A Six Point Solution for Structure and Motion

F. Schaffalitzky[1], A. Zisserman[1], and R. I. Hartley[2]

[1] Robotics Research Group, Oxford, UK
[2] G.E. CRD, Schenectady, NY

Abstract. This paper has three main contributions: (1) a "quasi-linear" method for computing structure and motion for $m \geq 3$ views of 6 points; (2) a "quasi-linear" method for computing consistent estimates of the multi-view tensors (fundamental matrix, trifocal tensor and quadrifocal tensor) from n image points; (3) an m view n point robust reconstruction algorithm which uses the 6 point method as a search engine.
The new algorithms are evaluated on synthetic and real image sequences, and compared to optimal estimation results (bundle adjustment).

1 Introduction

A large number of methods exist for obtaining 3D structure and motion from correspondences tracked through image sequences. Their characteristics vary from the so-called *minimal* methods [13, 14, 20] which work with the least data necessary to compute structure and motion, through intermediate methods [4, 16] which may perform mis-match (outlier) rejection as well, to the full-bore *bundle adjustment*.
The minimal solutions are used as search engines in robust estimation algorithms which automatically compute correspondences and tensors over multiple views. For example, the 2 view 7 point solution is used in the RANSAC estimation of the fundamental matrix in [20], and the 3 view 6 point solution in the RANSAC estimation of the trifocal tensor in [19]. It would seem natural then to use a minimal solution as a search engine in 4 or more views. The problem is that in 4 or more views a solution is forced to include a minimization to account for measurement error (noise). In the '2 view 7 point' and '3 view 6 point' cases there are the same number of measurement constraints as degrees of freedom in the tensor. In both cases 1 or 3 real solutions result (and the duality explanation for this equivalence was given by [2]). However, in four views six points provide one more constraint than the number of degrees of freedom in the four view geometry (the quadrifocal tensor). This means than unlike in the two and three view cases where a tensor can be computed which exactly relates the measured points (and also satisfies its internal constraints), this is not possible in the four view case. Instead it is neccesary to minimize a measurement error whether algebraic or geometric. The poor estimate which results by using an approach based on minimizing algebraic distance and a standard projective basis for the image is described and demonstrated in section 2.

Here we develop a novel quasi-linear solution for the 6 point $m \geq 3$ case. This solution involves only a SVD and the evaluation of a cubic polynomial in a single variable. This is described in 2. We also describe a sub-optimal (compared to bundle-adjustment) which minimizes geometric error at the cost of only a 3 parameter minimization.

1.1 Reconstruction for an image sequence

A second part of the paper describes yet another algorithm for computing a reconstruction of cameras and 3D scene points from a sequence of images. The objectives of such algorithms are now well established:

1. **Minimize reprojection error**. A common statistical noise model assumes that measurement error is isotropic and Gaussian in the image. The Maximum Likelihood Estimate in this case involves minimizing the total squared reprojection error over the cameras and 3D points. This is bundle-adjustment.
2. **Cope with missing data**. Structure-from-motion data often arises from tracking features through image sequences and any one track may persist only in few of the total frames.
3. **Cope with mis-matches**. Appearance-based tracking can produce tracks of non-features. A common example is a T-junction which generates a strong corner, but whose pre-image moves slowly between frames.

Bundle adjustment [7] is the most accurate and theoretically best justified technique. It can cope with missing data and, with suitable robust statistical cost function, can cope with mis-matches. However, it is expensive to carry out and most significantly requires a good initial estimate.

In the special case of affine cameras, factorization methods [17] minimize reprojection error [15] and so give the optimal solution found by bundle adjustment. However, factorization cannot cope with mis-matches, and methods to overcome missing data [10] lose the optimality of the solution.

In the general case of perspective projection iterative factorization methods have been successfully developed and have recently proved to produce excellent results [16, 9]. The problems of missing data and mis-matches remain though.

Bundle-adjustment will almost always be the final step of a reconstruction algorithm. However, achieving good sub-optimal estimates prior to bundle-adjustment is necessary for the latter to be effective (fewer iterations, and less likely to converge to local minimum.) For practical (in particular automated) applications, mismatches present a real problem. There exist effective methods for estimating structure and motion from data with mismatches for two [18] and three [19] views (based on RANSAC) and [21] based on LMS. These have been put to effective use [4] to compute structure and motion by starting from (very reliably) estimated three-view structures and hierarchically coalescing these into sub-sequences of the whole sequence. For four views there is the method in [6] for computing the quadrifocal tensor.

Current methods of initializating a bundle-adjustment include factorization [16], awf-segments [4], duality [1, 2] and the Variable State Dimension Filter (VSDF) [11].
In this paper we describe a novel algorithm for computing a reconstruction satisfying the 3 basic goals above (optimal, missing data, mismatches). It is based on using the 6-pt algorithm as a robust search engine, and is described in section 5.

1.2 Notation

The *standard basis* will refer to the five points in $\mathbb{P}^3$ whose homogeneous coordinates are :

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \mathbf{e}_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \mathbf{e}_* = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

For a 3-vector $\mathbf{v} = (x, y, z)^\top$, we use $[\mathbf{v}]_\times$ to denote the 3×3 skew matrix such that $[\mathbf{v}]_\times \mathbf{u} = \mathbf{v} \times \mathbf{u}$, where $\times$ denotes the vector cross product. For three points in the plane, represented in homogeneous coordinates by $\mathbf{x}, \mathbf{y}, \mathbf{z}$, the incidence relation of collinearity is the vanishing of the bracket $[\mathbf{x}, \mathbf{y}, \mathbf{z}]$ which denotes the determinant of the 3×3 matrix whose columns are $\mathbf{x}, \mathbf{y}, \mathbf{z}$. It equals $\mathbf{x} \cdot (\mathbf{y} \times \mathbf{z})$ where $\cdot$ is the vector dot product.

