# Metric Calibration of a Stereo Rig

Andrew Zisserman, Paul A Beardsley and Ian D Reid
Robotics Research Group
Department of Engineering Science
University of Oxford
Oxford OX1 3PJ, UK.

**Abstract**

*We describe a method to determine affine and metric calibration for a stereo rig. The method does not involve the use of calibration objects or special motions, but simply a single general motion of the rig with fixed parameters (i.e. camera parameters and relative orientation of the camera pair). The novel aspects of this work are: first, relating the distinguished objects of Euclidean geometry to fixed entities of a Euclidean transformation matrix; second, showing that these fixed entities are accessible from the conjugate Euclidean transformation arising from the projective transformation of the structure under a motion of the fixed stereo rig; third, a robust and automatic implementation of the method.*

*Results are included of affine and metric calibration and structure recovery using images of real scenes.*

## 1 Introduction

It is known that from two uncalibrated views, 3D structure can be recovered modulo a projectivity of 3-space [5, 8]. Furthermore, given three or more views acquired with a fixed camera, this projective ambiguity can be reduced to metric [7, 12] by computing the internal camera parameters (calibration). However, this *self-calibration* method requires solving a system of polynomial equations. An alternative strategy is to first recover affine structure, and thence solve for camera intrinsic parameters using relations generated by the affine calibration [9, 11]. The advantage of this two-stage (*stratified*) approach is that the equations on the internal camera parameters are linear.

For example, affine structure can be recovered from two uncalibrated images with pure (but unknown) translation between the views and unchanging (but unknown) internal parameters [13]. Subsequent rotated views can then be used to generate linear constraints on the internal camera parameters [1].

We investigate here methods for recovering metric structure from a single general motion of a fixed stereo rig (i.e. two stereo pairs of views, each with fixed — though different — internal parameters for the left and right cameras, and a fixed relation between the cameras). It is shown that:

1. Affine structure can be computed uniquely using only linear methods (eigenvectors).
2. Metric structure can be computed up to a one parameter family, again using linear methods, and reduced to a two fold ambiguity using a quadratic constraint.

From each uncalibrated stereo pair the recovered 3D structure is related by a collineation to the actual Euclidean 3D structure. Consequently, the two reconstructions are related to each other by a collineation (a projective transformation represented by a non-singular $4 \times 4$ matrix acting on homogeneous vectors). This projective transformation, in particular its eigenvectors, is the basis of the calibration method described here. The key result used is that this transformation is a special projectivity; it is conjugate to a Euclidean transformation.

In order to "upgrade" 3D projective structure to metric certain geometric objects have to be identified, namely the plane at infinity $\pi_\infty$ and the absolute conic $\Omega_\infty$. These are precisely the objects that are invariant to Euclidean transformations of 3-space. In the following we review these two objects (section 2), and describe their various manifestations in both a matrix representing Euclidean transformation (section 2.2) and the projective transformation which is conjugate to this (section 2.3). The algorithm for affine and metric calibration is covered in section 3. Implementation details and results are given in section 4.

### Notation

We will not distinguish between the Euclidean and similarity cases, both will be loosely referred to as metric.

Image coordinates are lower case 3-vectors, e.g. $\mathbf{x}$, world coordinates are upper case 4-vectors, e.g. $\mathbf{X}$. Lower case subscripts are used to distinguish left and right quantities, for example the $3 \times 4$ camera projection matrices are $\mathtt{P}_l$ and $\mathtt{P}_r$, e.g. $\mathbf{x}_l = \mathtt{P}_l\mathbf{X}$ is projection onto the left camera. Upper case subscripts are used to distinguish projective, affine and Euclidean, for example $\mathbf{X}_A$ $\mathbf{X}_E$ are affine and Euclidean structure respectively. Quantities measured after the motion (i.e. transformed quantities) are primed, e.g. $\mathbf{X}' = \mathtt{T}\mathbf{X}$ for a transformation by a $4 \times 4$ matrix $\mathtt{T}$, with $\mathbf{X}$, $\mathbf{X}'$ representing structure recovered from a pair of views before and after the transformation. The vector product $\mathbf{v} \wedge \mathbf{x}$ is represented as a matrix multiplication $\mathbf{v} \wedge \mathbf{x} = [\mathbf{v}]_\times \mathbf{x}$, where $[\mathbf{v}]_\times$ denotes the $3 \times 3$ skew matrix associated with the vector $\mathbf{v}$. For homogeneous quantities, $=$ indicates equality up to a non-zero scale factor.

# 2 Fixed Entities under Euclidean Actions

## 2.1 Conjugate transformations

The two images available from each position of the rig allow the recovery of 3D projective structure. After a (Euclidean) motion of the rig, the corresponding 3D points obtained from successive stereo pairs are related by a projective transformation of 3-space.

