

Reconstruction from six-point sequences

Richard I. Hartley and Nicolas Y. Dano
G.E. Corporate Research and Development
1 Research Circle,
Niskayuna, NY 12309

Abstract

An algorithm is given for computing projective structure from a set of six points seen in a sequence of many images. The method is based on the notion of duality between cameras and points first pointed out by Carlsson and Weinshall. The current implementation avoids the weakness inherent in previous implementations of this method in which numerical accuracy is compromised by the distortion of image point error distributions under projective transformation. It is shown in this paper that one may compute the dual fundamental matrix by minimizing a cost function giving a first-order approximation to geometric distance error in the original untransformed image measurements. This is done by a modification of a standard near-optimal method for computing the fundamental matrix. Subsequently, the error measurements are adjusted optimally to conform with exact imaging geometry by application of the triangulation method of Hartley-Sturm.

1 Introduction

An idea of Carlsson and Weinshall ([1, 12, 2]) allows the possibility of computing projective structure from a set of several images of a small number of points. In this paper the case of six points in many views is discussed (the three-view case is solved in [8]), and a practical, relatively simple algorithm is given. This algorithmic method was discussed in a previous paper ([4]) where it was mentioned that a straight-forward implementation of the Carlsson method suffers from instability because of the skewing of noise characteristics by application of a projective transformation to each image. In this paper, a method is given that largely avoids this problem. The central strategy of this paper, so-called Sampson iteration, minimizes a cost function that relates back to untransformed image measurements.

The Sampson method is named after the author of a well-known paper on conic fitting ([9]). It has been used (though not under this name) as one of the most successful method for computation of the fundamental matrix ([13]). The idea

behind the Sampson method is to minimize a first-order approximation to error in image measurements while satisfying the necessary constraints imposed by the imaging geometry. It is one of the principal observations of this paper that one can do this for the error measurement in the original untransformed images, while still working in the dual domain of transformed image coordinates necessary for the duality-based algorithmic method.

The value of a method in which one treats small numbers of points in many views is that it allows us to compute the camera motion for long sequences in a single algorithm. Previous methods with projective cameras have used various strategies such as pasting reconstructions from blocks of three views together, or factorization method ([10]) which require fundamental matrices to be computed for pairs of views. The method is also useful for cases where one has long sequences with relatively few matched points, as shown in this paper.

2 General Algorithm Outline

The simple observation that leads to Carlsson Duality is as follows. A camera of the form

$$P = \begin{bmatrix} a_i & & -d_i \\ & b_i & -d_i \\ & & c_i & -d_i \end{bmatrix} \quad (1)$$

is called a *reduced camera matrix*. If $\mathbf{X} = (x, y, z, t)^\top$ is a 3D point, then one verifies that

$$\begin{bmatrix} a & & -d \\ & b & -d \\ & & c & -d \end{bmatrix} \begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix} = \begin{bmatrix} x & & -t \\ & y & -t \\ & & z & -t \end{bmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} \quad (2)$$

One may denote the matrix in (1) by P^A , where \mathbf{A} is the vector $(a, b, c, d)^\top$. With this notation (2) may be written as $P^A \mathbf{X} = P^X \mathbf{A}$, and one sees that the roles of point and camera are swapped.

Now, consider a set of image correspondences \mathbf{x}_j^i , which one should read as the image of a j -th point in an i -th view¹. A set of 3D points \mathbf{X}_j and camera matrices \mathbf{P}^i are called a projective reconstruction of the correspondences if $\mathbf{x}_j^i = \mathbf{P}^i \mathbf{X}_j$ for all i, j . If all the matrices \mathbf{P}^i are reduced camera matrices, then the reconstruction is called a *reduced reconstruction*. If such a reduced reconstruction is possible, then the points \mathbf{x}_j^i are said to *allow a reduced reconstruction*.

Given a set of image correspondences \mathbf{x}_j^i , one defines a *dual* or transposed set of correspondences $\hat{\mathbf{x}}_i^j$ defined by $\hat{\mathbf{x}}_i^j = \mathbf{x}_j^i$. Note that if $i = 1, \dots, N$ and $j = 1, \dots, M$, then the original correspondences are for N views of M points, whereas the dual correspondences are for a set of M views of N points.

