# Linear Projective Reconstruction from Matching Tensors

**Bill Triggs**
INRIA Rhône-Alpes,
655 avenue de l'Europe, 38330 Montbonnot Saint-Martin, France.
*Bill.Triggs@inrialpes.fr*

**Abstract**

This paper describes initial work on a family of projective reconstruction techniques that extract projection matrices directly and linearly from matching tensors estimated from image data. The simplest methods use fundamental matrices and epipoles, alternative ones use trilinear tensors. All of the data is treated uniformly, and there is no reliance on 'privileged' images or tokens. The approach is based on 'joint image closure relations' — bilinear constraints between the projection matrices and their matching tensors, that express the fact that the latter derive from the former. The performance of the new techniques is quantified and compared with that of several existing methods.
**Keywords:** Multi-image Structure, Projective Reconstruction, Matching Tensors.

## 1 Introduction

Most existing stereo systems use two carefully calibrated cameras and attempt to reconstruct the scene from a single pair of static images. Recently, it has become clear that the redundancy offered by multiple images can significantly increase the quality and stability of visual reconstructions, as well as extending their coverage to previously unseen parts of the scene. Moreover, a great deal of useful information — in fact the complete 3D scene geometry modulo a single overall projective transformation — can be recovered without any camera calibration [3, 5, 10]. These factors are especially relevant to applications such as geometric reconstruction for virtual reality or robot navigation, where many images are needed to cover the scene and precise calibration is difficult owing to uncertain camera motions, changes in internal parameters (focus, zooming) or the use of several cameras.

In response to these needs, there has recently been a significant amount of theoretical work on the structure of multi-image projection and reconstruction [8, 7, 14, 13, 17, 1, 16, 6, 11, 12, 23, 20, 21, 2]. The problem turns out to have a surprisingly rich mathematical structure, and several complimentary approaches exist. The field is developing rapidly and there is no space for a complete survey here, so I will only mention a few isolated results. The epipolar constraint (the geometry of stereo pairs) has been understood for quite a while (*e.g. c.f.* [4]). Shashua [16] and Hartley [8] developed the theory of the trivalent tensor

Submitted to BMVC'96. This work was supported by an EC HCM grant and INRIA Rhône-Alpes.

(three view constraint). Faugeras and Mourrain [6] and I [21] systematically studied the complete family of multi-image constraints (only one was unknown: a quadrilinear one). As a means to this, I developed the 'joint image' picture and a tensorial approach to multi-image projection: these nicely unify the geometric and algebraic aspects of the subject. In retrospect, the joint image picture is closely related to Sparr's 'affine shape' approach [17, 11]. For this paper, all we will need are fundamental matrices, trivalent tensors, and a some basic joint image notions.

Our research group MOVI has recently been developing several families of projective reconstruction methods based on the joint image approach. The first set of methods [18, 22] used 'projective depth recovery' equations to recover a coherent set of projective scale factors for the image points. This gives an implicit reconstruction, which can be concreteized by factorizing the matrix of image points into projection and structure matrices by a process analogous to the Tomasi-Kanade-Poelman factorization for affine structure [19, 15]. Factorization-based methods give an implicit least squares fit to all of the image data. They are simple and extremely stable, but have the serious disadvantage that they require each point to be visible in every image (modulo 'hallucination' [19]). This is practically unrealistic when there are many images covering a wide range of viewing positions.

This paper represents an attempt to overcome this problem. It describes a new family of reconstruction methods that recover projection matrices directly and linearly from estimated matching tensors, after which the scene structure can be recovered linearly by back-projecting the image measurements. The projections are estimated using 'joint image closure relations' — bilinear constraints between projections and their matching tensors, analogous to the depth recovery equations used for projective factorization, but with projection matrices replacing image points.

The closure based reconstruction methods treat all of the images uniformly, so they have the potential to be significantly more stable than the commonly used approach of initially reconstructing from two images, and then reprojecting into the other ones to estimate the remaining projection matrices. On the other hand, because they only use the image data indirectly via the matching tensors, they are not as stable as factorization based methods. The suggestion is that they will prove good replacements for the 'stereo + reprojection' methods — whose main application is probably to initialize more refined nonlinear least squares iterations — but that when tokens are visible in every image factorization will still be the best linear method.

