

Minimizing Algebraic Error in Geometric Estimation Problems

Richard I. Hartley
G.E. Corporate Research and Development
PO Box 8, Schenectady, NY 12309
Email : hartley@crd.ge.com

Abstract

This paper gives a widely applicable technique for solving many of the parameter estimation problems encountered in geometric computer vision. A commonly used approach is to minimize an algebraic error function instead of a possibly preferable geometric error function. It is claimed in this paper that minimizing algebraic error will usually give excellent results, and in fact the main problem with most algorithms minimizing algebraic distance is that they do not take account of mathematical constraints that should be imposed on the quantity being estimated. This paper gives an efficient method of minimizing algebraic distance while taking account of the constraints. This provides new algorithms for the problems of resectioning a pin-hole camera, computing the fundamental matrix, and computing the tri-focal tensor. Evaluation results are given for the resectioning and tri-focal tensor estimation algorithms.

1 Introduction

For many problems related to camera calibration and scene reconstruction, linear algorithms are known for solving for the entity required. In the sort of problem that will be addressed in this paper, a set of data (such as point correspondences) is used to construct a set of linear equations, and solution of these equations, usually in the least-squares sense, provides an estimate of the entity being computed. As examples of such problems we have :

1. The DLT algorithm for computing a camera matrix given a set of points in space, and corresponding points in the image. Provided at least 6 correspondences are given (more precisely $5\frac{1}{2}$ correspondences), one can solve for the camera matrix.
2. Computation of the Fundamental Matrix. From 8 point correspondences $\mathbf{u}_i \leftrightarrow \mathbf{u}'_i$ between two images one can construct the fundamental matrix using equations $\mathbf{u}'_i{}^\top \mathbf{F} \mathbf{u}_i = 0$.

3. Computation of the trifocal tensor given a set of feature correspondences across three views.

These linear algorithms have been found to give poor quality results on occasions and much research has been expended in seeking more reliable, but complex methods. A previous paper ([1]) showed that normalization of the data in a systematic manner will improve the results immeasurably, and such normalization must be routinely carried out. Nevertheless, a common criticism of such linear algorithms is that they “do not minimize the right thing,” and we really should be minimizing “geometric error,” which is related to the actual means of error occurrence. Although this criticism is correct, it is the thesis of this paper that it really does not matter very significantly whether we minimize algebraic or geometric error. Usually, one may be content with minimizing algebraic error, as long as one enforces the constraints. This leads to simpler and more efficient algorithms than are possible minimizing geometric error.

In these three examples, and many others, the linear algorithm will lead to a solution that does not satisfy certain constraints that the estimated quantity must satisfy. In the cases considered here, the constraints are

1. The *skew* parameter of a camera matrix estimated using the DLT method will not generally be zero. This constraint, meaning the pixels are rectangular, should be enforced in cases where it is known to hold.
2. The fundamental matrix must satisfy a constraint $\det \mathbf{F} = 0$.
3. The trifocal tensor must satisfy 8 non-linear constraints. The form of these constraints is not easily determined, but it is essential to constrain the tensor to correspond to a valid set of camera matrices.

These constraints are not in general linear constraints, and it will be necessary to resort to iterative

techniques to enforce them. Since iterative techniques are slow and potentially unstable, it is important to use them sparingly. Further, the smaller the dimension of the minimization problem, the faster and generally more stable the solution will be. In this paper an iterative algorithm is used to solve the three problems posed above. In each case the algorithms are based on a common technique of data reduction, whereby the input data is condensed into a *reduced measurement matrix*. The size of the iteration problem is then independent of the size of the input set. In the case of estimation of the fundamental matrix, only three homogeneous parameters are used to parametrize the minimization problem, whereas for the trifocal tensor, just six parameters are used.

The problem of camera calibration solved using the DLT algorithm will be treated first. It will be used to illustrate the techniques that apply to the other problems. Details of the application of this technique to the computation of the fundamental matrix were given in [4], and are not repeated here. New evaluation results for trifocal tensor estimation are given in this paper.

2 Computing the Camera Matrix

We consider a set of point correspondences $\mathbf{x}_i \leftrightarrow \mathbf{u}_i$ between 3D points \mathbf{x}_i and image points \mathbf{u}_i . Our problem is to compute a 3×4 matrix \mathbf{P} such that $\mathbf{P}\mathbf{x}_i = \mathbf{u}_i$ for each i .

2.1 The Direct Linear Transformation (DLT) algorithm

We begin with a simple linear algorithm for determining \mathbf{P} given a set of 3D to 2D point correspondences, $\mathbf{x}_i \leftrightarrow \mathbf{u}_i$. The correspondence is given by the equation $\mathbf{u}_i = \mathbf{P}\mathbf{x}_i$. Note that this is an equation involving homogeneous vectors, thus \mathbf{u}_i and $\mathbf{P}\mathbf{x}_i$ may differ by a non-zero scale factor. One may, however write the equation in terms of the vector cross product as $\mathbf{u}_i \times \mathbf{P}\mathbf{x}_i = \mathbf{0}$.

