Self-Calibration from Image Triplets

Martin Armstrong[1], Andrew Zisserman[1] and Richard Hartley[2]

[1] Robotics Research Group, Department of Engineering Science, Oxford University, England
[2] The General Electric Corporate Research and Development Laboratory, Schenectady, NY, USA.

Abstract. We describe a method for determining affine and metric calibration of a camera with unchanging internal parameters undergoing planar motion. It is shown that affine calibration is recovered uniquely, and metric calibration up to a two fold ambiguity.
The novel aspects of this work are: first, relating the distinguished objects of 3D Euclidean geometry to fixed entities in the image; second, showing that these fixed entities can be computed uniquely via the trifocal tensor between image triplets; 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

From an image sequence acquired with an uncalibrated camera, structure of 3-space can be recovered up to a projective ambiguity [5, 7]. However, if the camera is constrained to have unchanging internal parameters then this ambiguity can be reduced to metric by calibrating the camera using only image correspondences (no calibration grid). This process is termed "self-calibration" [5, 11]. Previous attempts to make use of the constraint for image *pairs* have generated sets of polynomial equations that are solved by homotopy continuation [8] or iteratively [8, 17] over a sequence. In this paper we demonstrate the advantages of utilizing image *triplets* directly in the case of planar motion, both in the reduced complexity of the equations, and in a practical and robust implementation.

To reduce the ambiguity of reconstruction from projective to affine it is necessary to identify the plane at infinity, π_∞, and to reduce further to a metric ambiguity the absolute conic Ω_∞ on π_∞ must also be identified [4, 13]. Both π_∞ and Ω_∞ are fixed entities under Euclidean motions of 3-space. The key idea in this paper is that these fixed entities can be *accessed* via fixed entities (points, lines, conics) in the image.

To determine the fixed image entities we utilise geometric relations between images that are independent of three dimensional structure. The fundamental geometric relation between two views is the epipolar geometry, represented by the fundamental matrix [3]. This provides a mapping from points in one image to lines in the other, and consequently is not a suitable mapping for determining fixed entities directly. However, between three views the fundamental geometric

relation is the trifocal tensor [6, 14, 15], which provides a mapping of points to points, and lines to lines. It is therefore possible to solve directly for fixed image entities as fixed points and lines under transfer by the trifocal tensor.

In the following we obtain these fixed image entities, and thence the camera calibration, from a triplet of images acquired by a camera with unchanging internal parameters undergoing "planar motion". Planar motion is the typical motion undergone by a vehicle moving on a plane — the camera translates in a plane and rotates about an axis perpendicular to that plane. This extends the work of Moons *et al.* who showed that affine structure can be obtained in the case of purely translational motion [12]. We show that

1. Affine structure is computed uniquely.
2. Metric structure can be computed up to a one parameter family, and this ambiguity resolved using additional constraints.

Section 2, describes the fixed image entities and their relation to π_∞ and Ω_∞, and describes how these are related to affine and metric structure recovery. Section 3 gives an algorithm for computing the image fixed points and lines uniquely using the trifocal tensor. Section 4.1 describes results of an implementation of this algorithm, and section 4.2 results for affine and metric structure recovery based on these fixed points from image triplets. All results are for real image sequences.

Notation We will not distinguish between the Euclidean and similarity cases, both will be loosely referred to as metric. Generally vectors will be denoted by $\mathbf{x}$, matrices as $\mathtt{H}$, and tensors as $\mathtt{T}_i^{jk}$. Image coordinates are lower case 3-vectors, e.g. $\mathbf{x}$, world coordinates are upper case 4-vectors, e.g. $\mathbf{X}$. For homogeneous quantities, $=$ indicates equality up to a non-zero scale factor.

2 Fixed Image Entities for Image Triplets

Planar Motion Any rigid transformation of space may be interpreted as a rotation about a *screw axis* and a simultaneous translation in the direction of the axis [2]. There are two special cases – pure translation and pure rotation. In this paper we consider the latter case. A planar motion of a camera consists of a rotation and a translation perpendicular to the rotation axis. This is equivalent to a pure rotation about a screw axis parallel to the rotation axis, but not in general passing through the camera centre. The plane through the camera centre and perpendicular to the rotation axis is the plane of motion of the camera. We consider sequences of planar motions of a camera, by which we mean a sequence of rotations about parallel but generally distinct rotation axes. The plane of motion is common to all the motions. For visualisation, we assume the plane of motion is horizontal and the rotation axes vertical.