Suppose $\mathbf{X}_E$ and $\mathbf{X}'_E$ are homogeneous 4-vectors representing the original and transformed Euclidean structure respectively. Then corresponding points are related by

$$\mathbf{X}'_E = \mathtt{T}_E \mathbf{X}_E \tag{1}$$

where $\mathtt{T}_E$ is a non-singular $4 \times 4$ matrix that represents the Euclidean action. If $\mathbf{X}$ and $\mathbf{X}'$ is the projective structure recovered before and after the motion, then

$$\mathbf{X}_E = \mathtt{T}_{EP}\mathbf{X} \quad \mathbf{X}'_E = \mathtt{T}_{EP}\mathbf{X}' \tag{2}$$

where $\mathtt{T}_{EP}$ is a non-singular $4 \times 4$ matrix which relates metric to projective structure. Note, the collineation $\mathtt{T}_{EP}$ is the same before and after the motion of the stereo rig since both the internal and relative external parameters of the cameras are fixed. From equations (1) and (2) it follows that:

$$\mathbf{X}' = \mathtt{T}\mathbf{X},$$

where $\mathtt{T}$ is a non-singular $4 \times 4$ matrix that relates the projective structure before and after the motion. The matrix $\mathtt{T}$ is not a general projective transformation, but conjugate to a Euclidean transformation matrix, i.e.

$$\mathtt{T} = \mathtt{T}_{EP}^{-1}\mathtt{T}_E\mathtt{T}_{EP} \tag{3}$$

We note two important properties of these transformations:

1. $\mathtt{T}$ and $\mathtt{T}_E$ have the same eigenvalues (since (3) is a similarity transformation).

2. If $\mathbf{E}$ is an eigenvector of $\mathtt{T}_E$ then the corresponding eigenvector of $\mathtt{T}$, with the same eigenvalue, is $(\mathtt{T}_{EP}^{-1}\mathbf{E})$, i.e. the eigenvectors of $\mathtt{T}_E$ undergo the same transformation (2) as the structure. This follows from (3) for if $\mathtt{T}_E\mathbf{E} = \lambda\mathbf{E}$, then $\mathtt{T}_{EP}\mathtt{T}\mathtt{T}_{EP}^{-1}\mathbf{E} = \lambda\mathbf{E}$, and pre-multiplying by $\mathtt{T}_{EP}^{-1}$ gives the desired result.

In the following we describe the geometric significance of these eigenvectors, and how they are used for calibration.

## 2.2 Fixed entities and eigenvectors of a Euclidean transformation

Any particular translation and rotation is equivalent to a rotation about a *screw axis* (the screw axis is parallel to the rotation axis) and translation along the screw axis [4]. Consequently, this is the most general Euclidean action we need consider. The plane at infinity and absolute conic are invariant under *all* Euclidean actions. Under a *particular* action, $\mathtt{T}_E$, there are additional fixed geometric entities. These arise because planes perpendicular to the rotation axis do not change orientation under $\mathtt{T}_E$ — the planes are rotated about, and translated along the axis. Similarly, the intersection of the rotation axis with $\pi_\infty$ is unchanged. There are therefore three extra fixed entities:

1. **A fixed line on** $\pi_\infty$ the axis of the pencil of planes perpendicular to the rotation axis.

2. **A fixed point on** $\pi_\infty$ the intersection of $\pi_\infty$ and the rotation axis.

3. **Two fixed points on** $\Omega_\infty$ the intersection of the fixed line and $\Omega_\infty$ — these are the circular points on the planes perpendicular to the rotation axis. Since the transformation induced in any plane of the pencil is Euclidean, the circular points are invariant [15].

These three fixed entities can be extracted directly from the fixed points of the $4 \times 4$ matrix representing $\mathtt{T}_E$ which, without loss of generality, can be chosen as:

$$\mathtt{T}_E = \begin{bmatrix} \mathtt{R} & \mathbf{T} \\ \mathbf{0}^\top & 1 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 & 0 \\ \sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{4}$$

where the rotation axis is the $Z$ axis, and the translation is along the $Z$ axis. The eigenvalues and eigenvectors of $\mathtt{T}_E$ are $\lambda = \{e^{i\theta}, e^{-i\theta}, 1, 1\}$ and

$$\mathbf{E}_1 = \begin{bmatrix} 1 \\ -i \\ 0 \\ 0 \end{bmatrix} \quad \mathbf{E}_2 = \begin{bmatrix} 1 \\ i \\ 0 \\ 0 \end{bmatrix} \quad \mathbf{E}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \quad \mathbf{E}_4 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

Note that the matrix has a repeated eigenvalue (algebraic multiplicity equal to two), but the corresponding eigenvectors are identical (geometric multiplicity equal to one). The matrix is therefore defective [14]. $\mathbf{E}_1$ and $\mathbf{E}_2$ are complex points on $\pi_\infty$ lying on the line of intersection of $\pi_\infty$ and the $XY$ plane. $\mathbf{E}_3 = \mathbf{E}_4$ is the intersection of the $Z$ axis with $\pi_\infty$. Note that $(\mathbf{E}_1+\mathbf{E}_2)$, $(\mathbf{E}_1 - \mathbf{E}_2)$, and $\mathbf{E}_3$ are an orthogonal triad.