If the j -th row of the matrix \mathbf{P} is denoted by \mathbf{p}^j , then we may write $\mathbf{P}\mathbf{x}_i = (\mathbf{p}^{1\top}\mathbf{x}_i, \mathbf{p}^{2\top}\mathbf{x}_i, \mathbf{p}^{3\top}\mathbf{x}_i)^\top$. Writing $\mathbf{u}_i = (u_i, v_i, w_i)^\top$, the cross product may be given explicitly as

$$\mathbf{u}_i \times \mathbf{P}\mathbf{x}_i = \begin{pmatrix} v_i\mathbf{p}^{3\top}\mathbf{x}_i - w_i\mathbf{p}^{2\top}\mathbf{x}_i \\ w_i\mathbf{p}^{1\top}\mathbf{x}_i - u_i\mathbf{p}^{3\top}\mathbf{x}_i \\ u_i\mathbf{p}^{2\top}\mathbf{x}_i - v_i\mathbf{p}^{1\top}\mathbf{x}_i \end{pmatrix}.$$

Since $\mathbf{p}^{j\top}\mathbf{x}_i = \mathbf{x}_i^\top\mathbf{p}^j$ for $j = 1, \dots, 3$, this gives a set of three equations, in the entries of \mathbf{P} , which may be written in the form

$$\begin{bmatrix} \mathbf{0}^\top & -w_i\mathbf{x}_i^\top & v_i\mathbf{x}_i^\top \\ w_i\mathbf{x}_i^\top & \mathbf{0}^\top & -u_i\mathbf{x}_i^\top \\ -v_i\mathbf{x}_i^\top & u_i\mathbf{x}_i^\top & \mathbf{0}^\top \end{bmatrix} \begin{pmatrix} \mathbf{p}^1 \\ \mathbf{p}^2 \\ \mathbf{p}^3 \end{pmatrix} = \mathbf{0}. \quad (1)$$

Note that $(\mathbf{p}^1, \mathbf{p}^2, \mathbf{p}^3)^\top$ which appears in (1) is a 12-vector made up of the entries of the matrix \mathbf{P} . Although there are three equations, only two of them are linearly independent. Thus each point correspondence gives two equations in the entries of \mathbf{P} . One may choose to omit the third equation, or else include all three equations, which may sometimes give a better conditioned set of equations. In future, we will assume that only the first two equations are used, namely

$$\begin{bmatrix} \mathbf{0}^\top & -w_i\mathbf{x}_i^\top & v_i\mathbf{x}_i^\top \\ w_i\mathbf{x}_i^\top & \mathbf{0}^\top & -u_i\mathbf{x}_i^\top \end{bmatrix} \begin{pmatrix} \mathbf{p}^1 \\ \mathbf{p}^2 \\ \mathbf{p}^3 \end{pmatrix} = \mathbf{0}. \quad (2)$$

Solving the Equations. The equations (2) may be denoted by $\mathbf{M}_i\mathbf{p} = \mathbf{0}$, where the vector \mathbf{p} is a 12-vector, corresponding to the 12 entries of \mathbf{P} . The set of all equations derived from several point correspondences may be written $\mathbf{M}\mathbf{p} = \mathbf{0}$ where \mathbf{M} is the matrix of equation coefficients. This matrix \mathbf{M} will be called the *measurement matrix*. The obvious solution $\mathbf{p} = \mathbf{0}$ is of no interest to us, so we seek a non-zero solution.

2.2 Data Normalization

One of the most important things to do in implementing an algorithm of this sort is to prenormalize the data. This type of data normalization was discussed in the paper [1]. Without this normalization, all these algorithms are guaranteed to perform extremely poorly. The appropriate transformation is to translate all data points so that their centroid is at the origin. Then the data should be scaled so that the average distance of any data point from the origin is equal to $\sqrt{2}$ for image points and $\sqrt{3}$ for 3D points. The algorithms are then carried out with the normalized data, and final transformations are applied to the result to compensate for the normalizing transforms.

2.3 Algebraic Error

In the presence of noise, one can not expect to obtain an exact solution to an overconstrained set of equations of the form $\mathbf{M}\mathbf{p} = \mathbf{0}$ such as those that arise in the DLT method.

The DLT algorithm instead finds the unit-norm vector \mathbf{p} that minimizes $\|\mathbf{M}\mathbf{p}\|$. The vector $\boldsymbol{\epsilon} = \mathbf{M}\mathbf{p}$ is the error vector and it is this error vector that is minimized. The solution is the unit singular vector corresponding to the smallest singular value of \mathbf{M} .

Define a vector $(\hat{u}_i, \hat{v}_i, \hat{w}_i)^\top = \hat{\mathbf{u}}_i = \mathbf{P}\mathbf{x}_i$. Using this notation, we may write

$$\mathbf{M}_i\mathbf{p} = \boldsymbol{\epsilon}_i = \begin{pmatrix} v_i\hat{w}_i - w_i\hat{v}_i \\ w_i\hat{u}_i - u_i\hat{w}_i \end{pmatrix} = \mathbf{0}. \quad (3)$$

This vector is the *algebraic error vector* associated with the point correspondence $\mathbf{u}_i \leftrightarrow \mathbf{x}_i$ and the camera mapping \mathbf{P} . Thus,

$$d_{\text{alg}}(\mathbf{u}_i, \hat{\mathbf{u}}_i)^2 = (v_i \hat{w}_i - w_i \hat{v}_i)^2 + (w_i \hat{u}_i - u_i \hat{w}_i)^2. \quad (4)$$

