Invariant Fitting of Two View Geometry
or
"In Defiance of the 8 Point Algorithm"

P. H. S. Torr[1] and A. W. Fitzgibbon[2]

[1]Microsoft Research, 7 JJ Thomson Avenue, Cambridge, CB3 0FB,UK
http://research.microsoft.com/∼philtorr
[2] Department of Engineering Science, Oxford University, OX1 3PJ, UK
http://www.robots.ox.ac.uk/∼awf

Abstract

This paper describes the adaptation the Bookstein method for fitting conics to determination of epipolar geometry. The new method has the advantage that it exhibits the improved stability of previous methods for estimating the epipolar geometry, such as the preconditioning method of Hartley, whilst also being invariant to equiform transformations. Within this paper it is proven that there is only one invariant norm to the set of Euclidean transformations of the image, and that this norm gives rise to a quadratic form allowing eigenvector methods to be used to find the essential matrix **E**, the fundamental matrix **F**, or an arbitrary homography **H**. This is a surprising result, as previously it had been thought that there was no more to say on the matter of linear estimation of epipolar geometry. The improved performance is justified by theory and verified by experiments on real images.

1 Introduction

In a now classic paper, Hartley [9] proposed that the much maligned "8-point" algorithm of Longuet-Higgins [11], for computation of the Essential matrix, can be made far more accurate by the application of a simple preconditioning to the image coordinates. This was an important observation as the essential matrix and the Fundamental matrix [5, 7] (its uncalibrated analogue) are key concepts in structure and motion recovery problems, as they encapsulate epipolar geometry. Many authors had criticized the 8-point algorithm and have instead proposed much more computationally intensive iterative algorithms, e.g. [2, 8, 12]. However, as pointed out by Hartley, with suitable preconditioning the 8-point algorithm can be made quite accurate at a fraction of the computational expense of the more sophisticated algorithms. Furthermore, as most iterative algorithms use the 8-point algorithm as an initial starting point, it is of the utmost importance to ensure that this is as close to the global minimum as possible to speed convergence and circumvent local minima.

However preconditioning has some unattractive properties, foremost being that it does not render the estimation process invariant to choice of coordinate system within the image. Being hostage to the choice of coordinate system means that different answers will

BMVC 2003 doi:10.5244/C.17.8

be given, for the same point data, simply by cropping the image, or panning the image plane—a situation which is clearly undesirable. To quote Bookstein [3] in the context of curve and surface fitting: "When abscissa and ordinate represent an arbitrary coordinate system, or arbitrary orientation, placed on an underlying geometric object, such as the outline of the skull, the axes separately have no geometric meaning at all; they are wholly commensurate and may be rotated freely. We *must* get the same geometric result in all cases."

Within this paper we describe a new linear method for estimating the epipolar geometry, that has the pleasing stability properties of the Hartley method and yet is also invariant to equiform transformations[1] of the image planes, indeed the transformation can be different for each of the images. Furthermore the method can be readily extended to other two-image quantities such as homographies. This paper is laid out as follows: Section 2 reviews previous linear methods for determining **F**, pointing out the lack of invariance in these methods. Section 3 proves there is only one invariant normalization for **E**, **F** and homographies that are not affine. Section 3.1 shows that this quadratic normalization can be readily incorporated into the eigenvector solution used previously for linear estimation of **F**, indeed it involves only the solution of a 4×4 eigensystem. Section 3.3 discusses some further benefits of the formalism such as the ability to include arbitrary linear constraints on **F**. Finally in Section 4 results are given which show that the invariant method performs at least as well as previous linear methods, whilst yielding an invariant solution. Section 5 gives some directions for future research.

2 Estimating the Epipolar Geometry

Within this section we shall refer to estimation of the Fundamental matrix, hoping that the reader will take it as given that the method applies also the Essential matrix. It can also be shown that the method can be applied directly to the estimation of homographies **H**.