Now if a set of points \mathbf{x}_j^i allows a reduced reconstruction $\mathbf{x}_j^i = \mathbf{P}^{A_i} \mathbf{X}_j$, then the dual correspondences allow a reconstruction $\hat{\mathbf{x}}_i^j = \mathbf{P}^{\mathbf{X}_j} \mathbf{A}_i$, since $\mathbf{P}^{\mathbf{X}_j} \mathbf{A}_i = \mathbf{P}^{A_i} \mathbf{X}_j = \mathbf{x}_j^i = \hat{\mathbf{x}}_i^j$. In brief :

2.1. *If a set of correspondences \mathbf{x}_j^i allows a reduced reconstruction, then so does the dual set of correspondences $\hat{\mathbf{x}}_i^j$. Furthermore one may obtain one reconstruction from the other by interchanging the roles of point and camera.*

In solving for a reduced reconstruction from a set of point correspondences that allow it, one may solve directly from the original correspondences, or else transpose (dualize) the correspondences and find a reduced reconstruction in terms of the dual correspondences. To keep straight what we are doing, the reduced camera matrices and points obtained from the dual set of correspondences $\hat{\mathbf{x}}_i^j$ will be called *dual cameras* and *dual points*, together making up the *dual reconstruction*. As shown, dual cameras correspond to points in a non-dual reconstruction, and dual points correspond to cameras.

Transforming the data

A set of point correspondences that allow a projective reconstruction will not in general allow a reduced reconstruction. However, they may be transformed to such a set by transformation of the data.

Let $\mathbf{E}_1 = (1, 0, 0, 0)^\top$, $\mathbf{E}_2 = (0, 1, 0, 0)^\top$, $\mathbf{E}_3 = (0, 0, 1, 0)^\top$ and $\mathbf{E}_4 = (0, 0, 0, 1)^\top$ form part of a projective basis for \mathcal{P}^3 . Similarly, let $\mathbf{e}_1 = (1, 0, 0)^\top$, $\mathbf{e}_2 = (0, 1, 0)^\top$, $\mathbf{e}_3 = (0, 0, 1)^\top$, $\mathbf{e}_4 = (1, 1, 1)^\top$ be a projective basis for the projective image plane \mathcal{P}^2 .

Now, let $\mathbf{x}_j^i = \mathbf{P}^i \mathbf{X}_j$ for $i = 1, \dots, N$ and $j = 1, \dots, M$ be a set of image points in M views. Let \mathbf{T}^i be a set

¹The index i will be used throughout the paper to index the view, whereas j indexes the points. To remember this, observe that one views with one's eye (i). This is a pun – not a very good one but useful as a mnemonic.

of 2D projective transformations chosen to map the last four points in each view to the four canonical basis points $\mathbf{e}_1, \dots, \mathbf{e}_4$. Let $\mathbf{x}_j^{i'} = \mathbf{T}^i \mathbf{x}_j^i$ for all i and j . One may formulate the reconstruction problem in terms of these transformed image points by seeking transformed camera matrices $\mathbf{P}^{i'}$ that satisfy $\mathbf{x}_j^{i'} = \mathbf{P}^{i'} \mathbf{X}_j$ for all i, j . In a solution to the reconstruction problems, the cameras \mathbf{P}^i and $\mathbf{P}^{i'}$ will be related by $\mathbf{T}^i \mathbf{P}^i = \mathbf{P}^{i'}$, since then

$$\mathbf{P}^{i'} \mathbf{X}_j = \mathbf{T}^i \mathbf{P}^i \mathbf{X}_j = \mathbf{T}^i \mathbf{x}_j^i = \mathbf{x}_j^{i'}$$

as required.

Provided that no four of the 3D points \mathbf{X}_j are coplanar, it may be assumed that in a projective reconstruction, the last four 3D points are equal to \mathbf{E}_j for $j = 1, \dots, 4$. It follows that the camera matrices $\mathbf{P}^{i'}$ are in reduced form (1) since they must satisfy the condition $\mathbf{P}^{i'} \mathbf{E}_j = \mathbf{e}_j$ for $j = 1, \dots, 4$.

One may summarize this discussion as follows:

2.2. *Let \mathbf{x}_j^i be a set of image correspondences allowing a projective reconstruction. For each view, indexed by $i = 1, \dots, N$, let \mathbf{T}^i be a projective transformation that maps the last four points \mathbf{x}_j^i , $j = M - 3, \dots, M$ to the canonical basis $\mathbf{e}_1, \dots, \mathbf{e}_4$. Then the points $\mathbf{x}_j^{i'} = \mathbf{T}^i \mathbf{x}_j^i$ for $i = 1, \dots, N$ and $j = 1, \dots, M - 4$ allow a reduced reconstruction.*

3 Six points in N views

Now, let us specialize to the case of 6 points observed in N views. In this case the four last points $\mathbf{x}_3^i, \dots, \mathbf{x}_6^i$ are mapped to the canonical basis point. The remaining points $\mathbf{x}_1^{i'} = \mathbf{T}^i \mathbf{x}_1^i$ and $\mathbf{x}_2^{i'} = \mathbf{T}^i \mathbf{x}_2^i$ must now allow a reduced reconstruction. After transposition, we obtain a set of points $\hat{\mathbf{x}}_i^1 \leftrightarrow \hat{\mathbf{x}}_i^2$ that allow a reduced reconstruction. This is a set of correspondences of N (dual) points in two (dual) views. Such a reconstruction problem may (in principal) be solved by known techniques involving the fundamental matrix to obtain a dual reconstruction. From this a (non-dual) reduced reconstruction may be obtained by swapping points and cameras. A reconstruction of the original data is then obtained by applying the inverse transforms $(\mathbf{T}^i)^{-1}$ to each of the reduced camera matrices.

