1$ to n evaluate $\mathcal{F}(d_{\min}(i), i, n)$ and take $\mathbf{F}$ that minimizes this as the solution.
3. Re-estimate $\mathbf{F}$ using all the data identified as consistent. (A different more computationally expensive estimator may be used at this point e.g. iterative minimization.)

Table 4: *A brief summary of MINPRAN. d_i is the estimate errors for each datum.*

of $\underline{\sigma}$ is

$$\sigma = 1.4826 \left(1 + \frac{5}{d_f}\right) \text{med}_i |d_i|. \tag{28}$$

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 points are inlying. (The use of the EM algorithm to further improve the estimate of the standard deviation is described in [46].) The major problem with LMS is that it fails if more than half the data are outlying.

The third robust random sampling algorithm is MINPRAN as described by Stewart in [43]. Stewart initially assumes that all the data are outliers with a uniform distribution. Consequently he assumes that the absolute values of the errors have a uniform distribution in the range $[0..Z_0]$. Given a fit then the probability that at least k data are inliers (have error below threshold t) is

given by the Binomial distribution:

$$\mathcal{F}(t,k,n) = \Pr(k) = \sum_{i=1}^{n} \left(\begin{array}{c} n \\ i \end{array} \right) \frac{t}{Z_0}^{i} (1 - \frac{t}{Z_0})^{n-i} \quad . \tag{29}$$

If this value is small then it is unlikely that the data will all be outliers. Thus minimizing $\mathcal{F}$ over all fits and t will give the fit, together with associated residuals and inliers, that is least likely to have arisen under the assumption that all the data are outliers ('by chance'). Stewart's suggested algorithm for minimizing $\mathcal{F}$ is summarised in Table 4 and is based on random sampling. It is possible to use any distribution to model the outliers provided that the probability that the absolute value of a given residual lies below t can be evaluated.

MINPRAN shares with LMS the advantage that it does not require knowledge of the inlier distribution; it does however require knowledge of the outlier distribution. When estimating the fundamental matrix it is not clear what value Z_0 should be set to.

Because a reasonable estimate of the threshold for accepting data as inlying is usually known (either by specification for a particular application or by analysis) RANSAC is used to estimate **F**.

6.3 Robustly Estimating Other Constraints

The other models given in Table 1 are also fitted to the correspondences using RANSAC. The procedure is the same as for **F** except that the minimal number of correspondences needed for estimation varies according to the model, and is given in the Table. The threshold t^2 for determining error in (26) varies according to the dimension of the model. For dimension two varieties the squared error is a χ^2 with two degrees of freedom hence to achieve the same 95% confidence level the threshold must be increased to $t^2 = 5.99$. Later it will be seen that this is important when comparing the number of inliers for models of differing dimensions

7 The Model Selection Algorithm

Within this section the analysis of robust estimators and the analysis of degeneracy are brought together to develop an estimation algorithm that both rejects outliers and detects degeneracy. It has been stated in the last section that random sampling provides the best estimate of the model. The basic philosophy behind PLUNDER is to use RANSAC to estimate each model and then

develop a scoring function to select the most parsimonious estimated model. The new algorithm PLUNDER-DL[3] uses random sampling to determine the best fitting degenerate and non-degenerate models, and then uses MDL types ideas to select the correct model.

Before going into the details of PLUNDER-DL some preliminary approaches that were rejected are described. The comparison between these different approaches gives an insight into why the PLUNDER-DL approach was adopted.

7.1 Degeneracy and LMS

If over half the data are consistent with two or more fundamental matrices then LMS is unusable (to estimate $\mathbf{F}$). To test multiple solutions a lower order model (e.g. a quadratic transform $\mathbf{Q}$) is fitted as well. To detect degeneracy a test is made to see if the median error for the best fitting $\mathbf{Q}$ is below a user specified threshold. After some initial success this approach was rejected for the following reason. In many scenes, even those arising from cameras under general motion and with a wide range of scene structure, a large number, even a majority, of image correspondences are consistent with a projectivity. The crux of the matter is that half the data might lie on a plane (a degenerate configuration) *but* there may be enough points off the plane to resolve the degeneracy. Thus it is undesirable to base decisions about degeneracy on just half the data. Even worse, the majority of correspondences might be consistent with a plane-plane projectivity even when there are no planes in the scene. This often occurs with outdoor scenes, even under general motion, if points are so distant as to be indistinguishable from those on the plane at infinity for a given threshold.

In the general case of model fitting the median cannot be used safely as a means of comparing the fit of models, as the median of the fitted residuals can be made arbitrarily low by increasing the number of parameters in the model.

7.2 Degeneracy and MINPRAN

A second approach is to use the MINPRAN criterion. The assumption that the outliers are uniformly distributed, the basis of MINPRAN, leads to (29), which gives the probability that at least k data points fall within the threshold region given that the data consists entirely of uniformly

[3] In keeping with the original RANSAC acronym, PLUNDER stands for Pick Least UNDEgenerate Randomly. DL stands for description length

distributed outliers. One approach might be to estimate the best **F** and **Q** (from Table 1) for the data: evaluate (29) for each, and select the solution with the lowest probability. This is the basis of the initial PLUNDER approach set out in [48, 49], and is dubbed PLUNDER-MINPRAN here. The problem with this is that the value of Z_0 determines which model is favoured and there is no accurate way of estimating Z_0.

7.3 Using RANSAC to Detect Degeneracy—PLUNDER-DL

How might we incorporate the detection of degeneracy into RANSAC? First we use RANSAC to get a robust estimate of the parameters and inliers for each model from the set of data. Once this is done we would like to develop a method that is as general as possible, that will select the most likely model given only the number of inliers provided by RANSAC for each model. The original PLUNDER approach followed in [48] was to use hypothesis testing procedures. As noted in Lehman [29] hypothesis testing procedures are traditionally applied when multiple decision procedures are required (in this case model selection). If statistical identification is considered as a decision procedure, then the problem becomes that of determining a suitable loss function. In the Neyman-Pearson theory of statistical hypothesis testing only the probabilities of firstly rejecting or secondly not rejecting the null hypotheses, are considered [42]. The problem with this approach is that it is difficult to adapt to a situation where several models might be appropriate (rather than only two), as the test procedure for a multiple-decision problem involves a difficult choice of a number of dependent significance levels. Rather what is required is some sort of score function for each of the fitted models; and this score function must perforce be composed of some combination of the available information: the number of inliers and outliers, and the degree and dimensionality of the model. The choice of scoring function must be dictated by the application of the algorithm. In order to accomplish this within this paper we have adapted the ideas of Minimum Message Length (MML) [14] and Minimum Description Length (MDL) [38]. The general idea of these methods is to minimize the binary digits (*bits*) needed to encode a signal in order to send the signal over a communication channel and decode it on the other side of the channel. The idea that simpler descriptions are better than more complex ones is a strongly intuitive notion that was first enunciated as Occam's razor, which counsels us "not to multiply entities beyond necessity". Such a procedure might be useful if the image correspondences have to be transmitted for an application such as teleconferencing, where data compression is of interest.