It is straightforward to associate these eigenvectors with the additional fixed geometric entities:

1. $\mathbf{E}_1$ and $\mathbf{E}_2$ are the circular points, defining fixed points on $\Omega_\infty$ and the fixed line.

2. $\mathbf{E}_3$ is the other fixed point on $\pi_\infty$.

In a similar manner we can catalogue the fixed entities of the plane transformation matrix, the conic transformation on $\pi_\infty$ etc.

For the purposes of Euclidean calibration we need to identify $\pi_\infty$ and $\Omega_\infty$. These are invariant under *all* Euclidean transformations, but can they be *recovered* from a particular (i.e. single) transformation?

**The plane at infinity,** $\pi_\infty$ : From the above, we have three fixed points on $\pi_\infty$, this determines the plane. Furthermore, if the point transformation matrix is $\mathtt{T}_E$, then the plane transformation matrix is $\mathtt{T}_E^{-\top}$ [16]. The matrix $\mathtt{T}_E^{-\top}$ also has two equal, unit eigenvalues and a single eigenvector corresponding to these. This eigenvector corresponds to $\pi_\infty$ which is thus directly, and uniquely, computed.

**The absolute conic,** $\Omega_\infty$ : On $\pi_\infty$ the subordinate plane projective transformation [15] corresponding to $\mathtt{T}_E$ is the $3 \times 3$ rotation matrix $\mathtt{R}$. Under a point transformation $\mathtt{R}$, a conic $\mathtt{C}$ transforms as $\mathtt{C} \rightarrow \mathtt{RCR}^\top$. The fixed conics of this transformation (i.e. of $\mathtt{C} = \mathtt{RCR}^\top$) are the one-parameter family (see [6]):

$$\mathtt{C} = \lambda \mathtt{I} + \mu \mathbf{a}\mathbf{a}^\top$$

where $\mathtt{R}\mathbf{a} = \mathbf{a}$ i.e. there is one parameter family of conics which are preserved on $\pi_\infty$ by a particular Euclidean action on 3-space. This family includes $\Omega_\infty$, but is not restricted to it (of course only $\mathtt{C} = \mathtt{I}$ will be preserved under *all* rotations).

We now have the answer to the question posed above: $\pi_\infty$ is accessible from the fixed entities of a single transformation matrix (it is the eigenvector with real eigenvalue of $\mathtt{T}_E^{-\top}$). However, $\Omega_\infty$ is determined only up to a one parameter family.

It is worth noting that in the special case that the translation direction and rotation axis are orthogonal, there is a line of fixed points corresponding to the screw axis (rather than solely a fixed point of intersection between the rotation axis and $\pi_\infty$ as in the general case). However, in this case there is a two-dimensional space of eigenvectors corresponding to the repeated (real) eigenvalue. The consequence is that $\pi_\infty$ is recovered only up to a one parameter family, but the screw axis can be recovered uniquely. This case is discussed in more detail in [2].

### 2.3 Fixed entities and eigenvectors of a conjugate Euclidean transformation

From the discussion of section 2.1 it follows that the eigenvectors of the matrix $\mathtt{T}$ which relates the projective reconstructions are the transformed eigenvectors of the Euclidean transformation. Consequently, the eigenvectors of $\mathtt{T}$ (which are known) define $\pi_\infty$ and three orthogonal directions in the projective reconstruction. This clearly enables affine structure recovery.

## 3 Calibration of a stereo rig

We have shown that the real eigenvector of $\mathtt{T}_E^{-1}$ (i.e. the eigenvector corresponding to the real eigenvalue) is $\pi_\infty$ in Euclidean 3-space, and consequently the real eigenvector of $\mathtt{T}^{-\top}$ is $\pi_\infty$ in the projective frame of the structure $\mathbf{X}$.

### 3.1 Affine Calibration

In order for the reconstruction to be affine, rather than projective, it is necessary to know the infinite homography [11], $\mathtt{H}_\infty$, between the left and right cameras of the stereo pair. The infinite homography is the $3 \times 3$ point transformation matrix for vanishing points, i.e. $\mathbf{v}_r = \mathtt{H}_\infty \mathbf{v}_l$, where $\mathbf{v}_l$ and $\mathbf{v}_r$ are corresponding vanishing points in the left and right images respectively.