Given several point correspondences, the quantity $\boldsymbol{\epsilon} = \mathbf{M}\mathbf{p}$ is the algebraic error vector for the complete set, and one sees that

$$\sum_i d_{\text{alg}}(\mathbf{u}_i, \hat{\mathbf{u}}_i)^2 = \|\mathbf{M}\mathbf{p}\|^2 = \|\boldsymbol{\epsilon}\|^2 \quad (5)$$

The main lesson that we want to keep from this discussion is :

Proposition 2.1. *Given any set of 3D to image correspondences $\mathbf{u}_i \leftrightarrow \mathbf{x}_i$, let \mathbf{M} be the measurement matrix as in (2). For any camera matrix \mathbf{P} the vector $\mathbf{M}\mathbf{p}$ is the algebraic error vector, where \mathbf{p} is the vector of entries of \mathbf{P} .*

2.4 Geometric Distance

Under the assumption that measurement error is confined to image measurements, and an assumption of a Gaussian error model for the measurement of 2D image coordinates, the optimal estimate for the camera matrix \mathbf{P} is the one that minimizes the error function

$$\sum_i d(\mathbf{u}_i, \hat{\mathbf{u}}_i)^2 \quad (6)$$

where $d(\cdot, \cdot)$ represents Euclidean distance in the image. The quantity $d(\mathbf{u}_i, \hat{\mathbf{u}}_i)$ is known as the *geometric distance* between \mathbf{u}_i and $\hat{\mathbf{u}}_i$. Thus the error to be minimized is the sum of squares of geometric distances between measured and projected points.

For points $\mathbf{u}_i = (u_i, v_i, w_i)^\top$ and $\hat{\mathbf{u}}_i = (\hat{u}_i, \hat{v}_i, \hat{w}_i)^\top$, the geometric distance is

$$\begin{aligned} d(\mathbf{u}_i, \hat{\mathbf{u}}_i) &= \left((u_i/w_i - \hat{u}_i/\hat{w}_i)^2 + (v_i/w_i - \hat{v}_i/\hat{w}_i)^2 \right)^{1/2} \\ &= d_{\text{alg}}(\mathbf{u}_i, \hat{\mathbf{u}}_i)/w_i \hat{w}_i \end{aligned} \quad (7)$$

Thus, geometric distance is related to, but not quite the same as algebraic distance.

2.5 The Reduced Measurement Matrix

Let $\mathbf{u}_i \leftrightarrow \mathbf{x}_i$ be a set of correspondences, and let \mathbf{M} be the corresponding measurement matrix. Let \mathbf{P} be any camera matrix, and let \mathbf{p} be the vector containing its entries. The algebraic error vector corresponding to \mathbf{P} is $\mathbf{M}\mathbf{p}$, and its norm satisfies $\|\mathbf{M}\mathbf{p}\|^2 = \mathbf{p}^\top \mathbf{M}^\top \mathbf{M} \mathbf{p}$.

In general, the matrix \mathbf{M} may have a very large number of rows. It is possible to replace \mathbf{M} by a square matrix $\hat{\mathbf{M}}$ such that $\|\mathbf{M}\mathbf{p}\| = \|\hat{\mathbf{M}}\mathbf{p}\|$ for any vector \mathbf{p} .

Such a matrix $\hat{\mathbf{M}}$ is called a *reduced measurement matrix*. One way to do this is using the Singular Value Decomposition (SVD). Let $\mathbf{M} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$ be the SVD of \mathbf{M} , and define $\hat{\mathbf{M}} = \mathbf{D}\mathbf{V}^\top$. Then

$$\mathbf{M}^\top \mathbf{M} = (\mathbf{V}\mathbf{D}\mathbf{U}^\top)(\mathbf{U}\mathbf{D}\mathbf{V}^\top) = (\mathbf{V}\mathbf{D})(\mathbf{D}\mathbf{V}^\top) = \hat{\mathbf{M}}^\top \hat{\mathbf{M}}$$

as required. Another (faster) way of obtaining $\hat{\mathbf{M}}$ is to use the QR decomposition $\mathbf{M} = \mathbf{Q}\hat{\mathbf{M}}$, where \mathbf{Q} has orthogonal columns and $\hat{\mathbf{M}}$ is upper-triangular and square. This shows the following result.

Theorem 2.2. *Let $\mathbf{u}_i \leftrightarrow \mathbf{x}_i$ be a set of n world to image correspondences. Let \mathbf{M} be the measurement matrix derived from the point correspondences. Let $\hat{\mathbf{M}}$ be a reduced measurement matrix. Then, for any 3D to 2D projective transform \mathbf{P} and corresponding 3-vector \mathbf{p} , one has*

$$\sum_i d_{\text{alg}}(\mathbf{u}_i, \mathbf{P}\mathbf{x}_i)^2 = \|\hat{\mathbf{M}}\mathbf{p}\|^2$$

In this way, all the information we need to keep about the set of matched points $\mathbf{u}_i \leftrightarrow \mathbf{x}_i$ is contained in the single 12×12 matrix $\hat{\mathbf{M}}$. If we wish to minimize algebraic error as \mathbf{P} varies over some restricted set of transforms, then this is equivalent to minimizing the norm of the 12-vector $\|\hat{\mathbf{M}}\mathbf{p}\|$.