2 Linear estimation using a duality solution

This section (about 1.5 pages) should cover:

1. Outline duality algorithm for 6 points in $m \geq 4$ views (from 8 point algorithm).
2. Show using synthetic data that SVD solution in projectively transformed image space is a very poor estimate when pulled back to original image space.
3. Figure to illustrate this - az to draw - which shows that minimizing geometric error (as algebraic error minimization tries to approximate this) in very projectively transformed space pulls back to point away from ellipse centre.

3 Reconstruction from 6 points over m views

This section describes the main algebraic development of the 6 point method. In essence it is quite similar to the development given by Quan [13] for a reconstruction of 6 points from 3 views. The difference is that Quan used a standard projective basis for both the image and world points, whereas here the image coordinates are not transformed. As described in section 2 the use of a standard basis in the image severely distorts the error that is minimized. The numerical results that follow demonstrate that the method described here produces a near optimal solution.

In the following it will be assumed that we have 6 image points $\mathbf{x}_i$in correspondence over m views. The idea then is to compute cameras for each view such that the scene points $\mathbf{X}_i$ project exactly to their image $\mathbf{x}_i$ for the first five points. Any error minimization required is then restricted to the sixth point in the first instance.

3.1 Pencils of cameras

Each correspondence between a scene point $\mathbf{X}$ and its image $\mathbf{x}$ under a perspective camera $\mathtt{P}$ gives three linear equations for $\mathtt{P}$ whose combined rank is 2. These linear equations are obtained from

$$\mathbf{x} \times \mathtt{P}\mathbf{X} = \mathbf{0} \tag{1}$$

Given only five scene points, assumed to be in general position, it is possible to recover the camera up to a 1-parameter ambiguity. More precisely, the five points generate a linear system of equations for $\mathtt{P}$ which may be written $\mathtt{M}\mathbf{p} = \mathbf{0}$, where $\mathtt{M}$ is a 10×12 matrix formed from two of the linear equations (1) of each point correspondence, and $\mathbf{p}$ is $\mathtt{P}$ written as a 12-vector. This system of equations has a 2-dimensional null-space and thus results in a pencil of cameras.
Suppose that the five world points are the points of the standard projective frame so that both $\mathbf{X}_i$ and $\mathbf{x}_i$ $(i = 1, 2, 3, 4, 5)$are now known. Then the null-space of $\mathtt{M}$ can immediately be computed, and will be noted from here on by the basis of 3×4 matrices $[\mathbf{A}, \mathbf{B}]$. Then for any choice of the scalars $(s : t) \in \mathbb{P}^1$ the camera in the pencil $\mathtt{P} = s\mathtt{A} + t\mathtt{B}$ exactly projects the standard projective basis to the first five points.
Each camera $\mathtt{P}$ in the pencil has its optical centre located as the null-vector of $\mathtt{P}$ and thus a given pencil of camera gives rise to a 3D curve of possible camera centres. In general (there are degenerate cases) the locus of possible camera centres will be a twisted cubic passing through the five points of the standard projective basis.

3.2 The quadric constraints

Given m views of 6 points, suppose again that the first five world points are in known positions $\mathbf{X}_1, \ldots, \mathbf{X}_5$. To compute projective structure it suffices to find the sixth world point $\mathbf{X}_6$.
Let $[\mathbf{A}, \mathbf{B}]$ be the pencil of cameras consistent with the projections of the first five points (into the jth view, say). Since $\mathtt{P}$ lies in the pencil, there are scalars $(s : t) \in \mathbb{P}^1$ such that $\mathtt{P} = s\mathtt{A} + t\mathtt{B}$ and so the projection of the sixth world point $\mathbf{X}$ is $\mathbf{x}_6 = s\mathtt{A}\mathbf{X}_6 + t\mathtt{B}\mathbf{X}_6$. Eliminating s, t this means that the three points $\mathbf{x}_6, \mathtt{A}\mathbf{X}_6, \mathtt{B}\mathbf{X}_6$ are collinear in the image :

$$Q(\mathbf{X}_6) = [\mathbf{x}_6, \mathtt{A}\mathbf{X}_6, \mathtt{B}\mathbf{X}_6] = 0$$

, which is a quadratic constraint on $\mathbf{X}_6$. Each view thus provides a quadric on which $\mathbf{X}_6$ must lie. For two views the two associated quadrics intersect in a curve, and consequently there is a one parameter family of solutions for $\mathbf{X}_6$ in that case. The curve will meet a third quadric

in a finite number of points, so 3 views will determine a finite number (namely $2 \times 2 \times 2 = 8$ by Bézout's theorem) of solutions for $\mathbf{X}_6$.

Continuing with the case of three views for the moment. Suppose $\mathbf{X}_i$ is one of the five base points in the world. By definition of the pencil, $\mathbf{x}_i = \mathtt{A}\mathbf{X}_i$ and $\mathbf{x}_i = \mathtt{B}\mathbf{X}_i$ will be multiples of the corresponding image point (in homogeneous coordinates) and so $\mathtt{A}\mathbf{X} \times \mathtt{B}\mathbf{X}$ is $\mathbf{0}$ at such an $\mathbf{X}$. This means that each quadric contains each of the five base points so when solving for the sixth point we have to discard these five spurious solutions. We next develop a simple representation of the class of quadrics which vanish at the five base points, and use this to express the constraint arising from the sixth point in each view.