3D fixed entities The plane at infinity and absolute conic are invariant under *all* Euclidean actions. These are the entities that we desire to find in order

to compute respectively affine or metric structure. These entities can not be observed directly, however, so we attempt to find them indirectly. To this end we consider fixed points of a sequence of planar motions. A single planar motion has additional fixed entities, the screw axis (fixed pointwise), and the plane of motion (fixed setwise). In fact, any plane parallel to the plane of motion is fixed. The intersection of this pencil of planes with the plane at infinity is a line (fixed setwise). Although this line is fixed only as a set, its intersection with the absolute conic, Ω_∞, consisting of two points, is fixed pointwise by the motion. These two points are known as the *circular points*, denoted I and J, and lie on every plane parallel to the plane of motion. Knowledge of these two circular points is equivalent to knowing the metric structure in each of these planes ([13]).

The two circular points are fixed for all motions in a sequence of planar motions with common plane of motion. This is not true of the fixed screw axes, since we assume in general that the screw axis is not the same for all motions. However, since the screw axes are parallel, they all intersect at the plane at infinity at a point which we shall denote by V. The points I, J and V and their relation to Ω_∞ is shown in figure 1. They are fixed by all motions in the sequence.

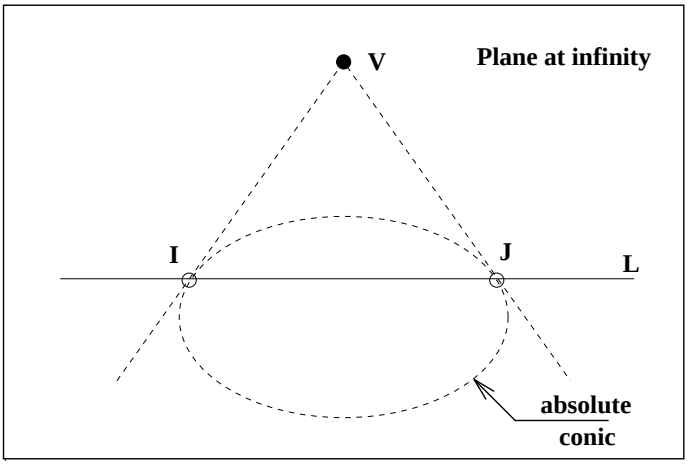

Fig. 1. The fixed entities on π_∞ of a sequence of Euclidean planar motions of 3-space. **V** is the ideal point of the screw axis, and **L** the ideal line of the pencil of planes, orthogonal to the screw axis. **I** and **J** are the circular points for these planes, defined by the intersection of **L** with Ω_∞. **V** and **L** are pole and polar with respect to the absolute conic.

If by some means we are able to find the locations of the points I, J and V in space, then we are able to determine the plane at infinity π_∞ as the unique plane passing through all three of them. This is equivalent to determining the affine structure of space. Although we do not know Ω_∞, and hence can not determine metric structure, we at least know two points on this absolute conic, and hence know the Euclidean geometry in every plane parallel to the plane of motion.

Fixed image entities Our goal is to find the three points I, J and V. Since they are fixed by the sequence of motions, their images will appear at the same location in all images taken by the moving camera (assuming fixed internal calibration). We are led to inquire which points are fixed in all images of a sequence. A fixed point in a pair of images is the image of a point in space that appears at the same location in the two images. It will be seen that apart from the images of I, J and V there are other fixed image entities. We will be led to consider both fixed points and lines.

The locus of all points in space that map to the same point in two images is known as the horopter curve. Generally this is a twisted cubic curve in 3-space passing through the two camera centres [10]. One can find the image of the horopter using the fundamental matrix of the pair of images, since a point on the horopter satisfies the equation $\mathbf{x}^\top \mathtt{F}\mathbf{x} = 0$. Hence, the image of the horopter is a conic defined by the symmetric part of $\mathtt{F}$, namely $\mathtt{F}_s = \mathtt{F} + \mathtt{F}^\top$.

In the case of planar motion, the horopter degenerates to a conic in the plane of motion, plus the screw axis. The conic passes through (and is hence defined by): the two camera centres, the two circular points, and the intersection of the screw axis with the plane of motion. It can be shown that for planar motion $\mathtt{F}_s$ is rank 2 [10], and the conic $\mathbf{x}^\top \mathtt{F}\mathbf{x} = \mathbf{x}^\top \mathtt{F}_s\mathbf{x} = 0$, which is the image of the horopter, degenerates to two lines. These lines are the image of the screw axis and the image of the plane of motion — the horizon line in the image [1]. The epipoles and imaged circular points lie on this horizon line. The apex (the vanishing point of the rotation axis) lies on the imaged screw axis. These points are shown in figure 2a. Although the lines can be computed from $\mathtt{F}_s$, and the imaged circular points and apex lie on these lines, we have not yet explained how to recover these points.