The rest of the paper outlines the theory of the closure relations, describes the resulting reconstruction algorithms and their implementation, reports on a comparative experimental study of their performance, and ends with a short discussion.

# 2 Theory

This section sketches the theoretical background of multi-image reconstruction, and discusses the 'joint image closure relations' on which the new reconstruction methods are based. The theory is not difficult, but when more than two images are involved the equations are hard to express without using tensorial notation. We will only use this for a few identities involving the trivalent tensor, so you should be able to follow the paper without it. An appendix provides an extremely brief sketch — see [21, 20] for more details.

We will work in homogeneous coordinates. Consider the image projection equations $\lambda_i \mathbf{x}_i = \mathbf{P}_i \mathbf{X}$. The $3 \times 4$ homogeneous projection matrices $\mathbf{P}_i$, $i = 1, \ldots, m$, project a homogeneous world point $\mathbf{X}$ to its $m$ homogeneous images $\mathbf{x}_i$, up to unknown scale factors $\lambda_i$. A useful way to view this process is to stack the projections into a big $3m \times 4$ 'joint projection matrix' $\mathbf{P}$ and treat the rescaled image points as a $3m$-component 'joint image space' vector

$$\mathbf{P} \equiv \begin{pmatrix} \mathbf{P}_1 \\ \vdots \\ \mathbf{P}_m \end{pmatrix} \qquad\qquad \mathbf{P}\mathbf{X} = \begin{pmatrix} \lambda_1 \mathbf{x}_1 \\ \vdots \\ \lambda_m \mathbf{x}_m \end{pmatrix}$$

The joint projection can be viewed as a projective injection mapping the 3D world to a 3D projective subspace of the joint image space called the 'joint image' [21, 20]. The joint image is a faithful projective copy of the world expressed entirely in image coordinates. Projection from it into the individual images is a trivial forgetting of coordinates and scale factors. Projective reconstruction reduces to recovering the missing scales ('projective depths'), and is a canonical process up to the single overall choice of scales for the projections $\mathbf{P}_i$. The four columns of the joint projection matrix form a spanning basis for the joint image. The coordinates of a rescaled joint image point with respect to this basis are also the reconstructed point's homogeneous world coordinates. But neither the basis nor the world coordinates are canonical: only the geometric position of the point in the joint image is recoverable from the image data.

The above joint image geometry can be converted directly to algebra. The set of all $4 \times 4$ minors (submatrix determinants) of the joint projection encode the location of the joint image — and hence the projective camera geometry — in a well-defined algebraic sense: they are its 'Grassmann-Plücker coordinates'. In fact, the minors turn out to be just the components of the **matching tensors** between the images. These generate the multilinear constraints that tokens in different images must satisfy if they are to be the projections of a single world token, and they can also be used to transfer tokens between images, for comparison or image synthesis. There are four basic types of matching tensors: **epipoles**, **fundamental matrices**, **trivalent tensors** and **quadrivalent tensors**. These are formed from minors with respectively 3+1, 2+2, 2+1+1, and 1+1+1+1 rows from 2, 2, 3 and 4 images [16, 6, 21].

The 'joint image closure relations' that underlie the new reconstruction methods are bilinear constraints between projection matrices and the corresponding matching tensors. They basically guarantee that the projections are coherent with the joint image subspace defined by the matching tensors. Alternatively, they are an algebraic expression of the four-dimensionality of the joint image. The simplest way to derive them is to append any column of the $3m \times 4$ joint projection matrix to the existing matrix, to form a rank deficient $3m \times 5$ matrix. The $5 \times 5$ minors of this matrix vanish. Choose a minor and expand it by cofactors in the appended column: the coefficients are $4 \times 4$ minors of the original joint projection matrix, or equivalently matching tensor components. The result is a 5 term bilinear tensorial relation between the entries of one column of the joint projection matrix and the corresponding matching tensors. Closer examination reveals five basic types of relation. We will only use the simplest two here:

$$\mathbf{F}_{21}\, \mathbf{P}_1 + [\mathbf{e}_{12}]_\times \mathbf{P}_2 = \mathbf{0} \qquad \mathbf{F}\text{-e closure} \qquad (1)$$

$$\mathbf{G}_{B_2}{}^{A_1 C_3}\, \mathbf{P}_a^{B_2} + \mathbf{e}_2^{A_1}\, \mathbf{P}_a^{C_3} - \mathbf{P}_a^{A_1}\, \mathbf{e}_2^{C_3} = \mathbf{0} \qquad \text{e-}\mathbf{G}\text{-e closure} \qquad (2)$$