To be more specific, the algorithm proceeds in the following steps.

1. From the dual image correspondence $\hat{\mathbf{x}}_i^1 \leftrightarrow \hat{\mathbf{x}}_i^2$ one computes a dual fundamental matrix $\hat{\mathbf{F}}$ by solving the set of equations $\hat{\mathbf{x}}_i^{2\top} \hat{\mathbf{F}} \hat{\mathbf{x}}_i^1 = 0$ for all i .
2. From the dual fundamental matrix, one may compute a (dual) projective reconstruction in the usual manner (for instance, see [3]). According to (2.2), a reduced reconstruction is possible, namely points and reduced

camera matrices satisfying $P^{X_1} A_i = \hat{x}_i^1$ and $P^{X_2} A_i = \hat{x}_i^2$.

3. The reconstruction for the original points x_j^i consists of the cameras $P^i = T^{i-1} P^{A_i}$ for all i and points X_1, X_6 where X_1 and X_2 were computed in the last step, and X_3, \dots, X_6 are the points E_1, \dots, E_4 . These satisfy $P^i X_j = x_j^i$ as required.

More details of this procedure are to be found in [4], and also [1]. The first two steps of this algorithm will be considered in more detail in the following sections.

4 Difficulties with the dual method

The method suggested above for solution from N views of six points may be thought of as consisting of three steps.

1. Apply a projective transformation T^i to the points in each view to generate a set of point correspondences allowing a reduced reconstruction.
2. Find a reduced reconstruction $(P^{i'}, X_j)$ by solving in the dual domain using the dual fundamental matrix.
3. Apply inverse transformations to the camera matrices to obtain $P^i = (T^i)^{-1} P^{i'}$. These along with the reconstructed points X_j form a projective reconstruction of the original data.

The main difficulty with this method is the distortion of points, and particularly their uncertainty regions by the projective transformations, T^i . Point measurements in an image are usually (though not quite correctly) assumed to be measured with independent isotropic Gaussian distributions. It is unusual for reconstruction algorithms to assume more general distributions. The common methods for solution of the two-view projective reconstruction problem, involving the fundamental matrix, have no way of taking into account a non-isotropic point distribution. The only way of doing this is to use an iterative scheme, such as bundle adjustment, that specifically accounts for non-isotropic error distributions. However circular (isotropic) error distributions do not remain circular when subjected to a projective transformation. Neither do they remain Gaussian. The projective transformations T^i , chosen to map four points in each view to a canonical basis may quite severely distort the error distribution. In applying a technique based on the fundamental matrix to the dual reduced reconstruction problem, one normally is obliged to ignore the particular form of the transformed error distribution and hope for the best. Frequently, one may work very hard to obtain a small residual in the reduced reconstruction problem (step 2 of the above outline) only to find that when one transforms back to the original

coordinate frame (step 3) the residual errors are very large for some points.

What is needed is a way of solving the reduced reconstruction problem while at the same time minimizing the residual error *in the original coordinate frame*. This is what is done in this paper.

5 The Reduced Fundamental Matrix

The fundamental matrix \hat{F} satisfying

$$\hat{x}_i^2 \top \hat{F} \hat{x}_i^1 = 0 \quad \text{for all } i \quad .$$

for a set of transformed points $\hat{x}_i^2 \leftrightarrow \hat{x}_i^1$ is of a special restricted form. Since the point correspondences allow a reduced reconstruction, according to (2.2), the fundamental matrix must correspond to a pair of reduced cameras P^{X_1} and P^{X_2} . A formula for the fundamental matrix corresponding to a pair of arbitrary camera matrices is given in [7] which allows us to compute the form of \hat{F} when the camera matrices are of this restricted form.

Since points X_3, \dots, X_6 are the points E_1, \dots, E_4 of a projective basis, we may choose $X_1 = E_5 = (1, 1, 1, 1)^\top$. If $X_2 = (x, y, z, t)^\top$, then

$$\hat{F} = \begin{bmatrix} 0 & -Y(Z - T) & Z(Y - T) \\ X(Z - T) & 0 & -Z(X - T) \\ -X(Y - T) & Y(X - T) & 0 \end{bmatrix} \quad . \quad (3)$$

Note that this matrix is singular, and satisfies the further constraint that the sum of its entries is zero. Such a matrix is called a *reduced fundamental matrix*. A reduced fundamental matrix is conveniently parametrized by the four parameters X, Y, Z and T , or by three parameters setting $T = 1$.