The class of all quadrics is a linear space of dimension 10 (regarding scale as significant for the moment). The subclass of quadrics which vanish at a given point is a linear subspace of codimension 1 and the class of quadrics which vanish at 5 could be expected to be a linear subspace of dimension $10 - 5 = 5$. While this is not always true (the dimension can be greater than 5) it is true for 5 points in general position. If a quadric is specified by a symmetric 4×4 matrix Q, then the vanishing conditions arising from world points at the standard projective basis positions are that $Q_{ii} = 0$ for $i = 1, 2, 3, 4$ and $\sum_{ij} Q_{ij} = 0$, which manifestly impose 5 (linearly) independent constraints on Q.

For definiteness, we choose the following basis for this 5D linear system :

$$W_1 = \begin{pmatrix} 0 & 1 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix} W_2 = \begin{pmatrix} 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix} W_3 = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}$$

$$W_4 = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \end{pmatrix} W_5 = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 1 & 0 \end{pmatrix}$$

Every quadric Q vanishing at the five base points must be a linear combination of $W_1, \ldots, W_5$, say $Q = \sum w_i W_i$, and the W_i consist of terms quadratic in the coordinates of $\mathbf{X}_6$. Explicitly, if $\mathbf{X}_6 = (p, q, r, s)^\top$ then $W_1 = pq - ps, W_2 = pr - ps, W_3 = qr - ps, W_4 = qs - ps, W_5 = rs - ps$. For each view the coefficients w_i are known since they are computed from $\mathtt{A}, \mathtt{B}$ and $\mathbf{x}_6$. This means that from $Q = \sum w_i W_i = 0$ for each view, a system of linear equations can be assembled for the unknowns W_i, and from the solution for W_i the coordinates of $\mathbf{X}_6$ may then be extracted.

In more abstract terms there is a map ψ

$$\psi : \mathbf{X} = \begin{pmatrix} p \\ q \\ r \\ s \end{pmatrix} - \rightarrow \begin{pmatrix} a \\ b \\ c \\ d \\ e \end{pmatrix} = \begin{pmatrix} pq - ps \\ pr - ps \\ qr - ps \\ qs - ps \\ rs - ps \end{pmatrix}$$

which is a (rational) transformation from $\mathbb{P}^3$ to $\mathbb{P}^4$, and maps the quadric $Q \subset \mathbb{P}^3$ into the hyperplane

$$w_1 a + w_2 b + w_3 c + w_4 d + w_5 e = 0 \tag{2}$$

where the (known) coefficients w_i are $Q_{12}, Q_{13}, Q_{23}, Q_{24}, Q_{34}$.
The basic method now is to solve for $\mathbf{W} = (a, b, c, d, e)^\top \in \mathbb{P}^4$ by intersecting hyperplanes in $\mathbb{P}^4$, rather than to solve directly for $\mathbf{X} \in \mathbb{P}^3$ by intersecting quadrics in $\mathbb{P}^3$.

3.3 Cubic constraint

The fact that $\dim \mathbb{P}^3 = 3 < 4 = \dim \mathbb{P}^4$ implies that the image of ψ is not all of $\mathbb{P}^4$. In fact the image is the hypersurface $\mathtt{S}$ cut out by the cubic equation

$$S(a, b, c, d, e) = abd - abe + ace - ade - bcd + bde = \begin{vmatrix} e & e & b \\ d & c & b \\ d & a & a \end{vmatrix} = 0$$

This can be verified by substitution, and a derivation in terms of determinants is sketched below.
The fact that the image $\psi(\mathbf{X})$ of $\mathbf{X}$ must lie on $\mathtt{S}$ introduces the problem of enforcing this constraint ($S = 0$) numerically. This will be dealt with below.
Having solved for $\mathbf{W} = (a, b, c, d, e)^\top$ we wish to recover $\mathbf{X} = (p, q, r, s)^\top$. By considering ratios of a, b, c, d, e and their differences, various form of solution can be obtained. In particular it can be shown that $\mathbf{X}$ is a nullvector of the following 6×4 design matrix :

$$\begin{pmatrix} e-d & 0 & 0 & a-b \\ e-c & 0 & a & 0 \\ d-c & b & 0 & 0 \\ 0 & e-b & a-d & 0 \\ 0 & e & 0 & a-c \\ 0 & 0 & d & b-c \end{pmatrix} \tag{3}$$

This will have nullity ≥ 1 if the point with coordinates a, b, c, d, e really does lie on the 3-fold $\mathtt{S}$. (In fact, imposing this degeneracy on 4×4 submatrices gives quartic algebraic expressions in a, b, c, d, e which are all multiples of the cubic expression S). At certain exceptional points the nullity will be greater than 1 (namely at the 10 singular points of the surface where the nullity will be exactly 2). When the point $\mathbf{W}$ doesn't lie exactly on $\mathtt{S}$, the matrix may have nullity 0 and more care has to be taken to recover a meaningful $\mathbf{X}$.

3 views : The linear constraints defined by the three hyperplanes (2) cut out a line in $\mathbb{P}^4$. The line intersects $\mathtt{S}$ in three points (generically) (see figure **). Thus there are three solutions for $\mathbf{X}$. This is a well-known [13] minimal solution. Our treatment gives a simpler (than the Quan [13] or Carlsson and Weinshall [2]) algorithm for computing a trifocal tensor from six points (from a projective reconstruction) because it does not involve changing basis in the images.