These relations provide constraints between matching tensors (which can be estimated from the image data) and columns of the joint projection matrix. Equation (1) contains 3 constraints of which 2 are linearly independent, while (2) contains $3 \times 3 = 9$ constraints of which 3 are linearly independent. By accumulating enough of these constraints, we can solve linearly for the four joint projection columns, up to an overall $4 \times 4$ linear transformation which amounts to a homography of the reconstructed world space. In other words, the constraints give the position of joint image — the 4D space spanned by the columns of the joint projection. Given the projections, the scene reconstruction can be completed by linearly back-projecting image structure into the world space, which amounts to solving the linear equations

$$\mathbf{x}_{ip} \wedge (\mathbf{P}_i \mathbf{X}_p) = 0 \tag{3}$$

for the world points $\mathbf{X}_p$ in terms of their images $\mathbf{x}_{ip}$ and the projection matrices $\mathbf{P}_i$.

A large number of closure constraints is available — relating the projections from any selection of 2-5 images — but these are highly redundant and only a small sample of them need be applied in practice. This paper considers only the simplest possible methods, based on minimal sets of constraints for the first two types of closure relation.

Note that the 'projective depth recovery' equations used for projective factorization [18, 22, 20] follow directly from the closure relations. Attaching a world point to each projection gives bilinear relations between matching tensors and *correctly scaled* image points:

$$\begin{aligned} \mathbf{F}_{21}\,(\lambda_1 \mathbf{x}_1) + \mathbf{e}_{12} \wedge (\lambda_2 \mathbf{x}_2) &= \mathbf{0} && (4) \\ \mathbf{G}_{B_2}{}^{A_1 C_3}\,(\lambda_2 \mathbf{x}^{B_2}) - (\lambda_1 \mathbf{x}^{A_1})\,\mathbf{e}_2^{C_3} + \mathbf{e}_2^{A_1}\,(\lambda_3 \mathbf{x}^{C_3}) &= \mathbf{0} && (5) \end{aligned}$$

Given the matching tensors and image points, the correct scale factors (projective depths) can be recovered using these relations.

Now consider the $\mathbf{F}$-$\mathbf{e}$ closure relations. In matrix form the complete set of constraints between $\mathbf{P}_i$ and $\mathbf{P}_j$ can be written

$$\begin{pmatrix} \mathbf{F}_{ij}^\top & [\mathbf{e}_{ij}]_\times \\ [\mathbf{e}_{ji}]_\times & \mathbf{F}_{ij} \end{pmatrix} \begin{pmatrix} \mathbf{P}_i \\ \mathbf{P}_j \end{pmatrix} = \mathbf{0}$$

However only two of these six constraint equations are linearly independent, so there is no need to include the second set of three equations. To fully constrain the projections, each image (except the first pair) must be related to *two* previous ones. Once this is done the $m$ image constraint matrix generically has rank $3m - 4$: its four dimensional null space is exactly the joint image (the 4D space spanned by the projection columns). Any basis for this null space provides four $3m$-component column vectors that can be regarded as the columns of a reconstructed joint projection matrix. The freedom in the choice of basis corresponds to a $4 \times 4$ nonsingular mixing of the four columns. This just amounts to a projective transformation of the reconstructed world coordinates.

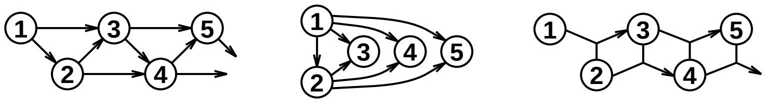
Numerically, there are several ways to calculate the null space. The experiments reported here find the singular value decomposition of the coefficient matrix and take the eigenvectors of the smallest four singular values. Forming normal equations and taking the smallest four eigenvectors gives similar results. These methods are very stable and can easily handle redundant constraints, but they are rather slow when there are many images.