2.6 Restricted Camera Mappings

The camera mapping expressed by a general 3D projective transformation is in some respects too general. A non-singular 3×4 matrix \mathbf{P} with center at a finite point may be decomposed as $\mathbf{P} = \mathbf{K}[\mathbf{R} \mid -\mathbf{R}\mathbf{t}]$ where \mathbf{R} is a 3×3 rotation matrix and

$$\mathbf{K} = \begin{bmatrix} \alpha_u & s & u_0 \\ & \alpha_v & v_0 \\ & & 1 \end{bmatrix}. \quad (8)$$

The non-zero entries of \mathbf{K} are geometrically meaningful quantities, the internal calibration parameters of \mathbf{P} . A common assumption is that $s = 0$, while for a true pinhole camera, $\alpha_u = \alpha_v$.

Given a set of world to image correspondences, one may wish to find a matrix \mathbf{P} that minimizes algebraic error, subject to a set of constraints on \mathbf{P} . Usually, this will require an iterative solution. For instance, suppose we wish to enforce the constraints $s = 0$ and $\alpha_u = \alpha_v$. One can parametrize the camera matrix using the remaining 9 parameters (p_u, p_v, α plus 6 parameters representing the orientation \mathbf{R} and location \mathbf{t} of the camera). Let this set of parameters be denoted collectively by \mathbf{q} . Then, one has a map $\mathbf{p} = g(\mathbf{q})$, where \mathbf{p} is as before the vector of entries of the matrix \mathbf{P} . According to Theorem 2.2, minimizing algebraic

error over all point matches is equivalent to minimizing $\|Mg(\mathbf{q})\|$. Note that the mapping $\mathbf{q} \mapsto Mg(\mathbf{q})$ is a mapping from R^9 to R^{12} . This is a simple parameter-minimization problem that may be solved using the Levenberg-Marquardt method. The important point to note is the following :

Given a set of n world-to-image correspondences, $\mathbf{x}_i \leftrightarrow \mathbf{u}_i$, the problem of finding a constrained camera matrix P that minimizes the sum of algebraic distances $\sum_i d_{\text{alg}}(\mathbf{u}_i, P\mathbf{x}_i)^2$ reduces to the minimization of a function $R^9 \rightarrow R^{12}$, independent of the number n of correspondences.

If this problem is solved using the Levenberg-Marquardt (LM) method, then an initial estimate of the parameters may be obtained by decomposing a camera matrix P found using the DLT algorithm. A central step in the LM method is the computation of the derivative matrix (Jacobian matrix) of the function being minimized, in this case $Mg(\mathbf{q})$. Note that $\partial Mg/\partial \mathbf{q} = M\partial g/\partial \mathbf{q}$. Thus, computation of the Jacobian reduces to computation of the Jacobian matrix of g , and subsequent multiplication by M .

Minimization of $\|Mg(\mathbf{q})\|$ takes place over all values of the parameters \mathbf{q} . Note, however, that if $P = K[R \mid -R\mathbf{t}]$ with K as in (8) then P satisfies the condition $p_{31}^2 + p_{32}^2 + p_3^2 = 1$, since these entries are the same as the last row of the rotation matrix R . Thus, minimizing $Mg(\mathbf{q})$ will lead to a matrix P satisfying the constraints $s = 0$ and $k_u = k_v$ and scaled such that $p_{31}^2 + p_{32}^2 + p_3^2 = 1$, and which in addition minimizes the algebraic error for all point correspondences.

2.7 Experimental Evaluation

Experiments were carried out with synthetic data to evaluate the performance of this algorithm. The data were created to simulate a standard 35mm camera with a 35mm focal length lens. A set of points were synthesized inside a sphere of radius 1m, and the camera was located at a distance of about 2.5m from the center of the sphere. The image is sampled so that the magnification factors are $\alpha_u = \alpha_v = 1000.0$, the same in each direction. This corresponds to a pixel size of $35\mu\text{m}$ for a 35mm camera.

Experiments were carried out to find the camera matrix with four different assumptions on known camera parameters.

1. There is no skew : $s = 0$. The number of remaining degrees of freedom d for the camera matrix is equal to 10.

2. There is no skew and the pixels are square : $s = 0$ and $\alpha_u = \alpha_v$. This corresponds to the situation for a true pinhole camera where image coordinates are measured in a Euclidean coordinate frame. In this case, $d = 9$.
3. In addition to the above assumptions, the principal point (u_0, v_0) is assumed to be known. There remain $d = 7$ degrees of freedom.
4. The complete internal calibration matrix K in (8) is assumed to be known. However, the pose of the camera is unknown. Thus $d = 6$.

To evaluate the performance of the algorithm, the result was compared with the optimal estimate with different degrees of noise. Thus, Gaussian noise with a given variance was added to the image coordinates of each point, and the camera matrix was estimated. The residual error was then computed, that is the difference between the measured and projected pixel. It is known (see [5]) that the expected lower bound on the root mean squared residual error is equal to

$$\epsilon_{\text{opt}} = \sigma(1 - d/N)^{1/2} \quad (9)$$

where σ is the standard deviation of the input noise, N is the number of measurements (in this case $2 \times$ number of points), and d is the number of degrees of freedom of the object being estimated. This represents the performance of an optimal estimation technique, and we can not do better. Since the residual error appeared to grow proportionally to injected noise (at least for noise levels less than about 10 pixels), a value of $v = 1$ pixel was used in the experiments. For each level of noise σ , the camera matrix was estimated 100 times with different noise. The residual error was RMS-averaged over all 100 runs and compared to the optimal value.