Once $\mathtt{H}_\infty$ is known, then an affine reconstruction can be achieved by using the $3 \times 4$ camera projection matrices $\mathbf{x}_l = \mathtt{P}_{lA}\mathbf{X} \quad \mathbf{x}_r = \mathtt{P}_{rA}\mathbf{X}$, where

$$\mathtt{P}_{lA} = [\mathtt{I} |\, \mathbf{0}] \quad \mathtt{P}_{rA} = [\mathtt{H}_\infty \,|\, \mathbf{e}_r]$$

and $\mathbf{e}_r$ is the epipole in the right camera, i.e. $\mathtt{F}^\top \mathbf{e}_r = 0$, with $\mathtt{F}$ the fundamental matrix of the rig.

The transformation $\mathtt{H}_\infty$, which is a plane projective transformation, can be computed directly or indirectly from the eigenvectors:

**Direct computation of $\mathtt{H}_\infty$** Suppose the plane $\pi_\infty$ defined by the real eigenvector of $\mathtt{T}^{-\top}$ has coefficients $\pi_\infty = (\mathbf{a}, 1)$ in projective 3-space, where $\mathbf{a}$ is a 3-vector. Then the infinite homography is:

$$\mathtt{H}_\infty = \mathtt{M} - \mathbf{e}_r \mathbf{a}^\top$$

To see this, consider the ray $(\mathbf{x}_l, s)^\top$ which is the back-projection of the point $\mathbf{x}_l$ in the left camera. This ray intersects $\pi_\infty$ at $(\mathbf{a}, 1)(\mathbf{x}_l, s)^\top = 0$, i.e. at $s = -\mathbf{a}.\mathbf{x}_l$. This intersection point $\mathbf{V} = (\mathbf{x}_l, -\mathbf{a}.\mathbf{x}_l)^\top$ on $\pi_\infty$, is imaged at $\mathbf{x}_r = \mathtt{P}_r\mathbf{V} = \mathtt{M}\mathbf{x}_l - \mathbf{e}_r\mathbf{a}^\top\mathbf{x}_l$. This defines the transformation between the left and right cameras induced by the plane at infinity, i.e. $\mathtt{H}_\infty$.

**Indirect computation of $\mathtt{H}_\infty$** The infinite homography can be computed from the correspondence of three vanishing points (and the epipole) between the left and right images. In this case three vanishing points are provided by the projected eigenvectors of $\mathtt{T}$. To avoid complex numbers ($\mathtt{H}_\infty$ is real), three real points $\mathbf{V}_i$ on $\pi_\infty$ are defined by:

$$\begin{aligned} \mathbf{V}_1 &= (\mathbf{E}_1 + \mathbf{E}_2) \\ \mathbf{V}_2 &= (\mathbf{E}_1 - \mathbf{E}_2) \\ \mathbf{V}_3 &= \mathbf{E}_3, \end{aligned}$$

and the corresponding image points are $\mathbf{v}_{il} = \mathtt{P}_l\mathbf{V}_i$ $\mathbf{v}_{ir} = \mathtt{P}_r\mathbf{V}_i$. $\mathtt{H}_\infty$ is then computed from the four corresponding image points:

$$\begin{aligned} \mathbf{v}_{ir} &= \mathtt{H}_\infty \mathbf{v}_{il}, i \in \{1, .., 3\} \\ \mathbf{e}_r &= \mathtt{H}_\infty \mathbf{e}_l \end{aligned}$$

#### 3.1.1 Summary - Affine Calibration

To conclude the discussion we summarise the result and the computational procedure.

**Theorem** *Suppose a fixed stereo rig moves under a general Euclidean action (i.e. both* $\mathtt{R}$ *and* $\mathbf{T}$ *are non-zero, and* $\mathbf{T}$ *not perpendicular to the axis of* $\mathbf{R}$*), and* $\mathtt{T}$ *is the projective transformation which relates the coordinates of corresponding 3D points computed at the original and final positions of the rig:*

$$\mathbf{X}' = \mathtt{T}\mathbf{X},$$

*where* $\mathbf{X}$ *and* $\mathbf{X}'$ *are homogeneous point coordinates computed in the fixed projective frame associated with the rig. Then:*

1. *The real eigenvector of* $\mathtt{T}^{-\top}$ *defines* $\pi_\infty$.
2. *The infinite homography* $\mathtt{H}_\infty$ *between any pairing of cameras is uniquely defined.*