The new PLUNDER-DL approach differs from MDL in that it does not assume knowledge of

the residuals or their probability distributions, as it encodes only the deterministic and not the stochastic part of the signal (which is assumed to be small). All that is needed is the set of inliers and the set of outliers to the model. In the results section we will argue that in the vast majority of cases further sophistry is unnecessary and adds an extra computational burden for little gain.

Consider a point in $\mathcal{R}^4$, in order to encode this point 4 coordinates are needed. Let the image size be $R \times R$ window, e.g. 512×512, and the coordinates are measured up to a resolution of ϵ, e.g. the Harris corner detector [15, 17] can achieve sub-pixel accuracy of 0.1 pixels in the location of a point. If a point is located within $\mathcal{R}^4$ then the minimum encoding length for each coordinate is $\log a = \log \frac{R}{\epsilon}$. Let the model parameters storage be $\log b$. As a fundamental matrix has 7 DOF then it may be encoded by seven parameters. As the variety defined by $\mathbf{F}$ has dimension three, each inlier may be encoded by 3 coordinates. Outliers to the fundamental matrix may not be so economically described being points in $\mathcal{R}^4$, and hence require 4 coordinates to describe them. Thus if the fit to the fundamental matrix provides n_i^F inliers and n_o^F outliers the PLUNDER-DL code length is

$$\mathbf{PL}(\mathbf{F}) = (3n_i^F + 4n_o^F)\log a + 7\log b \quad . \tag{30}$$

If for a homography there are n_i^H inliers and n_o^H outliers then the corresponding PLUNDER-DL score is (recalling that the dimension of $\mathbf{H}$ is 2 and there are 8 degrees of freedom)

$$\mathbf{PL}(\mathbf{H}) = (2n_i^H + 4n_o^F)\log a + 8\log b \quad . \tag{31}$$

The model with the least PLUNDER-DL score is deemed to encode the data most parsimoniously. The comparison between the models is based on the ratio of $\log a$ to $\log b$, further details of this are given below. The general PLUNDER-DL encoding score for a model is:

$$\mathbf{PL} = (n_i \text{ dimension of model } + 4n_o)\log a + \text{ DOF } \log b \ , \tag{32}$$

where there are n_i inliers and n_o outliers for a model with DOF degrees of freedom, 4 is the dimension of the data in this case. Thus the PLUNDER-DL criterion accounts for the dimension and degrees of freedom of the model whilst requiring minimal knowledge of the probability distribution of the data.

As mentioned in Section 6 it is important to set the threshold used for calculating inliers to give approximately the same chance that a 'true inlier' is accepted for each model; in order that the comparison between models of differing dimensions is valid.

Next the amount of storage $\log b$ required for the estimated parameters is determined. This is dependent on the storage required for the data $\log a$, and the precision of the estimated parameters. In the absence of *a priori* information we choose to store the parameters with the same amount of storage as the data, i.e. $\log b = \log a$.

$$\mathbf{PL} = (n_i \text{ dimension of model } + 4n_o) + \text{ DOF }, \tag{33}$$

and since $n = n_i + n_o$,

$$4n - \mathbf{PL} = n_i\text{codimension of model} - \text{ DOF}. \tag{34}$$

Thus minimizing the PLUNDER-DL score **PL** over all models selects the model which has the maximum number of degrees of freedom for its inliers (the degrees of freedom in the inliers being $n_i$codimension of model $-$ DOF, as used in the χ^2 test).

It can be seen that PLUNDER-DL is a generalisation of RANSAC to take into account multiple models.

Minimum description length approaches such as [28, 30] provide a more complex scoring function than (33), with additional terms for the sum of squares of residuals for the inliers (representing the log likelihood under Gaussian noise assumptions). Such methods are reviewed in Torr [50]. These more complex methods have two distinct disadvantages *vis* PLUNDER-DL. The first is that they require exact knowledge of the underlying noise distribution, whereas PLUNDER-DL simply requires a threshold to distinguish outliers. Secondly the MDL methods require exact computation of the residuals, usually entailing a non-linear gradient descent computation for each model, in addition to a RANSAC stage to eliminate outliers. The PLUNDER-DL score function can be used to eliminate obviously inappropriate models without ever having to do more computation than is involved in RANSAC.

PLUNDER-DL is the same as MDL under the limiting assumption that the noise on the data is negligible with respect to the effect of outliers. Under these circumstances it is much more computationally efficient to use PLUNDER-DL as the robust model selection criterion. The PLUNDER-DL method is summarised in Table 5.

1. Estimate the set of inliers for each model using RANSAC, as described in Section 6.
2. Compute PLUNDER-DL scores for each model using (33).
3. Accept the model with the lowest PLUNDER-DL score.

Table 5: *A brief summary of PLUNDER-DL.*

8 Use of Planes Within the Data

If any projectivity or affinity has a large proportion of inliers (for instance twenty percent) then we can use PLUNDER-DL to test whether this part of the image is the projection of a plane. In [6, 36] it is shown that given the images of four points on a plane and two off it we may compute the epipolar geometry. This is an extension of the work by Koenderink [27] where it is shown that for an affine camera the image of three points on the plane and one off it defines the epipolar geometry. Thus if we can identify planes within the image we can exploit their presence by repeatedly sampling just 1 (for affinity and $\mathbf{F}_A$) or 2 (for projectivity and $\mathbf{F}$) points off the plane and testing the estimated epipolar geometries as explained in Section 7. This approach is most efficacious when a large proportion of the data are imaged from a plane, for example as in Figure 4. A set of random samples might not include any correspondences off the plane and thus each sample would be degenerate. The benefit of this algorithm is that it will (once the plane is detected) allow a sampling of points on and off the plane to estimate epipolar geometry, even if there are only a few off the plane.

9 Results

Table 6 summarises the sizes of various degenerate sets and PLUNDER-DL scores found for the figures 3 and 5—9 using the algorithm outlined in Section 7. Scores are given for $\mathbf{F}$, $\mathbf{F}_A$, $\mathbf{F}_T$, $\mathbf{H}$, $\mathbf{Q}$ and $\mathbf{A}$. The standard deviation of the error was found by fitting the most general model, ($\mathbf{F}$), as we know that this model will always fit the data, and then using the method in Section 6.2. Using the most general model causes a slight under estimation of the standard deviation in some cases, but the effect is not severe as we found a good deal of insensitivity to the choice of σ.

Model	corr	Figure 3	Figure 5	Figure 6	Figure 7	Figure 8	Figure 9
Number of matches n		36	142	406	190	80	332
Fundamental	7	31/120	134/471	305/1326	131/645	75/252	241/1094
Translation	2	27/119	88/482	305/1322	69/693	14/308	241/1089
Affine Fundamental	4	28/120	128/444	304/1324	64/700	46/278	240/1092
Quadratic	18	31/100	131/324	270/1102	127/524	38/262	240/866
Projectivity	8	27/98	127/318	268/1096	71/626	36/256	239/858
Affinity	6	27/96	128/315	239/1152	53/600	32/262	239/856

Table 6: *For each model the number of inliers/PLUNDER-DL score (n_i/**PL**) is given. The model selected is underlined. corr refers to the minimum number of correspondences needed to instantiate a model.*

Truck data. In the truck data, shown in Figure 3, the model selected is an affinity, established by 3 point correspondences; 27 points are consistent with the affinity and 31 with a full fundamental matrix, which has $7 - 3 = 4$ more degrees of freedom. The data are thus clearly degenerate with respect to $\mathbf{F}$, with most of the points lying on the plane of the truck.