The epipolar constraint is represented by the Fundamental matrix [5, 7]. This relation applies for general motion and structure with uncalibrated cameras. Consider the movement of a set of point image projections from an object which undergoes a rotation and non-zero translation between views. After the motion, the set of homogeneous image points $\{\underline{\mathbf{x}}_i\}, i = 1, \ldots n$, as viewed in the first image is transformed to the set $\{\underline{\mathbf{x}}'_i\}$ in the second image, with the positions related by

$$\underline{\mathbf{x}}_i'^{\top} \mathbf{F} \underline{\mathbf{x}}_i = 0 \quad \text{where} \quad \mathbf{F} = \begin{bmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \\ f_7 & f_8 & f_9 \end{bmatrix} \tag{1}$$

where $\underline{\mathbf{x}} = (\underline{x}, \underline{y}, \zeta)^{\top}$ is a homogeneous image coordinate and **F** is the fundamental Matrix. Throughout, underlining a symbol $\underline{x}$ indicates the perfect or noise-free quantity, distinguishing it from $x = \underline{x} + \Delta x$, the value corrupted by noise (assumed Gaussian).

The fundamental matrix has 9 elements, but only 7 degrees of freedom. Thus if the fundamental matrix is parametrized by the 3×3 matrix **F** then it is overparametrized: the matrix elements are not independent but must satisfy two additional constraints. This is

[1] Equiform transformations comprise any combination of Euclidean transformations in the image together with change of scale.

because the elements are only defined up to a scale, and the determinant of $\mathbf{F}$ is zero. The first constraint is the scale constraint, which must be imposed to obtain a unique solution. Tsai and Huang [17] propose setting $f_9 = 1$. However, this normalization has the undesirable characteristic that it forbids certain solutions—specifically those with $f_9 = 0$—and produces a biased solution. Most current practitioners propose $\sum f_i^2 = 1$ and solve for $\mathbf{F}$ as an eigenproblem, which is now described.

To reformulate as an eigenproblem, let $\mathbf{f}$ be the 9×1 vector of elements of $\mathbf{F}$, such that $\mathbf{f} = (f_1 \dots f_9)$. The algebraic residual is then

$$\mathbf{f}^\top \mathbf{z}_i = r_i$$

where each

$$\mathbf{z}_i = \left(x'_i x_i \; x'_i y_i \; x'_i \zeta \; y'_i x_i \; y'_i y_i \; y'_i \zeta \; x_i \zeta \; y_i \zeta \; \zeta^2\right)^\top .$$

The eigenproblem will minimize the sum of squares of algebraic residuals. Form $\mathbf{Z}$, the $n \times 9$ design matrix with rows $\mathbf{z}_i$, so that

$$\mathbf{Z} = \begin{bmatrix} x'_1 x_1 & x'_1 y_1 & x'_1 \zeta & y'_1 x_1 & y'_1 y_1 & y'_1 \zeta & x_1 \zeta & y_1 \zeta & \zeta^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x'_n x_n & x'_n y_n & x'_n \zeta & y'_n x_n & y'_n y_n & y'_n \zeta & x_n \zeta & y_n \zeta & \zeta^2 \end{bmatrix}. \tag{2}$$

Let $\mathbf{M} = \mathbf{Z}^\top \mathbf{Z}$ be the 9×9 moment matrix, with eigenvalues in increasing order $\lambda_1 \dots \lambda_9$ and with $\mathbf{u}_1 \dots \mathbf{u}_9$ the corresponding eigenvectors forming an orthonormal system. The $\mathbf{F}$ that minimizes the sum of squares of algebraic residuals, $\sum_i r_i^2$, is given by the eigenvector $\mathbf{u}_1$ corresponding to the minimum eigenvalue λ_1 of the moment matrix; i.e. the estimate $\mathbf{f} = \mathbf{u}_1$ minimizes $\mathbf{f}^\top \mathbf{M} \mathbf{f}$ subject to the constraint, or normalization, $\mathbf{f}^\top \mathbf{J} \mathbf{f} = 1$, where $\mathbf{J} = \mathrm{diag}(1, 1, 1, \dots, 1)$. This normalization chooses a specific solution from the equivalence class of solutions with different scalings. In this case the quantity being minimized is

$$\sum e_i = \sum_i^n \frac{r_i^2}{\mathbf{f}^\top \mathbf{J} \mathbf{f}} = \sum_i^n \frac{\mathbf{x}_i'^\top \mathbf{F} \mathbf{x}_i}{\mathbf{f}^\top \mathbf{J} \mathbf{f}} = \frac{\mathbf{f}^\top \mathbf{M} \mathbf{f}}{\mathbf{f}^\top \mathbf{J} \mathbf{f}}. \tag{3}$$

It is well known that such a linear method can be very ill conditioned especially if the raw data is used with $\zeta = 1$. This has been pointed out in [9, 15]. To overcome the ill conditioning of the linear method Hartley suggests that the data should be preconditioned in the following manner: homographies $\mathbf{H}$ and $\mathbf{H}'$ should be applied to the image coordinates in the first and second image in order to

1. Translate the points in each image so that their centroid is the origin.
2. Scale the points so that their average distance to the origin is $\sqrt{2}$.