If a minimal set of constraints is used (as here), the projections can also be evaluated recursively starting from two projections obtained by (say) SVD. This is essentially equivalent to the full SVD-based method, but much faster.

As an alternative to fundamental matrix based reconstruction, the $\mathbf{e}$-$\mathbf{G}$-$\mathbf{e}$ closure relation gives a trivalent tensor based reconstruction method. Each instance of this relation gives nine linear constraints relating three image projections. Three of these are linearly independent, so it is enough to connect each image to two previous ones via a single trivalent tensor. Once the constraint matrix has been built, null space estimation is done just as for the fundamental matrix case.

The trivalent tensor based method is slower than the fundamental matrix based one, but significantly more stable. In particular, for a camera moving in a straight line, the fundamental matrices and epipoles of different images coincide. This is a well-known singular case for epipolar line based token transfer, and the fundamental matrix based reconstruction method fails here too. In contrast, trivalent tensor based transfer and reconstruction are well behaved, as is reconstruction by fundamental matrix based factorization. Basically, some sort of information about position along epipolar lines is needed to stabilize the reconstruction: this can be provided by trivalent transfer, or by anchoring onto explicit image correspondences.

The closure based reconstruction methods implemented here connect each image to exactly two previous ones in a chain. The following types of chain have been considered



**F-e serial** **F-e parallel** **e-G-e serial**

The serial $\mathbf{F}$-$\mathbf{e}$ chain connects each image to the two immediately preceding ones, while the parallel $\mathbf{F}$-$\mathbf{e}$ chain connects each image to the first two (and hence uses images 1 and 2 as 'key frames'). The serial $\mathbf{e}$-$\mathbf{G}$-$\mathbf{e}$ chain connects each image to the two preceding ones using a trivalent tensor and two epipoles. The trivalent tensor used is the one based in (with its covariant index in) the middle image of the triplet.

Finally, I have glossed over an important issue. The matching tensors have 'natural' scalings resulting from their definition as minors of projection matrices, but these can not be recovered when the tensors are estimated from image data. The closure relations depend critically on having a coherent relative scaling for the matching tensors, so at least part of the missing scale information must be recovered.

This is not conceptually difficult, but the implementation details are a little messy. Basically there is one scale consistency constraint for each closed loop in the above chains. There are two possible lines of attack: (*i*) use the bilinear Grassmann relations between the matching tensors [20]; or (*ii*) enforce consistent projective depths using image measurements. I have preferred the second approach. It is simpler, and on general grounds it ought to be more stable as it is lower order, closer to the image data, and has built in redundancy. The basic idea is to use the depth recovery equations (4) or (5) to propagate the projective depths of image points around each closed loop, and then to adjust the scale factor of (one of) the matching tensors in the loop so that the closed-loop gain is unity (as it must be for self-consistency). The relative scales of each fundamental matrix and its epipoles must also be internally self-consistent, but this can be enforced by simple normalization $|\mathbf{e}_{ij}||\mathbf{e}_{ji}| = 1$, $|\mathbf{F}_{ij}|^2 = 2$ (with a suitable choice of signs).

# 3 Implementation

Now we summarize the reconstruction algorithms, and discuss a few important implementation details. The $\mathbf{F}$-$\mathbf{e}$ closure based algorithm has the following steps:

0) Extract and match features between images.

1) Standardize the image coordinates (see below).

2) Estimate fundamental matrices and epipoles connecting each image to at least two previous ones.

3) Correct the scales of the fundamental matrices and epipoles using (4), as discussed in section 2.

4) Build the matrix of equations (1) and use SVD to find its 4D null space.

5) Extract the projection matrices from the null space column vectors.

6) Back-project and solve for 3D structure using (3).

7) Unstandardize the projection matrices (see below).

The trivalent tensor based algorithm follows the same pattern, except that: (*i*) both point and line features can be used to estimate the trivalent tensors; (*ii*) equation 5 is used to correct the trivalent scaling, and equation (2) to build the constraint matrix.

