A Comparison of Projective Reconstruction Methods for Pairs of Views

Charlie Rothwell, Gabriella Csurka, Olivier Faugeras

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

A Comparison of Projective Reconstruction Methods for Pairs of Views

Charlie ROTHWELL

Gabriella CSURKA

Olivier FAUGERAS

N° 2538

Avril 1995

PROGRAMME 4

ISSN 0249-6399

INRIA
SOPHIA ANTIPOLIS

A Comparison of Projective Reconstruction Methods for Pairs of Views

Charlie ROTHWELL
Gabriella CSURKA
Olivier FAUGERAS

Programme 4 — Robotique, image et vision
Projet Robotvis

Rapport de recherche n ° 2538 — Avril 1995 — 42 pages

Abstract: Recently, different approaches for uncalibrated stereo have been suggested which permit projective reconstructions from multiple views. These use weak calibration which is represented by the epipolar geometry, and so we require no knowledge of the intrinsic or extrinsic camera parameters. In this paper we consider projective reconstructions from pairs of views, and compare a number of the available methods.

Projective stereo algorithms can be categorized by the way in which the 3D coordinates are computed. The first class is similar to traditional stereo algorithms in that the 3D world geometry is made explicit; the initial phase of the processing always involves the estimation of the camera matrices from which the 3D coordinates are computed. We show how the camera matrices can be computed either from point correspondences, or how they are constrained by the fundamental matrices. The second class of algorithms are based on implicit image measurements which are used to compute projective invariants from image correspondences. The invariants are based on the Cayley algebra and on cross ratios. In all cases, the invariants are functionally dependent on the 3D coordinates.

We report on the stabilities of the different methods using a range of meaningful synthetic and real images. From these we can conclude which methods are most likely to be of use in applications that are dependent on 3D uncalibrated reconstructions.

Key-words: Stereo, Shape and object recognition, Invariants and geometry

(Résumé : tsvp)

Unité de recherche INRIA Sophia-Antipolis
2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex (France)
Téléphone : (33) 93 65 77 77 – Télécopie : (33) 93 65 77 65

Une Etude Comparative des Différentes Méthodes de Reconstruction Projective à Partir d'une Paire de Vues

Résumé : En vision stéréoscopique non calibrée, différentes approches permettant la reconstruction projective à partir de plusieurs images ont été suggérées. Ces approches utilisent la calibration faible, équivalente à la géométrie épipolaire, et donc la connaissance des paramètres intrinsèques ou extrinsèques des caméras n'est pas nécessaire. Dans cet article, nous étudions le problème de la reconstruction projective d'une scène à partir de deux images et nous comparons plusieurs méthodes disponibles.

Les algorithmes de stéréoscopie projective peuvent être classés selon les méthodes utilisées pour le calcul des coordonnées projectives des points reconstruits.

La première classe, contient les algorithmes traditionnels de stéréoscopie où la reconstruction est faite explicitement, c'est-à-dire le calcul des coordonnées tridimensionnelles s'effectue à l'aide des matrices des projections précédemment estimées. Ces matrices des projections sont contraintes par la matrice fondamentale qui représente la géométrie épipolaire et peuvent être calculées à partir de correspondances de points.

La deuxième classe d'algorithmes est basée sur des mesures implicites dans les images. L'idée est que les invariants projectifs de configurations de points et droites dans l'espace peuvent être calculés à partir d'une paire d'images en utilisant soit l'algèbre de Grassmann-Cayley soit des birapports. Dans les deux cas nous donnons les fonctions qui lient les coordonnées tridimensionnelles à ces invariants.

A l'aide de données synthétiques et réelles, les différentes méthodes ont été comparées afin de voir lesquelles sont les plus stables et les plus adaptées pour les applications où une reconstruction projective tridimensionnelle est suffisante.

Mots-clé : Stéréoscopie, Reconnaissance d'objets, Invariants géométriques

Contents

1 Introduction

Since the introductory papers of Faugeras [7] and Hartley, *et al.* [14] on computing projective structure using uncalibrated cameras, there has been a keen interest in developing reliable algorithms for uncalibrated stereo. Examples of this type of work are the treatise by Mohr, *et al.* [20, 21], Beardsley, *et al.* [2], and Shashua [30]. In this paper we collect together some of the different approaches that use pairs of views of a scene and compare their effectiveness at computing three-dimensional structure. Consequently, we are able to determine which algorithms are likely to be of most use to a scene reconstruction or to a 3D object recognition system.

It was shown in [7, 14] that one is able to compute a projective representation of the world from point correspondences in pairs of images when *no* initial assumption is made about either the intrinsic or extrinsic parameters of the cameras. This approach has been called *uncalibrated stereo* and is a projective extension of the well understood classical stereo techniques of Grimson [11], Ayache and Faugeras [1] and Pollard, *et al.* [24]. For the uncalibrated case there is also an equivalence to the *epipolar structure* [17] which is called *weak calibration* and is represented by `F`, the *fundamental matrix* [5, 8, 19]. `F` provides a correspondence structure between pairs of images which we discuss in more detail later. However, for the purpose of this paper we will assume that `F` is known.[1] We believe that computing structure without explicit camera calibration is more robust than using calibration because we need not make any (possibly incorrect) assumptions about the Euclidean geometry (remembering that calibration is itself often erroneous).

Given `F`, there are a number of ways to proceed towards the recovery of three-dimensional structure. It is these methods which are discussed in this paper. At first we develop the reconstruction algorithms for point correspondences, though we also demonstrate that the algorithms work for lines. In all, we examine five different approaches that can be divided into two distinct classes: *explicit reconstruction* and *implicit reconstruction*. The former compute the three dimensional coordinates of points directly using geometric arguments made within a three dimensional frame. As will be seen later, conventional stereo algorithms fall into this class. The implicit methods compute the structure purely from image measurements; in fact the approach is essentially algebraic and uses functions only of two dimensional images measurements (though of course both methods have very strong geometric foundations). More precisely, the functions are *three-dimensional invariants* of the point sets and the camera configuration. The invariants are functionally dependent on the 3D coordinates of the points and we actually compute invariants that are precisely equal to the 3D coordinates. In this manner we demonstrate a degree of mathematical equivalence between the explicit and implicit approaches.

In summary:

[1] The `F` matrix we use is actually computed automatically using the algorithm of Deriche, et al. [5]. Alternatively we could use the similar method of Torr [33].

1. **Explicit reconstruction:** we examine three different methods that rely on the computation of the projective camera matrices of the camera configuration. These cameras are represented by 3×4 matrices $\mathtt{P}_i$ similar to those used by Roberts [26]. We demonstrate two independent routes to the computation of the camera matrices. Knowledge of the cameras immediately facilitates the estimates of 3D structure.
2. **Implicit reconstruction:**
   (a) We compute projective invariants of three-dimensional points using the *Cayley algebra*, or the *Double algebra*, described by Carlsson [3]. These invariants are expressible as functions of the image measurements.
   (b) We measure projective invariants (cross ratios [6]) of sets of planes that pass through the three-dimensional point set. Again, these invariants are measured directly in the images.

In Section 2 we introduce the basic framework that is common to all of the reconstruction methods, and discuss some of the problems associated with working in a projective space. Then, in Section 3 we discuss the theory behind the explicit reconstruction approaches. The details of the Cayley algebra and the cross ratio algorithms are given in Sections 4 and 5 respectively. In Section 6 we compare the five methods on both real and synthetic data with additive noise. The reason for evaluating the performance on synthetic data is because we can do a very large number of tests with precise knowledge of the correct reconstruction results (i.e. we have a ground truth world structure). We actually see that the move from synthetic to real data makes no significant difference to our confidence in each method.

2 Projective reconstruction

In this section we summarize the camera geometry we use and introduce aspects of projective geometry. However, due to the limitations of space we omit many fundamentals of projective geometry and linear projection which form the framework of our computations; the reader is advised to refer to either the book by Mundy and Zisserman [22] or the one by Faugeras [9] for a complete introduction.

2.1 The camera model and the standard basis

A pinhole camera model is used which is represented mathematically by a linear projection from three-dimensional space into each two-dimensional image. For convenience, these spaces are represented projectively as $\mathcal{P}^3$ and $\mathcal{P}^2$. The actual projection is represented by:

$$\mathbf{x} \quad = \quad \mathtt{P}\,\mathbf{X}, \tag{1}$$

where $\mathbf{x}$ is the planar image point $(x, y, z)^\top$ and $\mathbf{X}$ the three-dimensional world point $(p, q, r, s)^\top$. The projection matrix $\mathtt{P}$ (which is 3×4), is known as the *camera matrix*.

Note that in a Euclidean frame we set $z = 1$ and $s = 1$; in the sequel we work in general projective space and so the only restriction we place on the coordinates is that $\mathbf{x} \neq \mathbf{0}$ and $\mathbf{X} \neq \mathbf{0}$. For the explicit reconstructions our goal is to derive a set of constraints on $\mathbf{X}$ based around eqn (1) once we have computed estimates for the unknown camera matrices.

The weak calibration between the images is represented by fundamental matrices and epipoles [5, 8, 19]. The image of a point $\mathbf{x}_1$ observed in the first image is the point $\mathbf{x}_2$ in the second image. $\mathbf{x}_2$ is constrained to lie on the line $\mathtt{F}_{12}\mathbf{x}_1$ where $\mathtt{F}_{12}$ is a 3×3 rank 2 matrix and is called the fundamental matrix. The epipolar constraint thus takes the form of:

$$\mathbf{x}_2^\top \, \mathtt{F}_{12} \, \mathbf{x}_1 \quad = \quad 0.$$

There is an equivalent relationship of the form $\mathbf{x}_1^\top \mathtt{F}_{21} \mathbf{x}_2 = 0$ mapping points from the second image to the first (where $\mathtt{F}_{21} = \mathtt{F}_{12}^\top$). There is a special point in the first image called $\mathbf{e}_{12}$ which is the epipole. This point is defined by $\mathtt{F}_{12}\mathbf{e}_{12} = 0$. Likewise, an epipole exists in the second image such that $\mathtt{F}_{21}\mathbf{e}_{21} = 0$.

Throughout this paper we make use of the standard bases in $\mathcal{P}^2$ and $\mathcal{P}^3$. All coordinates within 2D or 3D projective space can be represented a linear combinations of the basis points. The *image standard basis* is composed of the four points:

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad \mathbf{e}_4 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

The *world standard basis* consists of:

$$\mathbf{E}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{E}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{E}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{E}_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \quad \mathbf{E}_5 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

2.2 Projective constraints

Previous stereo algorithms have been associated with Euclidean frames and their implicit metrics. Projective spaces do not immediately have metrics associated with them.[2] The lack of a metric means that it is hard to formulate minimization procedures for the solution of over-constrained systems. In short, this will mean that any projective reconstruction process we employ that involves a minimization is unlikely to be invariant to a change of projective frame. Due to the fact that in general we pick an arbitrary frame in which to

[2] A metric can be associated with a projective space if we have knowledge of an object such as a quadric [28, 31].

do the reconstruction, we might find that that quality of the reconstructions vary dramatically. Beardsley, *et al.* [2] overcome the lack of a metric through the introduction of a *quasi-Euclidean* frame which constitutes the use of an estimate of the camera calibration. However, we find that a variety of different minimization procedures actually produce results comparable to those of [2] *without making any assumption* about the camera calibration; these methods are discussed in Section 3.

Rather than ignoring the problem caused by the lack of a metric, we can temporarily overcome it by making sure that all of the sets of corresponding points do actually lie in projective correspondence (and so the constraint systems we solve have exact solutions). This can be done by adjusting the locations of the image observations so that all points actually lie on their corresponding epipolar lines. Obviously there is then the difficulty of choosing how to move the points. Perhaps one could select a pair of points $\{\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2\}$ that are representative of the original image points $\{\mathbf{x}_1, \mathbf{x}_2\}$, such that the following distance is minimized:

$$\min \left(|\mathbf{x}_1 - \hat{\mathbf{x}}_1|^2 + |\mathbf{x}_2 - \hat{\mathbf{x}}_2|^2 \right),$$

where we also impose $\hat{\mathbf{x}}_2^\top \, \mathtt{F}_{12} \, \hat{\mathbf{x}}_1$, that is, the epipolar constraints are satisfied. One may again question whether this is a suitable criterion to use for minimization as its relationship to the three-dimensional structure is unclear. However, considering how poorly actual errors are modelled, minimizing such a measure is not *so* unwise as it does contain at least some meaningful metric information.

2.2.1 Minimizing image error

A first approach to estimating structure for a point set would be to choose the 3D coordinates so that the distances between the projection of each point in the images and their actual observation are minimized. For each point this would take the form:

$$\min \left(\sum_i^2 [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] \right), \tag{2}$$

where the 3D point is $\hat{\mathbf{X}}$ and its projections to each image $(\hat{x}_i, \hat{y}_i)$. This type of minimization is intuitively correct because without knowledge of the camera parameters we know that the image is the only place in which we can the recover metric information needed for the minimization.[3] The minimization may be done explicitly with respect to the 3D coordinates of the point (represented affinely by $\hat{\mathbf{X}} = (\hat{X}, \hat{Y}, \hat{Z}, 1)^\top$). If we denote the element of $\mathtt{P}$ in the i^{th} row and j^{th} column by p_{ij}, $\hat{x}_i$ and $\hat{y}_i$ take on the non-linear forms:

[3] Though this assumes calibration to the extent of understanding how the signal is extracted from the CCD due to the effect of a possible non-isotropic scaling of the image.

$$\hat{x} = \frac{p_{11}\hat{X} + p_{12}\hat{Y} + p_{13}\hat{Z} + p_{14}}{p_{31}\hat{X} + p_{32}\hat{Y} + p_{33}\hat{Z} + p_{34}} \qquad \text{and} \qquad \hat{y} = \frac{p_{21}\hat{X} + p_{22}\hat{Y} + p_{23}\hat{Z} + p_{24}}{p_{31}\hat{X} + p_{32}\hat{Y} + p_{33}\hat{Z} + p_{34}}.$$

In practice, these expression make the minimization of the constraint given in eqn (2) too difficult to use, except for certain special camera geometries. For instance, if we are able to make p_{31}, p_{32} and p_{33} all equal to zero (that is an affine camera), then we can reduce the minimization to being linear. However, one is generically unable to project a pair of cameras into this configuration.[4]

However, the process can be simplified dramatically using the observation that the distance constraint can be interpreted as the requirement to locate points in each image that satisfy the epipolar geometry and lie as close as possible to each of the image observations. In other words we need only minimize eqn (2) subject to $\hat{\mathbf{x}}_2^\top \mathtt{F}_{12}\hat{\mathbf{x}}_1 = 0$. The minimization is much easier to handle and has been considered in detail in the recent paper by Hartley and Sturm [16]. There they show how the optimal points can be found by solving sixth order polynomials in a single variable (which can be solved exactly). Although not discussed in detail in this paper, we have re-implemented this approach and found that it performs well, though in general it is not the best reconstruction method. What we do discuss in this report is a similar method that performs the same minimisation numerically, and so produces equivalent results except in cases in which local minima are encountered (which we try to prevent using a good initialisation to the minimisation process). This minimization procedure is discussed in Section 3.2 (as reconstruction method 2), and is again only one route to the computation of structure which in practice the one with the best performance.