In [15] a simpler mechanism is proposed, setting $\zeta = 256$, approximately the centre of the image. Both methods have been found to produce equal improvements over the original unnormalized 8-point algorithm.

Unfortunately, the normalization $\sum f_i^2 = 1$ is not invariant to change of coordinate system. For example, the epipolar geometry induced by the translation $[0, 0, 1]$ means that equation (1) becomes

$$\begin{bmatrix} x' & y' & 1 \end{bmatrix} \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = 0$$

or $y'x - x'y = 0$, with $\sum f_i^2 = 2$. This can be transformed into $y'x - \tilde{x}'y + y' = 0$, with $\sum f_i^2 = 3$, by a simple translation of the coordinate system, just that $\tilde{x}' = x' + 1$, in one of the images. Because this transformation increases the denominator in (3), the second fit may be preferred to the first, even if the residual errors are higher.

To some extent Hartley guards against the choice of coordinate system by choosing a data-dependent coordinate system, however this is not satisfactory in practice as the data may include outliers. Incorrect matches (outliers) which lie along the epipolar lines (corresponding to a wrong depth estimate) cannot be detected and will skew the choice of coordinate system. In many applications the choice of coordinate system is assigned in some arbitrary manner, and it would be desirable to have a linear method that is invariant to choice of coordinate system. To quote Bookstein [3] "These authors must forcibly standardize their data before conic fitting in any of various ad hoc ways. Translation to centre of mass and rotation to principal axes are the most common. This approach seems unsatisfactory. Data must be fit in a manner invariant under the Euclidean group, not arbitrarily constrained with respect to it."

In the next section a method for imposing the scaling invariant to Euclidean transformation of the image coordinates is considered.

3 Invariant Scaling for F

We seek an estimation rule which is *general, simple to compute* and *invariant*. Simplicity suggests we seek a quadratic norm, $\mathbf{f}^\top \mathbf{J}\mathbf{f} = \mathit{constant}$, on the parameters of $\mathbf{F}$ to enforce the scaling constraint as this will lead to a eigenvector solution. Invariance is to be with respect to Euclidean transformations of both image planes (possibly different transformations to different planes) i.e. if the coordinate system is changed in one or both of the images, then the best fitting $\tilde{\mathbf{F}}$ to the transformed points must be exactly the result of the same transformation(s) applied to the best fitting $\mathbf{F}$ of the original points.

Bookstein[3] suggested an invariant norm for conics under Euclidean transformations. It has been observed that the fundamental matrix is like a conic in the four dimensions of the joint image space $\mathbb{R}^4$ [15]. Following Bookstein we seek a parametrization of $\mathbf{F}$ invariant to Euclidean transformations in the image planes (which is a subgroup of the Euclidean transformations in the joint image space $\mathbb{R}^4$). Fortunately the construction of these invariants is a well studied problem [13].

Consider the transformations of the image coordinates $\mathbf{G}$ in image one such that $\mathbf{G}\tilde{\mathbf{x}} = \mathbf{x}$, and image two, $\mathbf{G}'\tilde{\mathbf{x}}' = \mathbf{x}'$, which leads to a transformation on $\mathbf{F}$ such that, $\tilde{\mathbf{F}} = \mathbf{G}'^\top \mathbf{F}\mathbf{G}$ with

$$\begin{array}{ll} \mathbf{G} = \left[\begin{array}{cc} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{array}\right], & \mathbf{G}' = \left[\begin{array}{cc} \mathbf{R}' & \mathbf{t}' \\ \mathbf{0}^\top & 1 \end{array}\right], \\ \mathbf{F} = \left[\begin{array}{cc} \mathbf{A} & \mathbf{b} \\ \mathbf{c}^\top & d \end{array}\right], & \tilde{\mathbf{F}} = \left[\begin{array}{cc} \tilde{\mathbf{A}} & \tilde{\mathbf{b}} \\ \tilde{\mathbf{c}}^\top & \tilde{d} \end{array}\right]. \end{array} \tag{4}$$