The current implementation uses linear methods to estimate fundamental matrices [9] and trivalent tensors. With properly standardized coordinates these turn out to be very stable and surprisingly accurate. Using a nonlinear least squares iteration to refine the estimates marginally improves the stability of (for example) the long serial chains of the $\mathbf{F}$-$\mathbf{e}$ method, but not enough to change the basic conclusions. The linear method for $\mathbf{F}$ includes a final $3 \times 3$ SVD to enforce det$(\mathbf{F}) = 0$ and calculate the epipoles. The epipoles for the $\mathbf{e}$-$\mathbf{G}$-$\mathbf{e}$ method are found from $\mathbf{G}_{B_2}{}^{A_1 C_3}$ and the image data by using the $\mathbf{e}$-$\mathbf{G}$-$\mathbf{e}$ depth recovery relation 5, taking the vector product with $\mathbf{x}^{A_1}$ to kill one of the three terms, and factorizing. (The epipoles can also be found bilinearly from $\mathbf{G}$ alone, but for maximum stability I prefer to use a linear method based on the image data).

## 3.1 Image Standardization

To get accurate results from the above reconstruction algorithms, it is *essential* to work in a well-adapted coordinate system. This is standard practice for numerical methods, but it is particularly important when there are implicit least-squares trade-offs between redundant constraints, as here. If some components of the input vectors are typically much larger than others — for example when homogeneous pixel coordinates $(x, y, z) \sim (10^2, 10^2, 1)$ are used — some constraints have a much higher implicit weight than others and this significantly distorts the estimated solution. Hartley has underlined the importance of this for fundamental matrix estimation [9], and it is equally true for reconstruction. Figure 1 shows the effect of standardization: pixel coordinates (scale $\sim 256$) give reconstructions hundreds of times worse than well standardized ones (scale $\sim 1$). The error also rises rapidly at scales below $10^{-1}$, owing to (32 bit) floating point truncation error.

There are many possible standardization schemes, but in practice it makes little difference which is used. Here, the homogeneous pixel coordinates $[0, 512] \times [0, 512] \times [1]$ are scaled uniformly into the unit square $[-1, 1] \times [-1, 1] \times [1]$, and then the resulting 3-component vectors are normalized to norm 1. This is easy to apply, independent of the

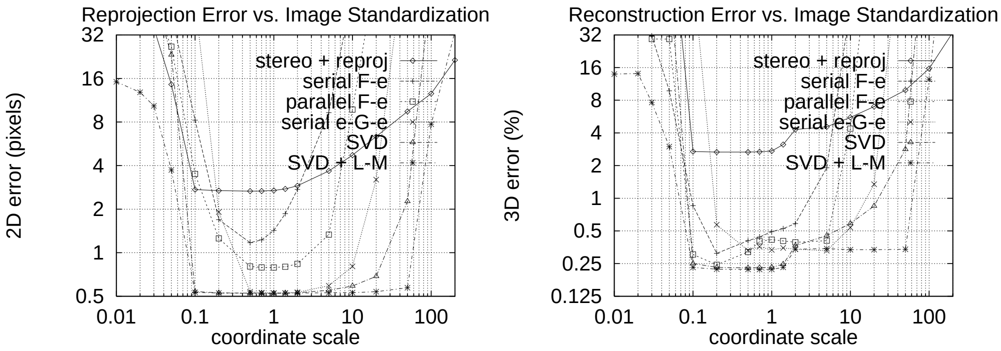



*Figure 1*: Mean reprojection and reconstruction error *vs.* image coordinate standardization.

image, and works equally well for visible points, line vectors and points (*e.g.* epipoles) at infinity.

# 4 Experiments

To quantify the performance of the algorithms, I have run a series of simulations using synthetic data. The algorithms have also been run successfully on hand-matched points extracted from real images. An implementation on 'live' images is in progress.