Four or more views : In this case the linear constraints from the hyperplanes alone will (generally) determine a unique solution for $\mathbf{W}$. In the presence of noise, though, this solution will not satisfy the cubic constraint. That is, it does not lie on $\mathtt{S}$; its coordinates do not satisfy $S = 0$. We would like to coerce it to do so. The problem is to perform a "manifold projection" in a non-Euclidean space, with the usual associated problem that we don't know in which direction to project. We will now give a novel solution to this problem.
An (over)determined linear system of equations is often solved using Singular Value Decomposition, by taking as null-vector the singular vector with the smallest singular value. The justification for this is that the SVD elicits the "directions" of space in which the solution is well determined (small singular values) and those in which it is poorly determined (large singular values). Taking the singular vector with smallest singular value is the usual "linear" solution, but as pointed out, it does not in general lie on $\mathtt{S}$. However, there may still be some information left in the second-smallest singular vector, and taking the space spanned by the two smallest singular vectors gives a line in $\mathbb{P}^4$, which passes through the "linear" solution and must also intersect $\mathtt{S}$ in three points (S is cubic). We use these three intersections as our candidates for $\mathbf{W}$. Since they lie exactly on $\mathtt{S}$, recovering their preimages $\mathbf{X}$ under ψ is not a problem.
This, then, is our heuristic. We overcome our manifold projection problems by projecting in the direction of the singular vector with second-smallest singular value. Note that in the case of 4 views, the smallest singular value will actually be 0.

Degeneracies : it is worth noting that if the sixth point in 3-space lies on the twisted cubic through the first five basis points then there is a one parameter family of cameras for each view which will exactly project the six space points to their images. This situation can be detected (in principle) because if the space point lies on the twisted cubic then all 6 image points lie on a conic. (In practice the problem shows up when intersecting the hyperplanes in $\mathbb{P}^4$; they will intersect in a line, namely the transform of the twisted cubic under ψ.)

3.4 Minimizing reprojection error

The previous sub-section has described a quasi-linear method involving the following two steps: first, a linear SVD decomposition of a matrix composed of one hyperplane from each view; second, intersecting the line in $\mathbb{P}^4$ (computed from two of the singular vectors) with a cubic surface. The best (most accurate) use of the given data is to minimize total squared image reprojection error over all camera and structure parameters, but that amounts to a full bundle adjustment.
In the current case, we have computed cameras which map the first five points exactly to their measured image points, and rather than jump directly to bundle adjustment, an intermediate case is to minimize total squared reprojection error for the sixth point $\mathbf{X}$ over $\mathbb{P}^3$. This fits in the middle of a spectrum of possible estimates :

1. **Algebraic fit**. The quasi-linear solution minimizes an "algebraic" error by a direct least squares fit on homogeneous coordinates in $\mathbb{P}^4$.
2. **Sub-optimal fit**. Minimizes total squared reprojection error for the sixth point over its position in $\mathbb{P}^3$, mapping the first five points exactly (3 dofs).
3. **Optimal fit (bundle adjustment)**. Minimizes total squared reprojection error for all points, over all structure and camera parameters ($11m + 3$ dofs).

The model fitted by the second item is clearly a reduced form of the model fitted by the third item. The cost of executing minimization is negligible (it has only 3 degrees of freedom), which can be seen as follows. To fit a model with a non-linear Levenberg-Marquardt type minimizer, we need to calculate at the current estimate, $\mathbf{X}$, the fitting residuals and the jacobian of these wrt the current estimate. The latter is obtained (if tediously) from the former, so let us concentrate on the fitting residuals. In each image, fitting error is the distance from the reprojected point $\mathbf{y} = \mathtt{P}\mathbf{X}$ to the measured image point $\mathbf{x} = (u, v, 1)^\top$. The reprojected point will depend both on the position of the sixth world point and on the choice of camera in the pencil for that image. But for a given world point $\mathbf{X}$, and choice of camera $\mathtt{P} = s\mathtt{A} + t\mathtt{B}$ in the pencil, the residual is the 2D image vector from $\mathbf{x}$ to the point $\mathbf{y} = \mathtt{P}\mathbf{X} = s\mathtt{A}\mathbf{X} + t\mathtt{B}\mathbf{X}$ on the line $\mathbf{l}$ joining $\mathtt{A}\mathbf{X}$ and $\mathtt{B}\mathbf{X}$. The optimal choice of s, t for given $\mathbf{X}$ is thus easy to deduce; it must be such as to make $\mathbf{y}$ the perpendicular projection of $\mathbf{x}$ onto this line (figure 3.4). What this means is that explicit minimization over camera parameters is unnecessary and so only the 3 dofs for $\mathbf{X}$ remain.

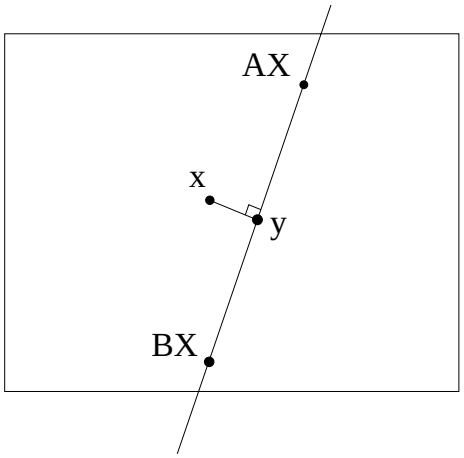

Fig. 1. Minimizing reprojection in the reduced model. For a given $\mathbf{X}$, the best choice $\mathtt{P} = s\mathtt{A} + t\mathtt{B}$ of camera in the pencil corresponds to the point $y = s\mathtt{A}\mathbf{X} + t\mathtt{B}\mathbf{X}$ on the line closest to the measured image point $\mathbf{x}$. Hence the image residual is the vector joining x and y.