Now we have discussed the preliminaries and noted some of the problems, we are free to review the different reconstruction methods in detail.

3 The constraints provided by the camera matrices

3.1 Using the standard basis — method 1

Given two images and a set of image correspondences, our preliminary goal for all of the explicit reconstruction methods is to compute the camera matrices corresponding to the projections from the world to each image; these are represented by $\mathtt{P}_1$ and $\mathtt{P}_2$. *A priori* we have *no* knowledge of the form of these two matrices, though they can be derived quite simply from a series of correspondences. As noted by [7, 14], we can only expect to recover a projective representation of space, and so we can choose a totally arbitrary projective frame in which to work (so long as we satisfy certain genericity constraints, that is, none three of the points that define the projective frame are collinear, and no four coplanar).

[4] Remember that the reconstructions we achieve are up to a projective transformation of space. Even with this, we are in the first instance still left with a difficult minimization.

The three-dimensional projective frame is determined by fixing five world points (which are observed through known correspondences in each image) to the standard 3D basis (the standard basis represents a generic configuration of points). In each image the representations are defined so that the first four points project to the image standard image basis:[5]

$$\left.\begin{array}{rcl} \rho_i\ \mathbf{e}_i & = & \mathtt{P}_1\ \mathbf{E}_i \\ \sigma_i\ \mathbf{e}_i & = & \mathtt{P}_2\ \mathbf{E}_i \end{array}\right\} \quad \text{for } i \in \{1,\ldots,4\},$$

where the ρ_i and σ_i are non-zero scalars that account for the use of homogeneous coordinates. This places the following constraints on the camera matrices:

$$\mathtt{P}_1 = \begin{bmatrix} \rho_1 & 0 & 0 & \rho_4 \\ 0 & \rho_2 & 0 & \rho_4 \\ 0 & 0 & \rho_3 & \rho_4 \end{bmatrix} \quad \text{and} \quad \mathtt{P}_2 = \begin{bmatrix} \sigma_1 & 0 & 0 & \sigma_4 \\ 0 & \sigma_2 & 0 & \sigma_4 \\ 0 & 0 & \sigma_3 & \sigma_4 \end{bmatrix}. \tag{3}$$

Using eqn (3) for the projection of the fifth point into the first image $\mathbf{x}_1 = (\alpha_1, \beta_1, \gamma_1)^\top$, and comparing with $\rho_5 \mathbf{x}_1 = \mathtt{P}_1\ \mathbf{E}_5$, provides the constraints:

$$\rho_1 + \rho_4 = \rho_5 \alpha_1, \qquad \rho_2 + \rho_4 = \rho_5 \beta_1, \qquad \rho_3 + \rho_4 = \rho_5 \gamma_1.$$

We can do likewise for the projection of the fifth point into the second image, $\mathbf{x}_2 = (\alpha_2, \beta_2, \gamma_2)^\top$. Then, making the substitutions $\mu_1 = \rho_5$ and $\nu_1 = \rho_4$, and $\mu_2 = \sigma_5$ and $\nu_2 = \sigma_4$, yields:

$$\mathtt{P}_i = \begin{bmatrix} \mu_i\alpha_i - \nu_i & 0 & 0 & \nu_i \\ 0 & \mu_i\beta_i - \nu_i & 0 & \nu_i \\ 0 & 0 & \mu_i\gamma_i - \nu_i & \nu_i \end{bmatrix}, \qquad i \in \{1,2\}. \tag{4}$$

Subsequently, we see that the $\mathtt{P}_i$ depend only of the four unknown parameters μ_i and ν_i. However, the camera matrices are defined only up to a scale factor, and so there are really only two unknown parameters $x_i = \mu_i/\nu_i$. These can be computed straightforwardly when we take into account the epipolar geometry and the corresponding epipoles $\mathbf{e}_{12}$ and $\mathbf{e}_{21}$ in each image:

$$x_1 = \frac{\mathbf{e}_{21} \cdot (\mathbf{x}_1 \times \mathbf{x}_2)}{\mathbf{v}_1 \cdot (\mathbf{x}_1 \times \mathbf{x}_2)} \quad \text{and} \quad x_2 = \frac{\mathbf{x}_1 \cdot \mathbf{v}_2}{\mathbf{x}_2 \cdot \mathbf{v}_2}\, x,$$

where:

[5] Doing this is not so easy. We must actually test that no four points are coplanar using a construction described by Robert [25] or Gros [12]. This test assumes only information about weak calibration.

$$\mathbf{v}_1 = \begin{pmatrix} \alpha_1 \ \mathbf{e}_{21_x} \\ \beta_1 \ \mathbf{e}_{21_y} \\ \gamma_1 \ \mathbf{e}_{21_z} \end{pmatrix} \qquad \text{and} \qquad \mathbf{v}_2 = \begin{pmatrix} (\gamma_1 - \beta_1) \ \mathbf{e}_{21_y} \ \mathbf{e}_{21_z} \\ (\alpha_1 - \gamma_1) \ \mathbf{e}_{21_z} \ \mathbf{e}_{21_x} \\ (\beta_1 - \alpha_1) \ \mathbf{e}_{21_x} \ \mathbf{e}_{21_y} \end{pmatrix}.$$

Full details of the computation are given in [7].

Computing x_1 and x_2 in this way constrains the two camera matrices entirely. However, we derived the camera forms above assuming that the image measurements were made in a frame such that four reference points have pre-specified coordinates (the planar standard basis). For convenience we can represent the mapping from the actual image coordinates to this frame in each images by the matrices $\mathtt{A}_1$ and $\mathtt{A}_2$, and in the sequel work with the following cameras and actual coordinates:

$$\mathtt{P}'_i \quad = \quad \mathtt{A}_i^{-1} \ \mathtt{P}_i, \qquad i \in \{1, 2\}.$$

We shall in face also drop the primes and assume that $\mathtt{P}_1$ and $\mathtt{P}_2$ represent the above cameras mapping from the three-dimensional standard basis into the images.

3.1.1 Least-square estimate through the pseudo-inverse

From the above estimates for the camera matrices we can compute the structure for a three-dimensional point through the construction of a linear system of constraints on the three-dimensional coordinates, and then we solve them using the pseudo-inverse. Given each image constraint of the form $k_i \mathbf{x}_i = \mathtt{P}_i \mathbf{X}$, we may eliminate the k_i to give a pair of constraints on the elements of $\mathbf{X}$:

$$\begin{bmatrix} \lambda_i & 0 \\ 0 & \mu_i \end{bmatrix} \begin{bmatrix} p_{11} - p_{31}x_i & p_{12} - p_{32}x_i & p_{13} - p_{33}x_i \\ p_{21} - p_{31}y_i & p_{22} - p_{32}y_i & p_{23} - p_{33}y_i \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} p_{34}x_i - p_{14} \\ p_{34}y_i - p_{24} \end{pmatrix}. \quad (5)$$

Here the λ_i and μ_i are weights that may be chosen freely for each point (both are set to unity in all of the examples that we give). Then, from a pair of images we form the constraint system:

$$\mathtt{A} \ \mathbf{X}_3 = \mathbf{b}, \qquad \text{where } \mathbf{X}_3 = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}.$$

$\mathtt{A}$ is a 4×3 matrix. The system is over constrained because there are only three unknowns. We solve for the 3D coordinates by minimizing $|\mathbf{b} - \mathtt{A} \ \mathbf{X}_3|^2$, which results in the use of the left pseudo-inverse:

$$\mathbf{X}_3 \quad = \quad (\mathtt{A}^\top \mathtt{A})^{-1} \mathtt{A}^\top \mathbf{b},$$

or $\mathbf{X}_3 = \mathtt{A}^+ \mathbf{b}$. Note that using the pseudo-inverse provides no good understanding of the measure is being minimized, except for certain special cases. For instance, if as discussed previously both cameras are affine, then the cost function equates to the minimization of the projected image distances. However, as noted before, it is rare that both cameras can be put into an affine frame (though the distance can be made approximately valid when the cameras tend towards such a configuration).

Up to this point we have not indicated how the basis points $\mathbf{e}_i$, $i \in \{1, \ldots, 4\}$ are chosen in the image, we have just assumed that the initial correspondences are given and then used to constrain the rest of the solution. Proceeding in this manner is actually seldom successful as we are placing extreme reliance on the correct localization of these basis points; every error in their coordinates is carried forward to the coordinates of the reconstructed data set through the badly formed camera matrices. Later on (in Section 4) we discuss how to choose a basis which has far superior noise properties using the weak calibration. The process makes use of a *virtual basis* that consists of a set of points that satisfy all of the pinhole camera imaging constraints, but are not actually observed in either image (and are thus free from any imaging error). All of the methods we describe in this paper can make use of a virtual basis.

3.2 Intersecting camera rays — method 2

An alternative way to recover the three-dimensional structure of a point is through a geometric construction based on intersecting pairs of rays which emerge from the projective cameras on which the 3D points must lie. These rays may be parametrized by their intersections with the ideal plane in projective three-space. Due to image noise the rays are likely to be skew, and so do not intersect. We must thus find the point in space that best represents the intersection point.

Given the pair of correspondences $\mathbf{x}_1$ and $\mathbf{x}_2$ we wish to find the intersections of the lines $\mathbf{d} =< \mathbf{C}_1, \mathbf{x}_1 >$ and $\mathbf{d}_2 =< \mathbf{C}_2, \mathbf{x}_2 >$ with the ideal plane (see Fig. 1); these points are represented by the Euclidean points $\mathbf{X}_{1_\infty}$ and $\mathbf{X}_{2_\infty}$ respectively (that is the projective coordinates of these points are $(\mathbf{X}_{1_\infty}^\top, 0)^\top$ and $(\mathbf{X}_{2_\infty}^\top, 0)^\top$). The point $\mathbf{X}$ is thus doubly constrained:

$$\mathbf{X} = \alpha_i \begin{pmatrix} \mathbf{C}_i \\ 1 \end{pmatrix} + \beta_i \begin{pmatrix} \mathbf{X}_{i_\infty} \\ 0 \end{pmatrix}, \qquad\qquad i \in \{1, 2\}.$$

However, we also know that:

$$\mathtt{P}_i \begin{pmatrix} \mathbf{X}_{i_\infty} \\ 0 \end{pmatrix} = \mathbf{x}_i,$$

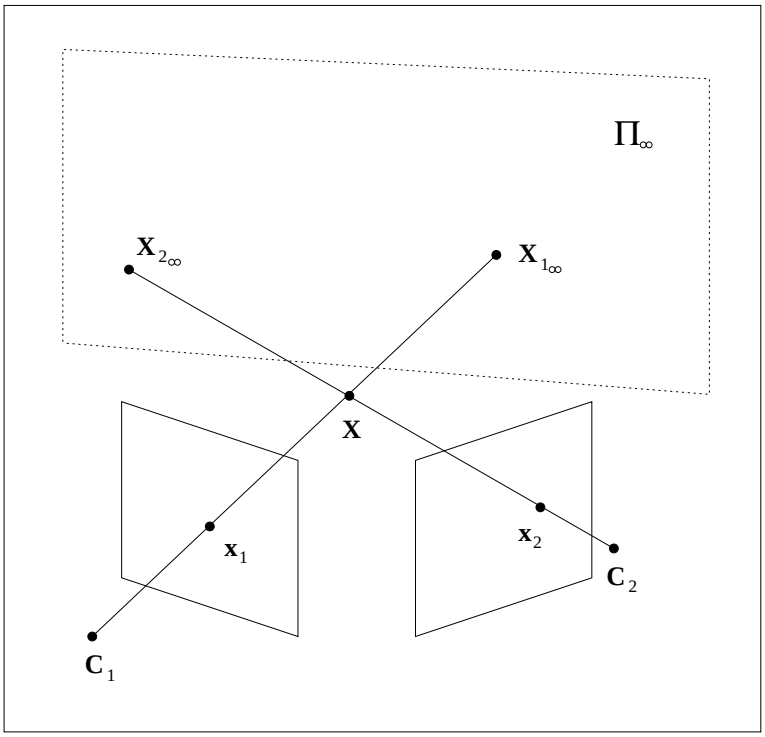

Figure 1: $\mathbf{X}$ *is the intersection of the rays passing through the camera centres* $\mathbf{C}_1$ *and* $\mathbf{C}_2$*, and the image points* $\mathbf{x}_1$ *and* $\mathbf{x}_2$*. The rays are parametrized by their intersection with the ideal plane:* $\mathbf{d}_1 =< \mathbf{C}_1, \mathbf{X}_{1_\infty} >$ *and* $\mathbf{d}_2 =< \mathbf{C}_2, \mathbf{X}_{2_\infty} >$*.*

Therefore, decomposing the camera matrices into $\mathtt{P}_i = [\mathtt{M}_i|\mathbf{t}_i]$, where each $\mathtt{M}_i$ are 3×3 matrices (both chosen to be invertible) and the $\mathbf{t}_i$ 3×1 vectors, yields $\mathtt{M}_i\ \mathbf{X}_{i_\infty} = \mathbf{x}_i$. Thus $\mathbf{X}_{i_\infty} = \mathtt{M}_i^{-1}\ \mathbf{x}_i$. We may also express the camera centres using the above decomposition using the fact that $\mathtt{P}_i(\mathbf{C}_i^\top, 1)^\top = \mathbf{0}$, and so $\mathbf{C}_i = -\mathtt{M}_i^{-1}\ \mathbf{t}_i$. For convenience we now write $\mathbf{D}_i = (\mathbf{X}_{i_\infty}^\top, 0)^\top$, and therefore:

$$\mathbf{X} \quad = \quad \alpha_i \begin{pmatrix} \mathbf{C}_i \\ 1 \end{pmatrix} + \beta_i\ \mathbf{D}_i. \tag{6}$$

Taking these expressions and dividing through by β_2 (which is zero if and only if $\mathbf{X}$ lies on the ideal plane), provides a set of four constraint equations:

$$\frac{\alpha_1}{\beta_2} \begin{pmatrix} \mathbf{C}_1 \\ 1 \end{pmatrix} + \frac{\beta_1}{\beta_2}\mathbf{D}_1 - \frac{\alpha_2}{\beta_2} \begin{pmatrix} \mathbf{C}_2 \\ 1 \end{pmatrix} = \mathbf{D}_2. \tag{7}$$

Equation (7) represents four equations and yet has only the three unknowns $(\frac{\alpha_1}{\beta_2}, \frac{\beta_1}{\beta_2}, \frac{\alpha_2}{\beta_2})$. Consequently, the system is over-constrained and can be solved using least-median-squares. The solution for $\mathbf{X}$ is obtained by back-substitutions into eqn (6).

This process provides a first estimate of the three dimensional coordinates of the point $\mathbf{X}$. We can refine the estimate using a minimization procedure based on the image distances between the projections of the estimated world point and the measured image projections. The minimization uses the following cost function:

$$C(\mathbf{x}_1, \mathbf{x}_2, \mathbf{X}) = (\mathtt{P}_1\ \mathbf{X} - \mathbf{x}_1)^2 + (\mathtt{P}_2\ \mathbf{X} - \mathbf{x}_2)^2,$$

where the $(\mathtt{P}i\ \mathbf{X})$ and $\mathbf{x}_i$ are normalised to have Euclidean coordinates. This constraint makes the method equivalent (in solution, though not in the path taken) to that of Hartley and Sturm [16].