The cyclotorsion data. In Figure 5 the camera motion combines a translation perpendicular to the image plane with a cyclotorsion. The best fitting degenerate model is an affinity with 128 points, as expected.

Gun fighter data. Figure 6 (a) (b) shows a mobile gun fighter viewed by a tracking camera. The gunfighter is moving to the right as the camera moves in parallel to keep him in the centre of the image. Points in independent motion on the gun fighter are identified if they are inconsistent with the motion model chosen for the background motion. The data support an projectivity as most of the correspondences are on the saloon wall.

Model house data 1. Figure 7 shows a scene in which a camera rotates and translates whilst fixating on a model house. The important thing about this image pair is that structure recovery for this image pair proved highly unstable. The reason for this instability is not immediately apparent until the good fit of the quadratic transformation is witnessed, indicating that the structure is close to a critical surface.

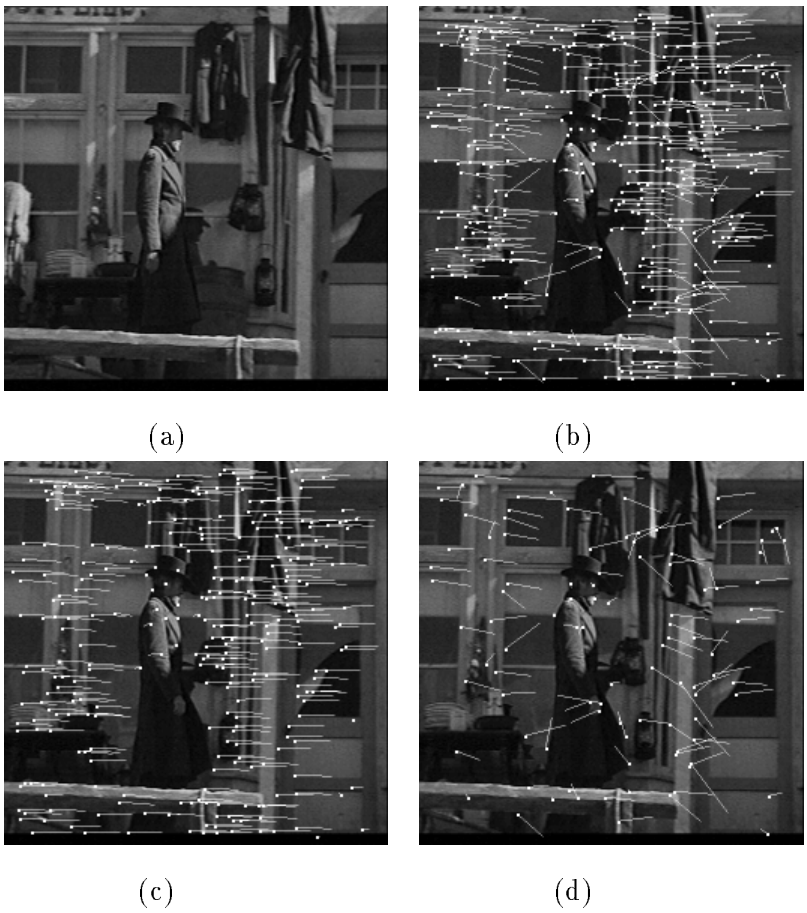


Figure 6: *Two images from the gun fighter sequence. The camera is tracking to the right and the gun fighter raises his gun. (a) first image, (b) second image and matches, (c) inliers, and (d) outliers.*

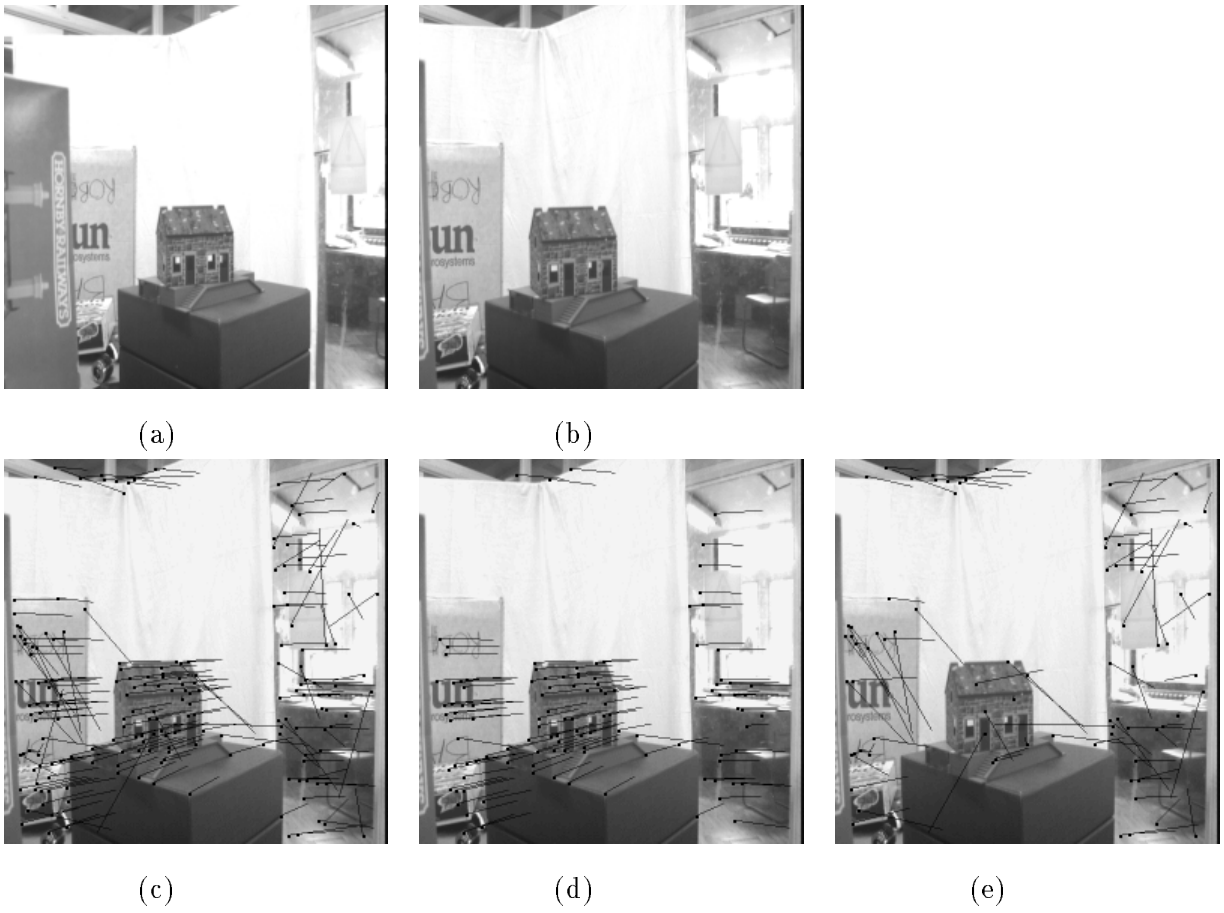
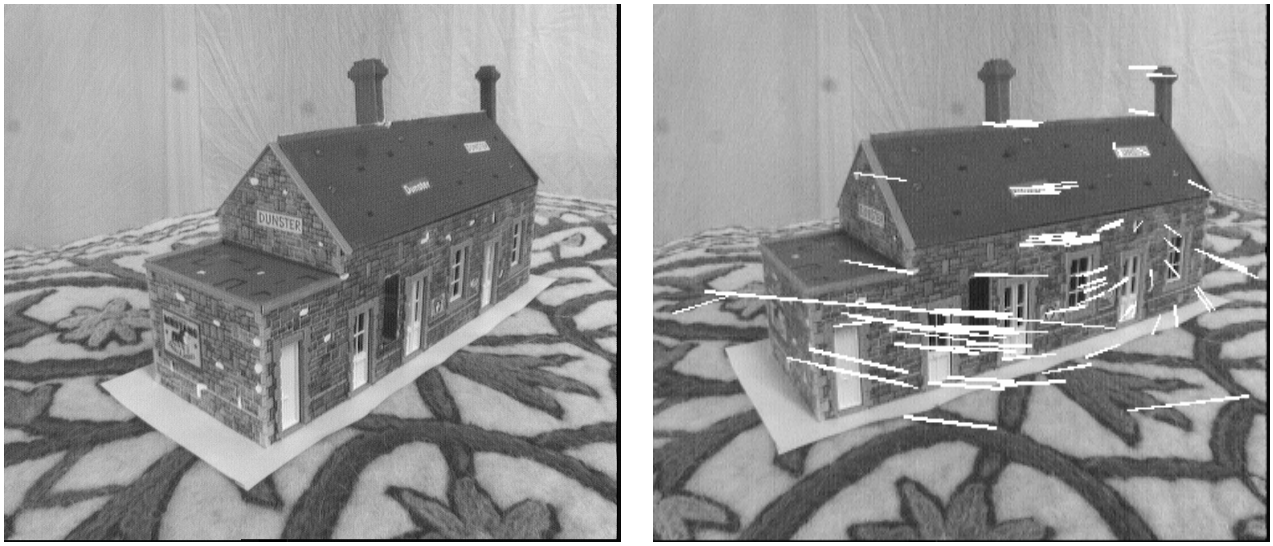