**Computational Algorithm**

1. Compute the fundamental matrix $\mathtt{F}$ for the stereo rig from image point correspondences $\mathbf{x}_l \leftrightarrow \mathbf{x}_r$, i.e. $\mathbf{x}_r^\top\mathtt{F}\mathbf{x}_l = 0$.
2. Compute $\mathbf{X}$, the 3D structure modulo a projectivity, for these correspondences: The 3D coordinates are obtained by backprojecting and intersecting rays through the image points $\mathbf{x}_l, \mathbf{x}_r$ using the projection matrices:

   $$\mathtt{P}_l = [\mathtt{I}|\ \mathbf{0}] \quad \mathtt{P}_r = [\mathtt{M}|\ \mathbf{e}_r]$$

   where $\mathtt{M}$ is obtained from the decomposition $\mathtt{F} = [\mathbf{e}_r]_\times\mathtt{M}$, e.g. $\mathtt{M} = [\mathbf{e}_r]_\times\mathtt{F}$.
3. After the motion, compute the projective coordinates $\mathbf{X}'$ using the same projection matrices $\mathtt{P}_l, \mathtt{P}_r$.
4. Compute the $4 \times 4$ transformation matrix $\mathtt{T}$ where

   $$\mathbf{X}' = \mathtt{T}\mathbf{X},$$
5. Compute $\pi_\infty = (\mathbf{a}, 1)$ from the real eigenvector of $\mathtt{T}^{-\top}$.
6. Compute the infinite homography as

   $$\mathtt{H}_\infty = \mathtt{M} - \mathbf{e}_r\mathbf{a}^\top$$
7. The projective structure $\mathbf{X}$ is transformed to affine by the projective transformation $\mathbf{X}_A = \mathtt{T}_{AP}\mathbf{X}$ where

   $$\mathtt{T}_{AP}^{-1} = \begin{bmatrix} \mathtt{I} & \mathbf{0} \\ -\mathbf{a}^\top & 1 \end{bmatrix}$$

### 3.2 Metric Calibration

In order for the reconstruction to be metric, rather than affine, it is necessary to know the image $\mathtt{K}$ of the dual (i.e. the inverse) of the absolute conic in the left and right cameras. This is equivalent to knowing the internal parameters of the cameras, i.e. the $3 \times 3$ upper triangular camera calibration matrix $\mathtt{C}$, since $\mathtt{K} = \mathtt{C}\mathtt{C}^\top$. In section 2.2 it was shown that $\Omega_\infty$ is determined only up to a one parameter family as the fixed conic of the Euclidean transformation. This result is reflected here in that, in the absence of further constraints, $\mathtt{K}$ is only determined up to a one parameter family.

The infinite homography between two views constrains the camera calibration matrix [9, 11] by:

$$\mathtt{K}_2 = \mathtt{H}_{21\infty}\mathtt{K}_1\mathtt{H}_{21\infty}^{\top} \tag{5}$$

where $\mathbf{v}_2 = \mathtt{H}_{21\infty}\mathbf{v}_1$. This infinite homography can be computed between any pair of views as described in the previous section.

In the case of the left camera before and after the motion:

$$\mathtt{K}_l = \mathtt{H}_{l'l\infty}\mathtt{K}_l\mathtt{H}_{l'l\infty}^{\top} \tag{6}$$

since $\mathtt{K}_l'$ (after the motion) equals $\mathtt{K}_l$ (before the motion). Similarly for the right camera

$$\mathtt{K}_r = \mathtt{H}_{r'r\infty}\mathtt{K}_r\mathtt{H}_{r'r\infty}^{\top} \tag{7}$$

Equations (6) and (7) determine the calibration matrices, $\mathtt{C}_l$ and $\mathtt{C}_r$ respectively up to a one parameter family [11]. The one-parameter family of solutions has the form $\mathtt{K} \rightarrow \mathtt{K} + \mu\mathbf{Ca}(\mathbf{Ca})^{\top}$ where $\mathtt{R}\mathbf{a} = \mathbf{a}$.

There is an additional relation that might be used, namely the infinite homography between the left and right cameras of the stereo pair:

$$\mathtt{K}_r = \mathtt{H}_{rl\infty}\mathtt{K}_l\mathtt{H}_{rl\infty}^{\top} \tag{8}$$

However, the three homographies are related by:

$$\mathtt{H}_{r'r\infty} = \mathtt{H}_{rl\infty}\mathtt{H}_{l'l\infty}\mathtt{H}_{rl\infty}^{-1}$$

(another conjugate relation) and equation (8) provides no additional constraints.