Results of the experiments are shown in Fig 1. The results show that minimizing algebraic error gives an almost optimal estimate of the camera matrix. In fact, the residual error is scarcely distinguishable from the optimal value. This is true in each of the four calibration problem types tried.

3 Computation of the Trifocal Tensor

The trifocal tensor ([3, 2]), relates the coordinates of points or lines seen in three views in a similar way to that in which the fundamental matrix relates points in two views.

The basic formula relates a point \mathbf{u} in one image and a pair of lines λ' and λ'' in the other two images. Provided there is a point \mathbf{x} in space that maps to \mathbf{u} in the first image and a point on the lines λ' and λ'' in

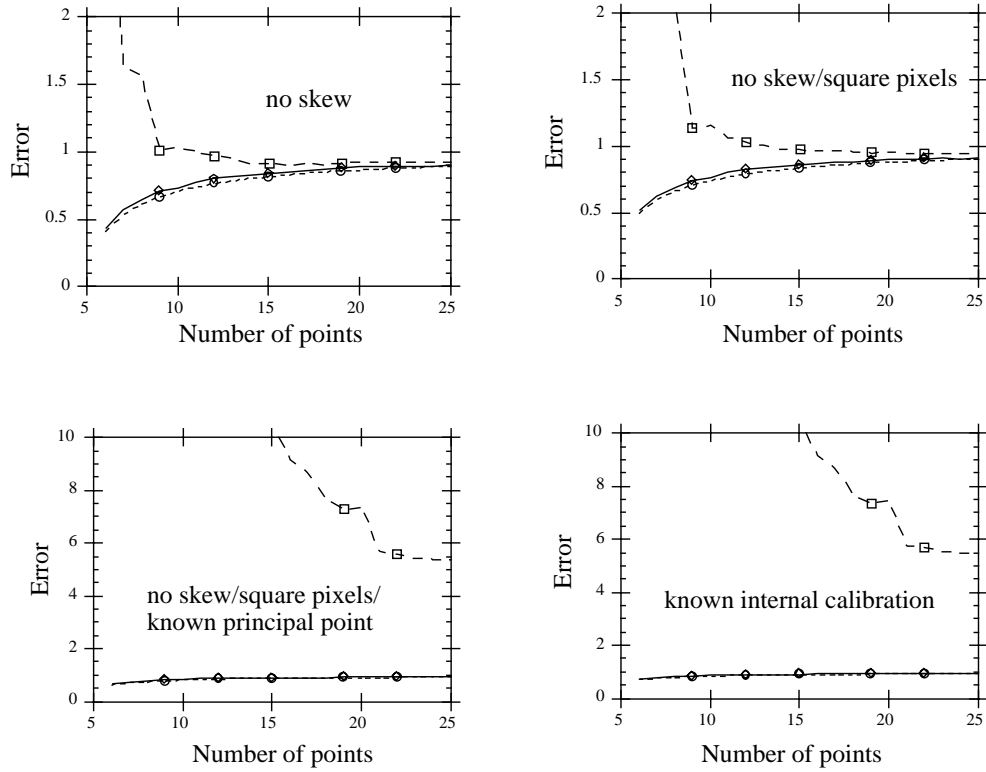


Figure 1: *The residual error was RMS-averaged over 100 runs for each $n = 6, \dots, 25$, where n is the number of points used to estimate the camera matrix. Four different levels of knowledge of the internal camera matrix were tried, corresponding to the four different graphs. In each of the graphs, the solid line represents the result of our iterative DLT algorithm, and the almost identical dotted line is the optimal estimate. In all four graphs, these two lines are barely distinguishable. For comparison, the results of a further method are also plotted. In this method, the complete calibration matrix K in (8) is estimated using the QR decomposition, and the known internal parameters are subsequently set to their known values. This method performs very poorly for small numbers of points, lying well off the graph, and is markedly inferior to the optimal estimation method even for larger numbers of points.*

the other two images, the following identity is satisfied :

$$u^i \lambda_j^i \lambda_k^i T_i^{jk} = 0 . \quad (10)$$

Here we are using tensor notation, in which a repeated index appearing in covariant (lower) and contravariant (upper) positions implies summation over the range of indices (namely, $1, \dots, 3$).

This equation may be used to generate equations given either point or line correspondences across three images. In the case of a line correspondence, $\lambda \leftrightarrow \lambda' \leftrightarrow \lambda''$ one selects two points \mathbf{u}_0 and \mathbf{u}_1 on the line λ , and for each of these points one obtains an equation of the form (10). In the case of a point correspondence $\mathbf{u} \leftrightarrow \mathbf{u}' \leftrightarrow \mathbf{u}''$ one selects any lines λ' and λ'' passing through \mathbf{u}' and \mathbf{u}'' respectively. Then (10) provides one equation. Four equations are generated from a single 3-view point correspondence by choosing two lines through each of \mathbf{u}' and \mathbf{u}'' , each pair of lines giving rise to a single equation.

The equations (10) give rise to a set of equations of the form $\mathbf{M}\mathbf{t} = \mathbf{0}$ in the 27 entries of the trifocal tensor. From these equations, one may solve for the entries of the tensor. As before, for any tensor T_i^{jk} the value of $\mathbf{M}\mathbf{t}$ is the algebraic error vector associated with the input data.