Figure 7: *Indoor sequence, camera translating and rotating while fixated on the model house; (a) first image, (b) second image, (c) matches, (d) inliers, and (e) outliers. The data are found to be degenerate with respect to the fundamental matrix.*



(a) (b)

Figure 8: *Indoor sequence, camera translating and rotating to fixate on the house. (a) Left images, (b) Right image and inliers.*

Model house data 2. Figure 8 show a scene in which a camera rotates and translates whilst fixating on a model house. The scene is correctly flagged as a general motion, as the translation and rotational components of the camera motion were both significant, and full perspective structure can be recovered.

Independent motion data. Finally, Figure 9 are taken by a moving camera. The scene contains an independently moving robot. The correspondences are shown in (c). The aim is to cluster all the points consistent with the background whilst excluding points on the robot and outliers. To demonstrate the advantage of the new degeneracy detection algorithm we compare our results with those obtained from a more conventional robust estimator, RANSAC. Figure 9 (d) shows the inliers and Figure 9 (e) shows the outliers predicted using a full fundamental matrix estimated by RANSAC. It can be seen in (d) that not all the points on the robot are flagged as inconsistent with the background. However, the PLUNDER-DL selected model (an affinity) does eliminate these matches.

10 Discussion

This paper has investigated something long overlooked in parameter estimation, namely the strong intertwining of the problems of outlier detection and degenerate model selection. The methodology