3.5 Approximating geometric error

We now compare the first item with the second. We have already seen that the components of the line $\mathbf{l}(\mathbf{X}) = \mathtt{A}\mathbf{X} \times \mathtt{B}\mathbf{X}$ are expressible as quadrics in $\mathbf{X}$, and moreover as linear functions of $W = \psi(\mathbf{X})$:

$$\mathbf{l}(\mathbf{X}) = \mathtt{A}\mathbf{X} \times \mathtt{B}\mathbf{X} = \begin{pmatrix} \mathbf{q}_0\psi(\mathbf{X}) \\ \mathbf{q}_1\psi(\mathbf{X}) \\ \mathbf{q}_2\psi(\mathbf{X}) \end{pmatrix} = \begin{pmatrix} \cdots \mathbf{q}_0 \cdots \\ \cdots \mathbf{q}_1 \cdots \\ \cdots \mathbf{q}_2 \cdots \end{pmatrix} W$$

for some 3×5 matrix with rows $\mathbf{q}_i$ whose coefficients can be determined from those of $\mathtt{A}$ and $\mathtt{B}$. If the sixth image point is $\mathbf{x} = (u, v, 1)^\top$ then the squared residual is

$$d(\mathbf{x}, l(\mathbf{X}))^2 = \frac{|u\mathbf{q}_0 W + v\mathbf{q}_1 W + \mathbf{q}_2 W|^2}{|\mathbf{q}_0 W|^2 + |\mathbf{q}_1 W|^2}$$

which we note is a rank 1 quadratic form in W divided by a rank 2 quadratic form in W. The conclusion we draw is that the minimization of image error over $\mathbf{X} \in \mathbb{P}^3$ can be carried out as a minimization over $W \in \mathtt{S}$ instead. (Note also that this form of the error is amenable to reweighted least squares because, given an initial estimate of $\mathbf{X}$, we can adjust the scale so as to make the denominator close to 1, while putting the numerator into a least squares problem.)

The algebraic fitting algorithm which we propose consists of first forming the linear least squares problem which minimizes the sum of squares of $\mathbf{q}_2 W$ over the images. We intersect the 2D SVD nullspace with $\mathtt{S}$ to impose constraints.

This fitting scheme gives unit-norm vector estimates for $W = \psi(\mathbf{X})$, but since ψ is singular at the five basis points, $\psi(\mathbf{X})$ is zero there and what this amounts to is that these five points are singularities of the fitting scheme in the sense of [12] as being points that the scheme cannot fit. This is a good, not a bad, thing. (Actually, since ψ is regular on any smooth curve through the base points, the method *can* fit points "infinitely near" the base points.)

As we have presented the algorithm so far, there is an arbitrary choice of scale for each quadric $Q_{\mathtt{A},\mathtt{B}}$, corresponding to the arbitrariness in the choice of representation $[\mathbf{A}, \mathbf{B}]$ of the pencil of cameras (in terms of the equation above the algebraic fitting scheme neglects the denominator and just minimizes the residuals defined by the $u\mathbf{q}_0 + v\mathbf{q}_1 + \mathbf{q}_2$), the scale of which depends on the scale of $\mathtt{A}, \mathtt{B}$. Which normalization is used matters, and we addres that issue now.

Firstly, by translating coordinates, we may assume that the sixth point is at the origin. This amounts to (pre)multiplying $\mathtt{A}, \mathtt{B}$ by a 3×3 translation homography and we assume this has been done (so $u, v = 0$ in the above derivation). Thus the geometric error we want to approximate is

$$\frac{|\mathbf{q}_2 W|^2}{|\mathbf{q}_0 W|^2 + |\mathbf{q}_1 W|^2}$$

Making this assumption on the position of the sixth image point means that the normalization is independent of (*ie* is invariant to) translations

of image coordinates. It is desirable that the normalization should be invariant to scaling and rotation as well since these are the transformations which preserve our error model (isotropic Gaussian noise). This requirement rules out many obvious candidates, like normalizing the Frobenius norms of $\mathtt{A}, \mathtt{B}$ to 1 or normalizing $\mathbf{q}_2$ to unit norm. To describe our choice of normalization, we introduce a dot product $(\mathtt{A}, \mathtt{B})_*$ on 3×4 matrices, defined by :

$$(\mathtt{A}, \mathtt{B})_* = \sum_{\substack{i=0,1 \\ j=0,1,2,3}} A_{ij} B_{ij}$$

Our normalization can now be described by saying that the choice of basis of the pencil $[\mathbf{A}, \mathbf{B}]$ must be an orthonormal basis wrt $(\cdot, \cdot)_*$. Scaling image coordinates corresponds to scaling the first two rows of the basis elements, which just scales our dot product. Rotating image coordinates corresponds to applying an orthogonal transformation to the first two rows of the basis elements, and this preserves our dot product. Finally, choosing a different orthonormal basis corresponds to a certain linear basis change in the pencil and the effect on the $\mathbf{q}_i$ is a scaling by the determinant of that basis change. But that basis change must be orthogonal, so it has determinant 1.

3.6 Summary and Results I

It has been demonstrated how to pass from m views of six points in the world to a projective reconstruction in a few steps. The positions of the six world points as well as the camera for each view have been computed. The reconstruction obtained is not the MLE (assuming isotropic Gaussian point localization noise), which optimally distributes measurement error over all the points, but an approximation which puts all the errors on the sixth point.