Characterization of Reduced Fundamental Matrix

A reduced fundamental matrix is a fundamental matrix (that is a singular matrix) that satisfies the additional conditions $e_i^\top \hat{F} e_i = 0$ for $i = 1, \dots, 4$. The condition $e_i^\top \hat{F} e_i = 0$ for $i = 1, \dots, 3$, implies that the diagonal entries of \hat{F} are zero. The requirement that $(1, 1, 1) \hat{F} (1, 1, 1)^\top = 0$ gives the additional condition that the sum of entries of \hat{F} is zero. A singular homogeneous 3×3 matrix has 7 degrees of freedom. The additional constraints $e_i^\top \hat{F} e_i = 0$ for $i = 1, \dots, 4$ reduce the number of degrees of freedom by 4, so that a reduced fundamental matrix has only 3 degrees of freedom.

Now, consider the reduced fundamental matrix of (3). It may be verified that this matrix is singular (the left null-vector is $(x - 1, y - 1, z - 1)^\top$). One may also verify that the sum of its entries is zero, and so are the diagonals. Finally, the matrix of (3) has three degrees of freedom (being defined by 3 parameters), and so one deduces that this is the general form of a reduced fundamental matrix.

Linearly solution for the reduced fundamental matrix

One may write \hat{F} in the form

$$\hat{F} = \begin{bmatrix} 0 & a & b \\ c & 0 & d \\ e & -(a+b+c+d+e) & 0 \end{bmatrix}$$

thereby parametrizing a fundamental matrix satisfying all the linear constraints (though not necessarily the condition $\det \hat{F} = 0$). Now, a pair of points \mathbf{x}_2 and \mathbf{x}_1 satisfying $\mathbf{x}_2^\top \hat{F} \mathbf{x}_1 = 0$ is easily seen to provide a linear equation in the parameters a, \dots, f of \hat{F} . Given at least 4 such correspondences, one may solve for these parameters, up to an inconsequential scale.

The matrix that results from the above method does not in general satisfy the condition $\det \hat{F} = 0$, which is required for a fundamental matrix. To find a matrix that does we tried various methods. An adaptation of the ‘‘algebraic’’ method of [5] works well.

6 Minimization of Sampson Error

Under the assumption of isotropic Gaussian error distribution, the Maximum Likelihood (usually considered the optimal) estimate of the fundamental matrix involves minimization of a cost function

$$\text{Cost}(\mathbf{F}) = \sum_i d(\mathbf{x}_1^i, \bar{\mathbf{x}}_1^i)^2 + d(\mathbf{x}_2^i, \bar{\mathbf{x}}_2^i)^2 \quad (4)$$

where $\bar{\mathbf{x}}_1^i \leftrightarrow \bar{\mathbf{x}}_2^i$ are estimated points that satisfy the fundamental matrix relationship $\bar{\mathbf{x}}_2^i \mathbf{F} \bar{\mathbf{x}}_1^i = 0$ exactly, and $d(\cdot, \cdot)$ represents Euclidean distance in the image. The cost function is to be minimized over all choices of \mathbf{F} and points $\bar{\mathbf{x}}_1^i \leftrightarrow \bar{\mathbf{x}}_2^i$. However, although optimal, this method is a little complex to implement. A method that works almost as well involves minimization of a first-order approximation to the geometric distance ([13]). In the above cost function, the cost of a given choice of \mathbf{F} is found by minimizing (4) subject to the condition $\bar{\mathbf{x}}_2^i \mathbf{F} \bar{\mathbf{x}}_1^i = 0$ – a sort of constrained minimization problem. The first-order approximation to this cost is given by

$$\text{Cost}^\perp(\mathbf{F}) = \sum_i \frac{(\mathbf{x}_2^i \mathbf{F} \mathbf{x}_1^i)^2}{(\mathbf{F} \mathbf{x}_1^i)_1^2 + (\mathbf{F} \mathbf{x}_1^i)_2^2 + (\mathbf{F}^\top \mathbf{x}_2^i)_1^2 + (\mathbf{F}^\top \mathbf{x}_2^i)_2^2} \quad (5)$$

where $(\cdot)_1$ and $(\cdot)_2$ mean the first and second components of the vector. The general strategy is to parametrize the fundamental matrix in some manner and then minimize this cost function over all \mathbf{F} by parameter minimization (usually the Levenberg-Marquardt method). This method is theoretically more sound than a commonly quoted method of minimizing symmetric epipolar distance, which has a similar-looking cost function.