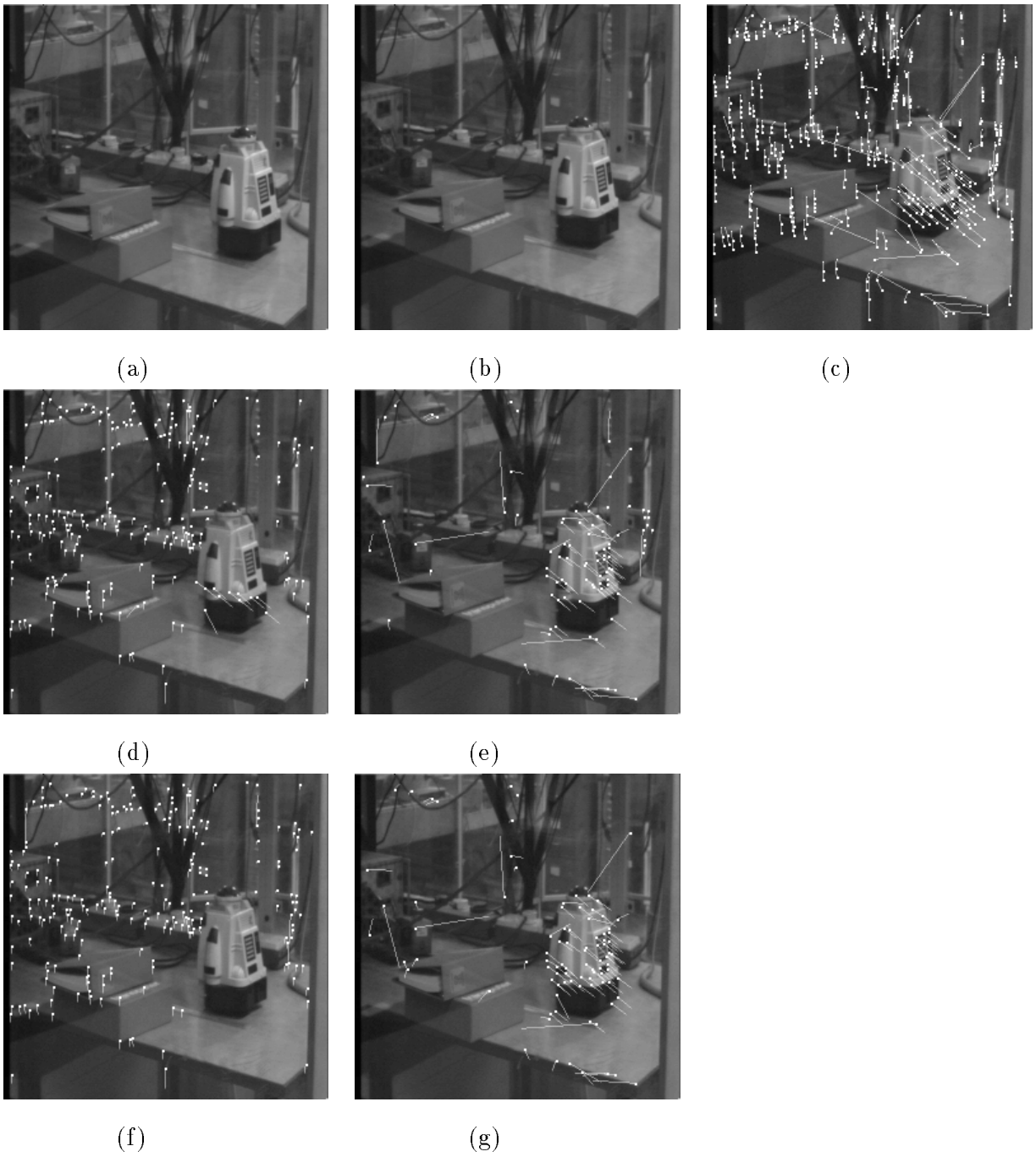


Figure 9: *First (a) and second (b) image and (c) matches of the sequence with an independently moving robot viewed from a camera, translating vertically down (parallel to the image plane). The robot moves along the table. Fundamental matrix inliers (d) and outliers (e) for* $\mathbf{F}$*. It can be seen that the independently moving robot is not fully segmented. PLUNDER-DL indicated an affinity fit, with inliers (f) and outliers (g). The independently moving robot is fully segmented.*

is general. Although it has been illustrated here only for the problem of estimating the fundamental matrix, it could be used on a much wider class of problems.

The PLUNDER-DL score is both simple and easy to implement, and it generalises the robust estimator RANSAC to the problem of model selection. The nature of the degenerate cases for the fundamental matrix have been clearly described and the relationships between the plethora of models used have been revealed by considering pairs of image coordinates as points in a four dimensional space. The sundry models describe varieties of varying degree and dimension in the measurement space. The PLUNDER-DL score function rates the suitability of each model taking into account the number of inliers estimated, the dimension, degrees of freedom and encoding length for each model.

10.1 Future Work

It is not difficult to extend the algorithm to deal with more than two views or to new models.

The fact that the data might be consistent with $\mathbf{F}_T$ and $\mathbf{F}_A$ suggests that it might be worth including a new model in our taxonomy: the skew symmetric affine camera model $\mathbf{F}_S$ which could arise either from affine image conditions or from general image conditions and pure translation parallel to the image plane.

$$\mathbf{x}'^{\top}\mathbf{F}_S\mathbf{x} = 0 \quad \text{where} \quad \mathbf{F}_S = \begin{bmatrix} 0 & 0 & -s_2 \\ 0 & 0 & s_1 \\ s_2 & -s_1 & 0 \end{bmatrix}, \quad s_1^2 + s_2^2 = 1. \tag{35}$$

This model would have dimension three and 1 degree of freedom, and will typically arise for stereo rigs with the cameras pointing in the same direction. Another model that is useful for camera calibration is the planar motion fundamental matrix $\mathbf{F}_P$ [7, 3], which arises when the rotation axis and direction of translation of the camera are perpendicular.

A problem with PLUNDER-DL is that the cost function favours segmentations with too many fragments i.e. it is biased towards under fitting. This problem might be overcome by using a more complex scoring function which involves the probability that an outlier might be misidentified as an inlier, and estimates of the accuracy of the parameters. This means sacrificing the simplicity of the score function along with the ease of implementation and computational efficiency of PLUNDER-DL, and is beyond the scope of this current work. Related work [47] explores the development of a score for each model based on maximum a posteriori estimation and the AIC criterion.

Often, however, the fundamental matrix is computed in the context of a task e.g. using the fundamental matrix as a means of obtaining corner correspondences through a sequence [4]. This has the important consequence that it is not necessary to decide from only a single image pair whether a particular set of correspondences is degenerate (for example, arise from a plane). Instead, several possibilities can be explored, and the decision postponed until later in the sequence when more evidence has accumulated (Multiple Hypothesis Tracking). In this case, it is better to avoid a definitive choice of model, but to assign probabilities to each of several possible models. What is important is that the model is sufficient to propagate correct correspondences, and not introduce mis-matches. Thus the work in this paper can only be considered as a first stab at a solution—a setting of the scene.

Acknowledgements

Financial support for this work was provided by EU ACTS Project VANGUARD. We are very grateful to Professor K. Kanatani, Dr D. Murray for interesting input; Dr. A. Fitzgibbon and Dr. P. Beardsley both for technical and software advice.

A Relation Between Projectivities and Degenerate configurations

Within this appendix we shall show that a) if $\dim(N) = 3$ then there must be three linearly independent $\mathbf{F}$ consistent with the data, and b) that if there are three linearly independent $\mathbf{F}$ consistent with the data then the data must also be consistent with an image-image projectivity $\mathbf{H}$.

Theorem. If $\dim(N) = 3$ then there must be three linearly independent $\mathbf{F}$ consistent with the data.

Proof. Suppose $\dim(N) = 3$ and that N is spanned by basis vectors $\mathbf{u}_1$, $\mathbf{u}_2$ and $\mathbf{u}_3$, corresponding to matrices $\mathbf{U}_1$, $\mathbf{U}_2$ and $\mathbf{U}_3$ then three linearly independent fundamental matrices may be constructed as follows $\mathbf{F}_1 = \alpha_1\mathbf{U}_1 + (1 - \alpha_1)\mathbf{U}_2$, $\mathbf{F}_2 = \alpha_2\mathbf{U}_1 + (1 - \alpha_2)\mathbf{U}_3$, and $\mathbf{F}_3 = \alpha_3\mathbf{U}_2 + (1 - \alpha_3)\mathbf{U}_3$, such that $\det \mathbf{F}_1 = 0$, $\det \mathbf{F}_2 = 0$, and $\det \mathbf{F}_3 = 0$.

Comment In fact there is a family of consistent fundamental matrices but for the purposes of proof we need only show that least three are linearly independent.

Theorem. Let $\mathbf{F}_1$, $\mathbf{F}_2$, $\mathbf{F}_3$ be three fundamental matrices compatible with a general set of image correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}_i^{'}$, $1 \leq i \leq n$. Then if $\mathbf{F}_1$, $\mathbf{F}_2$, $\mathbf{F}_3$ are linearly independent; then the image correspondences are related by a projectivity $\mathbf{x}_i^{'} = \mathbf{H}\mathbf{x}_i$, $1 \leq i \leq n$.

Proof. We shall use proof by contradiction, we shall show that the correspondences must be linked by a projectivity $\mathbf{x}_i^{'} = \mathbf{H}\mathbf{x}_i$, or else two of them must be linearly dependent giving rise to a contradiction. In order to do this we

consider two cases, either **case 1** two of the fundamental matrices $\mathbf{F}_1$, $\mathbf{F}_2$, $\mathbf{F}_3$ share a common epipole, or **case 2** they do not.

Case 1. In the first case, suppose that two of the $\mathbf{F}_i$, for example $\mathbf{F}_1$ and $\mathbf{F}_2$, have the same epipole $\mathbf{e}'$. Hence $\mathbf{F}_1\mathbf{x}_i = \mathbf{F}_2\mathbf{x}_i$, $1 \leq i \leq n$, since the epipolar line corresponding to $\mathbf{x}_i$ passes through $\mathbf{x}_i'$ and $\mathbf{e}'$. It follows that $\mathbf{F}_1$, $\mathbf{F}_2$ define the same epipolar transform, hence $\mathbf{F}_1 = \lambda\mathbf{F}_2$ for some $\lambda \neq 0$, and the $\mathbf{F}_i$ are linearly dependent which leads to a contradiction. Hence the $\mathbf{F}$ must either have distinct epipoles (the next case we consider) or the $\mathbf{F}_i$ are linearly dependent.