The steps of the algorithm are :

1. Compute, for each image, the pencil of cameras which map the five standard basis points in the world to the first five image points, using the recommended normalization to achieve invariance to image coordinate changes.
2. Form, from each pencil $[\mathbf{A}, \mathbf{B}]$ the quadric constraint on the sixth world point $\mathbf{X}$ as described in section 3.2.
3. Using the transformation $\psi : \mathbb{P}^3 - \rightarrow \mathbb{P}^4$, convert the quadric intersection problem to a hyperplane intersection problem. Use the SVD to compute a pencil of possible values for $\mathbf{W} = \psi(\mathbf{X})$.
4. Intersect that line with the cubic constraint $S = 0$ to get (up to) three solutions for $\mathbf{W} = \psi(\mathbf{X})$ satisfying the constraint.
5. Use (3) to recover values for the sixth point $\mathbf{X}$ from $\mathbf{W}$. Keep the solution (if there are 3) with the lowest residual.
6. (optional) Minimize reprojection error over the 3 dofs in the position of $\mathbf{X}$.

We will now give results on synthetic and real image sequences of 6 points in m views. The objective is to evaluate the performance of three algorithms: quasi-linear; minimizing on 6th point only; bundle adjustment. The performance measures are *** In practice, for a given set of six

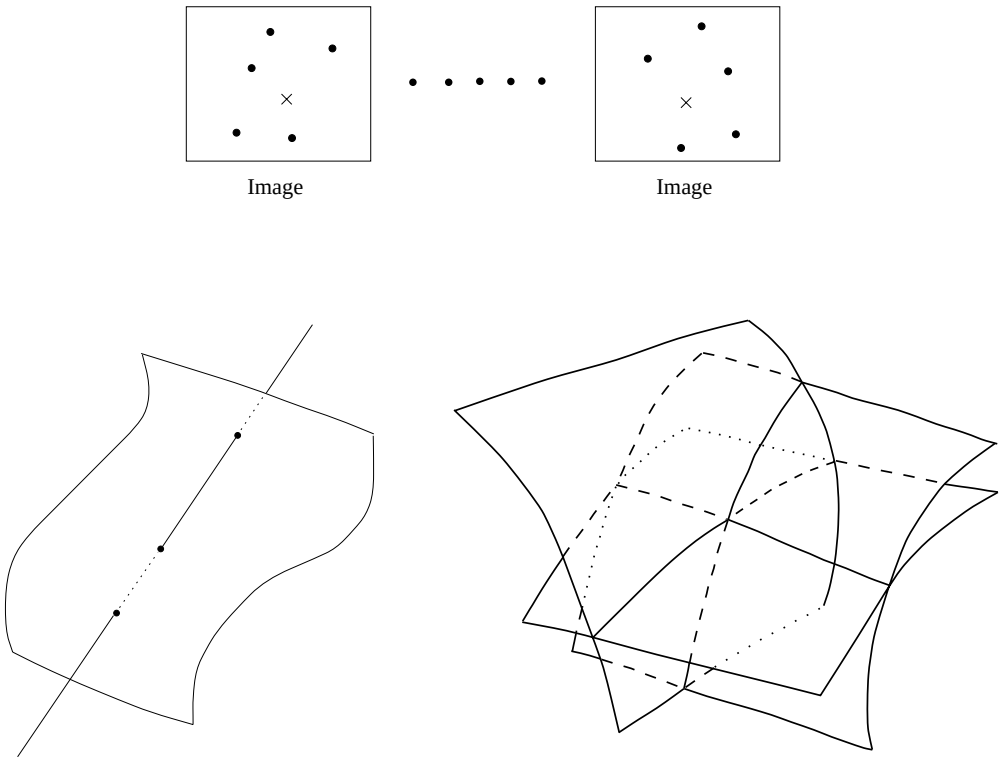

Fig. 2. The 6-point algorithm.

points, the quality of reconstruction can vary depending on which point is chosen to minimize over. We try all six in turn.
We show results here for an image sequence consisting of 9 colour images (JPEG, 768×1024) of a turntable (but the camera motion is not a single axis rotation!) with 24 tracks entered manually (by eye). The visibility matrix of the tracks is as follows, where each row corresponds to a tracked point and each column to an image. A cross in position (i, j) means that the ith point was seen in the jth image :

```
   0 1 2 3 4 5 6 7 8
 0 x x . x . . . . x
 1 x x . x . . . . x
 2 . x x x x x . . . -
 3 . x x . x x . . . -
 4 . . . . x x x x x
 5 x x . . . . x x x
 6 x x . . . . . x . x
 7 . . . . x x x x x
 8 x x x x . . . x x -
 9 x . . . . x x x x
10 x x x x x x x x x *
11 x x x x x x x x x *
12 x x x x x x . . . *
13 x x x x x x x . x -
14 . . . x x x x x .
15 . x x x x x x x . -
16 x x x x x x x x x *
17 x x x x x x x x x *
18 x x x x x x x x x *
19 x x x x x x x x x *
20 x . . . . x x x .
21 x x x x . . x x x -
22 x x x x x x . . . *
23 . . x x x x x x . -
```

We ran the algorithm described below on the subsequence consisting of images 0 to 5. Tracks $18, 19, 10, 11, 12, 17$ (in that order) were used to compute a six-point reconstruction over these views. Any remaining tracks seen in 4 or more views were then backprojected using the computed cameras to get structure for 15 points (marked with dashes) and 6 cameras. To evaluate the accuracy of reconstruction, we consider both image residuals and error of registration into a ground truth reconstruction obtained by means of calipers (estimated accuracy $0.5mm$). The following table compares our reconstruction with its bundle adjusted version (residuals reported as rms/max) :
Bundle adjustment achieves the smallest reprojection error over all residuals, because it has greater freedom in distributing the error. Our method minimizes error on the sixth point of a six point basis. Thus it is no surprise that the effect of applying bundle adjustment is to increase the

	basis residuals (pixels)	all residuals (pixels)	registration error (mm)
6 points no optimization	0.145 /0.569	1.18 /3.05	0.540/0.712
6 points with optimization	0.143 /0.526	1.27 /3.22	0.545/0.758
6 points bundled	0.0612/0.156	1.64 /6.94	0.574/0.881
All points bundled	0.609 /1.71	0.577/1.71	0.609/0.975

Fig. 3. Evaluating the reconstruction. Residuals are shown for the 6 points which formed the basis (first column) and for all reconstructed points taken as a whole (second column). There were 15 points and 6 views.

error in column 1 and to decrease the error in column 2. What *is* surprising is the rise in registration error (column 3). These figures support our claim that the linear method gives a very good approximation to the optimized method.
Figure 4 shows the reprojected reconstruction in the first and fourth views of the sequence. The large white dots are the input (measured) points and the smaller, darker dots are the reprojected points.