The simulations reported here are based on trial scenes consisting of random 3D points in the unit cube. These are viewed by identical perspective cameras spaced evenly along a $90^\circ$ arc of radius 2, looking directly at the centre of the scene. (These are ideal conditions for accurate reconstruction, but many other configurations have also been tested, including infinitesimal viewing angles and distant scenes with negligible perspective). When cameras are added, their spacing is decreased so that the total range of viewing angles remains the same. To model uncertainty, the positions of the projected image points are perturbed by uniform random noise. Mean-square (and median and maximum) 2D reprojection and 3D reconstruction errors are accumulated over 50 trials. The 3D error is the residual after projective alignment of the reconstruction with the scene. Unless otherwise stated, default values of 10 views, 50 points and $\pm 1$ pixel noise are used.

The new techniques under test are the serial and parallel chain $\mathbf{F}$-$\mathbf{e}$ closure, and the serial chain $\mathbf{e}$-$\mathbf{G}$-$\mathbf{e}$ closure based methods. For comparison, several existing techniques are also shown. The 'stereo + reprojection' method simply reconstructs the scene from the first two images, and then reprojects to estimate the projection matrices for the remaining images. This can be very inaccurate, but it is often used in the literature to initialize more refined reconstruction techniques (*e.g. c.f.* [14, 7]). The other techniques shown are SVD-based projective factorization [18, 22] and a Levenberg-Marquardt-like nonlinear least squares algorithm initialized from this.

Figure 2 summarizes the results. All of the methods scale linearly with noise level and improve as more points are added. The simple 'stereo + reprojection' method rapidly worsens as images are added, while the others initially improve. But only the factorization-based methods continue to perform well for 50 or more images. The serial chain $\mathbf{F}$-$\mathbf{e}$ method can only handle about 10 images, the parallel chain one 20, and the serial chain