Thus it can be seen that

$$\begin{aligned} \tilde{\mathbf{F}} &= \mathbf{G}'^\top \mathbf{F}\mathbf{G} = \left[\begin{array}{cc} \tilde{\mathbf{A}} & \tilde{\mathbf{b}} \\ \tilde{\mathbf{c}}^\top & \tilde{d} \end{array}\right] \\ &= \left[\begin{array}{cc} \mathbf{R}'^\top \mathbf{A}\mathbf{R} & \mathbf{R}'^\top \mathbf{A}\mathbf{t} + \mathbf{R}'^\top \mathbf{b} \\ \mathbf{t}'^\top \mathbf{A}\mathbf{R} + \mathbf{c}^\top \mathbf{R} & \mathbf{t}'^\top \mathbf{A}\mathbf{t} + \mathbf{t}'^\top \mathbf{b} + \mathbf{c}^\top \mathbf{t} + d \end{array}\right]. \end{aligned}$$

From this it can be seen that the norm cannot be any combination of f_3, f_6, f_7, f_8, f_9 as these can be transformed to arbitrary values by translations of the image coordinates. In the special case where both $\mathbf{t}$ and $\mathbf{t}'$ are known to be zero, the normalization $\sum_{i=1}^{i=9} f_i^2 = 1$ (amongst others) is invariant to rotations of the image plane. Discounting this non-generic case, this leaves the elements of the upper left 2×2 submatrix of $\mathbf{F}$ to define the norm. Because $\mathbf{R}$ and $\mathbf{R}'$ are rotation matrices it can be immediately seen that

$$\begin{aligned} \det(\tilde{\mathbf{A}}) &= \det(\mathbf{R}'^{\top}\mathbf{A}\mathbf{R}) = \det(\mathbf{A}) \\ \|\tilde{\mathbf{A}}\|_F &= \|\mathbf{R}'^{\top}\mathbf{A}\mathbf{R}\|_F = \|\mathbf{A}\|_F \end{aligned}$$

where $\|\cdot\|_F$ denotes the Frobenius norm of the matrix. Thus we have the choice of the following invariants, the determinant $\det(\mathbf{A}) = (f_1 f_5 - f_2 f_4)$, and the Frobenius norm $\|\mathbf{A}\|_F = (f_1^2 + f_2^2 + f_4^2 + f_5^2)^{\frac{1}{2}}$. How many invariants can there be? Referring to [13] the counting argument states: "suppose there is a configuration space $\mathcal{S}$, on which a group G acts, then the number of functionally independent primitive scalar invariants is greater than or equal to $\dim\mathcal{S} - \dim G$". In this case $\dim\mathbf{A} = 4$ and $\dim(\mathbf{R},\mathbf{R}') = 2$, thus we would expect at least two invariants.[2]

Which of these norms is most appropriate? In order to deduce this another desideratum is introduced: that the norm is positive definite. This is desirable because epipolar geometries for whose $\mathbf{F}$ the norm is zero can never be fitted at all, even if the data lie exactly upon them. Therefore we must bid goodbye to the determinant norm $\det(\mathbf{A})$, which excludes all $\mathbf{F}$ for which $\det(\mathbf{A}) = 0$. The square of the Frobenius norm $\|\mathbf{A}\|_F^2 = (f_1^2 + f_2^2 + f_4^2 + f_5^2)$ does not exclude general $\mathbf{F}$, rather it will fit all $\mathbf{F}$ except for data for which a linear or affine fundamental matrix $\mathbf{F}_A$ [13] is more suited;

$$\underline{\mathbf{x}}_i'^{\top}\mathbf{F}_A\underline{\mathbf{x}}_i = 0 \quad \text{where} \quad \mathbf{F}_A = \begin{bmatrix} 0 & 0 & g_1 \\ 0 & 0 & g_2 \\ g_3 & g_4 & g_5 \end{bmatrix}. \tag{5}$$

Whether or not $\mathbf{F}_A$ is the more appropriate model can be determined by model selection methods [16]. If $\mathbf{F}_A$ is the better model then an exact eigenvector solution [14] exists for $\mathbf{F}_A$ that minimizes reprojection error which should always be used rather than a more general algorithm for fitting $\mathbf{F}$. Thus we propose to minimize $\mathbf{f}^{\top}\mathbf{M}\mathbf{f}$ subject to $\mathbf{f}^{\top}\mathbf{J}\mathbf{f} =$ constant, where $\mathbf{J} = \mathrm{diag}(1,1,0,1,1,0,0,0,0)$, is the square of Frobenius normalization.

