Analysis and Computation of an Affine Trifocal Tensor

P. R. S. Mendonça and R. Cipolla
Department of Engineering
University of Cambridge
Cambridge, UK, CB2 1PZ
[prdsm2|cipolla]@eng.cam.ac.uk

Abstract

This paper investigates the trifocal tensor for an affine trinocular rig and defines an *affine trifocal tensor*. The question of the degrees of freedom of the tensor entries will be addressed, and a novel algorithm to compute the tensor by a linear technique is presented. It will be shown that 4 point or 8 line correspondences along the three images are enough for a reliable computation of the trifocal tensor in the affine case (uncalibrated, weak perspective camera model), in contrast to the projective case, where at least 7 point or 13 line correspondences are needed for a linear computation (and usually with poor results). An analysis of the error in the approximation of a generic trifocal tensor by the affine trifocal tensor is carried out and preliminary experiments with synthetic and real data show the reliability and robustness of the approximation under a wide range of conditions.

1 Introduction

The trifocal tensor is an important tool for several tasks in Computer Vision, like reconstruction [1, 14], self-calibration [6] and motion segmentation [17], and its computation is still a research topic [5, 18]. The linear algorithms [10, 9] have simple conception and execution, but they do not take the appropriate constraints into account. This results in a poor behaviour when the number of point or line correspondences available is small, as demonstrated in [5]. On the other hand, the nonlinear methods are more sensitive to noise and do not make use of all data available.

The concepts of affine cameras (uncalibrated cameras under weak perspective) and affine fundamental matrices are well established [11, 16], and their importance is clear. In situations where the depth variation ΔZ of the scene is small compared with the average distance Z the scene features to the camera centre, e. g. $\Delta Z/Z < 0.1$, the approximation has many advantages, like robustness. Also less points are needed for the computation of both the affine camera and the affine fundamental matrix. For a generic fundamental matrix $\mathbf{F}$, the constraint that $\det(\mathbf{F}) = 0$ must be ensured somehow, while in the affine case this constraint is automatically satisfied. As will be shown, all these features are automatically valid for the here introduced *affine trifocal tensor*, that extends the results presented in [17] about affine trilinear constraints. The problems of affine reconstruction and self-calibration for affine cameras in multiple views, topics related to the contents of

this paper, are addressed in [13, 15, 14]. In [13], Euclidean motion/structure recovery is done using weak perspective cameras, while in [15] the cameras are assumed to have the same intrinsic parameters, except for a scale factor, what is consistent with the context of self-calibration. In [14] only line correspondences are used, since the method presented is based in one-dimensional cameras.

A review of the trifocal tensor is presented in the next section, where tensorial notation is used. A good introduction to this subject can be found in [3]. In Section 3 the affine fundamental matrix [16] will be revisited, leading to a natural development of the affine trifocal tensor. We verify that the 16 entries of this tensor must satisfy only 4 constraints, in contrast with the general case where the 27 entries of the tensor must satisfy 9 constraints. A linear algorithm for its computation, adapted from [10], is developed in Section 4, and experiments in the transfer of points and lines are presented in Section 5.

2 Review of the Trifocal Tensor

The material presented in this section is based on [9, 10]. No proofs of the results will be presented, as they can be found in these references.

Let $\mathbf{M} = [\mathbb{I}|\mathbf{0}]$, $\mathbf{M}' = [a^i_j]$ and $\mathbf{M}'' = [b^i_j]$ be three camera matrices, and let $\boldsymbol{\lambda}$, $\boldsymbol{\lambda}'$ and $\boldsymbol{\lambda}''$be three homogeneous lines associated with the images of a line in 3D space, taken from the camera with corresponding superscript. The constraint that the planes defined by such lines and their respective camera centres must intersect in the same line in space is stated in the relation

$$\lambda_i \approx \lambda'_j \lambda''_k T_i^{jk}, \tag{1}$$

where $\approx$ means equality up to a scale factor and where

$$T_i^{jk} = a_i^j b_4^k - a_4^j b_i^k \tag{2}$$

is defined as the trifocal tensor.

Let $\mathbf{u}$, $\mathbf{u}'$ and $\mathbf{u}''$ be the images of a point in 3D space, taken from the cameras with corresponding superscripts. The points are related through the trifocal tensor by the expression

$$u''^l \approx u^k (u'^i T_k^{jl} - u'^j T_k^{il}). \tag{3}$$

(1) and (3) execute a transfer of lines and points through the three images, i. e., given two images of a line or a point, a third image can be determined (see fig. 1).

2.1 Degrees of Freedom in the Entries of the Trifocal Tensor

A triplet of cameras has 33 degrees of freedom (dof) [4] (11 dof for each one). Since the trifocal tensor is able to determine 3D structure up to a 3D homography [9] (15 dof), the correspondent trifocal tensor must have $18 = 33 - 15$ degrees of freedom, so there are 9 constraints over its 27 entries. This is an important issue in its computation, since a method that does not take these constraints into account may produce a "non-valid" tensor, i. e. a $3 \times 3 \times 3$ tensor that does not hold the properties of the trifocal tensor. In the affine case, the triplet of cameras has only 24 dof, and the 3D affine transformation has 12. Thus the affine trifocal tensor holds only 12 dof. As will be shown in the next section, the affine trifocal tensor has 16 non-zero entries, and thus there are only 3 constraints to be imposed, since the overall scale is not important.