3.2.1 Intersection of camera rays — minimal distance

An alternative description of the above method is to find the point that simultaneously lies closest to the two rays $\mathbf{d}_1$ and $\mathbf{d}_2$. This obviously assumes that the cameras are represented using a metric description of space. This is not true in our case as we compute only the projective forms of the cameras, but this path has been followed by those doing a *quasi-Euclidean* reconstruction, Beardsley, *et al.* [2]. If the reconstruction really is Euclidean, then this method leads to the most desirable solution. However, should it be anything like projective (in cases in which we can make no stable estimate of the camera parameters), then the intersection of camera rays is most likely to be one of the worst algorithms that could be used. In this we are in agreement with Hartley and Sturm [16].

The process involves localizing the points in three-dimensional space by the intersection of the two rays that pass through the camera centres and the respective image points. In the presence of image noise we would not expect the image rays from two images to intersect exactly, but rather they would be skew. In this situation we must use a point that in some manner lies closest to the two lines. If we had a metric space we could compute the location of the point that lies closest and equi-distant to the two rays and return this as the estimate of the location of the three-dimensional point. Again, we do not generally work in such a metric space, but we pursue the example as it is of interest.

Given the pair of projective cameras $\mathtt{P}_i = [\mathtt{M}_i | -\mathtt{M}_i\mathbf{C}_i]$ we know that:

$$\hat{\mathbf{X}}'_3 = k_1\ \mathtt{M}_1^{-1}\ \mathbf{x}_1 + \mathbf{C}_1 \quad \text{and} \quad \hat{\mathbf{X}}''_3 \quad = \quad k_2\ \mathtt{M}_2^{-1}\ \mathbf{x}_2 + \mathbf{C}_2. \tag{8}$$

where $\hat{\mathbf{X}}'_3$ and $\hat{\mathbf{X}}''_3$ are the two estimates of the points location if it is constrained to lie on the camera rays. The goal is to choose k_1 and k_2 to minimize the distance between $\hat{\mathbf{X}}'_3$ and $\hat{\mathbf{X}}''_3$, and then set $\hat{\mathbf{X}}_3 = (\hat{\mathbf{X}}'_3 + \hat{\mathbf{X}}''_3)/2$. More completely, we know that $\mathbf{v} = \hat{\mathbf{X}}'_3 - \hat{\mathbf{X}}''_3$ is perpendicular to both of the line directions, $\mathtt{M}_1^{-1}\ \mathbf{x}_1$ and $\mathtt{M}_2^{-1}\ \mathbf{x}_2$ (if the Euclidean distance is to minimized), and so $\mathbf{v} = (\mathtt{M}_1^{-1}\ \mathbf{x}_1) \times (\mathtt{M}_2^{-1}\ \mathbf{x}_2)$. Subsequently, we derive a third constraint:

$$\hat{\mathbf{X}}'_3 = \hat{\mathbf{X}}''_3 + k_3\ \mathbf{v}, \tag{9}$$

and use eqns (8) and (9) to solve for $\hat{\mathbf{X}}_3$.

As we are working in a projective frame, the notions of midpoint and orthogonality disappear, thus we cannot use precisely this form of constraint to determine $\hat{\mathbf{X}}_3$. However, if we permit ourselves the liberty of assigning approximate Euclidean parameters to each $\mathtt{P}_i$, then we derive a reasonable estimate to the location of the 3D point (or at least bound it). This approach has been taken by Beardsley, *et al.* [2].

3.3 Using singular value decomposition — method 3

The two reconstruction methods above have made use of cameras derived from individual points correspondences. In the first instance, proceeding from raw image measurements tends to lead to significant instability in the results as the camera matrices are unlikely to be projectively equivalent to the real cameras. In fact, we are placing our entire trust on the correct measurement of the first five image points. Here we introduce a method that defines the basis by taking every image measurement into account, and so for uncorrelated noise we would expect to average out any measurement errors. The camera matrices are computed directly from the weak calibration, which was itself computed using hundreds of image correspondences, and so can be assumed to be reliable.

From the fundamental matrices and epipoles between a pair of images we can derive a solution set for the cameras which is consistent with the epipolar geometry (the proof is given in Appendix A and is related to the result given by Luong and Viéville [18]):

$$\mathtt{P}_1 = [\mathtt{I}|\mathbf{0}]\ \mathtt{G} \qquad \text{and} \qquad \mathtt{P}_2 = [\,[\mathbf{e}_{21}]_\times\ \mathtt{F}_{12} \mid \mathbf{e}_{21}]\ \begin{bmatrix} \mathtt{I} & \mathbf{0} \\ \alpha^\top & \alpha_4 \end{bmatrix}\ \mathtt{G}, \tag{10}$$

where we have used the notation $[\mathbf{a}]_\times$ to represent the asymmetric matrix derived from the vector $\mathbf{a}$ that represents the cross product ($[\mathbf{a}]_\times \mathbf{b} = \mathbf{a} \times \mathbf{b}$). The matrix $\mathtt{G}$ is an arbitrary 4×4 projection matrix. However, as the reconstructions are embedded within the projective world we can immediately ignore the presence of this deformation and set $\mathtt{G}$ equal to the 4×4 identity matrix:

$$\begin{aligned} \mathtt{P}_1 &= [\mathtt{I}|\mathbf{0}], \\ \mathtt{P}_2 &= [\,[\mathbf{e}_{21}]_\times\ \mathtt{F}_{12} \mid \mathbf{e}_{21}]\ \begin{bmatrix} \mathtt{I} & \mathbf{0} \\ \alpha^\top & \alpha_4 \end{bmatrix}. \end{aligned} \tag{11}$$

Note also that the projection composed of the α_i, $i \in \{1, \ldots, 4\}$, has no effect on the form of $\mathtt{P}_1$; this is why we can use the above decomposition. The solution space represented by eqn (10) is four dimensional, and in theory we may choose any of the solutions such that α_4

is non-zero (so that the right hand projectivity is not singular). In practice we choose the α_i so that the camera matrices are numerically well conditioned, and so avoid any numerical difficulties later on during the computation of the three-dimensional structure.

3.3.1 Computing the structure

As we saw in Section 3.1, one way to compute the structure for a three-dimensional point is by the construction of a system of linear of constraints on the three-dimensional coordinates, and then solving them through the application of the pseudo-inverse. The problem with this approach is that it assumes that *none* the points lie on the ideal plane as we have made the assumption that *all* of the points can be parametrized by $\mathbf{X} = (X, Y, Z, 1)^\top$. This can be trouble some if the ideal plane passes through the data set, which is possible due to our goal of computing only projective reconstructions. Consequently, the pseudo-inverse method can give poor estimates of structure for certain choices of the camera matrices.

If we instead consider $\mathbf{X}$ to the be projective point $(p, q, r, s)^\top$, which is totally generic, we are led to a constraint system of the form:

$$\begin{bmatrix} \lambda_i & 0 \\ 0 & \mu_i \end{bmatrix} \begin{bmatrix} p_{11} - p_{31}x_i & p_{12} - p_{32}x_i & p_{13} - p_{33}x_i & p_{14} - p_{34}x_i \\ p_{21} - p_{31}y_i & p_{22} - p_{32}y_i & p_{23} - p_{33}y_i & p_{24} - p_{34}y_i \end{bmatrix} \mathbf{X} = \mathbf{0}.$$

Stacking the constraints for a pair of images leads to a constraint $\mathtt{A}\mathbf{X} = 0$, where $\mathtt{A}$ is a 4×4 matrix; without any prior information about the errors associated with individual points we assign both the λ_i and μ_i to unity. The reconstruction is then found from the null space of $\mathtt{A}$: we would in theory simply look for the eigenvector corresponding to the zero eigenvalue of $\mathtt{A}$, but in practice noise and numerical errors mean that the eigenvectors of $\mathtt{A}$ do not generically take on real values. We therefore use singular-value decomposition to ensure a real solution and take the solution vector corresponding to the smallest singular value. As will be shown in the results section later on, this method provides far more stable reconstruction than using the pseudo-inverse.

It is worth commenting on the geometric interpretation of this reconstruction method, and certainly worth noting its significance. Each row of $\mathtt{A}$ constrains the reconstructed point to lie on a plane in space (each row in fact is just composed of the plane's coefficients). The first two rows come from the first camera, and so constrain the point to lie on the line define by the intersection of the two corresponding planes. This line of course passes through the centre of the first camera. Likewise the third and fourth rows of $\mathtt{A}$ constrain the three-dimensional point to a line passing through the second camera centre. Consequently, the SVD minimisation is just a disguised form of the ray intersection method for computing three-dimensional structure, though we should be reminded that the measure being minimised is not Euclidean distance, but something more abstract.

4 Using the Cayley algebra — method 4

The methods discussed above for computing three-dimensional projective structure make the reconstruction process explicit. In the rest of the theoretical part of this paper we review two *implicit* reconstruction methods that allow the computation of structure (or identically invariants) from measurements that are confined to the images.

Carlsson promoted interest in measuring invariants between pairs of images without the need for the explicit reconstruction [3]. The process makes use of the *Cayley* or the *double algebra* [32]. As a brief summary of the process, projective invariants can be formed from the ratios of determinants of matrices composed of sets of homogeneous 3D points, and under projection these functions become rational functions of the fundamental matrix and certain other image measurements. Although we omit most of the mathematical details, a new interpretation of some of the measures is provided here that make use of the standard projective basis. This development makes clear the relationship between three-dimensional structure and its associated invariants (an equivalence relationship).

Principally, we place the first five points in the standard basis, and parametrize a sixth point by $\mathbf{X}_6 = (p, q, r, s)^\top$. Then we can show that the different invariants of the six point configuration are simply functions of $\{p, q, r, s\}$. The three-dimensional invariants of a set of six points can be expressed as a ratio of determinants:

$$I = \frac{|\mathbf{X}_a \mathbf{X}_b \mathbf{X}_c \mathbf{X}_d|}{|\mathbf{X}_a \mathbf{X}_b \mathbf{X}_d \mathbf{X}_e|} \frac{|\mathbf{X}_d \mathbf{X}_e \mathbf{X}_b \mathbf{X}_f|}{|\mathbf{X}_c \mathbf{X}_d \mathbf{X}_b \mathbf{X}_f|}. \tag{12}$$

Under projection to a pair of images, Carlsson demonstrates that this invariant is equivalent to:

$$I = \frac{\alpha_{ab\_cd}^\top \, \mathtt{F}_{21} \, \beta_{ab\_cd}}{\alpha_{ab\_de}^\top \, \mathtt{F}_{21} \, \beta_{ab\_de}} \frac{\alpha_{de\_bf}^\top \, \mathtt{F}_{21} \, \beta_{de\_bf}}{\alpha_{cd\_bf}^\top \, \mathtt{F}_{21} \, \beta_{cd\_bf}}, \tag{13}$$

where all of the parameters are based on image measurements and the fundamental matrix. The actual form of the expression requires some explanation: the projection of $\mathbf{X}_i$ into the first image is α_i and into the second image is β_i. The point α_{ab_cd} is the intersection of the *image* lines $< \alpha_a, \alpha_b >$ and $< \alpha_c, \alpha_d >$. Formally we see that $\alpha_{ab\_cd} = (\alpha_a \times \alpha_b) \times (\alpha_c \times \alpha_d)$ for points in the first image, and likewise for points in the second one. Although the expression for the two view invariant in eqn (12) does not have a ready geometric interpretation, we see that it is homogeneous in the points considered, and in the structure of the image-line configurations.

Using the standard basis and the equivalences described by eqns (12) and (13), we can derive the following projective equivalences:[6]

[6] These expression are in fact significantly different to those given by Carlsson [3] in that they take account of the homogeneous scaling of the image observations. See [4] for a detailed reasoning.

$$\begin{aligned}
\frac{\alpha_{24_36}^{\top}\ \mathtt{F}_{21}\ \beta_{24_36}}{\alpha_{12_36}^{\top}\ \mathtt{F}_{21}\ \beta_{12_36}}\ \frac{\alpha_{12_35}^{\top}\ \mathtt{F}_{21}\ \beta_{12_35}}{\alpha_{24_35}^{\top}\ \mathtt{F}_{21}\ \beta_{24_35}}\ k_{ps} &= \frac{p}{s}, \\
\frac{\alpha_{14_36}^{\top}\ \mathtt{F}_{21}\ \beta_{14_36}}{\alpha_{12_36}^{\top}\ \mathtt{F}_{21}\ \beta_{12_36}}\ \frac{\alpha_{12_35}^{\top}\ \mathtt{F}_{21}\ \beta_{12_35}}{\alpha_{14_35}^{\top}\ \mathtt{F}_{21}\ \beta_{14_35}}\ k_{qs} &= \frac{q}{s}, \\
\frac{\alpha_{24_16}^{\top}\ \mathtt{F}_{21}\ \beta_{24_16}}{\alpha_{23_16}^{\top}\ \mathtt{F}_{21}\ \beta_{23_16}}\ \frac{\alpha_{23_15}^{\top}\ \mathtt{F}_{21}\ \beta_{23_15}}{\alpha_{24_15}^{\top}\ \mathtt{F}_{21}\ \beta_{24_15}}\ k_{rs} &= \frac{r}{s}, \qquad (14)
\end{aligned}$$

where, for instance:

$$k_{ps} = \frac{(\overline{\alpha_2^T\ \mathtt{F}_{21}\ \beta_4})(\overline{\alpha_3^T\ \mathtt{F}_{21}\ \beta_6})(\overline{\alpha_1^T\ \mathtt{F}_{21}\ \beta_2})(\overline{\alpha_3^T\ \mathtt{F}_{21}\ \beta_5})}{(\overline{\alpha_1^T\ \mathtt{F}_{21}\ \beta_2})(\overline{\alpha_3^T\ \mathtt{F}_{21}\ \beta_6})(\overline{\alpha_2^T\ \mathtt{F}_{21}\ \beta_4})(\overline{\alpha_3^T\ \mathtt{F}_{21}\ \beta_5})},$$

with

$$(\overline{\alpha_i^T\ \mathtt{F}_{21}\ \beta_j}) = \sqrt{(\alpha_i^T\ \mathtt{F}_{21}\ \beta_j)(\alpha_j^T\ \mathtt{F}_{21}\ \beta_i)}$$

The importance of these relationships is quite profound: without having to compute a three dimensional reconstruction of the world, or even having to derive camera matrices, we are able to compute the three-dimensional coordinates of any point with respect to a five-point basis. Nominally the basis is defined by a set of five observed points, and the same basis is used for the reconstruction of all the image points. Note that we can compute the coordinates projectively (rather than restricting ourselves to affine representations where $s = 1$) by multiplying the expressions in eqn (14) through by their denominators: if the three invariants are expressed affinely as $(\frac{n_p}{d_p}, \frac{n_q}{d_q}, \frac{n_r}{d_r})^{\top}$, we compute the projective coordinates $(n_p d_q d_r, n_q d_p d_r, n_r d_p d_q, d_p d_q d_r)^{\top}$.