Application to reduced reconstruction

In the solution of the dual reduced reconstruction problem, one seeks a reduced dual fundamental matrix that satisfies the equation $\hat{\mathbf{x}}_2^{i\top} \hat{\mathbf{F}} \hat{\mathbf{x}}_1^i = 0$ for all i . However, recall that $\hat{\mathbf{x}}_i^j = \mathbf{x}_j^i = \mathbf{T}^i \mathbf{x}_j^i$. Thus, the dual fundamental matrix \mathbf{F} must satisfy

$$\mathbf{x}_2^i (\mathbf{T}^{i\top} \hat{\mathbf{F}} \mathbf{T}^i) \mathbf{x}_1^i = 0 \quad (6)$$

Note that this is expressed in terms of the original point measurements. The cost associated with a given choice of \mathbf{F} is found by minimizing the distance

$$\sum_i d(\mathbf{x}_1^i, \bar{\mathbf{x}}_1^i)^2 + d(\mathbf{x}_2^i, \bar{\mathbf{x}}_2^i)^2 \quad (7)$$

subject to the condition (6). Setting $\mathbf{F}^{(i)} = \mathbf{T}^{i\top} \hat{\mathbf{F}} \mathbf{T}^i$ the cost function may be written simply as

$$\text{Cost}^\perp(\hat{\mathbf{F}}) = \sum_i \frac{(\mathbf{x}_2^i \mathbf{F}^{(i)} \mathbf{x}_1^i)^2}{(\mathbf{F}^{(i)} \mathbf{x}_1^i)_1^2 + (\mathbf{F}^{(i)} \mathbf{x}_1^i)_2^2 + (\mathbf{F}^{(i)\top} \mathbf{x}_2^i)_1^2 + (\mathbf{F}^{(i)\top} \mathbf{x}_2^i)_2^2} \quad (8)$$

This cost function is (as before) a first-order approximation to the geometric distance (7) in terms of the **original untransformed** data. To the extent that the first-order approximation to geometric distance is a good approximation, this error function is near-optimal in that it finds $\hat{\mathbf{F}}$ (and subsequently a reconstruction) that requires the smallest possible variation to the original image points \mathbf{x}_1^i and \mathbf{x}_2^i . Note that it is computed without actually transforming the points \mathbf{x}_j^i , which refer to original image coordinates. To estimate the near-optimal reduced fundamental matrix, therefore, one proceeds to minimize this cost function by a Levenberg-Marquardt process, using a parametrization of the reduced fundamental matrix given in section 5.

For comparison, an inferior method that minimizes residual error in the transformed data would minimize the required variation in the points $\hat{\mathbf{x}}_1^i = \mathbf{T}^i \mathbf{x}_1^i$ and $\hat{\mathbf{x}}_2^i = \mathbf{T}^i \mathbf{x}_2^i$ would minimize the cost function

$$\text{Cost}^\perp(\hat{\mathbf{F}}) = \sum_i \frac{(\mathbf{x}_2^i \mathbf{F}^{(i)} \mathbf{x}_1^i)^2}{(\hat{\mathbf{F}} \mathbf{T}^i \mathbf{x}_1^i)_1^2 + (\hat{\mathbf{F}} \mathbf{T}^i \mathbf{x}_1^i)_2^2 + (\hat{\mathbf{F}}^\top \mathbf{T}^i \mathbf{x}_2^i)_1^2 + (\hat{\mathbf{F}}^\top \mathbf{T}^i \mathbf{x}_2^i)_2^2} \quad (9)$$

This cost function is subtly but importantly different.

7 Completing the reduced reconstruction

In the outline of a reconstruction algorithm given in section 3, the second step involves computing the dual camera matrices $\mathbf{P}^{\mathbf{x}_j}$ and the dual points \mathbf{A}_i once the reduced fundamental matrix is known. The first step is to find the dual

camera matrices from the reduced fundamental matrix. This is easy, since the Sampson minimization involved iteration over the coordinates of the point \mathbf{X}_2 while fixing the points $\mathbf{X}_1 = (1, 1, 1, 1)^\top$. Thus the two dual camera matrices $\mathbf{P}^{\mathbf{X}_j}$ arise as the parameters that minimize the cost function.