Following [11], the ambiguity can be removed by assuming the skew parameter is zero i.e. that the image axes are perpendicular. This is a very good approximation in practice. This constraint results in a quadratic, $K_{12}K_{33} = K_{23}K_{13}$, giving two solutions for $\mathtt{K}$. Then $\mathtt{K}$ can be decomposed into $\mathtt{C}$ by Choleski factorisation. Often one of the solutions for $\mathtt{K}$ is not positive definite and so can be eliminated.

Once $\mathtt{C}_l$ is known, the structure is upgraded from affine to metric by the transformation: $\mathbf{X}_E = \mathtt{T}_{EA}\mathbf{X}_A$ where

$$\mathtt{T}_{EA}^{-1} = \begin{bmatrix} \mathtt{C}_l & \mathbf{0} \\ \mathbf{0}^{\top} & 1 \end{bmatrix}$$

i.e. the transformation from projective to metric structure is $\mathtt{T}_{EP} = \mathtt{T}_{EA}\mathtt{T}_{AP}$.

# 4 Results

The stereo rig is shown in figure 1. The C-mount lenses have focal lengths of 8.5mm. Images are corrected for radial distortion. All processing: corner detection, robust point correspondence, computation of the fundamental matrix, computation of 3D structure etc, is automatic. A calibration grid is not required at any stage.

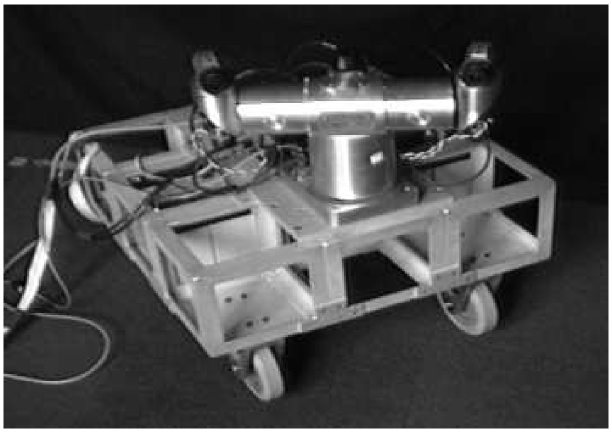

Figure 1: The stereo rig mounted on a mobile vehicle. Two degrees of freedom of the rig are used here: rotation about an elevation axis common to the left and right cameras, and panning of the rig about the central vertical axis.

## 4.1 Implementation

The projective structure in the two frames $\mathbf{X}$ and $\mathbf{X}'$ is computed via the fundamental matrix in a manner similar to that described in [3, 5, 10]. Proper account is taken of mis-matches (outliers) by the use of robust statistics.

In principle $\mathtt{T}$ could be computed directly from the correspondence of five or more points $\mathbf{X}_i \leftrightarrow \mathbf{X}_i'$. In practice, $\mathtt{T}$ is not computed in this fashion because of the absence of a suitable least squares error in projective 3-space. Instead, $\mathtt{T}$ is computed by minimising an image error as follows: the $3 \times 4$ projection matrices which project the structure to the image plane are $\mathtt{P}_{L,R}$ for the original stereo pair, and $\mathtt{P}_{L,R}'$ for the pair after the motion. The projection matrices $\mathtt{P}_{L,R}$ are known for the original stereo pair, and the new projection matrices are related to these by $\mathtt{P}_l' = \mathtt{P}_l\mathtt{T}^{-1}$ $\mathtt{P}_r' = \mathtt{P}_r\mathtt{T}^{-1}$. This enables $\mathtt{T}$ to be computed by a least squares minimisation of the image distances between projected 3D points, i.e. $\mathtt{P}_{L,R}'\mathbf{X}$, and their corresponding image positions in the new images.

## 4.2 Affine Calibration

Figure 2 shows two typical stereo pairs from a sequence used to generate affine structure.

The accuracy of the infinite homography $\mathtt{H}_\infty$ recovered in the affine calibration is demonstrated in Figure 3. Vanishing points in the right image are computed in two ways: first, as the intersection of imaged parallel lines; second, by determining the corresponding vanishing point in the left image (from images of the same

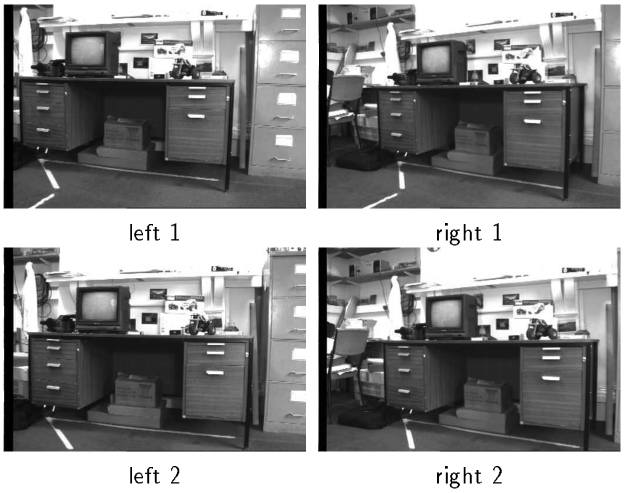


Figure 2: Successive stereo pairs from a sequence. From the upper to the lower the stereo rig moves left by about 20 cm, pans by about $10^o$ and changes elevation by about $10^o$.