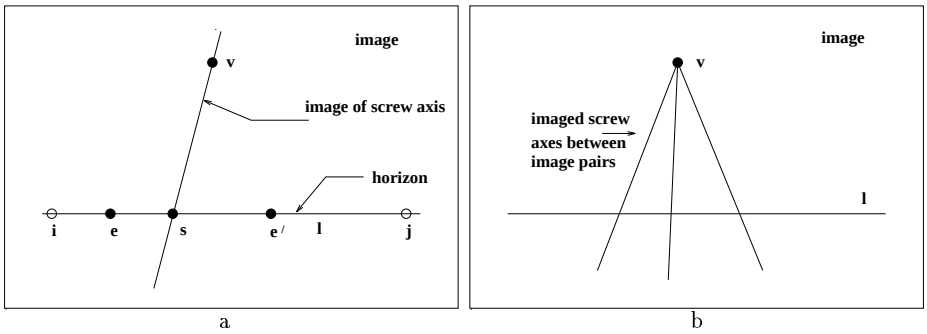

Fig. 2. Fixed image entities for planar motion. (a) For two views the imaged screw axis is a line of fixed points in the image under the motion. The horizon is a fixed line under the motion. (b) The relation between the fixed lines obtained pairwise for three images under planar motion. The image horizon lines for each pair are coincident, and the imaged screw axes for each pair intersect in the apex. All the epipoles lie on the horizon line.

We now consider fixed points in three views connected via planar motions. To do this, we need to consider the intersection of the horopter for cameras 1 and 2 with that for cameras 2 and 3. Each horopter consists of a conic in the plane of motion, plus the vertical axis. The two vertical axes, supposed distinct, meet at infinity at the point V. The two conics meet in 4 points, namely the circular points I and J, the centre of the second camera, plus one further point that is fixed in all three views. The horopter for cameras 1 and 3 will not pass through the second camera centre. Thus we are left with 4 points that are fixed in all three images. These are the circular points I and J, a third point X lying on the plane of motion, and the ideal point V.

Any two fixed points define a fixed line, the line passing through them. Since three of the points, namely the images of points I, J and the third point are collinear, there are just 4 fixed lines. There can be no others, since the intersection of two fixed lines must be a fixed point. We have sketched a geometric proof of the following theorem.

Theorem For three views from a camera with fixed internal parameters undergoing general planar motion, there are four fixed points, three of which are collinear:

1. *The vanishing point of the rotation axes,* $\mathbf{v}$ *(the apex).*
2. *Two complex points, the images of the two circular points* I, J *on the horizon line.*
3. *A third point* $\mathbf{x}$ *on the horizon line and peculiar to the image triplet.*

There are four fixed lines passing through pairs of fixed points.

3D Structure Determination A method for determining affine and metric structure is as follows. One determines the fixed points in the three images using the trifocal tensor as described in the following section. The third real collinear fixed point $\mathbf{x}$ can be distinguished from the complex circular points, the images of I and J. This third point is discarded. The 3-D points I, J and V corresponding to these fixed image points may be reconstructed. These three points define the plane at infinity, and hence affine structure. Planar metric structure is determined by the circular points I and J. Thus, in the absence of other constraints, 3D structure is determined up to a Euclidean transformation in planes parallel to the plane of motion, and up to a one dimensional affine transformation perpendicular to the plane of motion.

Following Luong and Vieville [9] an additional constraint is provided 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 polynomial giving two solutions for the internal calibration matrix, and hence for the recovery of metric structure. Alternatively, an assumption of equal scale factors in the two coordinate axis directions will allow for unique metric reconstruction.

We have now described the structure ambiguity once the fixed image entities are identified. The next section describes a method of identifying the fixed image entities using the trifocal tensor.

3 Fixed image entities via the trifocal tensor

Suppose the 3×4 camera projection matrices for the three views are $\mathtt{P}$, $\mathtt{P}'$ and $\mathtt{P}''$. Let a line in space be mapped to lines $\mathbf{l}$, $\mathbf{l}'$ and $\mathbf{l}''$ in three images. A trilinear relationship exists between the coordinates of the three lines, as follows :

$$l_i = l'_j l''_k \mathtt{T}_i^{jk} \tag{1}$$

where $\mathtt{T}_i^{jk}$ is the trifocal tensor [6]. Here and elsewhere we observe the convention that indices repeated in the contravariant and covariant positions imply summation over the range $(1, \ldots, 3)$ of the index. A similar relationship holds between coordinates of corresponding points in three images.

The trifocal tensor can be computed directly from point and line matches over three views. It can also be directly constructed from the camera projection matrices $\mathtt{P}$, $\mathtt{P}'$ and $\mathtt{P}''$ as follows. Assuming that $\mathtt{P} = [\mathtt{I} \mid \mathbf{0}]$, we have the formula

$$\mathtt{T}_i^{jk} = p'^{ij}_i p''^{k}_4 - p'^{ij}_4 p''^{k}_i \tag{2}$$

where p'^{i}_j and p''^{i}_j are the (ij)-th entry of the respective camera matrices, index i being the contravariant (row) index and j being the covariant (column) index.