It is very important to see that the expressions rely heavily on the the exact knowledge of the locations of the projections of the basis points ($\mathbf{X}_i$, $i \in \{1, \ldots, 5\}$). In practice errors in the measurement of these points can lead to substantial instabilities in the invariant estimates. This problem has previously affected other methods of estimating three-dimensional structures, and is no less a problem here. When we derived the structure using an explicit use of the camera matrices (method 3), we overcame the reliance on using a specific basis through the use of an implicit basis that we derived directly from the epipolar geometry. We exploit the same technique here.

The process involves the estimation of a set of virtual points in the images, that is, points that are not actually observed but that we know will be consistent with any measurement that we have or will make. These points are the projections of the standard three-dimensional projective basis. There are two equivalent approaches that we may follow to this end:

- Given the fundamental matrices, derive the two cameras using the techniques of Section 3.3. This provides an implicit camera frame dependent on the canonical choice of the first camera and the four degrees of freedom available in the second camera. Compute the projections of the standard basis in the given camera frame, and use these points for the subsequent computation of the Cayley invariants. This process is guaranteed to provide the coordinates in a known non-degenerate three-dimensional basis.

- Alternatively, we proceed by choosing five points in each image that are in correspondence and that satisfy the epipolar constraints. These points can be considered to be the projections of a chosen basis (whose three-dimensional coordinates are those of the standard basis), and are guaranteed to be consistent with actual three-dimensional points through agreement with the epipolar constraints. The only drawback with this approach is that we must also test that no four of the chosen basis points are actually coplanar in space (an arbitrary choice of points could be coplanar).

Given that the Cayley invariants can be expressed in closed form, we can easily determine the likely error characteristics of a measurement given an assumed error model. These are omitted, but the process of analysis is relatively straightforward.

4.1 A range of Cayley invariants

Using only the three invariants for the computation of the three-dimensional point locations enforces only a minimal degree of constraint; this means that the world structure is susceptible to image noise. Ideally, we should attempt a least-squares approach that averages the measurements and hence minimizes the errors. Note that this is always possible because our imaging system provides an over-constrained solution: there are only three unknowns, the projective coordinates of a point in space, but there are a total of four constraints (the two coordinates in each image). Although the three invariants given above are the only independent ones in the noise-free case, after the addition of image noise, a large number of other invariants can be computed which are not strictly dependent. This section investigates the range of invariants that can be measured.

For six points in space there are fifteen different 4×4 determinants that can be measured. If we again ensure that all of the analysis is done using the standard basis, and define $\mathbf{X}_6 = (p, q, r, s)^\top$, then we get a range of different rational polynomial forms for the measures in the parameters $\{p, q, r, s\}$; all of these are given in Table 1. Invariants are computed using sets of four of the determinants, subject to the constraint that the expressions be homogeneous in the point coordinates. Given a proposed invariant, we make use of eqns (12) and (13) to write the invariants in the form associated with the Cayley Algebra. For instance, we can derive the invariant equivalences of the form:

$$\frac{\alpha_{26\_35}\ \mathtt{F}_{21}\ \beta_{26\_35}}{\alpha_{26\_13}\ \mathtt{F}_{21}\ \beta_{26\_13}} \frac{\alpha_{24\_13}\ \mathtt{F}_{21}\ \beta_{24\_13}}{\alpha_{24\_35}\ \mathtt{F}_{21}\ \beta_{24\_35}}\ k_{1-ps} \quad = \quad 1 - \frac{p}{s}. \tag{15}$$

$\lvert\mathbf{X}_3\mathbf{X}_4\mathbf{X}_5\mathbf{X}_6\rvert = q-p$	$\lvert\mathbf{X}_2\mathbf{X}_4\mathbf{X}_5\mathbf{X}_6\rvert = p-r$	$\lvert\mathbf{X}_2\mathbf{X}_3\mathbf{X}_5\mathbf{X}_6\rvert = s-p$
$\lvert\mathbf{X}_2\mathbf{X}_3\mathbf{X}_4\mathbf{X}_6\rvert = -p$	$\lvert\mathbf{X}_2\mathbf{X}_3\mathbf{X}_4\mathbf{X}_5\rvert = -1$	$\lvert\mathbf{X}_1\mathbf{X}_4\mathbf{X}_5\mathbf{X}_6\rvert = r-q$
$\lvert\mathbf{X}_1\mathbf{X}_3\mathbf{X}_5\mathbf{X}_6\rvert = q-s$	$\lvert\mathbf{X}_1\mathbf{X}_3\mathbf{X}_4\mathbf{X}_6\rvert = q$	$\lvert\mathbf{X}_1\mathbf{X}_3\mathbf{X}_4\mathbf{X}_5\rvert = 1$
$\lvert\mathbf{X}_1\mathbf{X}_2\mathbf{X}_5\mathbf{X}_6\rvert = s-r$	$\lvert\mathbf{X}_1\mathbf{X}_2\mathbf{X}_4\mathbf{X}_6\rvert = -r$	$\lvert\mathbf{X}_1\mathbf{X}_2\mathbf{X}_4\mathbf{X}_5\rvert = -1$
$\lvert\mathbf{X}_1\mathbf{X}_2\mathbf{X}_3\mathbf{X}_6\rvert = s$	$\lvert\mathbf{X}_1\mathbf{X}_2\mathbf{X}_3\mathbf{X}_5\rvert = 1$	$\lvert\mathbf{X}_1\mathbf{X}_2\mathbf{X}_3\mathbf{X}_4\rvert = 1$

Table 1: *For six points in space we can compute fifteen different determinants. When the standard basis is used they take the above forms (ignoring the scaling of the homogeneous points).*

This provides multiple constraints on the invariant parameters of the 3D points, and so over-constrain their estimation.

Note also that we are likely to experience difficulties in the computations should certain sets of four points become close to coplanar. Of course we will design our basis to ensure that none of the points $\mathbf{X}_i$, $i \in \{1,\ldots,5\}$ become spatially dependent, though we cannot hope to prevent $\mathbf{X}_6$ from becoming coplanar with any triple of the sets of basis points (for instance we may find that a world point may genuinely have the coordinate $s = 0$). Algebraically, the coplanarity event manifests itself as one of the determinants becoming zero, or as an expression of the form $\alpha_{ab\_cd}\ \mathtt{F}_{21}\ \beta_{ab\_cd}$ becoming zero. Interestingly, if $\{a,b,c,d\}$ are genuinely coplanar, then the projections of the intersection of the three-dimensional lines $< \mathbf{X}_a, \mathbf{X}_b >$ and $< \mathbf{X}_c, \mathbf{X}_d >$ in each image are the points α_{ab_cd} and β_{ab_cd}; these should mutually satisfy the epipolar constraint, and so $\alpha_{ab\_cd}\ \mathtt{F}_{21}\ \beta_{ab\_cd} = 0$.

In some cases in which the degeneracies occur, there will be poles in the invariant functions and so their measurement will be unreliable (a zero in the function does not cause the same type of problem). In these cases we make use of the other invariant formulations to recover meaningful invariants, and then use an averaging process between all of the invariants to judge which structure is the most meaningful. For instance, if p should become small, it is numerically more prudent to rely on an invariant based on p rather than one proportional to $1/p$. Another example is if q and p become approximately equal: any invariant with a denominator of $q - p$ will be unstable under such a condition, and so should not be used.

5 Using the cross ratio — method 5

A second implicit reconstruction process exploits invariance through the cross ratio [6]. The method involves constructing the images of a pencil of planes that form a projective description of the three-dimensional configuration (again a set of six points); Fig. 2 demonstrates the case.

We first derive a relationship between the cross ratios measured between various sets of planes and the 3D coordinates of the point $\mathbf{X}$ in the standard basis, and then show how to compute the invariants in the images. Consider first a pencil of planes whose axis is the line $< \mathbf{X}_1, \mathbf{X}_2 >$. Taking four members of this pencil Π_i, $i \in \{3,\ldots,6\}$, such that the plane Π_i

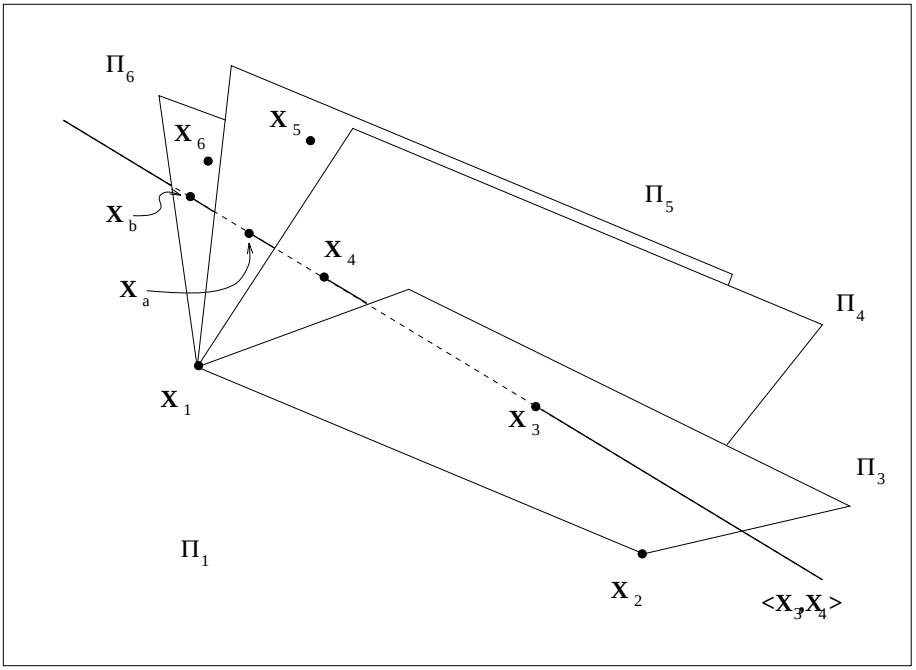

Figure 2: *There are a number of projective measurements that can be computed for a generic configuration of six points in 3-space. For instance, we can form a pencil of planes whose axis is* $< \mathbf{X}_1, \mathbf{X}_2 >$*, and compute cross ratios between planes in the pencil. One such cross ratio is for the set of planes* Π_i*,* $i \in \{3, \ldots, 6\}$*, where each* Π_i *contains the points* $\{1, 2, i\}$*. The cross ratio could be measured from the points of intersection of the line* $< \mathbf{X}_3, \mathbf{X}_4 >$ *with the four planes.*

contains the points $\{1, 2, i\}$, yields a projective description of the six point configuration if we employ the cross ratio $\{\Pi_3, \Pi_4; \Pi_5, \Pi_6\}$. In practice we might measure the actual cross ratio in 3-space by considering the four collinear points formed by the intersections of the line $< \mathbf{X}_3, \mathbf{X}_4 >$ with the set of four planes; this would be expressed as $\{\mathbf{X}_3, \mathbf{X}_4; \mathbf{X}_a, \mathbf{X}_b\}$.

The value of the cross ratio $\{\Pi_3, \Pi_4; \Pi_5, \Pi_6\}$ is equal to r/s when $\mathbf{X}_i$, $i \in \{1, \ldots, 5\}$, represent the standard basis, and $\mathbf{X}_6$ is again the point $(p, q, r, s)^\top$. We can see this by determining the equations of the four planes and then parametrizing the pencil by Π_3 and Π_4. First, we see that Π_3 passes through the points $(1, 0, 0, 0)^\top$, $(0, 1, 0, 0)^\top$ and $(0, 0, 1, 0)^\top$. Its equation is therefore[7] $\Pi_3 = (0, 0, 0, 1)^\top$. By similar reasoning we can derive the representations of all of the planes:

[7] Note the homogeneous representation of planes. A plane of the form $ap + bq + cr + ds = 0$ is represented by the projective point $(a, b, c, d)^\top$.

I_{123456}	=	$\frac{r}{s}$	I_{132456}	=	$\frac{q}{s}$	I_{142356}	=	$\frac{q}{r}$
I_{152346}	=	$\frac{q-s}{r-s}$	I_{162345}	=	$\frac{r(q-s)}{q(r-s)}$	I_{231456}	=	$\frac{p}{s}$
I_{241356}	=	$\frac{p}{r}$	I_{251346}	=	$\frac{p-s}{r-s}$	I_{261345}	=	$\frac{r(p-s)}{p(r-s)}$
I_{341256}	=	$\frac{p}{q}$	I_{351246}	=	$\frac{p-s}{q-s}$	I_{361245}	=	$\frac{q(p-s)}{p(q-s)}$
I_{451236}	=	$\frac{r-p}{r-q}$	I_{461235}	=	$\frac{q(r-p)}{p(r-q)}$	I_{561234}	=	$\frac{(q-s)(p-r)}{(p-s)(q-r)}$

Table 2: *Fifteen different cross ratios can be computed for a set of six points in a generic configuration in space.*

$$\Pi_3 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \quad \Pi_4 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad \Pi_5 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}, \quad \Pi_6 = \begin{pmatrix} 0 \\ 0 \\ s \\ -r \end{pmatrix}.$$

The pencil can be parametrized so that the planes takes the form $\Pi = \lambda\Pi_1 + \mu\Pi_2$, and then we can use the projective parameter $\theta = \lambda/\mu$. Doing this gives the four planes the following projective coordinates: $\theta_3 = \infty$; $\theta_4 = 0$; $\theta_5 = -1$; and $\theta_6 = -r/s$. From these we compute the cross ratio:

$$\tau = \frac{\theta_1 - \theta_3}{\theta_1 - \theta_4} \frac{\theta_2 - \theta_4}{\theta_2 - \theta_3} = \frac{r}{s}.$$

For an alternative derivation of this result, see Faugeras [10]. Different invariants are computed for each of the different permutations of the points (fifteen invariants in all); these are given in Table 2. For convenience in notation, in the table we represent the plane configuration we have already studied by the invariant function $I_{123456}(\mathbf{X})$. Note that an affine point can be described exactly by $\mathbf{X}_6 = (I_{231456}, I_{132456}, I_{123456}, 1)$. Finally, we do not actually have direct access to the 3D coordinates of the points use in the cross ratio computation, but the invariants can be computed as easily from the images of the projections of the points. We now describe how to find the projections of the points required for the computation in the two images (these are the projections of the points $\mathbf{X}_a$ and $\mathbf{X}_b$ in Fig. 2 which are not directly observable but are located at the intersections of a line with different planes).

5.1 Estimating the cross ratios in the image

As we saw above the invariant I_{ijklmn} is nothing else than the cross ratio of the planes $\{\Pi_k, \Pi_l; \Pi_m, \Pi_n\}$, where Π_α is the plane $\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_\alpha$. This cross ratio is equal to $\{\mathbf{X}_k, \mathbf{X}_l; \mathbf{X}_a, \mathbf{X}_b\}$, where $\mathbf{X}_a$ (respectively $\mathbf{X}_b$) is the intersection of line $< \mathbf{X}_k, \mathbf{X}_l >$ with the plane Π_m (respectively Π_n). This cross ratio can be computed from a pair of images using the epipolar geometry in several different way. We give two of them here.