Consider the analogy with the 8-point algorithm for computing the fundamental matrix in the two-view case. The fundamental matrix has a constraint $\det F = 0$ that is not in general precisely satisfied by the solution found from linear algorithm. In the case of the trifocal tensor, there are 27 entries in the tensor, but the camera geometry that it encodes has only 18 degrees of freedom. This means that the trifocal tensor must satisfy 8 constraints, apart from scale ambiguity to make up the 27 degrees of freedom of a general $3 \times 3 \times 3$ tensor. The exact form of these constraints is not known precisely. Nevertheless, they must be enforced in order that the trifocal tensor should be well behaved. It will now be shown how this can be done, while minimizing algebraic error.

Formula for the Trifocal Tensor. We denote the three camera matrices \mathbf{P}' and \mathbf{P}'' by a_j^i and b_j^i respectively, instead of by p_j^i and $p_j''^i$. Thus, the three camera matrices \mathbf{P} , \mathbf{P}' and \mathbf{P}'' may be written in the form $\mathbf{P} = [\mathbf{I} \mid \mathbf{0}]$, $\mathbf{P}' = [a_j^i]$ and $\mathbf{P}'' = [b_j^i]$.

In this notation, the formula for the entries of the trifocal tensor is :

$$T_i^{jk} = a_i^j b_4^k - a_4^j b_i^k . \quad (11)$$

Our task will be to compute a trifocal tensor T_i^{jk} of this form from a set of image correspondences. The

tensor computed will minimize the algebraic error associated with the input data. The algorithm is similar to the one given in [4] for computation of the fundamental matrix. As with the fundamental matrix, the first step is the computation of the epipoles.

3.1 Retrieving the epipoles

We consider the task of retrieving the epipoles from the trifocal tensor. If the first camera has matrix $\mathbf{P} = [\mathbf{I} \mid \mathbf{0}]$, then the epipoles \mathbf{e}_{21} and \mathbf{e}_{31} are the last columns a_4^i and b_4^i of the two camera matrices $\mathbf{P}' = [a_j^i]$ and $\mathbf{P}'' = [b_j^i]$ respectively. These two epipoles may easily be computed from the tensor T_i^{jk} according to the following proposition.

Proposition 3.3. *For each $i = 1, \dots, 3$, the matrix T_i^{\cdot} is singular. Furthermore, the generators of the three left null-spaces have a common perpendicular, the epipole \mathbf{e}_{21} . Similarly epipole \mathbf{e}_{31} is the common perpendicular of the right null-spaces of the three matrices T_i^{\cdot} .*

This proposition translates easily into an algorithm for computing the epipoles ([3, 2]). This algorithm may be applied to the tensor T_i^{jk} obtained from the linear algorithm to obtain a reasonable approximation for the epipoles.

3.2 Constrained Estimation of the Trifocal Tensor

From the form (11) of the trifocal tensor, it may be seen that once the epipoles $\mathbf{e}_{21} = a_4^j$ and $\mathbf{e}_{31} = b_4^k$ are known, the trifocal tensor may be expressed linearly in terms of the remaining entries of the matrices $[a_i^j]$ and $[b_i^k]$.

Assuming the epipoles a_4^j and b_4^k to be known, we may write $\mathbf{t} = \mathbf{H}\mathbf{a}$ where \mathbf{a} is the vector of the remaining entries a_j^i and b_j^i , \mathbf{t} is the vector of entries of the trifocal tensor, and \mathbf{H} is the linear relationship expressed by (11). We wish to minimize the algebraic error $\|\mathbf{M}\mathbf{t}\| = \|\mathbf{M}\mathbf{H}\mathbf{a}\|$ over all choices of \mathbf{a} constrained such that $\|\mathbf{t}\| = \|\mathbf{H}\mathbf{a}\| = 1$.¹ Writing $\hat{\mathbf{t}} = \mathbf{H}\mathbf{a}$ where \mathbf{a} is the solution vector, we see that $\hat{\mathbf{t}}$ minimizes algebraic error $\|\hat{\mathbf{M}}\hat{\mathbf{t}}\|$ subject to the condition that T_i^{jk} is of the correct form (11), for the given choice of epipoles. Thus, finding the best T_i^{jk} is reduced to the solution of a simple minimization problem. Details of how to solve this problem are contained in the next section.

¹It does not do to minimize $\|\mathbf{M}\mathbf{H}\mathbf{a}\|$ subject to the condition $\|\mathbf{a}\| = 1$, since a solution to this occurs when \mathbf{a} is a unit vector in the right null-space of \mathbf{H} . In this case, $\mathbf{H}\mathbf{a} = \mathbf{0}$, and hence $\|\mathbf{M}\mathbf{H}\mathbf{a}\| = 0$.

3.3 A Constrained Minimization Problem

We consider the minimization problem : given matrices \mathbf{M} and \mathbf{H} , minimize $\|\mathbf{M}\mathbf{t}\|$ subject to $\|\mathbf{t}\| = 1$ and the additional constraint $\mathbf{t} = \mathbf{H}\mathbf{a}$ for some vector \mathbf{a} .