The square of the Frobenius norm $\|\mathbf{A}\|_F^2 = (f_1^2 + f_2^2 + f_4^2 + f_5^2)$ is also invariant to choice of scale. Without loss of generality consider only the change of scaling in one of the images. If the coordinates in one image are rescaled by k, $(x,y) \rightarrow (\tilde{x},\tilde{y}) = k(x,y)$, let $\mathbf{f}_k$ be the vector of coefficients of the best fitting $\mathbf{F}$ by this norm. The rescaling replaces the moment matrix $\mathbf{M}$ by $\mathbf{D}_k\mathbf{M}\mathbf{D}_k$ where $\mathbf{D}_k = \mathrm{diag}(k,k,k,k,k,k,1,1,1)$. Thus in the new coordinate system the minimization is of $\mathbf{f}^{\top}\mathbf{M}_k\mathbf{f}$ subject to $\mathbf{f}^{\top}\mathbf{J}\mathbf{f} =$ constant, where $\mathbf{J} = \mathrm{diag}(1,1,0,1,1,0,0,0,0)$ (because $\mathbf{D}_k\mathbf{J}\mathbf{D}_k = k^2\mathbf{J}$). Since $\mathbf{D}_k$ is not singular the extremum is given by $\mathbf{D}_k^{-1}\mathbf{f}$, where $\mathbf{f}$ is the extremum before rescaling. But $\mathbf{D}_k^{-1}\mathbf{f}$ is simply the transformed version of $\mathbf{f}$. Hence the method presented is invariant under equiform transformations (Euclidean and scaling transformation in the images), and consequently choice of the third projective coordinate ζ.

[2] If only one rotation was applied to both images i.e. we knew the common orientation of the two images, then we could expect another invariant, which would correspond to $\mathrm{trace}(\mathbf{A})$

3.1 Imposition of the quadratic constraint $f_1^2 + f_2^2 + f_4^2 + f_5^2 = K$

We wish to minimize $\mathbf{f}^\top \mathbf{M}_k \mathbf{f}$ subject to $\mathbf{f}^\top \mathbf{J}\mathbf{f} = \text{constant}$, where

$$\mathbf{J} = \text{diag}(1,1,0,1,1,0,0,0,0). \tag{6}$$

There are two eigenvector methods for conducting this minimization. The first is somewhat easier to implement (especially in MATLAB) and involves solving the generalized eigenvector problem:

$$\mathbf{J}\mathbf{f} - \lambda \mathbf{M}\mathbf{f} = 0. \tag{7}$$

The second is like that proposed by Bookstein for the conic fitting problem, and is faster and more stable: First partition $\mathbf{f}$ into two components, $\mathbf{f}_1 = (f_1, f_2, f_4, f_5)$ comprising the four elements of $\mathbf{A}$, the second $\mathbf{f}_2$, comprising the other five elements. Let $\mathbf{M}$ be partitioned correspondingly:

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{12}^\top & \mathbf{M}_{22} \end{bmatrix}$$

so that

$$\mathbf{f}^\top \mathbf{M}\mathbf{f} = \mathbf{f}_1^\top \mathbf{M}_{11}\mathbf{f}_1 + 2\mathbf{f}_1^\top \mathbf{M}_{12}\mathbf{f}_2 + \mathbf{f}_2^\top \mathbf{M}_{22}\mathbf{f}_2,$$

as $\mathbf{M}$ and its partitions are all symmetric. We must minimize this subject to $\mathbf{f}_1^\top \mathbf{J}_{11}\mathbf{f}_1 =$ constant, where $\mathbf{J}_{11} = \text{diag}(1,1,1,1) = \mathbf{I}$. For any fixed $\mathbf{f}_1$, $\mathbf{f}^\top \mathbf{M}\mathbf{f}$ is minimal when