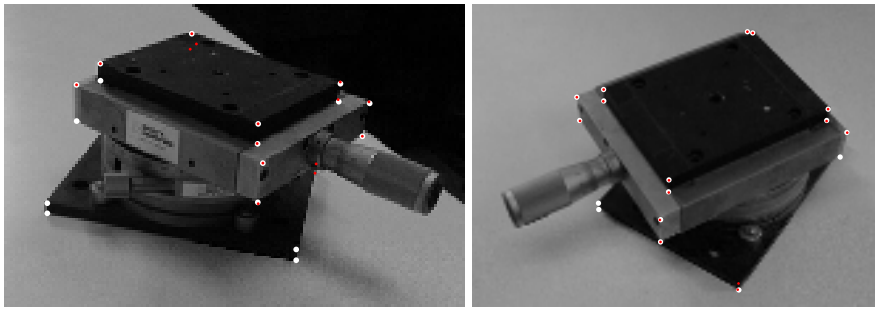

Fig. 4. Reprojected reconstruction in views 0 and 3. The smaller, dark points are the reprojected points.

4 Estimating multi-view tensors

For two views of 7 points there is a well-known method [20, **?**] for recovering the fundamental matrix between the two views. Essentially, each point correspondence $\mathbf{x} \leftrightarrow \mathbf{x}'$ between the two views imposes a single linear constraint $\mathbf{x}'^{\top}\mathtt{F}\mathbf{x} = 0$ and so seven points define a pencil of candidates for $\mathtt{F}$. The requirement that $\mathtt{F}$ be singular imposes a cubic constraint on this pencil and so there are up to three solutions. In geometric terms, the (linear) space of 3×3 matrices can be identified with $\mathbb{P}^8$ and the fundamental matrices lie in a subset of this, namely the locus of singular matrices. Singularity is characterized by the vanishing of the determinant

$\det(\mathtt{F}) = 0$, so that the locus of fundamental matrices lies on a cubic hypersurface in $\mathbb{P}^8$. This surface has three intersections with the line cut out in $\mathbb{P}^8$ by the 7 hyperplanes obtained from 7 correspondences.

Given many (n more than 7) correspondences the linear constraints alone will determine a solution, but as before, in the presence of noise, that solution will not satisfy the constraint, i.e. it will not lie on the cubic hypersurface defined by $\det(\mathtt{F}) = 0$). The method described in section 3.3 can be used to project the linear solution onto the constraint manifold as follows: Use the linear 8-point algorithm as described by Hartley [8] (with data normalization) to construct the $n \times 9$ design matrix $\mathtt{A}$. The linear estimate of $\mathtt{F}$ is obtained from $\mathtt{A}$ as the singular vector corresponding to the least singular value. In the orginal algorithm [8] Hartley then converts this matrix to one with rank 2 by using the SVD. The alternative proposed here is to compute the pencil of matrix solutions defined by the line joining the singular vectors corresponding to the *two* least singular values, and intersect this pencil with the cubic surface. The result is a rank 2 fundamental matrix "close to" the linear solution.

5 Robust Reconstruction Algorithm

Using our basic 6-point engine, we have constructed a robust algorithm for reconstruction from motion tracks. Robustness means that the algorithm is capable of rejecting mismatches, using the RANSAC [3] paradigm. It is a straightforward generalization of the corresponding algorithm for 7 points in 2 views [] and 6 points in 3 views [].

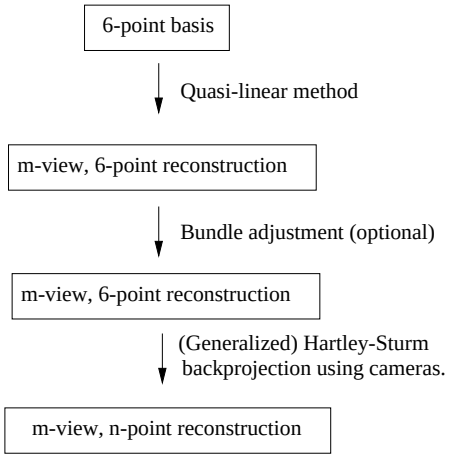

Fig. 5. Schematic of the algorithm.

5.1 Algorithm

The input is a set of measured image projections. A number of world points (usually thousands) have been tracked through a number of images. Some tracks may last for many images, some for only a few (*ie* there may be missing data). There may be mismatches.

1. From the set of tracks which appear in all images, select six at random. This set of tracks will be called a *basis*.
2. Initialize a projective reconstruction using those six tracks. This will provide the world coordinates (of the six points whose tracks we chose) and cameras for all the views.
3. For all remaining tracks, compute optimal world point positions using the computed cameras. This is a straightforward generalization of the Hartley-Sturm [] algorithm for two views. Reject tracks whose image reprojection errors exceed a threshold. The number of tracks which pass this criterion is used to score the reconstruction.
4. Repeat the above steps as required.

As we have already pointed out, the ordering of the six point basis can sometimes make a difference to the quality of reconstruction, so if at any point a basis achieves a score of 90% or more of the best basis so far, we try that basis again with the different choices of sixth point (the ordering of the first five points makes no difference).