This problem may be posed in terms of finding the vector \mathbf{a} , in which case it becomes : minimize $\|\mathbf{M}\mathbf{H}\mathbf{a}\|$ subject to the condition $\|\mathbf{H}\mathbf{a}\| = 1$. This problem is solved as follows. Let the Singular Value Decomposition of \mathbf{H} be $\mathbf{H} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$. Let n be the rank of matrix \mathbf{H} – in the tensor estimation problem being considered, $\text{rank}\mathbf{H} = 15$. Then \mathbf{D} has n non-zero diagonal entries. We assume that the non-zero entries of \mathbf{D} precede the zero entries along the diagonal. Let \mathbf{U}' be the matrix consisting of the first n columns of \mathbf{U} , let \mathbf{V}' consist of the first n columns of \mathbf{V} and let \mathbf{D}' be the top-left $n \times n$ minor of \mathbf{D} , containing the non-zero diagonal entries. The minimization problem then becomes : minimize $\|\mathbf{M}\mathbf{U}'\mathbf{D}'\mathbf{V}'^\top\mathbf{a}\|$ subject to $\|\mathbf{U}'\mathbf{D}'\mathbf{V}'^\top\mathbf{a}\| = 1$. This last condition is equivalent to $\|\mathbf{D}'\mathbf{V}'^\top\mathbf{a}\| = 1$, since \mathbf{U}' has orthogonal columns. Now writing $\mathbf{t}' = \mathbf{D}'\mathbf{V}'^\top\mathbf{a}$, the problem becomes : minimize $\|\mathbf{M}\mathbf{U}'\mathbf{t}'\|$ subject to $\|\mathbf{t}'\| = 1$, which is a standard minimization problem. The solution \mathbf{t}' is the singular vector corresponding to the smallest singular value of $\mathbf{M}\mathbf{U}'$. Next, one computes \mathbf{a} by solving $\mathbf{t}' = \mathbf{D}'\mathbf{V}'^\top\mathbf{a}$. A crucial point here is that $\mathbf{D}'\mathbf{V}'^\top$ has rank equal to its row dimension, n and \mathbf{t}' is an n -vector. Therefore, there exists a solution to the equations $\mathbf{t}' = \mathbf{D}'\mathbf{V}'^\top\mathbf{a}$. If \mathbf{H} does not have full column-rank, then the solution for \mathbf{a} is not unique. However, one solution is given by $\mathbf{a} = \mathbf{V}'\mathbf{D}'^{-1}\mathbf{t}'$. To see this, one verifies $\mathbf{D}'\mathbf{V}'^\top\mathbf{a} = \mathbf{D}'\mathbf{V}'^\top\mathbf{V}'\mathbf{D}'^{-1}\mathbf{t}' = \mathbf{D}'\mathbf{D}'^{-1}\mathbf{t}' = \mathbf{t}'$ as required. Finally, $\mathbf{t} = \mathbf{H}\mathbf{a}$.

Note that if the value of \mathbf{a} is not specifically required, then \mathbf{t} may be computed without explicitly computing \mathbf{a} , as follows : $\mathbf{H}\mathbf{a} = \mathbf{U}'\mathbf{D}'\mathbf{V}'^\top\mathbf{a} = \mathbf{U}'\mathbf{t}'$. The complete algorithm is follows.

Algorithm : Given matrices \mathbf{M} and \mathbf{H} , find the vector \mathbf{t} that minimizes $\|\mathbf{M}\mathbf{t}\|$ subject to $\|\mathbf{t}\| = 1$ and the additional condition $\mathbf{t} = \mathbf{H}\mathbf{a}$ for some vector \mathbf{a} .

Solution :

1. Compute the SVD $\mathbf{H} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$ such that the non-zero values of \mathbf{D} appear first down the diagonal.
2. Let \mathbf{U}' be the matrix comprising the first n columns of \mathbf{U} , where n is the rank of \mathbf{H} . Further, let \mathbf{V}' consist of the first n columns of \mathbf{V} and \mathbf{D}' consist of the n first rows and columns of \mathbf{D} .
3. Find the unit vector \mathbf{t}' that minimizes $\|\mathbf{M}\mathbf{U}'\mathbf{t}'\|$. This is the singular vector corresponding to the smallest singular value of $\mathbf{M}\mathbf{U}'$.

4. The required vector \mathbf{t} is given by $\mathbf{t} = \mathbf{U}'\mathbf{t}'$, A vector \mathbf{a} such that $\mathbf{t} = \mathbf{H}\mathbf{a}$ is given by $\mathbf{a} = \mathbf{V}'\mathbf{D}'^{-1}\mathbf{t}'$.

3.4 Iterative Solution.

We have just seen how in terms of algebraic error, the optimal T_i^{jk} may be computed if the two epipoles are known. The two epipoles used to compute a correct constrained tensor T_i^{jk} are computed using the estimate of T_i^{jk} obtained from the linear algorithm.