$$\frac{\partial \mathbf{f}^\top \mathbf{M}\mathbf{f}}{\partial \mathbf{f}_2} = 2\mathbf{M}_{12}^\top \mathbf{f}_1 + 2\mathbf{M}_{22}\mathbf{f}_2 = 0 \tag{8}$$

which implies

$$\mathbf{f}_2 = -\mathbf{M}_{22}^{-1}\mathbf{M}_{12}^\top \mathbf{f}_1 \tag{9}$$

Then

$$\mathbf{f}^\top \mathbf{M}\mathbf{f} = \mathbf{f}_1^\top (\mathbf{M}_{11} - \mathbf{M}_{12}\mathbf{M}_{22}^{-1}\mathbf{M}_{12}^\top)\mathbf{f}_1 = \mathbf{f}_1^\top \mathbf{Q}\mathbf{f}_1. \tag{10}$$

To minimize this for $\mathbf{f}_1^\top \mathbf{J}_{11}\mathbf{f}_1 =$ constant, let λ be a Lagrangian multiplier for the constraint. Then we must set the derivative with respect to $\mathbf{f}_1$ of $\mathbf{f}_1^\top \mathbf{Q}\mathbf{f}_1 - \lambda \mathbf{f}_1^\top \mathbf{f}_1$. This yields

$$\mathbf{Q}\mathbf{f}_1 = \lambda \mathbf{f}_1, \tag{11}$$

thus $\mathbf{f}_1$ may be recovered as the eigenvector solution of (11), and then $\mathbf{f}_2$ is obtained from (9).

3.2 Imposing the cubic constraint $\det(\mathbf{F}) = 0$

The second constraint which $\mathbf{F}$ should satisfy is a cubic polynomial in the matrix elements imposing $\det(\mathbf{F}) = 0$. If it is not imposed, then the epipolar lines do not all intersect in a single epipole. Assume that we have an estimate of the fundamental matrix, $\hat{\mathbf{F}}$. Typically, when performing only linear estimation to obtain $\hat{\mathbf{F}}$ this constraint is imposed by projecting the linear solution onto the space of Fundamental matrices such that $\det(\mathbf{F}) = 0$, such that the Frobenius norm $||\hat{\mathbf{F}} - \mathbf{F}||_F$ is a minimum.

A problem is that smaller elements will have a relatively greater perturbation in relation to their size. Thus the method of [9] is adopted. This is not entirely satisfactory as ideally we would like to choose the $\hat{\mathbf{f}}$ which minimizes some Mahalanobis distance $(\hat{\mathbf{f}} - \mathbf{f})^\top \mathbf{M}_f^{-1}(\hat{\mathbf{f}} - \mathbf{f})$, taking into account the covariance of $\mathbf{F}$. Work on this is in progress.

3.3 Imposing Linear Constraints on F

The computation remains tractable when we place arbitrary linear constraints on the parameters. If these linear constraints satisfy $\mathbf{Lf} = 0$, then $\mathbf{f}$ is the eigenvector corresponding to the largest eigenvalue of the system

$$(\mathbf{I} - \mathbf{L}(\mathbf{L}^\top \mathbf{M}^{-1}\mathbf{L})^{-}\mathbf{L}^\top)\mathbf{Jf} - \lambda\mathbf{Mf} = 0. \tag{12}$$

where $\mathbf{A}^-$ denotes the generalized inverse of an arbitrary square matrix $\mathbf{A}$. A more efficient solution to this generalized eigensystem is given by Golub and Underwood [6], which reduces the problem to a smaller, symmetric system.

Setting linear constraints erodes the available degrees of freedom in various ways, some possibly useful e.g. (1) $f_1 = 0,\ f_5 = 0$: the two cameras are mounted on a lateral stereo rig [4], with the cameras free only to vary their angle of d_i or e_i; (2) skew symmetry $f_2 = -f_4,\ f_3 = -f_7$ and $f_6 = -f_8$, this occurs under a pure translation, generally it would be preferable to fit a two parameter model directly. If the camera has both epipoles in the centre of the image (forward translation and cylcorotation) then $f_3, f_6, f_7, f_8 = 0$.