To find the dual 3D points, \mathbf{A}_i a triangulation method is required to find the \mathbf{A}_i that satisfy $\mathbf{P}^{\mathbf{X}_j} \mathbf{A}_i = \hat{\mathbf{x}}_j^i$ for $j = 1, 2$. The computation may be carried out separately for each i . As in the usual projective reconstruction method, because of noise in the measurements, such \mathbf{A}_i do not exist precisely, the two rays from the two camera centres not intersecting in space. As expounded in [6], one should find the 3D point that minimizes image error – in fact a method is given in the cited paper for doing just that. In the present case, we want to minimize the error in the original image measurements, and **not** in the transformed data. This may be accomplished as follows. The method of [6] starts with a point pair $\mathbf{x} \leftrightarrow \mathbf{x}'$ and a fundamental matrix \mathbf{F} and finds the pair of corrected points $\bar{\mathbf{x}} \leftrightarrow \bar{\mathbf{x}}'$ that satisfy $\bar{\mathbf{x}}^\top \mathbf{F} \bar{\mathbf{x}} = 0$ exactly, while minimizing the distance from the estimated points $\bar{\mathbf{x}}$ and $\bar{\mathbf{x}}'$ to the measured points \mathbf{x} and \mathbf{x}' . In the current situation, the required corrected points must satisfy the condition $\hat{\mathbf{x}}_i^2 \hat{\mathbf{F}} \hat{\mathbf{x}}_i^1 = 0$. However, since $\hat{\mathbf{x}}_i^j = \mathbf{x}_j^i = \mathbf{T}^i \mathbf{x}_j^i$, this condition becomes $\mathbf{x}_1^i (\mathbf{T}^i \mathbf{T}^\top \hat{\mathbf{F}} \mathbf{T}^i) \mathbf{x}_2^i = 0$. This leads to the observation :

- To find the optimal corrected points $\bar{\mathbf{x}}_1^i$ and $\bar{\mathbf{x}}_2^i$, one applies the Hartley-Sturm algorithm ([6]) to the original measured points $\mathbf{x}_j^i; j = 1, 2$ using $\mathbf{T}^i \mathbf{T}^\top \hat{\mathbf{F}} \mathbf{T}^i$ for a fundamental matrix.

Note that once more we are finding the optimally corrected measurements using untransformed points in the original images. Once the points have been corrected, triangulation will be exact, and computation of the dual 3D space points is easy. Alternatively, observe that the dual 3D space points $\mathbf{P}^{\mathbf{A}_i}$ are ultimately used to find the camera matrices \mathbf{P}^i in the third step of the algorithm of *section 3*. An alternative approach is to estimate the cameras \mathbf{P}^i directly using the well known DLT algorithm ([11]) in terms of the now known 3D points \mathbf{X}_j and corrected image points $\bar{\mathbf{x}}_j^i$. The solution is guaranteed to be exact.

8 Results

The algorithm was tested on both synthetic and real data. For the synthetic data, a set of six points were chosen at random in a unit sphere, and a sequence of images were generated. The images parameters were chosen to approximate those taken with a 35mm camera with a total width of the image equal to 1000 pixels.

Five algorithms were compared

1. Minimization using transformed image coordinates. The algorithm used was an adaptation of the method of

[5] adapted to estimation of reduced fundamental matrix. The algorithm minimizes algebraic error instead of geometric error. However for estimation of the usual fundamental matrix it is one of the best algorithms we know, usually performing within a few percent of optimal in terms of residual error. It is a non-iterative algorithm, but the iteration is over the three coordinates of an epipole only, and is very fast. This will be called the **algebraic** method.

2. The results of the Sampson algorithm described in this paper. This minimizes a first-order approximation to geometric error in the original untransformed coordinates. This is the **Sampson** method.
3. Bundle adjustment to minimize residual error in the untransformed coordinates, initialized with the results of the algebraic method. This is the **refined algebraic** method.
4. Bundle adjustment initialized with the Sampson method (the **refined Sampson** method).
5. The theoretical minimum achievable residual error, computed based on the known amount of noise and the number of degrees of freedom of the system (the **ideal** method).

In all cases, the measured error residual is the residual error in the original image coordinates resulting from the reconstruction. This is true even in the algebraic method which works with the transformed image coordinates, and this is the essential weakness of all algorithms working in the transformed domain. From the results the advantage of the Sampson method is clear.

Although the Sampson method used here does not require the transformation of the point, it still requires the computation of the transformation, and its use in obtaining an initial estimate to use as a basis for the Sampson transformation. In some cases such as where three of the points are nearly collinear, this transformation will be unstable. As a result, the Sampson method still sometimes fails. This failure is characterized by a large residual error, which is easily detected. Various strategies are being developed for avoiding this failure. In particular, it is not necessary to include all the views in the Sampson estimate. We are experimenting with least-median squares methods for eliminating critical views from those used to compute the structure. Once the position of points \mathbf{X}_1 and \mathbf{X}_2 are computed from the other views the omitted cameras can be computed in a separate step. At present we are experiencing a failure rate of less than 2% at a noise level of 1 pixel in each coordinate. To avoid skewing of the results by these failures, essentially outliers, the worst 2% of cases are omitted from all the graphs shown here.