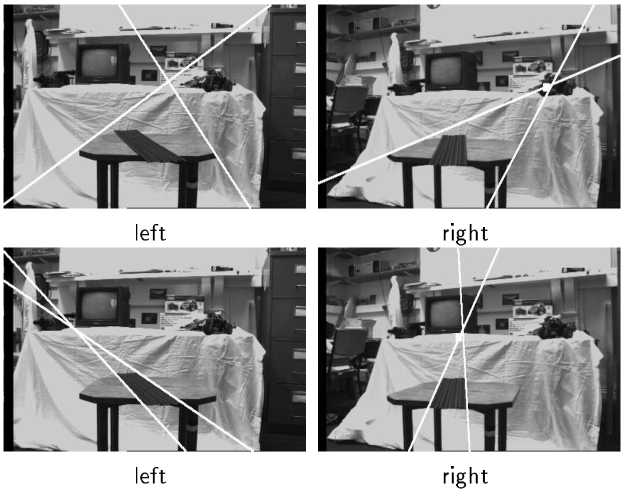


Figure 3: Accuracy of the computed infinite homography. The top stereopair shows (in white) image lines corresponding to the parallel sides of the table, and intersecting at a vanishing point for that direction. For the right images, the square is the result of transferring the vanishing point from the left image using the computed infinite homography. The lower stereopair repeats this computation for a different direction (and hence a different vanishing point) using the parallel sides of a rail track on the table. In both cases the agreement between the transferred and constructed vanishing points is excellent, indicating the accuracy of $\mathtt{H}_\infty$.

parallel lines), and transferring this vanishing point to the right image using the infinite homography. The discrepancy between the vanishing points is a measure of the accuracy of $\mathtt{H}_\infty$. As shown in the figure, there is close agreement between transferred and intersection vanishing points.

A more quantitative assessment of the affine calibration is achieved by first establishing an accurate 3D affine frame using an alternative method (described below), and then examining how close the coordinates of the recovered plane at infinity are to the veridical coordinates $(0,0,0,1)^\top$. The accurate frame is established by computing the infinite homography of corresponding vanishing points in the left and right images. The vanishing points are obtained by intersecting imaged parallel lines from an angled pair of Tsai [17] calibration grids. Table 1 gives the eigenvectors and computed plane at infinity in this frame for various motions of the rig.

We have found that the direct computation of $\mathtt{H}_\infty$ produces superior results to the indirect, but these results are preliminary.

### 4.3 Metric Calibration

Metric calibration is computed using the quadratic constraint for the larger motion of table 1. Figure 4 shows an example of the angle subsequently measured between two reconstructed world lines. The angle is accurately recovered. However, the computation is extremely sensitive to the accuracy of $\mathtt{H}_\infty$. If $\mathtt{H}_\infty$ is inaccurate it is not always possible to obtain a positive definite matrix $\mathtt{K}$. This sensitivity is removed if further motions are made, and $\mathtt{K}$ obtained from the combined constraints of a number of $\mathtt{H}_\infty$s [1].

## 5 Discussion

Moons *et al.* [13] demonstrated a method for attaining affine structure based on a pure translation of a camera (with fixed internal parameters). We have described a more general alternative constraint for a fixed stereo rig — where metric structure is attained by a general motion of the rig.

It may be possible to resolve the two fold ambiguity for each camera matrix by using cheirality constraints [9], though this has not been investigated yet. Of course, the ambiguity can be removed by making additional movements of the fixed rig and obtaining a unique solution using linear methods from the additional $\mathtt{H}_\infty$s generated. Generally, we would aim to update both $\mathtt{H}_\infty$ and camera calibration over a sequence.