Case 2. It is now assumed that the $\mathbf{F}_i$ have distinct epipoles. The image correspondences $\mathbf{x} \mapsto \mathbf{x}'$, $1 \leq i \leq n$, are compatible with the map

$$\mathbf{x} \mapsto \Phi(\mathbf{x}) \equiv \mathbf{F}_1\mathbf{x} \times \mathbf{F}_2\mathbf{x} = \mathbf{F}_2\mathbf{x} \times \mathbf{F}_3\mathbf{x} = \mathbf{F}_3\mathbf{x} \times \mathbf{F}_1\mathbf{x} \ . \tag{36}$$

Which must either be a quadratic transformation or a projectivity. The components of $\Phi(\mathbf{x})$ are polynomials of degree two in the coordinates of $\mathbf{x}$. If the components of $\Phi(\mathbf{x})$ have a common factor then (36) defines a projectivity. If there is no common factor; then Φ is a quadratic transformation with three distinct fundamental points[4] namely the epipoles of the $\mathbf{F}_i$ in the first image. It will now be demonstrated that if the correspondences are consistent with a quadratic transform then they are consistent with at most two *linearly independent* fundamental matrices.

Let coordinates be chosen in both images such that the epipoles $\mathbf{e}_i$, $\mathbf{e}_i'$ have coordinates

$$\begin{array}{lllllll} \mathbf{e}_1 & = & (1,0,0)^\top & \mathbf{e}_2 & = & (0,1,0)^\top & \mathbf{e}_3 & = & (0,0,1)^\top \\ \mathbf{e}_1' & = & (1,0,0)^\top & \mathbf{e}_2' & = & (0,1,0)^\top & \mathbf{e}_3' & = & (0,0,1)^\top \end{array}$$

and such that $(1,1,1)^\top \leftrightarrow (1,1,1)^\top$. The matrix $\mathbf{F}_1$ satisfies $\mathbf{F}_1\mathbf{e}_1 = 0$ and $\mathbf{e}_1'^\top\mathbf{F}_1 = 0$. In addition the image $\mathbf{F}_1\mathbf{e}_2 \times \mathbf{F}_3\mathbf{e}_2$ of $\mathbf{e}_2$ under Φ is undefined, thus $\mathbf{F}_1\mathbf{e}_2$ is proportional to $\mathbf{F}_3\mathbf{e}_2$. It follows that $\mathbf{e}_3'^\top\mathbf{F}_1\mathbf{e}_2 = 0$; similarly, $\mathbf{e}_3^\top\mathbf{F}_1\mathbf{e}_2' = 0$. As a consequence of all these equations,

$$\mathbf{F}_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix} \tag{37}$$

where b, c are yet to be found. The condition $(1,1,1)\mathbf{F}_1(1,1,1)^\top = 0$ ensures that $b + c = 0$, thus after a rescaling,

$$\mathbf{F}_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \tag{38}$$

Similar arguments yield

$$\mathbf{F}_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \qquad \mathbf{F}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{39}$$

[4] A quadratic transformation is defined everywhere except at three points (where all 3 components of Φ evaluate to zero). The standard name for these points is 'fundamental points' unfortunately clashing with 'fundamental matrix'. The quadratic transformation is not defined at the epipoles thus the epipoles are included in the 'fundamental points' of Φ. There are three distinct epipoles, one for each $\mathbf{F}$, thus the epipoles are the fundamental points of Φ.

thus the $\mathbf{F}_i$ are linearly dependent, and there is a contradiction. Hence if there are three or more linearly independent fundamental matrices consistent with the data the data must conform to a projectivity. □

In [33] Longuet-Higgins describes cases in which three different epipolar geometries are compatible with the same set of image correspondences. It follows from the above theorem that in these cases the three fundamental matrices are linearly dependent.

B Testing the Distributions

Within this appendix a method for testing the distribution of errors for inliers and outlier is presented. If the data do not conform to the postulated distribution (Gaussian for inlier and uniform for outlier), then it is likely that there is a problem with the fitted model, in this case there may be some unexpected event, such as an independently moving object.

It has been postulated that the distribution of errors on the inliers to the optimal fitting fundamental matrix is well approximated by a Gaussian distribution, and that the distribution of errors on the outliers is well approximated by a uniform distribution.

These assumptions may be automatically tested within the algorithm by an appropriate χ^2 test as follows [26]: For the inliers the null and alternative hypotheses are

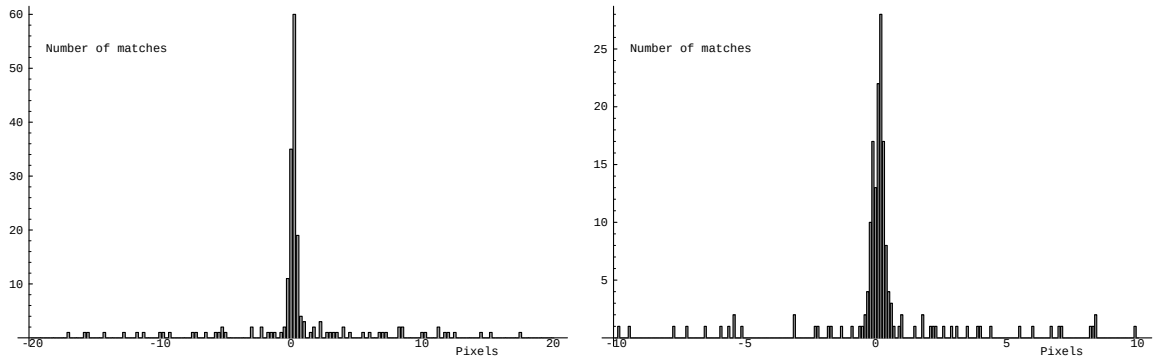
Hypothesis 1 *H_0 The distribution of the errors on the inliers is Gaussian against H_1 The distribution of the errors on the inliers is not Gaussian*

The test is to compare the sample frequencies with the frequency one would expect if the null hypothesis H_0 is true. The absolute value of the residuals are sorted into bins and the sum of difference between observed and expected (given Gaussian (or uniform) assumptions) frequencies in the bins calculated. The comparison is achieved by calculating a χ^2 statistic:

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} \tag{40}$$

where O_i = the observed frequency for bin i, and E_i = the expected frequency for bin i. If this calculated χ^2 value is large in comparison with that expected under the null hypothesis, it is held to be significant and H_0 is rejected. In order to calculate E; the mean and standard deviation of the errors of the inliers are first computed, and the errors normalized to have zero mean and unit variance. Ten bins are chosen such that the area beneath the Gaussian curve within each bin is 0.1, i.e. 10% of the observations are expected to lie in each bin. This corresponds to cutoffs at 0.0, 0.1257, 0.2533, 0.3853, 0.5244, 0.6745, 0.8416, 1.0364, 1.2816 and 1.6449 (obtained from [31]). The corresponding χ^2 given by equation (40) has 10-3 = 7 degrees of freedom (two of the degrees of freedom are lost because the mean, and the standard deviation are estimated from the data, and a third is lost because it is necessary to know the number of observations [54]). If the χ^2 value exceeds a user defined threshold then H_0 is rejected. In this case the threshold is $\chi^2_{0.01} = 18.48$. Under H_0 the test statistic exceeds this threshold only 1% of the time.