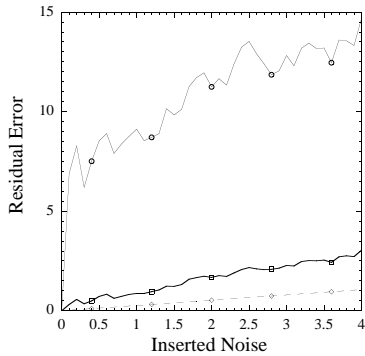


Figure 1. Comparison of the algebraic and Sampson algorithms. Because the algebraic algorithm (top) carries out minimization with transformed coordinates it works quite poorly. Sampson method (middle) performs quite close to the ideal result (bottom). Note that we can not expect the Sampson method to achieve the same residual error as the ideal, since in the Sampson method the residual is concentrated in the image of the last two points, whereas in the ideal case it is spread over all points.

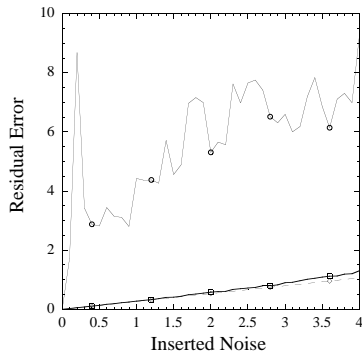


Figure 2. Results obtained by using the algebraic and Sampson methods to initialize a bundle adjustment. Notice that the refined Sampson method obtains almost optimal results, whereas the refined algebraic method performs quite poorly. The initial estimate obtained from the algebraic method is just not good enough to work effectively as an initial estimate for full-scale bundle adjustment.

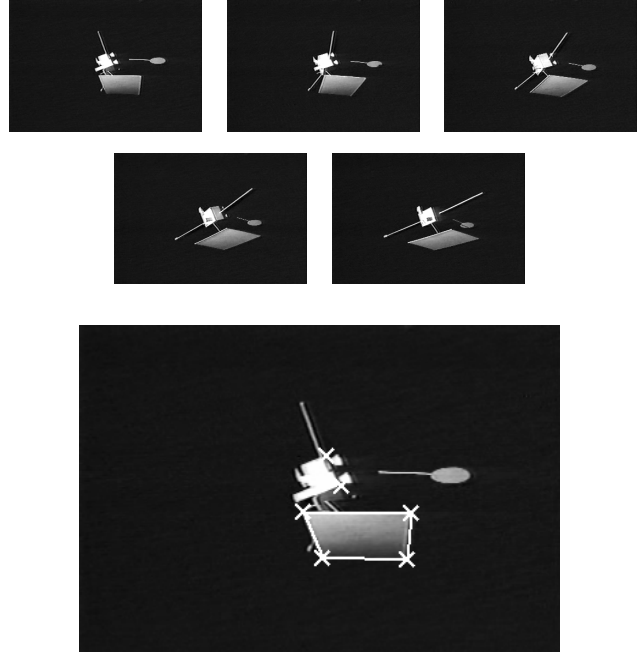


Figure 3. Frames from a sequence of images of a model satellite rotating and moving in front of a camera. The images were digitized at 352×240 pixels. The result of the algorithm was a residual pixel error of 0.15 pixels. After reconstruction, a wireframe model was reprojected onto the images, and observed to align well. Shown in the images are the reprojected points and solar panel reconstructed model.

Real image sequences

The algorithm was tested with a set of real images of a model satellite. A 20-frame sequence was used to test the algorithm. The sequence is shown in Fig 3.

9 Conclusion

Although the general outline of a method for computing projective structure from a sequence of six points in N view has been known since the original papers of Carlsson and Weinshall, a naïve implementation of the algorithm has been known to suffer from numeric difficulties because of the necessity to apply a projective transformation to the image points. This transformation can severely distort the error distribution of the image points. Any method that estimates structure by minimizing residual error in the transformed image points will perform relatively poorly. The present paper avoids minimization in the transformed domain by proposing methods for estimation of the reduced fundamental matrix and subsequent correction of the image

measurements while minimizing error functions that are directly related to Euclidean distance in the original untransformed image coordinates. The algorithm has been evaluated with both synthetic and real images and has been shown to perform well.

Acknowledgement

Thanks to Gilles Debunne for discussions and code related to six-point reconstruction.

Appendix - Sampson error.

Given a fundamental matrix F and a pair of noisy correspondences $\mathbf{x}' \leftrightarrow \mathbf{x}$, generally the coplanarity condition $\mathbf{x}'^\top F \mathbf{x} = 0$ will not be satisfied exactly, but there will exist a pair of points $\bar{\mathbf{x}}', \bar{\mathbf{x}}$ close to \mathbf{x}', \mathbf{x} that satisfy this equation precisely. We wish to find the points $(\bar{\mathbf{x}}', \bar{\mathbf{x}})$ closest to $(\mathbf{x}', \mathbf{x})$ that satisfies this relation. For simplicity, let \mathbf{X} represent the 4-vector obtained by concatenating the inhomogeneous representations of \mathbf{x} and \mathbf{x}' . Further denote by $C_F(\mathbf{X})$ the value $\mathbf{x}'^\top F \mathbf{x}$.