4 Experiments

Within this section tests on synthetic and real data are described. For synthetic data, where the ground truth is known, an empirical measure of the goodness of fit is achieved by calculating the reprojection error of the *actual* noise free projections of the synthetic world points to $\mathbf{F}$ provided by each estimator. Traditionally the goodness of fit has been assessed by seeing how well the parameters fit the *observed* data. But we point out that this is the wrong criterion as the aim is to find the set of parameters that best fit the (unknown) *true data*. The parameters of the fundamental matrix themselves are not of primary importance, rather it is the structure of the corresponding epipolar geometry. Consequently it makes little sense to compare two solutions by directly comparing corresponding parameters in their fundamental matrices; one must rather compare the the difference in the associated epipolar geometry weighted by the density of the given matching points. The inadequacy of using the fit to the *observed* data to assess efficiency, in the presence of outliers, will be demonstrated in the results section. This error metric is the first order approximation of the reprojection error of the noise free points to $\mathbf{F}$: $\mathcal{E}_1 = \sum_{i=1}^{n}(\underline{w}_i\mathbf{f}^\top\underline{\mathbf{z}}_i)^2$. The second statistic $\mathcal{E}_2$ is the average distance in pixels from the true epipole in each image to that yielded by the estimate of $\mathbf{F}$.

Data $\underline{\mathbf{X}}$ are randomly generated in the region of $\mathbb{R}^3$ visible to two positions of a synthetic camera having intrinsic coordinates equivalent to an aspect ratio of 1.5, an optic centre at the image centre $(256, 256)$, and a focal length of $f = 703$ (notionally pixels), giving a field of view of 40°, and giving $0 \leq x, y \leq 512$. These values were chosen to be similar to the camera used for capturing real imagery. The projection of a point $\underline{\mathbf{X}}$ in the first position is $\underline{\mathbf{x}} = \mathbf{C}[\mathbf{I}|\mathbf{0}]\underline{\mathbf{X}}$ and in the second is $\underline{\mathbf{x}}' = \mathbf{C}[\mathbf{R}|\mathbf{t}]\underline{\mathbf{X}}$ where the camera makes a rotation $[\mathbf{R}]$ and translation $\mathbf{t}$. The motion is random and different in each test. In order to simulate the effects of the search window commonly employed in feature matchers, and to limit the range of depths in 3D, correspondences were accepted only if the disparity lay between $4 \leq \delta \leq 30$ pixels. (Some notion of the limits on depth Z can be obtained for pure translation as $|\mathbf{t}|f/\delta_{\text{max}} \leq Z \leq |\mathbf{t}|f/\delta_{\text{min}}$.)

(a) (b) (c) (d)

Sequence	Length	Bundle	8-point ($\|f\| = 1$)	Invariant ($\|F_{22}\|_F^2 = 1$)
(a)	11	582	583	583
(b)	100	752	751	751
(c)	100	297	297	297
(d)	108	695	657	695

Figure 1: (Top) Example images from each of four test sequences. (Bottom) The table shows tracking performance for each algorithm as average number of inlying tracks over the sequence. “Length” is the number of image pairs tested. The general conclusion is that the invariant algorithm and the 8-point algorithm have similar average perfomance, but that the 8-point algorithm depends on an ad-hoc scaling of the data, which the invariant algorithm avoids. Also, for certain sequences (such as (d) here), the 8-point algorithm can produce biased fits, resulting in significant loss of tracks.

Figure 1 shows some images from real sequences which were used to test the new algorithm. In this test, as no ground truth is available, an estimate of success will be the number of matches recovered which are consistent with the final recovered epipolar geometry. Consistency is measured using the exact reprojection error [10]. The procedure follows that of Beardsley [1]: putative pairwise point matches are found by cross-correlating image patches centered on Harris corners, yielding a few hundred matches. These are pruned using RANSAC, and refined using the algorithm under test. Results are shown for (1) the baseline bundle adjustment algorithm, (2) the 8-point algorithm and (3) the new invariant algorithm. The general conclusion is that the new algorithm and the 8-point algorithm have similar performance, but that the 8-point algorithm occasionally fails to fit the track data, resulting in significant loss of tracks. In addition, the comparison with bundle adjustment shows that the linear algorithms are competitive in terms of number of tracks maintained. This is an interesting result as they are orders of magnitude faster.