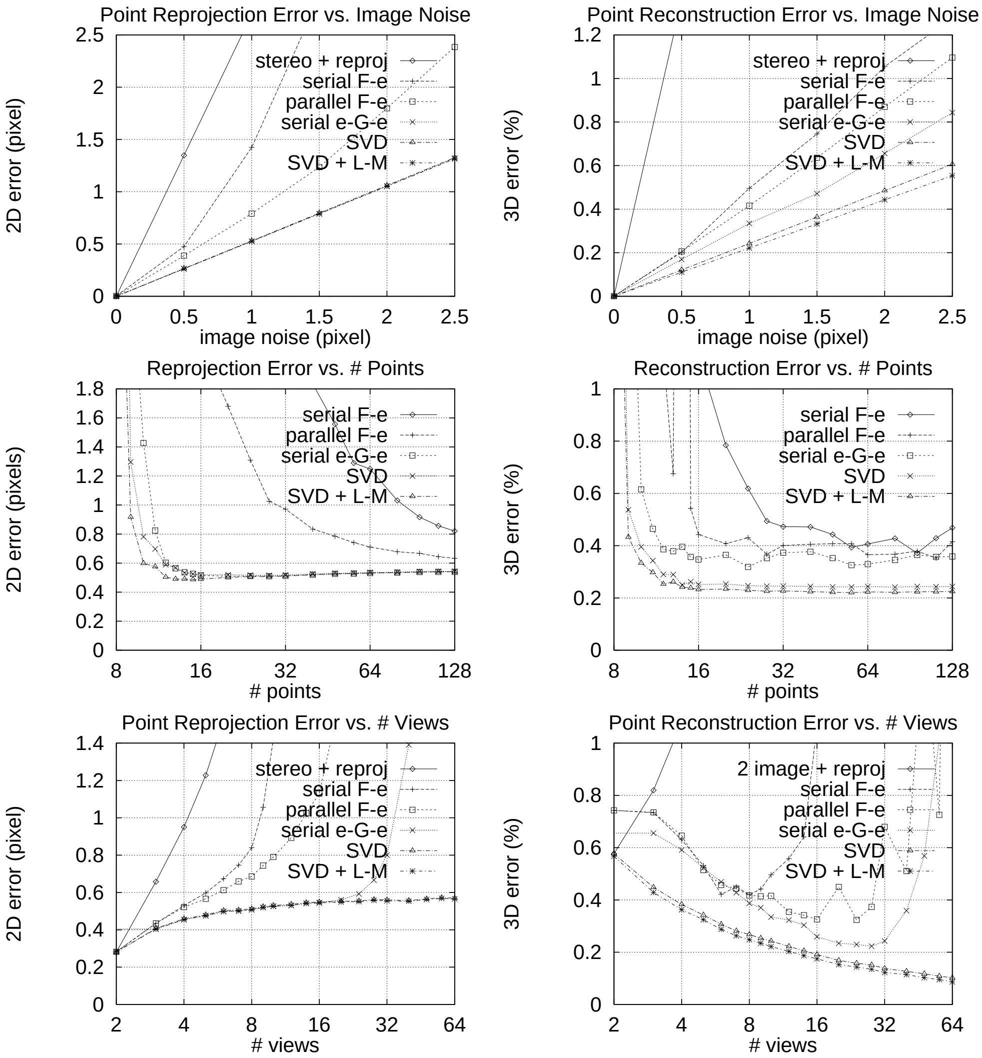



*Figure 2*: Mean reprojection and reconstruction error *vs.* noise, number of points and number of views.

trivalent method about 30-40. The trivalent method is significantly better than the fundamental matrix ones, and parallel methods are consistently more stable than serial ones. This suggests that the instabilities arising from solving long chains of equations are far greater than any biases introduced by working from a few 'key' images.

The factorization method is notable for being linear yet close to optimal. It is based on $\mathbf{F}$-$\mathbf{e}$ depth recovery (4) — essentially the same equations as the $\mathbf{F}$-$\mathbf{e}$ closure based method, but applied directly to the image points rather than to the projections. Clearly, the direct use of image data gives a significant improvement in accuracy. Unfortunately, the factorization method is practically rather limited as it requires every token to be visible in every image: this is why the closure-based methods were developed.

The $\mathbf{F}$-$\mathbf{e}$ closure based techniques fail when all of the camera centres lie on a line, while the $\mathbf{e}$-$\mathbf{G}$-$\mathbf{e}$ closure method seems to be stable whenever the camera centres are distinct. However, all of the techniques (but especially the fundamental matrix based ones) give very poor results for points near the axis of fronto-parallel motion, since there is no stereo baseline on which to base reconstruction there.

# 5 Discussion

The closure relation based projective reconstruction techniques work reasonably well in practice. They are more accurate than the common 'reconstruct from 2 images and reproject for the other projections' paradigm, but can not compete with projective factorization when that is feasible. However they do not require features to be tracked over all the images. They do not need to single out 'privileged' features or images. However, short chains of closure relations turn out to be significantly more stable than long ones, so in practice it is probably sensible to relate all of the images to a few 'key' ones. Trivalent techniques are significantly more stable than fundamental matrix based ones, but somewhat slower.

Future investigation will include the implementation of the methods on 'live' image data with fast recursive solutions of the reconstruction equations, and study the stabilizing effect of including redundant constraints.

**Appendix: Tensors**

Tensors are just multidimensional arrays of components, direct analogues of vectors (1D arrays) and matrices (2D arrays). For multi-index objects it is easiest to write out the indices explicitly, rather than eliding them as in the usual matrix-vector notation. Indices can belong to different spaces: world-space indices are denoted $a, b, c \ldots$ and image $i$ indices $A_i, B_i, C_i \ldots$ A basic operation is 'contraction' — summing a corresponding pair of indices over all their values. Contractions are so common that we omit the summation sign: any repeated indices in a term are implicitly summed over. In projective space each index comes in two varieties — covariant (subscript) and contravariant (superscript). These have opposite transformation laws under changes of coordinates, and only contractions over a covariant-contravariant pair are meaningful (coordinate invariant). World and image points have contravariant indices — $\mathbf{x}^a$ and $\mathbf{x}^{A_i}$. Image projections are $3 \times 4$ matrices $\mathbf{P}_a^{A_i}$. The epipole of camera 1 in image 2 is denoted $\mathbf{e}_1^{A_2}$. The fundamental matrix between images 1 and 2 is $\mathbf{F}_{A_1 B_2}$, and the trivalent tensor between images 2 and 3 based in image 1 is $\mathbf{G}_{A_1}{}^{B_2 C_3}$.


**References**

[1] S. Carlsson. Multiple image invariance using the double algebra. In J. Mundy, A. Zissermann, and D. Forsyth, editors, *Applications of Invariance in Computer Vision*, volume 825 of *Lecture Notes in Computer Science*. Springer-Verlag, 1994.