Now in order to find fixed lines, we seek solutions $\mathbf{l}$ to the equations (from (1))

$$l_i = l_j l_k \mathtt{T}_i^{jk} \tag{3}$$

In (1) as well as (3) the equality sign represents equality up to a non-zero scale factor. We may remove the unknown scale factor in (3) by taking the cross product of the two sides and equating the result to the zero vector. This results in three simultaneous homogeneous cubic equations for the components of $\mathbf{l}$. In the following we discuss methods for obtaining the solutions to these cubics. First we describe the general case, and then show that this can be transformed to a special case where the solution reduces to a single cubic in one variable. The transformation required is a plane projective transformation of the images. Finally, we arrive at a two step algorithm, tailored to real images, for determining the fixed image points and lines.

3.1 General Planar Motion

We consider three views taken by a camera undergoing planar motion. Without loss of generality, we may assume that the camera is moving in the plane $Y = 0$. The rotation axes are perpendicular to this plane, and meet at the point at infinity $(0, 1, 0, 0)^\top$. We assume that the camera has fixed, but unknown calibration. The origin of coordinates may be chosen at the location of the first camera, which means that the camera has matrix $\mathtt{P} = \mathtt{H}[\mathtt{I} \mid \mathbf{0}]$, for some matrix $\mathtt{H}$. The other two camera differ by a planar motion from this first camera, which means that the three camera have the form

$$\mathtt{P} = \mathtt{H}[\mathtt{I} \mid \mathbf{0}] \qquad \mathtt{P}' = \mathtt{H}[\mathtt{R}' \mid \mathbf{t}'] \qquad \mathtt{P}'' = \mathtt{H}[\mathtt{R}'' \mid \mathbf{t}''] \tag{4}$$

where $\mathtt{R}'$ and $\mathtt{R}''$ are rotations about the Y axis, and $\mathbf{t}'$ and $\mathbf{t}''$ are translations in the plane $Y = 0$.

One may solve for the fixed lines using the trifocal tensor $\mathtt{T}_i^{jk}$. Denoting a fixed line $\mathbf{l}$ for convenience as $\mathbf{l} = (x, y, z)$ instead of (l_1, l_2, l_3), the fixed line equation $l_i = l_j l_k \mathtt{T}_i^{jk}$ may be written as

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \approx \begin{pmatrix} h^{(2)}(x, y, z) \\ h'^{(2)}(x, y, z) \\ h''^{(2)}(x, y, z) \end{pmatrix} \tag{5}$$

where the superscript (2) denotes the degree of the polynomial. Setting the cross-product of the two sides of this equation to zero, one obtains a set of three cubic equations in x, y and z. By the discussion of section 2, there should be four fixed lines as solutions to this set of equations.

The first thing to note, however, is that the three equations derived from (5) are not linearly independent. There are just two linearly independent cubics. Inevitably, for a trifocal tensor computed from real image correspondences, the solutions obtained depend on just which pair of the three equations one chooses. Furthermore, if there is noise present in the image measurements, then the number of solutions to these equations increases. In general, two simultaneous cubics can have up to 9 solutions. What happens is that one obtains a number of different solutions close to the four ideal solutions. Thus, for instance, there are a number of solutions close to the ideal horizon line. Generally speaking, proceeding in this way will lead into a mire of unpleasant numerical computation.

3.2 Normalized Planar Motion

One can simplify the problem by applying a projective transformation to each image before attempting to find the fixed lines. The transformation to be applied will be the same for each of the images, and hence will map the fixed lines to fixed lines of the transformed images. The transformation that we apply will have the effect of mapping the apex point $\mathbf{v}$ to the point at infinity $(0, 1, 0)^\top$ in the direction of the y-axis. In addition, it will map the horizon line to the x-axis, which has coordinates $(0, 1, 0)^\top$. The transformed images will correspond to camera matrices

$$\tilde{\mathtt{P}} = \mathtt{GH}[\mathtt{I} \mid \mathbf{0}] \qquad \tilde{\mathtt{P}}' = \mathtt{GH}[\mathtt{R}' \mid \mathbf{t}'] \qquad \tilde{\mathtt{P}}'' = \mathtt{GH}[\mathtt{R}'' \mid \mathbf{t}'']$$

where $\mathtt{G}$ represents the applied image transformation. We considering now the first camera matrix $\tilde{\mathtt{P}}$. This matrix maps $(0, 1, 0, 0)^\top$, the vanishing point of the Y axis, to the apex $(0, 1, 0)^\top$ in the image. Furthermore the plane $Y = 0$ with coordinates $(0, 1, 0, 0)$ is mapped to the horizon line $(0, 1, 0)^\top$ as required. This constrains the camera matrix $\tilde{\mathtt{P}} = [\mathtt{GH} \mid \mathbf{0}]$ to be of the form

$$\tilde{\mathtt{P}} = [\mathtt{GH} \mid \mathbf{0}] = \begin{bmatrix} \times & 0 & \times & 0 \\ 0 & \times & 0 & 0 \\ \times & 0 & \times & 0 \end{bmatrix} \tag{6}$$

where 0 represents a zero entry and $\times$ represents a non-zero entry.