For the outliers a similar procedure may be followed. The hypotheses are



(a) (b)

Figure 10: *Histograms of all the errors for the optimal fitting* $\mathbf{F}$ *to the data shown in Figure 7, shown with two different bin sizes. (a) bin size is 0.25; the height of the middle block indicates the number of correspondences with errors 0-0.25, the second with errors 0.25-0.5 etc. (b) bin size 0.1; the number of correspondences shown in blocks of 0.1 pixels.*

Hypothesis 2 *H_0 The distribution of the errors for the outliers is uniform; against H_1 The distribution of the errors for the outliers is not uniform.*

This may be tested by calculating the frequencies of the errors in ten equal bins between $-b$ and b, b being the side of the search window as described in Section 5.1, and following the procedure described above for the inliers—comparing the expected to observed frequencies.

The frequencies of the errors for the optimal fit to the data given in Figure 7 are shown in Figure 10, as histograms. The χ^2 for both the inliers and outliers falls below the threshold $\chi^2_{0.01} = 18.48$ which indicates that the assumptions made on the distributions may be correct (or at least that there is not sufficient evidence to indicate that they are incorrect).

The test can only be a rough guide as it depends upon achieving a near optimal estimate of the fundamental matrix, but nevertheless failure of the test strongly indicates that one of the assumptions underpinning the estimation of $\mathbf{F}$ is invalid, and that further investigation is necessary. Such a failure might indicate that the estimate of the fundamental matrix is incorrect, that the model fitted is incorrect or that there are several independently moving objects. The last case is illustrated in Figure 11. In Figure 11 (a) the errors to the fit of affine camera fundamental matrix $\mathbf{F}_A$ are shown. As explained in Section 2.3, $\mathbf{F}_A$ defines a three dimensional hyperplane in $\mathcal{R}^4$ and consequently the errors are the distances of each correspondence orthogonal to the hyperplane. Examination of Figure 11 (a) reveals that there is some correlation between the points off the hyperplane[5]. The correlation in the outliers is revealed by the χ^2 test which indicates that the errors are *not* uniformly distributed. This is clearly seen in the histogram of the errors Figure 11 (b). The errors in the group to the right correspond to the independently moving robot. Of course the errors for independently moving objects need not be always be so visibly correlated.

C Derivation of First Order Approximation to the Error Term for Projectivities

The variety V for a projectivity in $\mathcal{R}^4$, is the intersection of two dimension three quadric hypersurfaces as shown schematically in Figure 12. The two quadric hypersurfaces are derived as follows: If $\underline{\mathbf{x}}' = \mathbf{H}\underline{\mathbf{x}}$ then for the elements of $\mathbf{H}$, h_i, $1 \leq i \leq 9$ such that

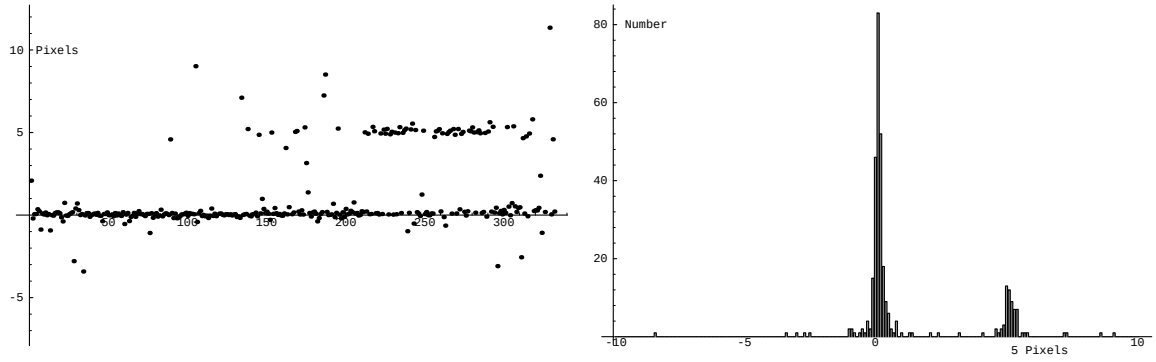
$$x' = \frac{h_1 x + h_2 y + h_3}{h_7 x + h_8 y + h_9} \tag{41}$$

$$y' = \frac{h_4 x + h_5 y + h_6}{h_7 x + h_8 y + h_9}. \tag{42}$$

Let $\mathbf{h} = (h_1, h_2, h_3, h_4, h_5, h_6, h_7, h_8, h_9)^\mathsf{T}$. It follows from (41) and (42) that

$$\mathbf{v}_1^\mathsf{T}\mathbf{h} = 0 \tag{43}$$

[5]In fact the idea of examining the projections of points in lower dimension ("projection pursuit") has long been proposed in the statistics literature as a means to identify clustering [22].



(a) (b)

Figure 11: *(a) Plot of the 332 errors in Figure 9, to a fit of* $\mathbf{F}_A$*. The residuals are of correspondences in the image working from the top left to the bottom right. The* Y *axis is the error in pixels. (b) Histograms of the errors in groups of 0.1.*

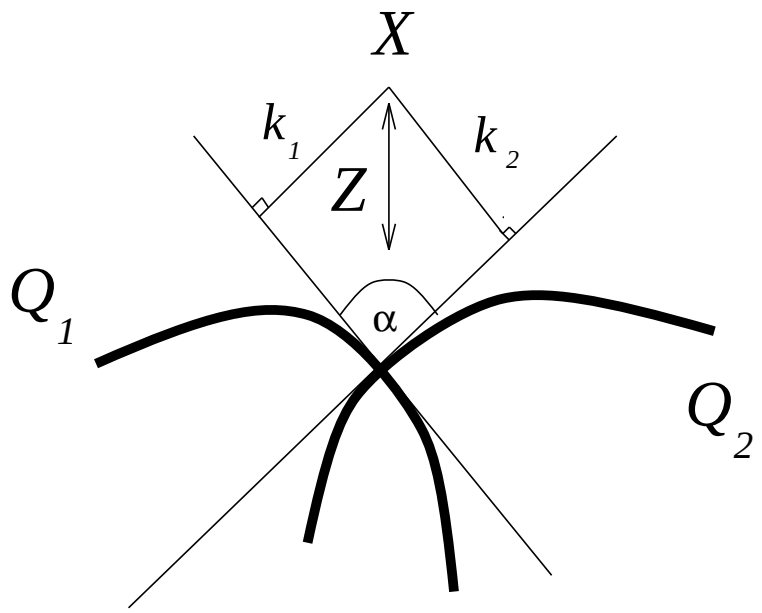


Figure 12: *The dimension 2 variety V is the intersection of two dimension 3 varieties Q_1, Q_2. To calculate a first order approximation of the distance of a given point $\mathbf{X}$ in $\mathcal{R}^4$ to V, calculate Sampson's distances from $\mathbf{X}$ to Q_1 and Q_2. Assuming Q_1 and Q_2 are locally linear, let the angle between Q_1 and Q_2 be α then the distance of $\mathbf{Z}$ to V may be calculated from elementary trigonometry, as explained in the text.*