The justification for this algorithm is, as always with RANSAC, that once a "good" basis is found it will (a) score highly and (b) provide a reconstruction against which other points can be tested (to reject mismatches).

5.2 Results II

The second sequence is a turntable sequence (*ie* the camera motion is a turntable motion) of a dinosaur model. The image size was 720×576. Motion tracks were kindly provided by Andrew Fitzgibbon. We ran the algorithm on the subsequence consisting of images 0 to 5. 100 samples were used.

Linear reconstructions were rejected if any reprojection error exceeded 10 pixels, the 3 dof optimization applied and a threshold of 5 pixels applied. These are very generous thresholds and are only intended to avoid spending computation on very bad initializations. The real criterion of quality is how much support an initialization has.

When backprojecting tracks to score the reconstruction, only tracks seen in 4 or more views were used and tracks were rejected as mismatches if any residual exceed 1.25 pixels after backprojection. To assess the performance of our algorithm, we tried three variations. The first mode just uses our quasi-linear algorithm. The second applies the optimization described in section 3.4. The third applies a full bundle adjustment to the 6-point reconstructions. The errors are summarized in the following table. The last row shows errors after applying bundle adjustment to the final reconstruction (many points, many cameras). Remarks entirely analogous to the ones made about the previous sequence apply to this one, but note specifically that optimizing makes no difference to the residuals at this level of precision (3 significant figures).

	basis residuals (pixels)	all residuals (pixels)	inlier count
6 points no optimization	0.0443/0.183	0.401/1.24	95
6 points with optimization	0.0443/0.183	0.401/1.24	95
6 points bundled	0.0422/0.127	0.383/1.181	97
All points bundled	0.313 /0.718	0.234/0.925	95

Fig. 6. There were 6 views. For each mode of operation, the number of points marked as inliers by the algorithm is shown in the third column. There were 127 tracks seen in four or more views.

6 Conclusion

1. Have shown how to use our 6-point engine to perform robust reconstruction for m views of n points. This reconstruction can now form the basis of a hierarchical method for extended image sequencess. [4] built a hierarchical reconstruction from image triplets. Now can proceed from extended sub-sequences over which at least 6 points tracked.
2. Other minimal cases involving points and lines over $m \geq 4$ views.

References

1. S. Carlsson. Duality of reconstruction and positioning from projective views. In *IEEE Workshop on Representation of Visual Scenes, Boston*, 1995.
2. S. Carlsson and D. Weinshall. Dual computation of projective shape and camera positions from multiple images. *IJCV*, 1998. in Press.
3. M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. ACM*, 24(6):381–395, 1981.
4. A. W. Fitzgibbon and A. Zisserman. Automatic camera recovery for closed or open image sequences. In *Proc. ECCV*, pages 311–326. Springer-Verlag, Jun 1998.
5. G.-M. Greuel, G. Pfister, and H. Schönemann. Singular version 1.2 user manual. In *Reports On Computer Algebra*, number 21 in Reports On Computer Algebra. Centre for Computer Algebra, University of Kaiserslautern, June 1998. `http://www.mathematik.uni-kl.de/~zca/Singular`
6. R. Hartley. Computation of the quadrifocal tensor. In *Proc. ECCV*, LNCS 1406, pages 20–35. Springer-Verlag, 1998.
7. R. I. Hartley. Euclidean reconstruction from uncalibrated views. In J.L. Mundy, A. Zisserman, and D. Forsyth, editors, *Proc. 2nd European-US Workshop on Invariance, Azores*, pages 187–202, 1993.
8. R. I. Hartley. In defence of the 8-point algorithm. In *Proc. ICCV*, pages 1064–1070, 1995.
9. A. Heyden. Projective structure and motion from image sequences using subspace methods. In *Scandinavian Conference on Image Analysis, Lappenraanta, 1997*, 1997.

10. D. Jacobs. Linear fitting with missing data: Applications to structure from motion and to characterizing intensity images. In *Proc. CVPR*, pages 206–212, 1997.
11. P. F. McLauchlan and D. W. Murray. A unifying framework for structure from motion recovery from image sequences. In *Proc. ICCV*, pages 314–320, 1995.
12. V. Pratt. Direct least-squares fitting of algebraic surfaces. *Computer Graphics*, 21(4):145–151, 1987.
13. L. Quan. Invariants of 6 points from 3 uncalibrated images. In J. O. Eckland, editor, *Proc. ECCV*, pages 459–469. Springer-Verlag, 1994.
14. L. Quan and F. K. A. Heyden. Minimal projective reconstruction with missing data. In *Proc. CVPR*, 1999.
15. I. D. Reid and D. W. Murray. Active tracking of foveated feature clusters using affine structure. *IJCV*, 18(1):41–60, 1996.
16. P. Sturm and W. Triggs. A factorization based algorithm for multi-image projective structure and motion. In *Proc. ECCV*, pages 709–720, 1996.
17. C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization approach. *IJCV*, 9(2):137–154, Nov 1992.
18. P. H. S. Torr and D. W. Murray. The development and comparison of robust methods for estimating the fundamental matrix. *IJCV*, 24(3):271–300, 1997.
19. P. H. S. Torr and A. Zisserman. Robust parameterization and computation of the trifocal tensor. *Image and Vision Computing*, 15:591–605, 1997.
20. P. H. S. Torr and A. Zisserman. Robust computation and parameterization of multiple view relations. In *Proc. ICCV*, pages 727–732, Jan 1998.
21. Z. Zhang, R. Deriche, O. D. Faugeras, and Q. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence*, 78:87–119, 1995.