5.1.1 Using the intersection of a line and plane

We wish to compute the intersection of the line $< \mathbf{X}_a, \mathbf{X}_b >$ and the plane defined by the three points $\{\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3\}$. This points is called $\mathbf{X}$. In the first image the line is observed as $< \alpha_a, \alpha_b >$, and the points as $\{\alpha_1, \alpha_2, \alpha_3\}$. The points are observed in the second image are represented by β_i. The intersection point is imaged as α and β in each image. The scheme is depicted in Fig. 3.

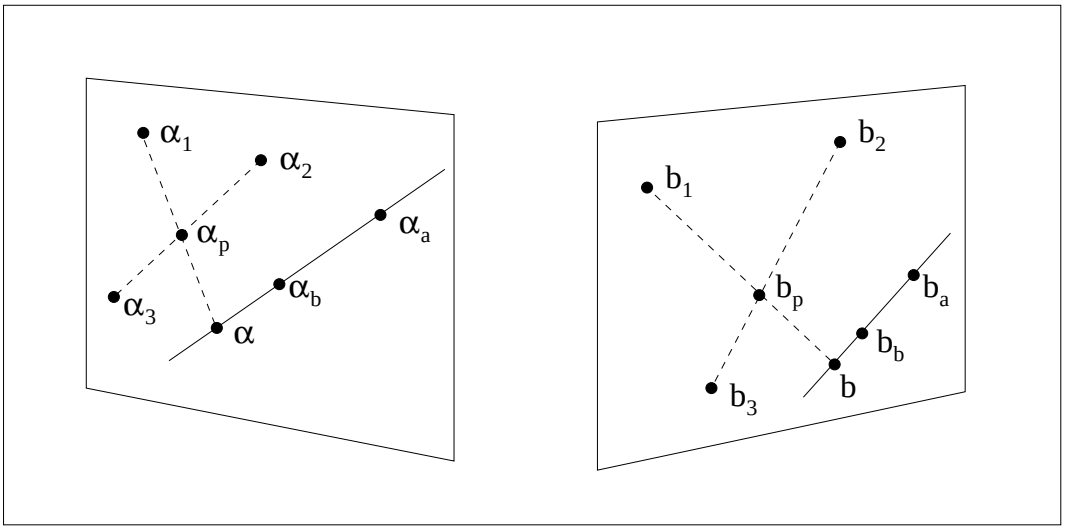

Figure 3: *The two images of the line* $< \mathbf{X}_a, \mathbf{X}_b >$ *and the plane* $\{\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3\}$ *are represented respectively by the letters* α_i *and* β_i*. The intersection point of the line and the plane is imaged as* α *and* β *in each image.*

The point α satisfies the image constraint:

$$\alpha^T \, . \, (\alpha_a \times \alpha_b) = 0. \tag{16}$$

β must lie on the epipolar line of this point in the second image, and also on the line $< \beta_a, \beta_b >$. Consequently $\beta = \mathtt{F}_{12}\alpha \times (\beta_a \times \beta_b)$. Consider now the point α_p (and respectively β_p), which is the intersection point of the lines $< \alpha_1, \alpha >$ and $< \alpha_2, \alpha_3 >$. This point, and its equivalent in the second image are:

$$\begin{aligned} \alpha_p &= (\alpha_2 \times \alpha_3) \times (\alpha_1 \times \alpha), \\ \beta_p &= (\beta_2 \times \beta_3) \times (\beta_1 \times \beta). \end{aligned}$$

We wish to enforce coplanarity of the points $\{\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3\}$ and $\mathbf{X}$. If they are coplanar, we know that α_p and β_p are in direct correspondence and satisfy the epipolar constraint $\beta_p^T \mathtt{F}_{12} \alpha_p = 0$. This can be written as:

$$((\beta_2 \times \beta_3) \times (\beta_1 \times \mathtt{F}_{12}\alpha \times (\beta_a \times \beta_b)))^\top \mathtt{F}_{12}((\alpha_2 \times \alpha_3) \times (\alpha_1 \times \alpha)) \quad = \quad 0,$$

which is quadratic in α. However, it may be decomposed into two parts:[8]

$$\begin{aligned} P &= (({}^1\mathtt{F}_{21} \times^2 \mathtt{F}_{21}) \times \alpha_1), \\ Q &= ((\alpha_2 \times \alpha_3).\alpha)((\beta_2 \times \beta_3).\beta)(({}^1\mathtt{F}_{12} \times^2 \mathtt{F}_{12}).(\alpha_a \times \alpha_b))(({}^1\mathtt{F}_{21} \times^2 \mathtt{F}_{21}) \times \alpha_1)_1 \\ &+ ((\alpha_2 \times \alpha_3).\alpha_1)((\beta_2 \times \beta_3).(\mathtt{F}_{12}\alpha))(\beta.{}^1\mathtt{F}_{12})({}^1\mathtt{F}_{21} \times^2 \mathtt{F}_{21})_3, \end{aligned}$$

such that $P\ Q = 0$. In general $P \neq 0$, and so we solve for $Q = 0$. This places a single linear constraint on α, which along with the eqn (16) (also linear), allows us to solve for the intersection point of the line and plane in the first image uniquely. We can similarly determine the position of β.

Given this method of intersecting a line with a plane we can simply estimate the values of the invariant I_{ijklmn} in the first image (or similarly in the second one) as follows:

$$I_{ijklmn} = \{\alpha_k, \alpha_l; \alpha_{ijm-kl}, \alpha_{ijn-kl}\}. \tag{17}$$

where α_{ijm-kl} (respectively α_{ijn-kl}) is the image of intersection point of the plane $\{\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_m\}$ (respectively $\{\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_n\}$) and the line $< \mathbf{X}_k, \mathbf{X}_l >$.

5.1.2 Computing the homography between the two images of a plane

The configuration of Fig. 2 projects to the pair of images shown in Fig. 4. In the first image, each of the $\mathbf{X}_i$ project to the points $\mathbf{p}_i$, and in the second image they project to the $\mathbf{q}_i$. Each point pair $\mathbf{p}_i$ and $\mathbf{q}_i$, $i \in \{1, \ldots, 6\}$ are therefore in correspondence. The epipoles are denoted by $\mathbf{e}_{12}$ and $\mathbf{e}_{21}$, and are also in correspondence. It is known that $\mathbf{l}_1$, which is the line $< \mathbf{p}_3, \mathbf{p}_4 >$, intersects planes Π_5 and Π_6; it is from the intersection points of this line with the planes that we form the cross ratio.[9] The corresponding line in the second image is $\mathbf{l}_2$.

First consider only the plane Π_5; this contains the points that project to $\mathbf{p}_1$, $\mathbf{p}_2$ and $\mathbf{p}_5$ in the first image, and their correspondences in the second image. Consider also the intersection of the line containing the two camera centres with Π_5. By construction, this intersection point projects to $\mathbf{e}_{12}$ in the first image and $\mathbf{e}_{21}$ in the second. Using the epipoles like this provides the images of a fourth points coplanar with $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_5$ in each image. Four coplanar points provide a projective basis in the plane and so we can estimate where the image of any fifth point lying in Π_5 and observed in the first image would be observed in the second. To do this we compute the projectivity $\mathtt{T}_5$ that maps the image of Π_5 between the first and second image.

[8] We use ${}^k\mathtt{F}_{ij}$ to denote the k^{th} column of $\mathtt{F}_{ij}$ and $(\mathbf{x})_k$ is the k^{th} element of the vector $\mathbf{x}$.

[9] It also intersects Π_3 and Π_4 at the points $\mathbf{p}_3$ and $\mathbf{p}_4$.

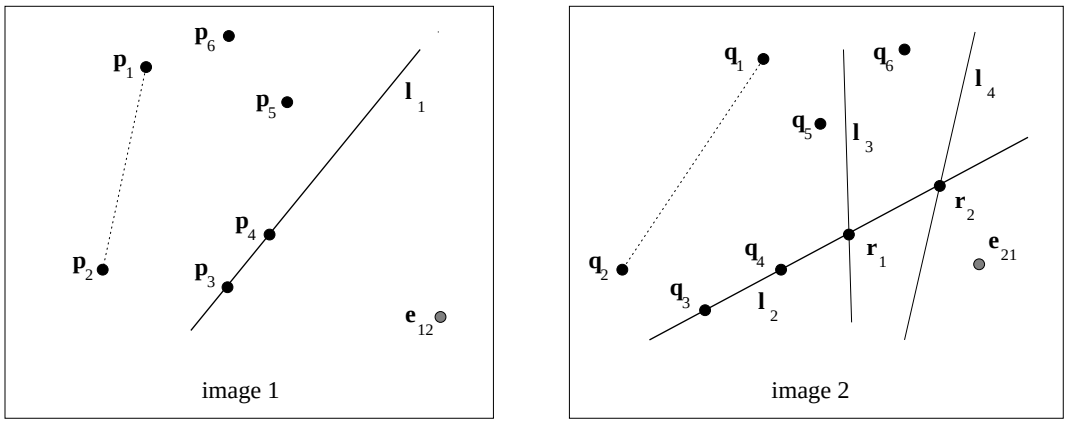

Figure 4: *The projections of the intersection of the line* $< \mathbf{X}_3, \mathbf{X}_4 >$ *,with the planes* Π_5 *and* Π_6 *can be computed from image measurements involving projecting lines observed in the images between the two images. Details are given in the text.*

We know that $k_i \mathbf{q}_i = \mathtt{T}_5 \mathbf{p}_i$, $i \in \{1, 2, 5\}$ and $k_e \mathbf{e}_{21} = \mathtt{T}_5 \mathbf{e}_{12}$. Then, considering t_{ij} to be the element of $\mathtt{T}_5$ in the i^{th} row and j^{th} column we derive the following constraints on $\mathtt{T}_5$:

$$\begin{bmatrix} p_x & p_y & 1 & 0 & 0 & 0 & -p_x q_x & -p_y q_x & -q_x \\ 0 & 0 & 0 & p_x & p_y & 1 & -p_x q_y & -p_y q_y & -q_y \end{bmatrix} \begin{pmatrix} t_{11} \\ t_{12} \\ \vdots \\ t_{33} \end{pmatrix} = 0,$$

where $\mathbf{p} = (p_x, p_y, 1)^\top$, $\mathbf{q} = (q_x, q_y, 1)^\top$, $\mathbf{p} \in \{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_5, \mathbf{e}_{12}\}$ and $\mathbf{q} \in \{\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_5, \mathbf{e}_{21}\}$. Combining all of the constraints for the four points yields a system of the form $\mathtt{A}\ \mathbf{t}_5 = 0$; $\mathbf{t}$ is found from the null space of $\mathtt{A}$ subject to a constraint such as $|\mathbf{t}_5| = 1$ (using singular value decomposition).

All of the points on Π_5 map from the first image to the second one via $\mathtt{T}_5$. There is also a single point on $\mathbf{l}_1$ that lies on Π_5, so if we apply $\mathtt{T}_5$ to map the image line $\mathbf{l}_1$ into the second image,[10] then the point actually lying on the plane will project correctly.[11] The image of $\mathbf{l}_1$ in the second image is $\mathbf{l}_3$. It is also known that only point $\mathbf{l}_2$ lies on Π_5, and this point corresponds to the point on $\mathbf{l}_1$ that lies on the same plane in 3-space. Consequently,

[10] In fact by using the line projectivity $\mathtt{T}_5^{-\top}$.

[11] The other points on the line do not project to where they would be observed in the second image, unless the focal point for the second camera is constrained to lie in the plane defined by the line $< \mathbf{X}_3, \mathbf{X}_4 >$ and the centre of the first camera. Anyway, our interest is not actually about where the non-coplanar points project, and so we ignore them in the sequel.

the point is constrained to lie on both $\mathbf{l}_2$ and $\mathbf{l}_3$ in the second image and is thus determined by their intersection. We have therefore located $\mathbf{r}_1$; the fourth point $\mathbf{r}_2$ can be computed in a similar manner using the projectivity $\mathtt{T}_6$ for Π_6 which contains the points $\mathbf{X}_i$, $i \in \{1, 2, 6\}$ and another point that also lies on the line containing the two camera centres. Given the four points $\mathbf{q}_3$, $\mathbf{q}_4$, $\mathbf{r}_1$, and $\mathbf{r}_2$ we can measure the cross ratio $\{\Pi_3, \Pi_4; \Pi_5, \Pi_6\} = \{\mathbf{q}_3, \mathbf{q}_4; \mathbf{r}_1, \mathbf{r}_2\}$ which is equal to the coordinate r/s of $\mathbf{X}_6$.

We are again faced with the problem of finding a stable basis of five points to use for the structure computation, though as before, we use a virtual basis as described in Section 3.3. An additional procedure we employ to improve noise stability is to form the cross ratio by a reciprocal construction in the first image (as all the measurement above were taken in the second image), and average between the different measures that we recover. Ideally the invariants measured in both of the images should be identical, but due to errors introduced during the computation they often differ slightly.

5.1.3 A closed form expression for the cross ratios

The computation given above for computing the invariants is rather complex and cannot readily be expressed in closed form. This makes it difficult to track errors throughout the calculation of the invariants using standard perturbation techniques (and so it is hard to improve our understanding of the stability of the process by theoretical analysis). However, we can derive a closed form expression for the invariants that make use of some special changes of projective frames to simplify the expressions. For the purposes of demonstration, we compute I_{123456}, though obviously method extends to all of the invariants in Table 2.

We first make use of an alternative coordinate frame in the image that is different to the origin frame produced by the camera. In each image we derive coordinates based on the projective frame that has the points $\mathbf{p}_1$, $\mathbf{p}_2$, $\mathbf{e}_{12}$ and $\mathbf{p}_5$ in the standard basis (also with points for the second image substituted):

$$\mathbf{p}'_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \qquad \mathbf{p}'_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \qquad \mathbf{e}'_{12} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \qquad \mathbf{p}'_5 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

The coordinates of all of the other points can be defined in this frame by computing the projectivity that maps these four points into the standard basis.[12] This projectivity, which that maps the points by $\mathbf{p}_i = \mathtt{T}_\alpha \mathbf{p}'_i$, is defined following the method of Sinclair, *et al.* [29] Using the fact that $[\mathbf{p}'_1, \mathbf{p}'_2, \mathbf{e}'_{12}] = \mathtt{I}$, and $[\alpha_1\mathbf{p}_1, \alpha_2\mathbf{p}_2, \alpha_3\mathbf{e}_{12}] = \mathtt{T}_\alpha\ [\mathbf{p}'_1, \mathbf{p}'_2, \mathbf{e}'_{12}]$ we see that:

$$\mathtt{T}_\alpha = [\alpha_1\mathbf{p}_1, \alpha_2\mathbf{p}_2, \alpha_3\mathbf{e}_{12}].$$

[12] Alternatively, but equivalently, we can design a set of planar projective invariants that directly return the coordinates in the new frame when supplied with a point and the image coordinates of the four-point basis.

The α_i are variables that are determined by constraining $\mathbf{p}_5$. This point maps by:

$$\alpha_4 \mathbf{p}_5 = \mathtt{T}_\alpha \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = [\mathbf{p}_1, \mathbf{p}_2, \mathbf{e}_{12}] \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}.$$

Thus,

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = [\mathbf{p}_1, \mathbf{p}_2, \mathbf{e}_{12}]^{-1} \ \alpha_4 \mathbf{p}_5.$$

Due to the freedom of the homogeneous scale of the above expression we are free to set α_4 to any non-trivial value. We therefore compute the values for α_i, $i \in \{1,2,3\}$ by setting $\alpha_4 = 1$, and by employing the adjoint of $[\mathbf{p}_1, \mathbf{p}_2, \mathbf{e}_{12}]$.[13] Then, given the α_i, we see that $\mathtt{T}_\alpha$ and more importantly its inverse (again scale means that we need compute only the adjoint of $\mathtt{T}_\alpha$) are derived. This mapping can be expressed in closed form, though even then it is too complicated to express here.