Consider now the other camera matrices $\tilde{\mathtt{P}}'$ and $\tilde{\mathtt{P}}''$. Since $\mathtt{R}'$ and $\mathtt{R}''$ are rotations about the Y axis, and $\mathbf{t}'$ and $\mathbf{t}''$ are translations in the plane $Y = 0$, both $[\mathtt{R}' \mid \mathbf{t}']$ and $[\mathtt{R}'' \mid \mathbf{t}'']$ are of the form [16]

$$\begin{bmatrix} \times & 0 & \times & \times \\ 0 & \times & 0 & 0 \\ \times & 0 & \times & \times \end{bmatrix} \tag{7}$$

Premultiplying by $\mathtt{GH}$, we find that both $\tilde{\mathtt{P}}'$ and $\tilde{\mathtt{P}}''$ are of the same form (7). This particularly simple form of the camera matrices allows us to find a simple form for the trifocal tensor as well. In order to apply formula (2), we require matrix $\mathtt{P}$ to be of the form $\mathtt{P} = [\mathtt{I} \mid \mathbf{0}]$. This can be achieved by right multiplication of all the camera matrices by the 3D transformation matrix $\begin{bmatrix} (\mathtt{GH})^{-1} & 0 \\ 0 & 1 \end{bmatrix}$. It may be observed that this multiplication does not change the format (7) of the matrices $\tilde{\mathtt{P}}'$ and $\tilde{\mathtt{P}}''$. Now, for $i = 1$ or 3, we see that $\tilde{\mathbf{p}}'_i$ and $\tilde{\mathbf{p}}''_i$ are of the form $(\times, 0, \times)^\top$, whereas for $i = 2$, they are of the form $(0, \times, 0)^\top$. Further, $\tilde{\mathbf{p}}'_4$ is of the form $(\times, 0, \times)^\top$. One easily computes the following form for $\tilde{\mathtt{T}}_i$.

$$\tilde{\mathtt{T}}_i = \begin{bmatrix} \times & 0 & \times \\ 0 & 0 & 0 \\ \times & 0 & \times \end{bmatrix} \quad \text{for } i = 1, 3 \qquad \tilde{\mathtt{T}}_2 = \begin{bmatrix} 0 & \times & 0 \\ \times & 0 & \times \\ 0 & \times & 0 \end{bmatrix} \tag{8}$$

Using this special form of the trifocal tensor, we see that (3) may be written as

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} a_1x^2 + b_1xz + c_1z^2 \\ d_2xy + e_2yz \\ a_3x^2 + b_3xz + c_3z^2 \end{pmatrix} \tag{9}$$

where $\mathbf{l} = (x, y, z)^\top$ represents a fixed line. This set of equations has eight parameters $\{a_1 \ldots c_3\}$. The fixed lines may be found by solving this system of equations. One fixed point in the three views is the apex, $\mathbf{v} = (0, 1, 0)^\top$. Let us consider only lines passing through the apex fixed in all three views. Such a line has coordinates $(x, 0, z)$. Thus, we may assume that $y = 0$. The equations (9) now reduce to the form

$$\begin{pmatrix} x \\ z \end{pmatrix} \approx \begin{pmatrix} a_1x^2 + b_1xz + c_1z^2 \\ a_3x^2 + b_3xz + c_3z^2 \end{pmatrix} \tag{10}$$

Cross-multiplying reduces this to a single equation

$$z(a_1x^2 + b_1xz + c_1z^2) = x(a_3x^2 + b_3xz + c_3z^2) \tag{11}$$

This is a homogeneous cubic, and may be easily solved for the ratio $x : z$. The solutions to this cubic are the three lines passing through the apex joining it to three points lying on the horizon line. These three points are the images of the two circular points, and the third fixed point. The third fixed point may be distinguished by the fact that it is a real solution, whereas the two circular

points are a pair of complex conjugate solutions. The third fixed point is of no special interest, and is discarded.

This analysis is an example of a generally useful technique of applying geometric transformations to simplify algebraic computation.

3.3 Algorithm Outline

We now put the parts of the algorithm together. The following algorithm determines the fixed lines in three views, and hence the apex and circular points on the horizon line. The first four steps reduce to the case of normalized planar motion. The fixed points and lines are then computed in steps 5 to 7, and the last step relates the fixed points back to the fixed points in the original images.