To first-order, C_F may be approximated by a Taylor expansion

$$C_F(\mathbf{X} + \delta_{\mathbf{X}}) = C_F(\mathbf{X}) + \frac{\partial C_F}{\partial \mathbf{X}} \delta_{\mathbf{X}} . \quad (10)$$

If we write $\delta_{\mathbf{X}} = \bar{\mathbf{X}} - \mathbf{X}$ and desire $\bar{\mathbf{X}}$ to satisfy $C_F(\bar{\mathbf{X}}) = 0$, then the result is $C_F(\mathbf{X}) + (\partial C_F / \partial \mathbf{X}) \delta_{\mathbf{X}} = 0$, which we will henceforth write as $J \delta_{\mathbf{X}} = -\epsilon$ where J is the partial-derivative matrix, and ϵ is the cost $C_F(\mathbf{X})$ associated with \mathbf{X} . The minimization problem that we now face is to find the smallest $\delta_{\mathbf{X}}$ that satisfies this equation, namely :

- Find the vector $\delta_{\mathbf{X}}$ that minimizes $\|\delta_{\mathbf{X}}\|$ subject to $J \delta_{\mathbf{X}} = -\epsilon$.

The standard way to solve problems of this type is to use Lagrange multipliers. By a fairly straight-forward computation, one finds that

$$\delta_{\mathbf{X}} = -J^\top (J J^\top)^{-1} \epsilon , \quad (11)$$

The norm $\|\delta_{\mathbf{X}}\|^2$ is the Sampson error :

$$\|\delta_{\mathbf{X}}\|^2 = \delta_{\mathbf{X}}^\top \delta_{\mathbf{X}} = \epsilon^\top (J J^\top)^{-1} \epsilon . \quad (12)$$

In the case of fundamental matrix estimation being considered here, $J J^\top$ is a scalar, equal to the denominator of (5) and $\epsilon^\top \epsilon$ is the numerator. Thus, the right hand side of (12) is the same as (5). Furthermore, the left hand side of (12) is the same as (4), and so (5) represents a first-order approximation to geometric error as claimed. This is already known from the work of [13], but the above discussion has been included to demonstrate that in just the same way (8) minimizes geometric error up to first order in the minimization of the dual fundamental matrix.

References

- [1] Stefan Carlsson. Duality of reconstruction and positioning from projective views. In *Workshop on Representations of Visual Scenes*, 1995.
- [2] Stefan Carlsson and Daphna Weinshall. Dual computation of projective shape and camera positions from multiple images. *International Journal of Computer Vision*, 27(3), 1998.
- [3] R. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 761–764, 1992.
- [4] Richard I. Hartley. Dualizing scene reconstruction algorithms. In *3D Structure from Multiple Images of Large Scale Environments : European Workshop, SMILE'98*, pages 14 – 31. LNCS-Series Vol. 1506, Springer Verlag, 1998.
- [5] Richard I. Hartley. Minimizing algebraic error. *Philosophical Transactions of the Royal Society, Series A*, 356(1740):1175 – 1192, May 1998.
- [6] Richard I. Hartley and Peter Sturm. Triangulation. In *Proc. Computer Analysis of Images and Patterns, Prague, LNCS-Series Vol. 970, Springer Verlag*, pages 190 – 197, September 1995.
- [7] Anders Heyden. A common framework for multiple view tensors. In *Computer Vision - ECCV '98, Volume I, LNCS-Series Vol. 1406, Springer-Verlag*, pages 3–19, 1998.
- [8] Long Quan. Invariants of 6 points from 3 uncalibrated images. In *Computer Vision - ECCV '94, Volume II, LNCS-Series Vol. 801, Springer-Verlag*, pages 459–470, 1994.
- [9] 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.
- [10] Peter Sturm and Bill Triggs. A factorization based algorithm for multi-image projective structure and motion. In *Computer Vision - ECCV '96, Volume II, LNCS-Series Vol. 1065, Springer-Verlag*, pages 709–720, 1996.
- [11] I.E. Sutherland. Sketchpad: A man-machine graphical communications system. Technical Report 296, MIT Lincoln Laboratories, 1963. Also published by Garland Publishing Inc, New York, 1980.
- [12] D. Weinshall, M. Werman, and A. Shashua. Shape descriptors : Bilinear, trilinear and quadrilinear relations for multi-point geometry and linear projective reconstruction algorithms. In *Workshop on Representations of Visual Scenes*, 1995.
- [13] Zhengyou Zhang. Determining the epipolar geometry and its uncertainty : A review. *International Journal of Computer Vision*, 27(2):161 – 195, 1998.