5 Discussion and Conclusion

There are several lines of work which we are following. First the analysis can be trivially extended to homographies and various other image relations, allowing invariant estimation of these quantities. For homographies $\mathbf{x}' = \mathbf{Hx}$, following the same notation as before, $\tilde{\mathbf{H}} = \mathbf{G}'^{-1}\mathbf{HG}$, yielding the form of the norm as for $\mathbf{F}$, such that the Frobenius norm of the top 2×2 of $\mathbf{H}$ should be held constant. Second the result may have some bearing on nonlinear methods for estimating $\mathbf{F}$, for instance in the past it has been suggested minimizing $\mathbf{F}$ with the determinant of the upper 2×2 held constant, it can now be seen that it makes more sense to hold the Frobenius norm of the top 2×2 constant instead. The linear estimation of **F** is a two part process, estimating a non-rank 2 **F** and then constraining it to be rank 2, in depth consideration of the second part is beyond the scope of this present paper.

Within this paper the important problem of epipolar geometry estimation has been revisited. It has been thought that there is nothing more to be said on the subject and yet within this paper we have demonstrated a linear algorithm, solvable by a simple 5×5 eigenvector method (smaller than for the 8 point algorithm which solves an 8×8 problem!), which has the desirable property of invariance to shifts in the image coordinate system. The algorithm is compared to the state of the art algorithms, specifically those involving preconditioning and has been found to be as good as those algorithms and to yield an invariant fit. Its adoption is recommended whenever the practitioner intends to use linear methods to fit **H**, **E** or **F**, and the situation is not affine. Code for the method is available at `http://www.robots.ox.ac.uk/~awf/bmvc03invariance`.

References

[1] P. A. Beardsley, P. H. S. Torr, and A. Zisserman. 3D model acquisition from extended image sequences. In *Proceedings of the 4th European Conference on Computer Vision, LNCS 1065, Cambridge*, pages 683–695, 1996.

[2] P. A. Beardsley, A. Zisserman, and D. W. Murray. Sequential update of projective and affine structure from motion. *International Journal of Computer Vision*, 23(3):235–259, 1997.

[3] F. Bookstein. Fitting conic sections to scattered data. *Computer Vision, Graphics and Image Processing*, 9:56–71, 1979.

[4] M. Brooks, L. De Agapito, D. Huynh, and L. Baumela. Towards robust metric reconstruction via a dynamic uncalibrated stereo head. *Image and Vision Computing*, 16(14):989–1002, 1998.

[5] O. D. Faugeras and M. Hebert. The representation, recognition, and locating of 3-D objects. *International Journal of Robotics Research*, 5(3):27–52, 1992.

[6] G.H. Golub and R. Underwood. Stationary values of the ratio of quadratic forms subject to linear constraints. *Z. Angew. Math. Phys.*, 21:318–326, 1970.

[7] R. I. Hartley. Estimation of relative camera positions for uncalibrated cameras. In *Proceedings of the European Conference on Computer Vision*, LNCS 588, pages 579–587. Springer-Verlag, 1992.

[8] R. I. Hartley. Cheirality invariants. In *Proc. D.A.R.P.A. Image Understanding Workshop*, pages 745–753, 1993.

[9] R. I. Hartley. A linear method for reconstruction from lines and points. In *Proceedings of the International Conference on Computer Vision*, pages 882–887, 1995.

[10] R. I. Hartley and P. Sturm. Triangulation. *Computer Vision and Image Understanding*, 68(2):146–157, November 1997.

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

[12] Q.-T. Luong and T. Viéville. Canonical representations for the geometries of multiple projective views. *Computer Vision and Image Understanding*, 64(2):193–229, September 1996.

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

[14] L. S. Shapiro, A. Zisserman, and M. Brady. 3D motion recovery via affine epipolar geometry. *International Journal of Computer Vision*, 16(2):147–182, 1995.

[15] P. H. S. Torr. *Motion segmentation and outlier detection*. PhD thesis, Dept. of Engineering Science, University of Oxford, 1995.

[16] P. H. S. Torr. An assessment of information criteria for motion model selection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Puerto Rico*, June 1997.

[17] 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.