1. Compute the fundamental matrix for all pairs of image, and obtain the epipoles.
2. Find the orthogonal regression line fit to the epipoles. This is the horizon line $\mathbf{l}$.
3. Decompose the symmetric part of the $\mathtt{F}$'s into two lines, this generates the image of the screw axis for each pair. Find the intersection of the imaged screw axes, or in the presence of noise, the point with minimum squared distance to all the imaged screw axes. This determines the apex $\mathbf{v}$.
4. Find a projective transformation $\mathtt{G}$ taking the horizon line to the line $(0,1,0)$ and the apex to the point $(0,1,0)$. Apply this projective transform to all images.
5. For three views, compute the trifocal tensor from point and line matches, enforcing the constraint that it be of the form described in (8).
6. Compute the cubic polynomial defined in (9) and (11), and solve for the ratio $x:z$. There will be two imaginary and one real solution. Discard the real solution. The imaginary solutions will be lines with coordinates $(1,0,z)$ and $(1,0,\bar{z})$ passing through the apex and the two circular points on the horizon line.
7. Compute the intersection of the horizon line $(0,1,0)$ and the line $(1,0,z)$. This is the point $(-z,0,1)$. Do the same for the other solution $(1,0,\bar{z})$.
8. Apply the inverse transform $\mathtt{G}^{-1}$ to the two circular points to find the image of the two circular points in the original images.

4 Results

Numerical results are improved significantly by enforcing that both $\mathtt{F}$ and $\mathtt{F}_s$ are rank 2 during the minimization to compute the fundamental matrix. Implementation details of the algorithms are given in [18].

4.1 Fixed image points and lines

In this section we describe the results of obtaining the fixed points/lines over image triplets. These points are used for affine and metric calibration, which is described in section 4.2.

The image sequences used are shown in figure 3 (sequence I) and figure 4 (sequence II). The first sequence is acquired by a camera mounted on an Adept robot, the second by a different camera mounted on an AGV. The latter sequence has considerably more camera shake, and consequently is not perfect planar motion.

Figure 5 shows the two view fixed lines obtained from the seven sequential image pairs from sequence I. The trifocal tensor is computed for the six sequential image triplets, and the circular points computed from the fixed points of the tensor. The results are given in table 1a. The circular points are certainly stable, but it is difficult to quantify their accuracy directly because they are complex. In the next section the circular points are used to upgrade projective structure to affine and metric. The accuracy of the circular points is hence measured indirectly by the accuracy of the recovered structure. For comparison, the estimated circular points, based on approximate internal parameters, are $(400 \pm 1100i, -255, 1)^\top$.

The camera undergoes a smaller rotation in sequence II and the images are noisier due to camera shake. Superior results are obtained by using fundamental matrices from image pairs which are separated by 2 time steps (i.e. pairs {1,4},{1,5},{2,4},...), rather than sequential image pairs. Figure 6 shows these results. The tensor is calculated using image triplets separated by one time step (i.e. triplets {1,3,5},{2,4,6},...). The circular points computed are shown in table 1b. The estimated circular points, based on approximate internal parameters, are $(257 \pm 800i, 196 \pm 45i, 1)^\top$.

4.2 Structure Recovery

Section 4.1 obtained the 3 fixed points of image triplets using the algorithm of section 3.3. These points define the position of the plane at infinity π_∞, which allows affine structure to be recovered. In this section we describe the results of an implementation of affine and metric structure recovery, and assess the accuracy by comparing with ground truth.

Affine Structure Using the image sequence in figure 3, and the circular points listed in table 1a, affine structure is recovered. We can quantify the accuracy of

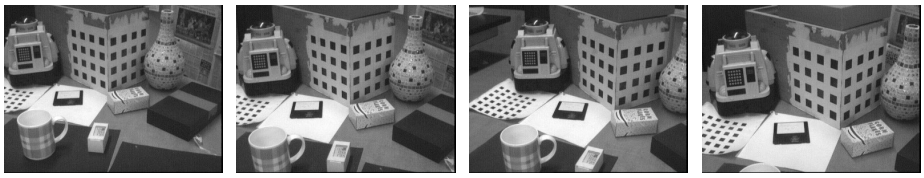

Fig. 3. Image sequence I: four images from an eight image sequence acquired by a camera mounted on an Adept robot arm. Planar motion with the rotation axis at approximately 25^o to the image y axis and perpendicular to the image x axis. 149 corners are automatically matched and tracked through the 8 images. The Tsai grids are used only to provide ground truth, not to calibrate.

Fig. 4. Image sequence II: four images from a nine image sequence. Planar motion with the rotation axis approximately aligned with the image y axis. 75 corners are automatically matched across the 9 images. The sequence was acquired by a camera mounted on an AGV.