### Acknowledgements

The possibility of metric calibration from Euclidean motion was first mentioned by Olivier Faugeras. This

| | | Real pts | | Circular pts | |
|---|---|---|---|---|---|
| Translation 10cm Rotation 5° | | | | | |
| Eigenvalues | Plane at infinity | $\mathbf{E}_3$ | $\mathbf{E}_4$ | $\mathbf{E}_1$ | $\mathbf{E}_2$ |
| 1.06 | 0.05 | 0.94 | 0.69 | 0.17 + i0.44 | 0.16 - i0.44 |
| 0.95 | 0.03 | 0.18 | 0.59 | 0.09 - i0.57 | 0.09 + i0.57 |
| 1 + i0.046 | -0.04 | 0.28 | -0.40 | 0.66 | 0.66 |
| 1 - i0.046 | 0.99 | -0.04 | 0.06 | 0.02 - i0.01 | 0.02 + i0.01 |
| Translation 20cm Rotation 10° | | | | | |
| Eigenvalues | Plane at infinity | $\mathbf{E}_3$ | $\mathbf{E}_4$ | $\mathbf{E}_1$ | $\mathbf{E}_2$ |
| 0.98 | 0.03 | 0.79 | 0.67 | 0.11 - i0.60 | 0.11 + i0.60 |
| 1.04 | -0.04 | -0.58 | -0.66 | -0.03 - i0.44 | -0.03 + i0.44 |
| 1 + i0.11 | -0.08 | 0.17 | -0.31 | 0.65 | 0.65 |
| 1 - i0.11 | 0.99 | -0.03 | 0.05 | 0.05 | 0.05 |

Table 1: Measurement of the coordinates of the plane at infinity, and eigenvectors (corresponding to points on the plane at infinity), in an affine coordinate frame. The veridical plane at infinity has coordinates $(0,0,0,1)^\top$, and veridical eigenvectors (of T) have the form $(X,Y,Z,0)$. There is a rotation about both the pan and elevation axes. The coordinates of the plane at infinity are computed from the real eigenvector of $\mathtt{T}^{-\top}$, the eigenvectors $\mathbf{E}_i$ are of the matrix T. The measured values are in close agreement with veridical, with agreement improving for the larger motion.

Figure 4: Following the metric calibration the computed angle between the desk sides (shown in white) from the 3D reconstruction is $90.7^o$, in very good agreement with the veridical value.

paper was improved by comments from Steve Maybank. Financial support was provided by EU Esprit Project 6448 'VIVA', UK EPSRC Grant GR/H77668, and a Glasstone Research Fellowship.

## References


[1] M. Armstrong, A. Zisserman, and P.A. Beardsley. Euclidean reconstruction from uncalibrated images. *Proc. British Machine Vision Conference*, 1994.

[2] P.A. Beardsley and A. Zisserman. Affine calibration of mobile vehicles. In R. Mohr and W. Chengke, editors, *Europe-China workshop on Geometrical Modelling and Invariants for Computer Vision, Xi'an, China*, 1995.

[3] P.A. Beardsley, A.P. Zisserman, and D.W. Murray. Navigation using affine structure and motion. In *Proc. 3rd European Conference on Computer Vision*, pages 85–96. Springer-Verlag, 1994.

[4] O. Bottema and B. Roth. *Theoretical kinematics.* Dover, New York, 1979.

[5] O.D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In *Proc. 2nd European Conference on Computer Vision*, pages 563–578. Springer-Verlag, 1992.

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

[7] O.D. Faugeras, Q.T. Luong, and S.J. Maybank. Camera self-calibration: theory and experiments. In *Proc. 2nd European Conference on Computer Vision*, pages 321–334. Springer-Verlag, 1992.

[8] R.I. Hartley. Estimation of relative camera positions for uncalibrated cameras. In *Proc. 2nd European Conference on Computer Vision*, pages 579–587. Springer-Verlag, 1992.

[9] R.I. Hartley. Euclidean reconstruction from uncalibrated views. In *Proc. 2nd European-US Workshop on Invariance, Azores*, pages 187–202, 1993.

[10] R.I. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. *Proc. Conference Computer Vision and Pattern Recognition*, 1992.

[11] Q.T. Luong and T. Vieville. Canonic representations for the geometries of multiple projective views. In *Proc. 3rd European Conference on Computer Vision*, pages 589–597. Springer-Verlag, 1994.

[12] S.J. Maybank and O. Faugeras. A theory of self-calibration of a moving camera. *International Journal of Computer Vision*, 8(2):123–151, 1992.

[13] T. Moons, L. van Gool, M. van Diest, and A. Oosterlinck. Affine structure from perspective image pairs under relative translations between object and camera. Technical report KUL/ESAT/M12/9306, Departement Elektrotechniek, Katholieke Universiteit Leuven, 1993.

[14] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling. *Numerical Recipes in C.* Cambridge University Press, 1988.

[15] J.G. Semple and G.T. Kneebone. *Algebraic projective geometry.* Oxford University Press, 1952.

[16] C.E. Springer. *Geometry and analysis of projective spaces.* Freeman, 1964.

[17] R.Y. Tsai. An efficient and accurate camera calibration technique for 3D machine vision. In *Proc. IEEE CVPR*, 1986.