and

$$\mathbf{v}_2^{\mathsf{T}}\mathbf{h} = 0 \tag{44}$$

where

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

For $\mathbf{h}$ fixed, (43) and (44) are quadric hypersurfaces $\mathbf{Q}_1$ and $\mathbf{Q}_2$ in $\mathcal{R}^4$. The distance from a point $\mathbf{X} = (x, y, x', y')$ to each variety can be estimated as follows: It is assumed that $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are both locally approximated by their tangent hyperplanes. Let the distance from $\mathbf{Z}$ to each hyperplane be k_i, $i = 1, 2$ and let the angle between the two hyperplanes be α. Then the squared distance of the $\mathbf{Z}$ to $\mathbf{Q}_1 \cap \mathbf{Q}_2$ is approximated by

$$\frac{k_1^2 + k_2^2 + 2k_1k_2\cos(\alpha)}{\sin^2(\alpha)} . \tag{45}$$

If each image coordinate is subject to zero mean Gaussian noise, with standard deviation σ, then the error term for the optimal fundamental matrix (codimension 1) has standard deviation σ and the error term for the optimal projectivity (assuming that the projectivity model-codimension 2-is the correct model) has standard deviation $\sqrt{2}\sigma$.

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] H. Akaike. Factor analysis and AIC. *Psychometrika*, 52(3):317–332, 1987.

[3] 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.

[4] 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.

[5] P. A. Beardsley, A. Zisserman, and D. W. Murray. Navigation using affine structure and motion. In J. O. Eklundh, editor, *Proc. 3rd European Conference on Computer Vision, LNCS 800/801, Stockholm*, pages 85–96. Springer-Verlag, 1994.

[6] P.A. Beardsley, D. Sinclair, and A. Zisserman. Ego-motion from six points. Insight meeting, Catholic University Leuven, Feb 1992.

[7] Beardsley, P. and Zisserman, A. Affine calibration of mobile vehicles. In Mohr, R. and Chengke, W., editors, *Europe-China workshop on Geometrical Modelling and Invariants for Computer Vision*, pages 214–221. Xidan University Press, Xi'an, China, 1995.

[8] H. Bozdogan. Model selection and Akaike's information criterion (aic): The general theory and its analytical extensions. *Psychometrika*, 52(3):345–370, 1987.

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

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

[11] 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.

[12] O.D. Faugeras and S.J. Maybank. Motion from point matches: multiplicity of solutions. *International Journal of Computer Vision*, 4:225–246, 1990.

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

[14] M. P. Georgeff and C. S. Wallace. A general selection criterion for inductive inference. In T. O'Shea, editor, *Proceedings ECAI, PISA*, pages 473–483. Springer–Verlag, 1984.

[15] C. Harris. The DROID 3D vision system. Technical Report 72/88/N488U, Plessey Research, Roke Manor, 1988.

[16] C. Harris. Structure-from-motion under orthographic projection. In O. Faugeras, editor, *Proc. 1st European Conference on Computer Vision, LNCS 427*, pages 118–128. Springer–Verlag, 1990.

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

[18] R. I. Hartley. Estimation of relative camera positions for uncalibrated cameras. In G. Sandini, editor, *Proc. 2nd European Conference on Computer Vision, LNCS 588, Santa Margherita Ligure*, pages 579–87. Springer–Verlag, 1992.

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

[20] R. I. Hartley. Projective reconstruction and invarients from mutiple images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 16(10):1036–1041, 1995.

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

[22] P. J. Huber. Projection pursuit. *Annals of Statistics*, 13:433–475, 1985.

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

[24] K. Kanatani. Automatic singularity test for motion analysis by an information criterion. In *Proc. 4th European Conference on Computer Vision, LNCS 1064, Cambridge*, pages 697–708, Springer–Verlag, 1996. Buxton, B. and Cipolla R.

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

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

[27] J.J. Koenderink and A.J. Van Doorn. Affine structure from motion. *Journal of Optical Society of America*, 8(2):377–385, 1991.

[28] Y.G. Leclerc. Constructing simple stable descriptions for image partitioning. *International Journal of Computer Vision*, 3:73–102, 1989.

[29] E. L. Lehman. *Testing Statistical Hypothesis*. Wiley, New York, 1959.

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

[31] D. V. Lindley and W. F. Scott. *New Cambridge Elementary Statistical Tables*. Cambridge University Press, 1984.

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

[33] H.C. Longuet-Higgins. Multiple interpretations of a pair of images of a surface. *Proc. R. Soc. London*, vol.418, Series A:1–15, 1988.

[34] 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.

[35] S.J. Maybank. Properties of essential matrices. *Int. J. of Imaging Systems and Technology*, 2:380–384, 1990.

[36] R. Mohr. Projective geometry and computer vision. In Pau Chen and Wang, editors, *Handbook of Pattern Recognition and Computer Vision*. 1992.

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

[38] J. Rissanen. *Encyclopedia of Statistical Sciences*, volume 5, chapter Minimum-description-length principle, pages 523–527. Wiley:New York, 1987.

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

[40] 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.

[41] L. S. Shapiro, A. Zisserman, and J. M. Brady. Motion from point matches using affine epipolar geometry. In J. O. Eklundh, editor, *Proc. 3rd European Conference on Computer Vision, LNCS 800/801, Stockholm*, pages 161–166. Springer–Verlag, 1994.

[42] S. D. Silvey. *Statistical Inference*. Penguin, Harmondsworth, Middlesex, 1970.

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

[44] 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.

[45] P. H. S. Torr and D. W. Murray. A review of robust methods to estimate the fundamental matrix. To Appear IJCV, 1997.

[46] P. H. S. Torr and A. Zisserman. Performance characterizaton of fundamental matrix estimation under image degradation. *Machine Vision and Applications*, 9:321–333, 1997.

[47] P. H. S. Torr and A Zisserman. Robust parameterization and computation of the trifocal tensor. To Appear in IVC, 1997.

[48] P. H. S. Torr, A. Zisserman, and S. Maybank. Robust detection of degeneracy. In E. Grimson, editor, *Proc. 5th Int'l Conf. on Computer Vision, Boston*, pages 1037–1044. Springer–Verlag, 1995.

[49] P. H. S. Torr, A Zisserman, and S. Maybank. Robust detection of degenerate configurations for the fundamental matrix. Technical Report 2090/96, University of Oxford, 1996.

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

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

[52] R.Y. Tsai and T.S. Huang. Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:13–27, 1984.

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

[54] K. A. Yeomans. *Applied Statistics, Statistics for Social Scientist: Volume Two*. Penguin Books, 1968.