Analogous to the case of the fundamental matrix described in [4], the mapping $(\mathbf{e}_{21}, \mathbf{e}_{31}) \mapsto \mathbf{M}\mathbf{H}\mathbf{a}$ is a mapping $R^6 \rightarrow R^{27}$. An application of the Levenberg-Marquardt algorithm to optimize the choice of the epipoles will result in an optimal (in terms of algebraic error) estimate of the trifocal tensor. Note that the iteration problem is of modest size, since only 6 parameters, the homogeneous coordinates of the epipoles, are involved in the iteration problem. Since the function being minimized is quite complex, derivatives are computed numerically. In general terms, the steps of the estimation problem are

1. Compute the reduced measurement matrix M from the given data.
2. Obtain an initial estimate of the epipoles \mathbf{e}_{21} and \mathbf{e}_{31} from an initial estimate of T_i^{jk} .
3. Compute the matrix \mathbf{H} expressing the trifocal tensor (represented by vector \mathbf{t}) in terms of the epipoles.
4. Use the constrained minimization method of section 3.3 to find the best value of $\mathbf{t} = \mathbf{H}\mathbf{a}$ and compute the algebraic error vector $\mathbf{M}\mathbf{t}$.
5. Iterate these last two steps with different values of the epipoles, so as to minimize the error vector.

The size of the iteration problem contrasts with an iterative estimation of the optimal trifocal tensor in terms of geometric error. This latter problem would require estimating the camera parameters of the three cameras, plus the coordinates of all the points – a large estimation problem.

3.5 Experimental Results

Once more, the iterative algorithm for computing the trifocal tensor was tested with synthetic data. The configuration of the points and cameras was similar to that used for the DLT algorithm, but in this case there were three cameras aimed at the point set from random angles. Data sets of 10, 15 and 20 points were used to test the algorithm. Residual errors were compared with the optimal values of the residual errors,

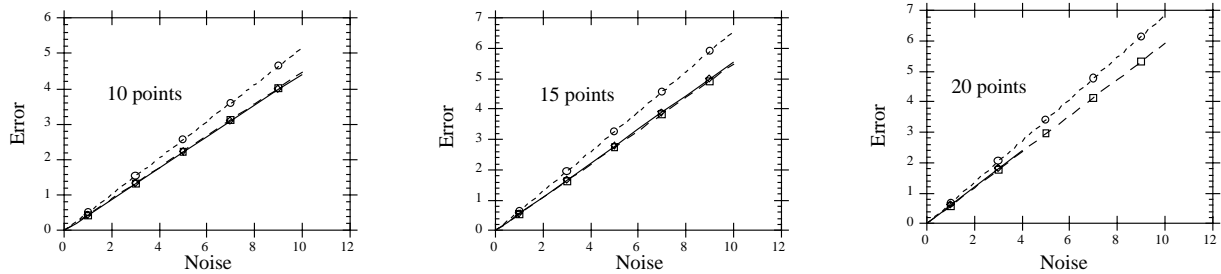


Figure 2: The residual error RMS-averaged over 100 runs is plotted against the number of points. Note that these plots are different from the plots for the DLT algorithm in that the horizontal axis represents the noise level, not the number of points as in the DLT case. The three plots are for 10, 15 and 20 points. Each graph contains three curves. The top curve is the result of the algebraic error minimization, whereas the lower two curves, actually indistinguishable in the graphs, represent the theoretical minimum error, and the error obtained by iteration to minimize geometric error, using the algebraic minimization as a starting point ([3, 2]). Note that the residual errors are almost exactly proportional to added noise. We learn two things from this. Minimization of the algebraic error achieves residual errors within about 15% of the optimal and using this estimate as a starting point for minimizing geometric error achieves a virtually optimal estimate. These results are notably better than results obtained previously in [2].

given by the formula (9). In this case, if n is the number of points, then the number of measurements is $N = 6n$, and the number of degrees of freedom in the fitting is $d = 18 + 3n$, where 18 represents the number of degrees of freedom of the three cameras (less 15 to account for projective ambiguity) and $3n$ represents the number of degrees of freedom of n points in space.

The results are shown in Fig 2.

4 Conclusion

Experimental evidence backs up the assertion that minimizing algebraic distance can usually give good results at a fraction of the computation cost associated with minimizing geometric distance. The great advantage of the method for minimizing algebraic error given in this paper is that even for problems that need an iterative solution the size of the iteration problem is very small. Consequently, the iteration is very rapid and there is reduced risk of falling into a local minimum, or otherwise failing to converge.

The method has been illustrated by applying it to three problems. For the computation of the fundamental matrix, iteration is over only three homogeneous parameters. (For details see [4].) For the trifocal tensor, iteration is over 6 parameters. This leads to more efficient methods than have been previously known.

The general technique is applicable to problems other than those treated here. It may be applied in a straight-forward manner to estimation of projective

transformations between 2 or 3-dimensional point sets. In these problems, iteration is necessary if one restricts the class of available transformations to a subgroup of the projective group, such as planar homologies conjugates of rotations.

References

- [1] R. I. Hartley. In defence of the 8-point algorithm. In *Proc. International Conference on Computer Vision*, pages 1064 – 1070, 1995.
- [2] R. I. Hartley. A linear method for reconstruction from lines and points. In *Proc. International Conference on Computer Vision*, pages 882 – 887, 1995.
- [3] Richard I. Hartley. Lines and points in three views and the trifocal tensor. *International Journal of Computer Vision*, 22(2):125–140, March 1997.
- [4] Richard I. Hartley. Minimizing algebraic distance. In *Proc. DARPA Image Understanding Workshop*, pages 631 – 637, 1997.
- [5] K. Kanatani. *Statistical Optimization for Geometric Computation : Theory and Practice*. North Holland, Amsterdam, 1996.