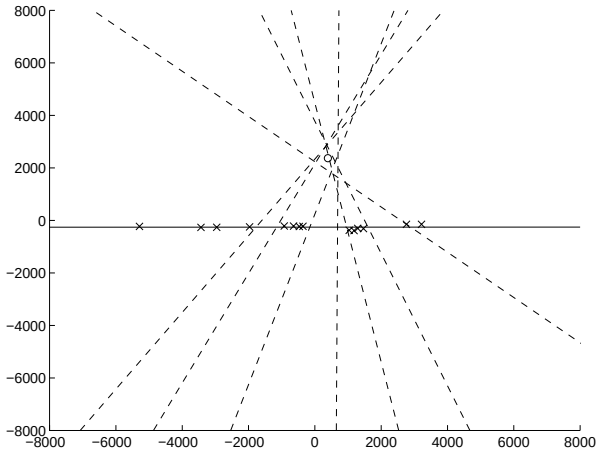

Fig. 5. The fixed points and lines obtained from all 7 possible sequential image pairs from sequence I (figure 3), with axes in pixels. $\times$ are the epipoles, dashed and solid lines are the screw axes and the horizon respectively, $\circ$ is the apex at (394,2370). The horizon line is at $y = -255$.

the affine structure by comparing the values of affine invariants measured on the recovered structure to their veridical values. The affine invariant used is the ratio of line segment lengths on parallel lines. The lines in the scene are defined by the corners on the partially obscured calibration grid shown in figure 3. The veridical value of these ratios is 1.0, and the results obtained are shown in table 2. Clearly, the projective skewing has been largely removed.

Metric Structure Metric structure in planes parallel to the motion plane (the ground plane here) is recovered for the sequence in figure 3. The accuracy of the metric structure is measured by computing an angle in the recovered structure with a known veridical value. We compute two angles for each image triplet. First is the angle between the planes of the calibration grid. We fit planes to 23 and 18 points on the left and right faces respectively and compare the interplane angle to a veridical value of 90^o. Second is the angle between the three computed camera centres for each image triplet which is known from the robot motion. The

Image Triplet	Circular points
123	$(415 \pm 1217i, -255, 1)^\top$
234	$(393 \pm 1229i, -255, 1)^\top$
345	$(407 \pm 1225i, -255, 1)^\top$
456	$(445 \pm 1222i, -255, 1)^\top$
567	$(372 \pm 1274i, -255, 1)^\top$
678	$(371 \pm 1232i, -255, 1)^\top$

a

Image Triplet	Circular points
135	$(301 \pm 698i, 199 \pm 39i, 1)^\top$
246	$(244 \pm 866i, 196 \pm 48i, 1)^\top$
357	$(265 \pm 714i, 197 \pm 40i, 1)^\top$
468	$(583 \pm 747i, 214 \pm 42, 1)^\top$
579	$(302 \pm 691i, 199 \pm 39, 1)^\top$

b

Table 1. The circular points obtained (complex conjugates) for (a) sequence I (figure 3) and (b) sequence II (figure 4). Note, the stability of the points estimated from different triplets.

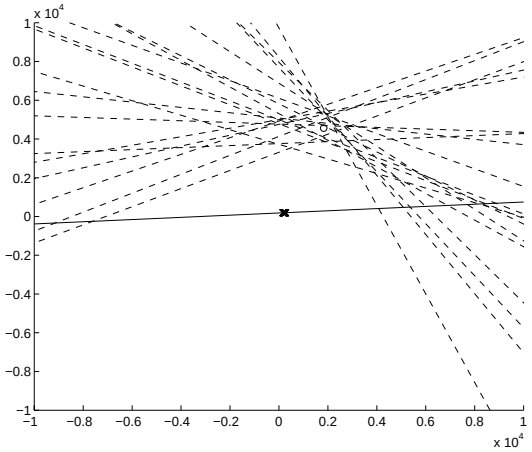

Fig. 6. The fixed points and lines sequence II, figure 4, with axes in pixels. $\times$ are the epipoles, dashed and solid lines are the screw axes and the horizon line respectively, $\circ$ is the image apex at (1430,4675). The horizon line passes through (0,182).

camera centres are computed from the camera projection matrices.

Table 2 shows the computed angles for the 6 image triplets from sequence I, while figure 7 shows a plan view of the recovered metric structure from the first image triplet.