Subsequently, the use of the canonical frame defined above means that $\mathtt{T}_5$ which is used to map $\mathbf{l}_1$ from the first image to the second is now the 3×3 identity $\mathtt{I}$. The overall mapping from the first image to the canonical frame of the second, which is $\mathtt{I}\ \mathtt{T}_\alpha$, can therefore be computed (trivially) for each image pair once the world basis has been set. Likewise the matrix $\mathtt{T}_\alpha^{-1}$ is computed only once for a pair of images; we also compute the projectivity mapping points in the second image into a similar frame, $\mathtt{T}_\beta^{-1}$. The invariant computation then has to determine only the matrix $\mathtt{T}_6$ for each point, and measure the associated cross ratio. $\mathtt{T}_6$ also takes on a special form in due to the nature of the canonical frames as the points $\mathbf{p}_1$, $\mathbf{p}_2$, $\mathbf{e}_{12}$ are fixed (as they have the same coordinates as $\mathbf{q}_1$, $\mathbf{q}_2$, $\mathbf{e}_{21}$). $\mathtt{T}_6$ and the associated line projectivity (after scaling) take the form:

$$\mathtt{T}_6 = \begin{bmatrix} \mathbf{q}'_{6_x}\mathbf{p}'_{6_y} & 0 & 0 \\ 0 & \mathbf{p}'_{6_x}\mathbf{q}'_{6_y} & \\ 0 & 0 & \mathbf{p}'_{6_x}\mathbf{p}'_{6_y} \end{bmatrix} \quad \text{and} \quad \mathtt{T}_6^{-\top} = \begin{bmatrix} \mathbf{p}'_{6_x}\mathbf{q}'_{6_y} & 0 & 0 \\ 0 & \mathbf{q}'_{6_x}\mathbf{p}'_{6_y} & \\ 0 & 0 & \mathbf{q}'_{6_x}\mathbf{q}'_{6_y} \end{bmatrix}.$$

The points $\mathbf{p}'_6$ and $\mathbf{q}'_6$ are computed from the matrices $\mathtt{T}_\alpha$ and $\mathtt{T}_\beta$. The invariant r/s is therefore computed from:

$$\frac{r}{s} = \{\mathbf{q}_3, \mathbf{q}_4; \mathbf{e}_1, \mathbf{e}_e\} = \{\mathbf{q}_3, \mathbf{q}_4; \mathbf{l}_2 \times (\mathtt{T}_\beta^{-\top}\mathtt{T}_\alpha^{\top}(\mathbf{p}_3 \times \mathbf{p}_4)), \mathbf{l}_2 \times (\mathtt{T}_\beta^{-\top}\mathtt{T}_6^{-\top}\mathtt{T}_\alpha^{\top}(\mathbf{p}_3 \times \mathbf{p}_4))\}.$$

When written out in full, this expression takes a relatively simple form which at first sight appears similar to the invariant expressions for the Cayley algebra, though is not quite as easy to express succinctly.

[13] The full inverse could be used, but the amount of computation is reduced by using only the cofactors. Again, the homogeneity of the expression makes the scale of the matrix irrelevant.

5.2 Over-constraining the system

As with the Cayley invariants we over-constrain the structure estimates through the use of variety of invariants from Table 2. The actual invariants used are:

$$\begin{aligned}
\frac{p}{s} &= I_{231456} = I_{241356}I_{123456} = 1+(I_{123456}-1)I_{251346} \\
&= I_{341256}I_{132456} = 1+(I_{132456}-1)I_{351246} = I_{123456}-I_{451236}(I_{123456}-I_{132456}) \ , \\
\frac{q}{s} &= I_{132456} = I_{142356}I_{123456} = 1+(I_{123456}-1)I_{152346} \\
&= I_{231456}/I_{341256} = 1+(I_{231456}-1)/I_{351246} = I_{123456}-(I_{123456}-I_{231456})/I_{451236} \ , \\
\frac{r}{s} &= I_{123456} = I_{132456}/I_{142356} = 1+(I_{132456}-1)/I_{152346} \\
&= I_{231456}/I_{241356} = 1+(I_{231456}-1)/I_{251346} = (I_{231456}-I_{132456}I_{451236})/(1-I_{451236}) \ .
\end{aligned}$$

This provides an enhanced tolerance to noise, and we simply employ a variety of averaging processes to improve the value of the invariants that are returned. Recalling the argument given in Section 3.3, where we stated that it is frequently difficult to find a realistic interpretation of the error handling under such an averaging process, we realize that we employing exactly the same degree of ignorance here to suppress errors in the measurements.

6 Results

The performances of the five reconstruction methods have been evaluated on a number of synthetic and real images. Overall we have found that the approaches have different abilities to cope with image measurement errors, though they are all potentially useful for applications requiring uncalibrated stereo-scopic reconstructions. For all of the examples given here we have assumed knowledge of the weak calibration, and hence the matrix $\mathtt{F}_{12}$. The results are presented in three series:

1. To provide a quantitative understanding of the effects of image noise, and hence to demonstrate which methods have superior error handling characteristics, the algorithms have been tested on synthetic images with different levels of added noise. We work with point sets (known with absolute accuracy) that are projected into images and then noise added. This provides a precise geometric model of the world, and we can hence compare reconstruction results with the ground-truth over a very large number of trials.

2. As assumed noise models are frequently uncharacteristic of real imaging situations, the algorithms have also been applied to images of real scenes for which features are extracted using conventional early visual methods. So that we can again recover a qualitative understanding of the performances, images of a precisely known object used. The object is actually a traditional camera calibration grid.

3. Finally, more general reconstruction examples are shown for a series of images containing a building and some cars. In this case we can determine only a qualitative measure of the performance, but in fact it is still quite clear which methods perform better. We also demonstrate how the same methods can be used to estimate structure for mixed point-and-line data sets which are more characteristic of reconstruction applications.

6.1 Mapping to a Euclidean frame

All of the algorithms recover the world structure projectively. Frequently this structure will be very different to the correct Euclidean shape of the scene (though the two are related by a 3D projectivity). To improve the visualisation of the results we map the reconstructions back into a Euclidean frame and then determine the error characteristics. For the synthetic data we can compute the map we require exactly; for the real images we are only able to determine approximate estimates.

6.2 Synthetic images

In these examples we have an exact three-dimensional model of the world that is projected into the images and then noise added. The noise model used is zero mean Gaussian with differing variances σ^2. The scene is reconstructed using the five different methods, and then mapped back into the correct Euclidean frame. In all of the cases we have exact knowledge of the camera geometries and a realistic projection model is used. In turn, this means that we know the fundamental matrices exactly.

Tables 3 and 4 show how the five different methods compare for two examples of sets of noisy image data (with five levels of noise). The error measure is the mean distance between the reconstructed points mapped to the correct Euclidean frame and the actual 3D locations. A bar-plot of the error values for $\sigma = 0.5$ for the first data set is shown in Fig 6. We have found that over a large number of image trials (both with real and synthetic images) that the singular value decomposition method performs better than the others, with the pseudo-inverse often having nearly comparable results. Generally, the cross ratio method has some very large errors. As can be seen from the visualisations of the reconstructions in Figure 5, the errors for this reconstruction method are generally caused by a few outliers with the bulk of the data being reliable.

As discussed in Sections 4 and 5 we are able to make a large number of estimates of both the Cayley invariants and the invariants based on the cross ratio. The structure results reported in Tables 3 and 4 for the two methods used measures based on the construction of a linear constraint system on the projective coordinates of the points of the form $\mathtt{B}\mathbf{X} = 0$, and then the solution of the system by singular value decomposition. For a different data set we show in Table 5 that computing an estimate based on the singular value decomposition is typically far more stable than if we just used a simple averaging process (computation of the mean). We have also found the singular value decomposition approach to perform better than various forms of the α-trimmed mean [23], even though these would be expected to be more capable of handling outliers.

Method	$\sigma = 0.1$	$\sigma = 0.25$	$\sigma = 0.5$	$\sigma = 1$	$\sigma = 2$
1	73.52011	73.77007	75.98249	73.98717	76.18643
2	73.52803	73.84400	76.21575	75.07321	78.94152
3	2.322487	5.277889	11.14911	23.42297	42.76225
4	73.74798	73.82166	77.10320	80.39079	90.75704
5	98.86071	105.7079	154.5469	190.5779	1376.891

Table 3: *A comparison of the five reconstruction methods for synthetic images. For a given projection Gaussian noise of standard deviation σ is added, and then each reconstruction method used to estimate the 3D structure.*

	$\sigma = 0.1$	$\sigma = 0.25$	$\sigma = 0.5$	$\sigma = 1$	$\sigma = 2$
1	74.12385	74.90703	75.29391	70.28763	80.46459
2	74.13678	74.98710	75.63947	72.53291	84.93975
3	2.177906	5.123861	11.81602	20.65025	43.21792
4	74.31306	74.89345	76.90256	76.86017	102.6320
5	92.54126	161.3369	168.5504	185.1871	1149.210

Table 4: *A second comparison of the reconstruction methods.*

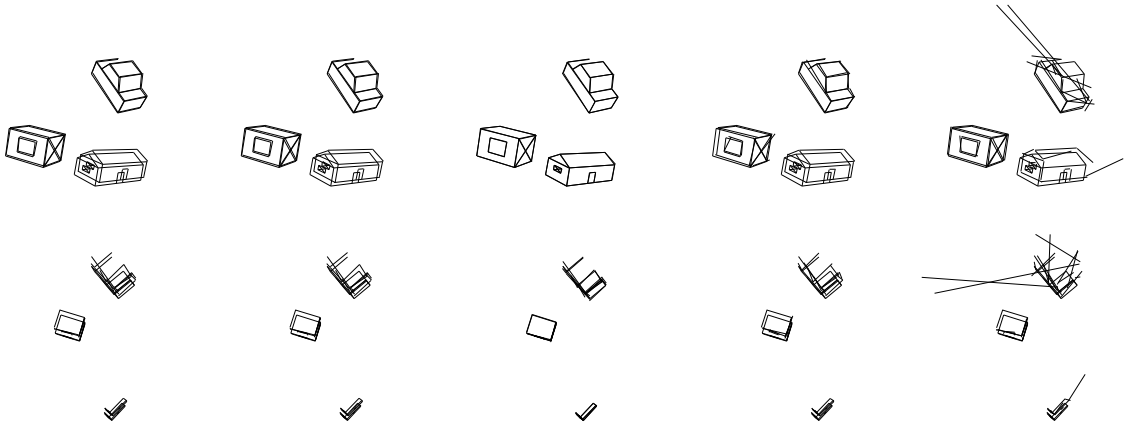

Figure 5: *The reconstructions for methods 1 to 5 for a noise level of $\sigma = 0.5$. The upper and lower rows show different views of the reconstructions. The third column shows that the explicit SVD approach is the most accurate, and the fifth column demonstrates that the problem with the cross ratio reconstruction approaches is generally the presence of outliers. The correct noise-free reconstructions are super-imposed in each case.*

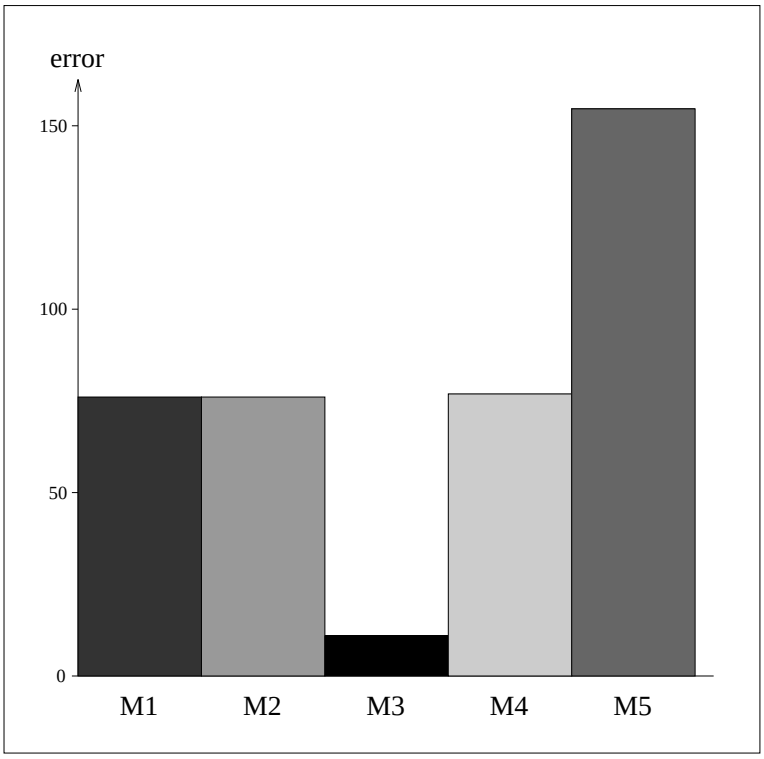

Figure 6: *A bar graph of the data set given in Table 3 for* $\sigma = 0.5$. *Note that method 3, which is the explicit reconstruction method based on the singular value decomposition, performs best.*

	$\sigma = 0.1$	$\sigma = 0.25$	$\sigma = 0.5$	$\sigma = 1$	$\sigma = 2$
M4 no SVD	938.5780	939.7471	1141.501	943.1791	2252.830
M4 with SVD	89.10260	174.0092	195.6364	247.9471	954.6233
M5 no SVD	76.90807	255.3177	376.0157	1523.277	1340.567
M5 with SVD	43.43694	57.02755	104.8106	352.7063	1164.160

Table 5: *Comparison for the methods M4 (Cayley invariants) and M5 (the cross ratios) with and without the use of singular value decomposition for averaging.*

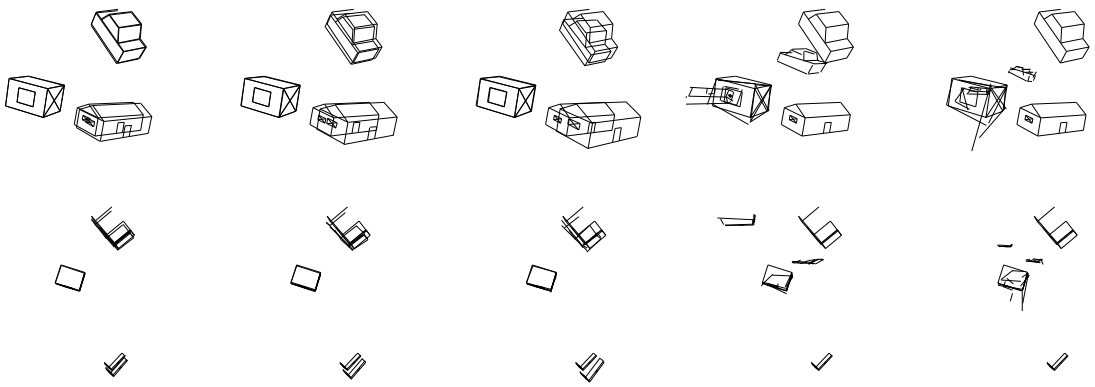

Figure 7: *The reconstructions found using singular value decomposition for different levels of noise ranging from $\sigma = 0.1$ at the left to $\sigma = 2$ on the right. The upper and low rows again show different views of the reconstructions.*

Finally, we show how reconstruction method 3, which is the one that has overall proved to be the most reliable, copes with difference levels of noise. In Figure 7 we show how this method degrades as noise is added to the synthetically projected image data.