[2] S. Carlsson. Duality of reconstruction and positioning from projective views. In P. Anandan, editor, *IEEE Workshop on Representation of Visual Scenes*. IEEE Press, 1995.

[3] O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In G. Sandini, editor, *European Conf. Computer Vision*, Santa Margherita Ligure, Italy, May 1992. Springer-Verlag.

[4] O. Faugeras. *Three-dimensional computer vision: a geometric viewpoint*. MIT Press, 1993.

[5] O. Faugeras, Q.-T. Luong, and S. J. Maybank. Camera self calibration: Theory and experiments. In G. Sandini, editor, *European Conf. Computer Vision*, Santa Margherita Ligure, Italy, May 1992. Springer-Verlag.

[6] O. Faugeras and B. Mourrain. On the geometry and algebra of the point and line correspondences between $n$ images. In E. Grimson, editor, *IEEE Int. Conf. Computer Vision*, pages 951–6, Cambridge, MA, June 1995.

[7] R. Hartley. Euclidean reconstruction from multiple views. In $2^{nd}$ *Europe-U.S. Workshop on Invariance*, pages 237–56, Ponta Delgada, Azores, October 1993.

[8] R. Hartley. Lines and points in three views – an integrated approach. In *Image Understanding Workshop*, Monterey, California, November 1994.

[9] R. Hartley. In defence of the 8-point algorithm. In E. Grimson, editor, *IEEE Int. Conf. Computer Vision*, pages 1064–70, Cambridge, MA, June 1995.

[10] R. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *IEEE Conf. Computer Vision & Pattern Recognition*, pages 761–4, Urbana-Champaign, Illinois, 1992.

[11] A. Heyden. Reconstruction from image sequences by means of relative depths. In E. Grimson, editor, *IEEE Int. Conf. Computer Vision*, pages 1058–63, Cambridge, MA, June 1995.

[12] A. Heyden and K. Åström. A canonical framework for sequences of images. In *IEEE Workshop on Representations of Visual Scenes*, Cambridge, MA, June 1995.

[13] P. F. McLauchlan and D. W. Murray. A unifying framework for structure and motion recovery from image sequences. In E. Grimson, editor, *IEEE Int. Conf. Computer Vision*, pages 314–20, Cambridge, MA, June 1995.

[14] R. Mohr, B. Boufama, and P. Brand. Accurate projective reconstruction. In $2^{nd}$ *Europe-U.S. Workshop on Invariance*, page 257, Ponta Delgada, Azores, October 1993.

[15] C. J. Poelman and T. Kanade. A parapersective factorization method for shape and motion recovery. In J-O. Eklundh, editor, *European Conf. Computer Vision*, pages 97–108, Stockholm, 1994. Springer-Verlag.

[16] A. Shashua. Algebraic functions for recognition. *IEEE Trans. Pattern Analysis & Machine Intelligence*, 1995.

[17] G. Sparr. A common framework for kinetic depth, reconstruction and motion for deformable objects. In J-O. Eklundh, editor, *European Conf. Computer Vision*, pages 471–82, Stockholm, 1994. Springer-Verlag.

[18] P. Sturm and B. Triggs. A factorization based algorithm for multi-image projective structure and motion. In *European Conf. Computer Vision*, Cambridge, England, 1996. Springer-Verlag.

[19] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *Int. J. Computer Vision*, 9(2):137–54, 1992.

[20] B. Triggs. The geometry of projective reconstruction I: Matching constraints and the joint image. Submitted to *Int. J. Computer Vision*, March 1995.

[21] B. Triggs. Matching constraints and the joint image. In E. Grimson, editor, *IEEE Int. Conf. Computer Vision*, pages 338–43, Cambridge, MA, June 1995.

[22] B. Triggs. Factorization methods for projective structure and motion. Submitted to *IEEE Conf. Computer Vision and Pattern Recognition.*, 1996.

[23] M. Werman and A. Shashua. The study of 3D-from-2D using elimination. In E. Grimson, editor, *IEEE Int. Conf. Computer Vision*, pages 473–9, Cambridge, MA, June 1995.