5 Conclusions and extensions

We have demonstrated the geometric importance of fixed points and lines in an image sequence as calibration tools. These fixed entities have been measured, and used to recover affine and partial metric structure from image sequences of real scenes. There are a number of outstanding questions, both numerical and theoretical:

1. We have demonstrated that estimates of the plane at infinity and camera internal parameters can be computed from image triplets. It now remains to derive the variance of these quantities. Then a recursive estimator can

	Affine Invariants				Metric Invariants		
				standard	Plane angle	Motion Angle	
Image Triplet	max	min	average	deviation	$(90^\circ \pm 1^\circ)$	Actual	Computed
123	1.156	0.938	1.034	0.070	86.6	63.8	52.4
234	1.113	0.896	0.994	0.072	90.7	139.1	137.8
345	1.080	0.872	0.980	0.070	92.3	75.0	81.4
456	1.112	0.948	1.015	0.049	85.7	101.7	92.3
567	1.072	0.938	1.010	0.040	89.1	33.0	28.5
678	1.100	0.976	1.022	0.037	88.7	76.1	66.0

Table 2. *Affine Invariants* The ratio of lengths of parallel lines measured on the recovered affine structure of the calibration grid. The veridical value is unity. *Metric Invariants* Angles measured in the ground plane. The interplane angle for the calibration grid, and the angle between the computed camera centres.

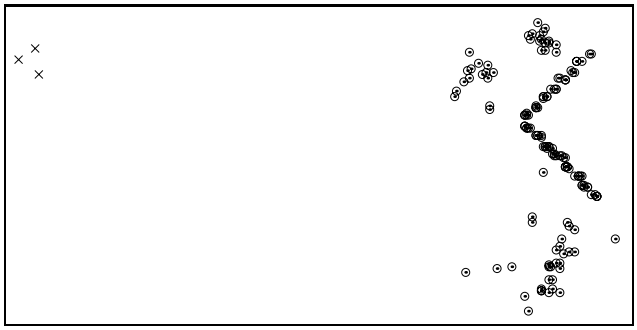

Fig. 7. Plan view of the structure recovered from the first image triplet of sequence I. The first three camera centres are marked with $\times$. The calibration grid and other objects are clearly shown.

be built, such as an Extended Kalman Filter, which updates the plane at infinity and camera calibration throughout an image sequence.

2. The image of the absolute conic is a fixed entity over all images with unchanging internal parameters. The study of fixed image entities opens up the possibility of solving for this directly as the fixed conic of a sequence.

Acknowledgements

Financial support for this work was provided by EU ACTS Project VANGUARD, the EPSRC and GE Research and Development.

References

1. Beardsley, P. and Zisserman, A. Affine calibration of mobile vehicles. In Mohr, R. and Chengke, W., editors, *Europe-China workshop on Geometrical Modelling and Invariants for Computer Vision*. Xi'an, China, 1995.
2. Bottema, O. and Roth, B. *Theoretical Kinematics*. Dover, New York, 1979.

3. Faugeras, O. What can be seen in three dimensions with an uncalibrated stereo rig? In *Proc. ECCV*, LNCS 588, pages 563–578. Springer-Verlag, 1992.
4. Faugeras, O. Stratification of three-dimensional vision: projective, affine, and metric representation. *J. Opt. Soc. Am.*, A12:465–484, 1995.
5. Faugeras, O., Luong, Q., and Maybank, S. Camera self-calibration: theory and experiments. In *Proc. ECCV*, LNCS 588, pages 321–334. Springer-Verlag, 1992.
6. Hartley, R. A linear method for reconstruction from lines and points. In *Proc. ICCV*, pages 882–887, 1995.
7. Hartley, R., Gupta, R., and Chang, T. Stereo from uncalibrated cameras. In *Proc. CVPR*, 1992.
8. Luong, Q. *Matrice Fondamentale et Autocalibration en Vision par Ordinateur.* PhD thesis, Université de Paris-Sud, France, 1992.
9. Luong, Q. and Vieville, T. Canonic representations for the geometries of multiple projective views. In *Proc. ECCV*, pages 589–597. Springer-Verlag, 1994.
10. Maybank, S. *Theory of reconstruction from image motion.* Springer-Verlag, Berlin, 1993.
11. Maybank, S. and Faugeras, O. A theory of self-calibration of a moving camera. *International Journal of Computer Vision*, 8:123–151, 1992.
12. Moons, T., Van Gool, L., Van Diest, M., and Pauwels, E. Affine reconstruction from perspective image pairs. In *Applications of Invariance in Computer Vision*, LNCS 825. Springer-Verlag, 1994.
13. Semple, J. and Kneebone, G. *Algebraic Projective Geometry.* Oxford University Press, 1979.
14. Shashua, A. Trilinearity in visual recognition by alignment. In *Proc. ECCV*, 1994.
15. Spetsakis, M.E. and Aloimonos, J. Structure from motion using line correspondences. *International Journal of Computer Vision*, pages 171–183, 1990.
16. Wiles, C and Brady, J.M. Ground plane motion camera models. In *Proc. ECCV*, 1996.
17. Zeller, C. and Faugeras, O. Camera self-calibration from video sequences: the kruppa equations revisited. Technical report, INRIA, 1995.
18. Zisserman, A., Armstrong, M., and Hartley, R. Self-calibration from image triplets. Technical report, Dept. of Engineering Science, Oxford University, 1996.

This article was processed using the LaTeX macro package with ECCV'96 style