6.3 Calibration grid

The study of the synthetic images has provided a fairly good understanding of the behaviours of the different reconstruction methods. We now proceed with a demonstration of the accuracy of the methods for images of a calibration grid whose geometry is known precisely. The object is somewhat artificial, but again we have ground-truth estimates of the three-dimensional structure and so can test out the stabilities of the methods with respect to errors that arise in the early visual processing. This time we use edge detection to provide the point features for the reconstructions.

A pair of sample images is given in Fig. 8 for which we compute the reconstructions shown in Fig. 10. Again, as shown in Table 6 and Fig. 9, the SVD approach provides stable results with the pseudo inverse being comparable. Both the Cayley invariants and those based on the cross ratio have much larger errors. The errors are given in millimetres, and so we find that the best methods have a reconstruction accuracy of about 1mm, which we believe to be good (each face of the calibration grid is 30cm by 30cm square).

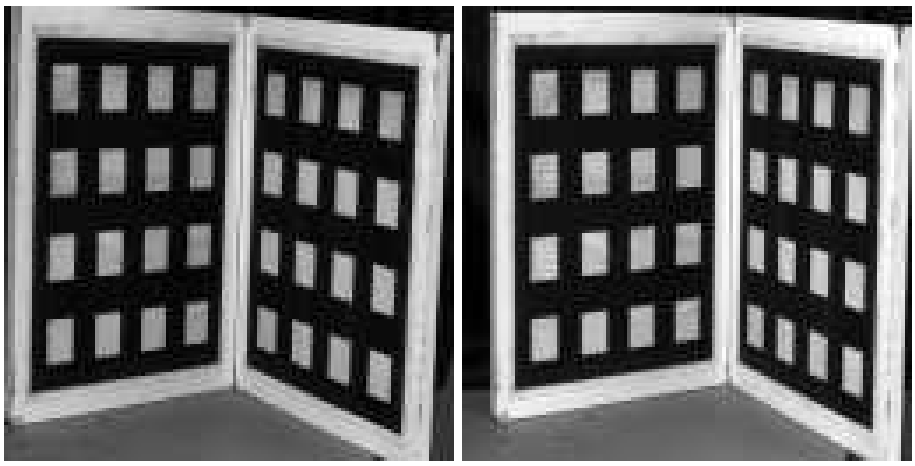

Figure 8: *The two images of the calibration grid used to test the different reconstruction methods.*

M1	M2	M3	M4	M5
1.129	3.465	1.145	6.745	25.533

Table 6: *The errors in millimetres associated with the reconstruction of the calibration grid in Fig. 8 when mapped to a Euclidean three-dimensional frame. Note that M1 and M3, the pseudo-inverse and SVD methods have the best performance, with M4 and M5, the Cayley and cross ratio invariants, having the worst.*

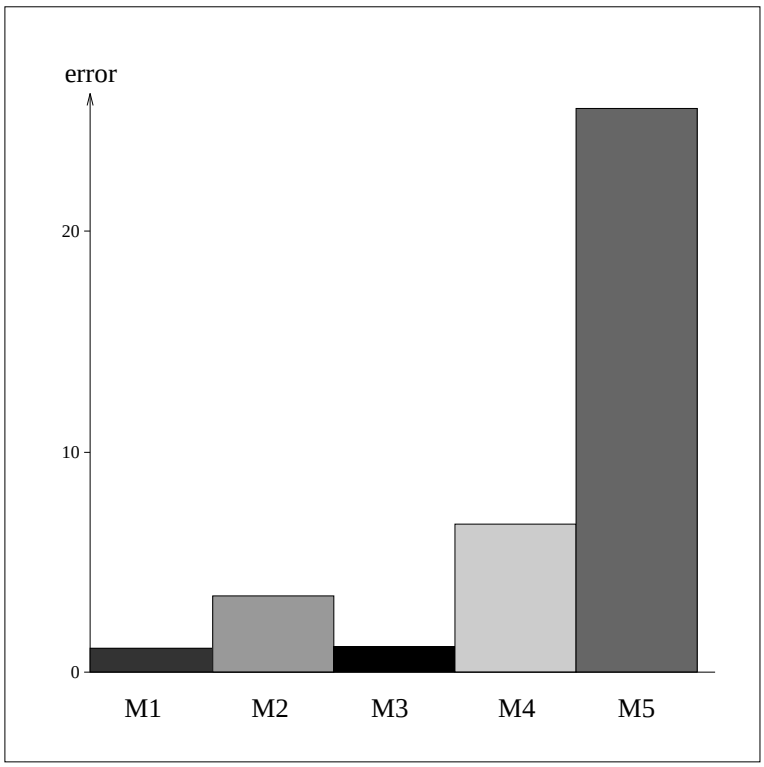

Figure 9: *A bar graph of the data represented in Table 6 for the reconstruction of the calibration grid. Note the similarity of the results with those given in Fig. 6 for the synthetic experiments.*

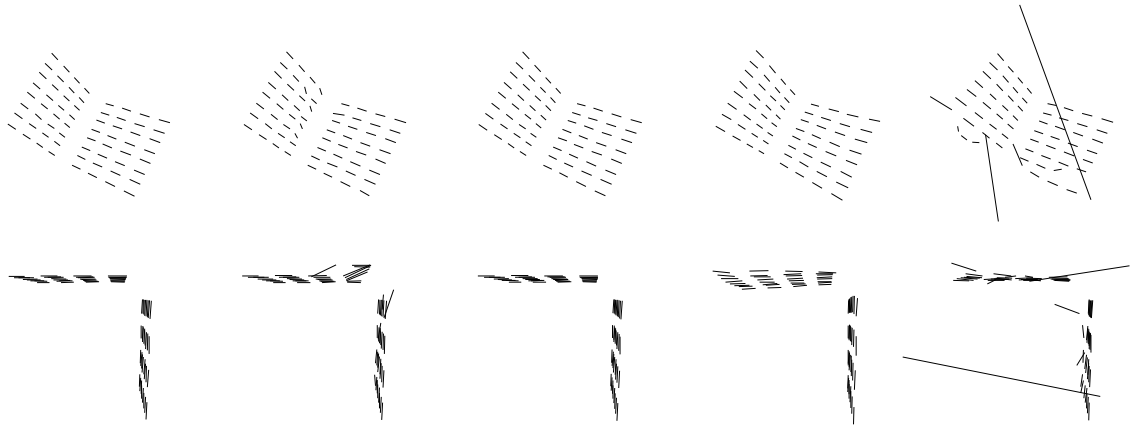

Figure 10: *Projections from two different viewpoints of the reconstructions of the calibration grid for the five different methods.*

6.4 Use of exact image distance

Although we have not discussed the method in detail throughout this report, there is another explicit reconstruction method which is of potential value. This method is based on the exact minimisation of image distance and is described in detail by Hartley and Sturm [16]. Should the noise we experience in images be Gaussian in nature (or in fact of a number of similar distributions), then minimising image distance will provide the optimal path to reconstruction. It also has the benefit of being invariant to the choice of projective frame we choose for the reconstruction.

Briefly, one can parametrize the distance of a pair of image observations in each image to the projection of a common epipolar plane in the same images by a sixth order polynomial in a single variable. Exhaustively finding the roots of this polynomial (or at least finding the three minima) provides an optimal choice of the respective epipolar lines in the images to which the reconstructed point should project. Then, all that needs to be done, is to choose the points in both images lying on the epipolar lines which lie closest to the original observations. As these points are in exact projective correspondence, the reconstruction can be done without any further need for minimisation.

Table 7 shows how the exact image minimisation approach performs against method 3 which computes structure using singular value decomposition. These results are characteristic of a number of similar experiments. Although the results are all fairly similar, we

image pair	SVD	Dist. Min.
1-2	34.93	34.98
2-3	35.45	35.98
3-4	24.24	24.30

Table 7: *The reconstruction errors for an image sequence in which we we compare the reconstruction qualities for the singular value decomposition approach and the minimisation of image distance by exact evaluation. Note that although the image distance approach is not much worse than that of the SVD, it definitely produces poorer quality results.*

see that the image minimisation approach always performs slightly worse than the SVD.[14] We therefore conclude that minimising image distance is in fact not an optimal approach and are led to infer that the noise distributions in the images cannot be Gaussian, and must be based around a different distribution. (Note that under projective projection a Gaussian distribution in three-dimensions does not project to equivalent distributions in the images. However, it also seems unlikely that the three-dimensional errors distributions are also Gaussian.)

6.5 More general objects

The main interest of our investigation into the differing stabilities of the reconstruction methods is so that we can determine which method will work best for a range of reconstruction, navigation, and recognition tasks. Consequently, we demonstrate how well the approaches deal with images that would be typical for such applications: a building which has cars parked around it and other features in the background (see Fig. 11).

We have also concentrated on using edge based data for the experiments rather than corner features as used by [2, 13]. There are two reasons for this:

1. Corner detector technology is currently far behind the abilities of edge detectors as far as accuracy and reliability of detection go. Edges are consequently far more robust to the degradations in image quality we experience in practice. We therefore prefer to compute correspondences on edge data (using the algorithms and implementations of [5, 25]) and then compute the three-dimensional structure from these.

2. Corner-based descriptions are relatively sparse, providing only a few hundred data points in a scene. They also ignore all of the topology that is present in the scene. We believe that recovering scene structure coherently is simpler when we make use of edgel chain topology [27], rather than by attempting to reconstruct it using methods such as Delauney triangulation [13].

[14] Note that the approach based on the numerical minimisation of image distance on average performed far worse than method 3. We are thus led to conclude that this method (method 2), sometimes becomes stuck within certain local minima and so produces erroneous reconstructions for *some* points.

Figure 11: *The two images used of the reconstruction. Note the similarity in viewpoint; we are in fact achieving remarkably good reconstructions using very small base-line stereo.*

Reconstructions for the different methods based on the image pair in Fig. 11 are given in Fig 12. Note how little the viewpoint has changed between the two views. As the singular value decomposition method performs better than the other extrinsic reconstruction methods,[15] we consider only methods 3, 4, and 5 for this example. Observe how the explicit reconstruction method based on singular value decomposition is again far more stable than both the Cayley and the cross ratio invariant methods. However, the figures show that a significant amount of the data for all three methods has their structure estimated correctly, and the errors tend to be associated with a number of outliers. The data points cluster together to form the walls of the buildings, and the sides of the cars. Certainly, all of the results are sufficiently good to be used as a starting point for an iterative process that estimates the structure over many images. This type of processing constitutes the next step in the development of the algorithms.

6.5.1 Integrating the structure

The above examples have considered only individual points. We can recover a much more robust scene reconstruction (which includes the topology vital to subsequent visual processing) if we represent straight line segments in the image by actual lines rather than sets of edgel points. Reconstruction using the lines is actually very simple. Even though we have pairs of corresponding line segments, we do not reconstruct using their endpoints as these are unlikely to correspond to exactly the same points in space (each line is likely to be shortened by differing amount due to segmentation problems). However, we are able to put the line endpoints in exact correspondence using the epipolar structure (one endpoint of a line in the second image should lie on the epipolar line of the endpoint of its corresponding

[15] In most of the results we have shown in this report, the method based on the use of the pseudo-inverse (method 1) performs nearly as well as that using singular value decomposition (method 3). This would only be expected when the projective cameras are chosen so that the ideal plane does not pass through the three-dimensional data set in the reconstruction. Although chance has meant that this event has not happened for the examples given, there is no reason for this to be the case, and so we prefer the generality of method 3 over method 1.

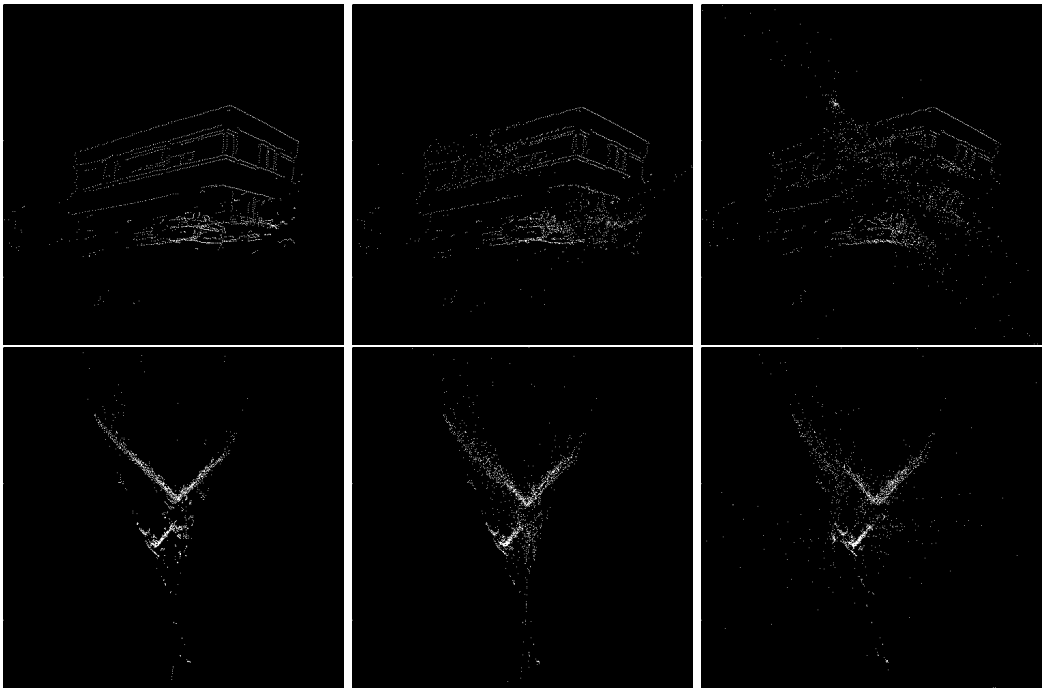

Figure 12: *The top row shows reconstructions for methods 3 through 5 when viewed from positions near to the original camera locations. From such a viewpoint all of the reconstructions are reasonable. The bottom row shows the same reconstruction from above the building, demonstrating the superiority of the explicit reconstruction method. Note however, that even for the two implicit methods, a significant proportion of the data set is reconstructed correctly, and that only the number of outliers is larger.*

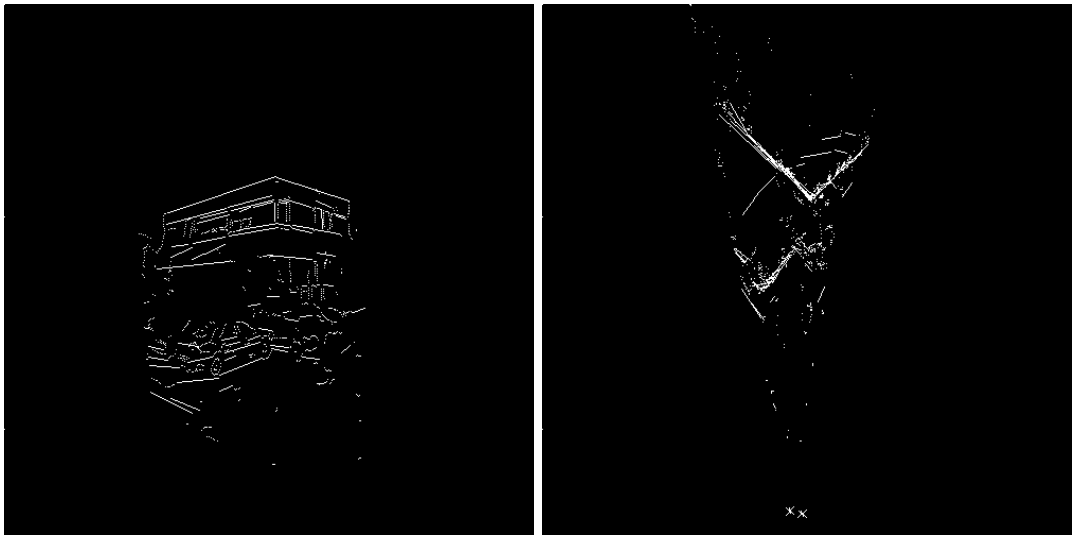

Figure 13: *If we fit lines to the image edge data and then reconstruct using method 3 we can recover a more accurate and reliable understanding of the three-dimensional geometry. Furthermore, we can begin to reason about the more general world structure such as searching for connectivity and coplanarity constraints. The two camera positions are marked in the lower part of the right hand image, and points that are not suitably represented by lines are reconstructed as isolated points.*

line). We therefore adjust the locations of the endpoints of the lines, and then reconstruct using the line endpoints. We demonstrate this in Fig 13.

7 Discussion

In this paper we have reviewed a number of stereo reconstruction algorithms which assume only knowledge of the weak calibration between the cameras; for all of the examples based in real images given in Section 6, the weak calibration was computed automatically using the algorithm of [5]. The reconstruction algorithms have been divided into two distinct classes, *explicit* and *implicit*. The former class are relatively familiar within the domain of calibrated stereo, and the latter have strong connections with invariant theory. In summary we have found that:

- Projective reconstructions are accurate. Although we have not tested the algorithms directly against conventional calibrated stereo approaches, it appears that much can be gained from freeing the reconstruction process from calibration. In effect, the use of calibration (which is very likely to be erroneous) introduces incorrect assumptions that

only detract from the qualities of the reconstructions (here we ignore non-pinhole camera distortions in the image). Perhaps introducing the calibration post-reconstruction (rather than pre-reconstruction) actually leads to more robust scene measurements, though for our purposes we are interested only in projective reconstructions.

- The explicit reconstruction methods tend to produce more reliable reconstructions over the entire data set, with an approach based on using singular value decomposition to solve the constraint system providing the best results. We do not recommend the use of method 1 based on the use of the pseudo inverse due to its fragility with respect to the positioning of the ideal plane.

- The implicit reconstruction methods suffer from having a few outliers that detract from the overall quality of the reconstructions. These outliers actually result from a breakdown of some of the assumptions used in the computation of the invariants. (For example, the cross ratio computation assumed that we could extract four different planes from a pencil, for some point configurations we find only three distinct planes. Although this can most likely fixed, it is at present a problem.)

- We have in practice found no great difficulty in computing general projective reconstructions. We are actually able to choose the camera matrices for method 3 (which is the one that works best) almost arbitrarily, and certainly without the need to use a *quasi-Euclidean* basis. This suggests that the uncalibrated stereo reconstruction task is in general solvable without requiring the assumptions made by Beardsley, *et al.* [2].

- The quality of the reconstructions is enhanced greatly when we use edge-based stereo rather than the output of a corner detector. Although we have not shown results for features derived from a corner detector, we have found that reconstructions that do use these features are far less accurate. Furthermore, fitting lines or in fact other features to the edge data sets improves the accuracy further. Through the use of the weak calibration we can almost trivially extend the point-based algorithms to line segments.

Finally, we emphasize the fact that this investigation has studied only reconstruction algorithms for pairs of views. In any major application we would expect to integrate information over a large number of views and hence derive very reliable structure estimates. Details of the benefits of using a Kalman filter to track the structure are given by Beardsley, *et al.* [2]. Our future intentions are to include such a notion of temporal filter for ameliorating the projective structure.

A Deriving the projective camera model

Here we derive the camera forms given in Section 3. According to the projective model given by [26], each camera $\mathtt{P}_i$, $i \in \{1,2\}$ has eleven independent parameters. Hartley gives a parametrization of the cameras based on $\mathtt{P} = [\mathtt{M}| - \mathtt{M}\mathbf{t}]$, where $\mathtt{M}$ is non-singular [15]. The reason for using this decomposition is so that the camera centre appears at the affine point $(\mathbf{t}^\top, 1)^\top$. However, if we wish to have the freedom to centre the camera on the ideal plane (as might be necessary for general projective reconstructions) we must resort to a more general form $\mathtt{p} = [\mathtt{M}|\mathbf{t}]$ (now $\mathtt{M}$ can be singular, though we can experience problems in the derivations of the epipolar structure given below, though not the actual cameras, should the optical centres actually lie on the ideal plane).

Now, consider projecting an affine point $(\mathbf{X}_3, 1)$ from the world into each image. On the elimination of $\mathbf{X}$ from the projection constraints $k_i\mathbf{x}_i = \mathtt{P}_i\mathbf{X}$, we find:

$$\lambda\ \mathbf{e}_{ij} = \mathbf{t}_i - \mathtt{M}_i\ \mathtt{M}_j^{-1}\ \mathbf{t}_j \qquad \text{and} \qquad \mu\ \mathtt{F}_{ij} = [\mathbf{e}_{ji}]_\times\ \mathtt{M}_j\ \mathtt{M}_i^{-1}. \tag{18}$$

We subsequently derive the constraints on the forms $\mathtt{M}_i$ and $\mathbf{t}_i$ so that the epipolar constraints are satisfied. This in turn leads to solutions for the $\mathtt{P}_i$. As we have freedom in the choice of the projective frame we may set $\mathtt{P}_1 = [\mathtt{I}|\mathbf{0}]$, where $\mathtt{I}$ is the 3×3 identity matrix. This is equivalent to fixing the form of $\mathtt{G}$ in eqn (10), and in effect aligns the first camera with the world coordinate frame and fixes its optical centre and the world origin. We must also consider which projectivities of space may be applied to our system so that the parametrization of the first camera remains unchanged. This is so that we know the number of degrees of freedom that would be required to map our assumed frame back to the actual frame (were it known). The family of projectivities is quite simply:

$$\mathtt{H} = \begin{bmatrix} \mathtt{I} & \mathbf{0} \\ \alpha^\top & \alpha_4 \end{bmatrix}.$$

Here the vector $\alpha = (\alpha_1, \alpha_2, \alpha_3)^\top$ has a full three degrees of freedom and so $\mathtt{H}$ has a total of four. Now, as $\mathtt{M}_1 = \mathtt{I}$ and $\mathbf{t}_1 = \mathbf{0}$ the epipolar constraints yield:

$$\lambda\ \mathbf{e}_{21} = \mathbf{t}_2 \qquad \text{and} \qquad \mu\ \mathtt{F}_{12} = [\mathbf{e}_{21}]_\times\ \mathtt{M}_j.$$

Expanding the second of these constraints by columns, and using the superscript prefix i to denote the i^{th} column of a matrix, gives up to scale ${}^i\mathtt{F} = \mathbf{e}_{21} \times {}^i\mathtt{M}$, $i \in \{1,2,3\}$. Therefore, given the fundamental matrix and the epipole:

$${}^i\mathtt{M} = \frac{{}^i\mathtt{F} \times \mathbf{e}_{21}}{|\mathbf{e}_{21}|^2} + \gamma_i\mathbf{e}_{21}.$$

Moving all of the γ_i to a projectivity on the right we derive the desired simple form for $\mathtt{P}_2$:

$$\begin{aligned}\mathtt{P}_2 &= \left[\frac{{}^1\mathtt{F}\times\mathbf{e}_{21}}{|\mathbf{e}_{21}|^2},\ \frac{{}^2\mathtt{F}\times\mathbf{e}_{21}}{|\mathbf{e}_{21}|^2},\ \frac{{}^3\mathtt{F}\times\mathbf{e}_{21}}{|\mathbf{e}_{21}|^2},\ \mathbf{e}_{21}\right]\begin{bmatrix}\mathtt{I} & \mathbf{0}\\ \gamma^\top & \gamma_4\end{bmatrix},\\ &= \left[[\mathbf{e}_{21}]_\times\,\mathtt{F}_{12}\mid\mathbf{e}_{21}\right]\begin{bmatrix}\mathtt{I} & \mathbf{0}\\ \alpha^\top & \alpha_4\end{bmatrix}.\end{aligned}$$

Note that the factor $1/|\mathbf{e}_{21}|^2$ has been absorbed into the overall scale of $\mathtt{P}_2$ (and is accounted for by the α_i), with the '$=$' denoting equivalence up to scale. This result means that given the weak calibration constraints, which can be represented entirely by the single matrix $\mathtt{F}_{12}$, we are able to find solutions for the two camera matrices in some projective frame.

References

[1] Ayache, N. and Faugeras, O.D. "Building a Consistent 3D Representation of a Mobile Robot Environment by Combining Multiple Stereo Views," Proceedings IJCAI, p.808-810, 1987.

[2] Beardsley, P., Zisserman, A. and Murray, D. "Sequential Update of Projective and Affine Structure from Motion," *TR OUEL 2012/94*, Department of Engineering Science, University of Oxford, 1994.

[3] Carlsson, S. "Multiple Image Invariants Using the Double Algebra," Proceedings 2nd ESPRIT-ARPA-NSF Workshop on Invariance, Azores, p.335-350, 1993.

[4] Csurka, G. and Faugeras, O.D. "Computing Three-dimensional Projective Invariants from a Pair of Images Using the Grassmann-Cayley Algebra," in Proceedings Europe-China Workshop on Geometrical Modeling & Invariants for Computer Vision, p.150-157, 1995.

[5] Deriche, R., Zhang, A., Luong, Q.-T. and Faugeras, O.D. "Robust Recovery of the Epipolar Geometry for an Uncalibrated Stereo Rig," Proceedings ECCV94, p.565-576, 1994.

[6] Duda, R.O. and Hart P.E. *Pattern Classification and Scene Analysis,* Wiley, 1973.

[7] Faugeras, O.D. "What can be Seen in Three Dimensions with an Uncalibrated Stereo Rig?" Proceedings ECCV2, p.563-578, 1992.

[8] Faugeras, O.D, Luong, Q.-T. and Maybank, S.J. "Camera Self-Calibration: Theory and Experiments," in Proceedings ECCV92, p.321-334, 1992.

[9] Faugeras, O.D. *Three-Dimensional Computer Vision: a Geometric Viewpoint,* MIT Press, 1993.

[10] Faugeras, O.D. "Stratification of 3D Vision: Projective, Affine and Metric Representations," to appear *JOSA,* 1995.

[11] Grimson, W.E.L. "Computational Experiments with a Feature Based Stereo Algorithm," *PAMI,* Vol. 7, No. 1, p.17-34, 1985.

[12] Gros, P. "Outils Geometriques pour la Modelisation et la Reconnaissance d'Objets Polyedriques," Ph.D. thesis, LIFIA-IMAG-INRIA, 1993.

[13] Harris, C.G. and Pike, J.M. "3D Position Integration from Image Sequences," Proceedings 3rd Alvey Vision Conf., p.233-236, 1987.

[14] Hartley, R.I., Gupta, R. and Chang, T. "Stereo from Uncalibrated Cameras," Proceedings CVPR92, p.761-764, 1992.

[15] Hartley, R. "Projective Reconstruction and Invariants from Multiple Images," *PAMI,* Vol. 16, No. 10, p.1036-1040, 1994.

[16] Hartley, R. and Sturm, P. "Triangulation," Proceedings ARPA IUW, 1994.

[17] Longuet-Higgins, H.C. "A Computer Algorithm for Reconstructing a Scene from Two Projections," *Nature,* No. 293, p.133-135, 1981.

[18] Luong, Q.-T. and Viéville, T. "Canonic Representations for the Geometries of Multiple Projective Views," Proceedings ECCV94, p.589-599, 1994.

[19] Luong, Q.-T. and Faugeras, O.D. "The Fundamental Matrix: Theory, Algorithms, and Stability Analysis," to appear *IJCV,* 1995

[20] Mohr, R., Veillon, F. and Quan, L. "Relative 3d Reconstruction using Multiple Uncalibrated Images," Proceedings CVPR93, p.543-548, 1993.

[21] Mohr, R., Boufama, B. and Brand, P. "Accurate Projective Reconstruction," Proceedings 2nd ESPRIT-ARPA-NSF Workshop on Invariance, Azores, p257-276, 1993.

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

[23] Pitas, I. and Venetsanopoulos, N. "Order Statistics in Digital Image Processing," *Proceedings IEEE,* Vol. 80, No. 12, p.1893-1921, 1992.

[24] Pollard, S.B, Pridmore, T.P, Porrill, J., Mayhew, J.E.W. and Frisby, J.P. "Geometrical Modelling from Multiple Stereo Views," *International Journal of Robotics Research,* Vol. 8, No. 4, p.132-138, 1989.

[25] Robert, L. "Perception Stereoscopique de Courbes et de Surfaces Tridimensionnelles. Applications à la Robotique Mobile," Ph.D. thesis, L'Ecole Polytechnique, 1993.

[26] Roberts, L.G. "Machine Perception of Three-Dimensional Solids," *Optical and Electro-optical Information Processing,* Tippett, *et al.* editors, MIT Press, p.159-197, 1965,

[27] Rothwell, C., Mundy, J., Hoffman, W. and Nguyen, V.-D. "Driving Vision by Topology," *TR-2444,* INRIA, 1994.

[28] Semple, J.G. and Kneebone, G.T. *Algebraic Projective Geometry,* Oxford University Press, 1952.

[29] Sinclair, D.A., Blake, A., Smith, S. and Rothwell, C.A. "Planar Region Detection and Motion Recovery," Proceedings BMVC92, p.59-68, 1992, and *Image and Vision Computing,* Vol. 11, No. 4, p.229-234, 1993.

[30] Shashua, A. "Algebraic Functions for Recognition," *PAMI,* in press, 1994.

[31] Springer, C.E. *Geometry and Analysis of Projective Spaces,* Freeman, 1964.

[32] Sturmfels, B. *Algorithms in Invariant Theory,* Springer-Verlag, 1992.

[33] Torr, P. and Murray, D. "Outlier Detection and Motion Segmentation," in Proceedings SPIE Sensor Fusion Conference, p.432-443, 1993.

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 46 avenue Félix Viallet, 38031 GRENOBLE Cedex 1
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
ISSN 0249-6399