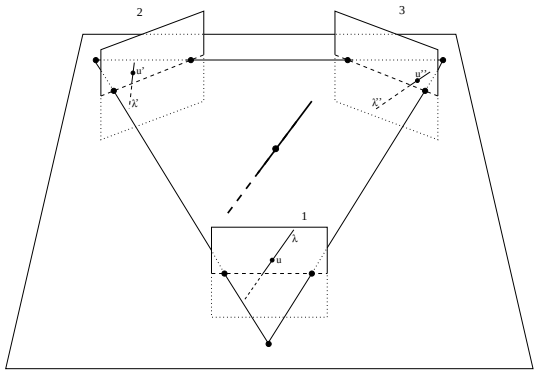

Figure 1: Images of a point and a line in a trinocular stereo rig.

3 The Affine Trifocal Tensor

The key idea in the derivation of the affine trifocal tensor is to transform, by the application of an appropriated 3D homography, a generic triplet of affine cameras to a new triplet that can be used in the computation of the tensor as defined in (2). This method can also be used in the derivation of the affine fundamental matrix, yielding, as will be shown, the same result found in [16]. This procedure also implies that the affine trifocal tensor will naturally inherits the stability and reliability of the affine camera.

3.1 The Affine Fundamental Matrix Revisited

Consider the camera $\tilde{\mathbf{P}}_2$ below, where the motivation for the choice of indexes will soon be clear:

$$\tilde{\mathbf{P}}_2 = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ 0 & 0 & a_{33} & 0 \end{bmatrix}. \tag{4}$$

It is shown in [7] that the fundamental matrix $\mathbf{F}_{12}$ related to the pair of cameras $\tilde{\mathbf{P}}_1 = [\mathbb{I}|0]$ and $\tilde{\mathbf{P}}_2 = [\mathbf{P}|\mathbf{p}]$ is given by $\tilde{\mathbf{F}}_{12} = [\mathbf{p}]_\times[\mathbf{P}]$, where $[\mathbf{p}]_\times$ is the skew-symmetric matrix such that $[\mathbf{p}]_\times \mathbf{a} = \mathbf{p} \times \mathbf{a}$ for all $\mathbf{a}$. The matrix $\tilde{\mathbf{P}}_1$ is not an affine matrix, so it is not possible to use it directly with in the computation of an affine fundamental matrix. Nevertheless, the pair of affine cameras $\tilde{\mathbf{P}}'_1$ and $\tilde{\mathbf{P}}'_2$ given by

$$\tilde{\mathbf{P}}'_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ and } \tilde{\mathbf{P}}'_2 = \begin{bmatrix} a_{11} & a_{12} & a_{14} & a_{13} \\ a_{21} & a_{22} & a_{24} & a_{23} \\ 0 & 0 & 0 & a_{33} \end{bmatrix} \tag{5}$$

are transformed to $\tilde{\mathbf{P}}_1$ and $\tilde{\mathbf{P}}_2$ by the 3D homography $\mathbf{H}$, i. e., $\tilde{\mathbf{P}}_1 = \tilde{\mathbf{P}}'_1\mathbf{H}$ and $\tilde{\mathbf{P}}_2 = \tilde{\mathbf{P}}'_2\mathbf{H}$, where

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \tag{6}$$

Since the fundamental matrix is invariant to a 3D homography applied to each camera [7], the fundamental matrix $\mathbf{F}$ for both pairs is the same, and thus,

$$\mathbf{F} = \begin{bmatrix} 0 & 0 & -a_{24}a_{33} \\ 0 & 0 & -a_{14}a_{33} \\ a_{24}a_{11} + a_{14}a_{21} & a_{24}a_{12} + a_{14}a_{22} & a_{24}a_{13} + a_{14}a_{23} \end{bmatrix}. \tag{7}$$

The affine fundamental matrix reconstructs the two cameras (16 dof) up to a 3D affine homography (12 dof), and thus it must have 4 dof, or 5 dof minus a scale factor, exactly how it is pointed in (7). So the affine fundamental matrix is, by definition, minimally parametrised, and four correspondences are enough for its computation. It is worth noting that $\det(\mathbf{F}) = 0$; this condition must be imposed *a posteriori* in many algorithms to compute a general fundamental matrix.

3.2 Derivation of the Affine Trifocal Tensor

The method presented in the previous section can also be extended to the computation of the trifocal tensor. Beginning with the affine cameras $\tilde{\mathbf{P}}_1$, $\tilde{\mathbf{P}}_2$ and $\tilde{\mathbf{P}}_3$, where

$$\tilde{\mathbf{P}}_3 = \begin{bmatrix} b_{11} & b_{12} & b_{14} & b_{13} \\ b_{21} & b_{22} & b_{24} & b_{23} \\ 0 & 0 & 0 & b_{33} \end{bmatrix}, \tag{8}$$

the application of the homography $\mathbf{H}$ to each of the matrices $\tilde{\mathbf{P}}_i$, $i = 1, 2, 3$, results in a new triplet of cameras with the same trifocal tensor T_i^{jk} (represented here by the matrices $\mathbf{T}_i^{\cdot\cdot}$, $i = 1, 2, 3$), which can be calculated by the definition in (2), resulting in

$$\mathbf{T}_1^{\cdot\cdot} = \begin{bmatrix} a_{11}b_{14} - a_{14}b_{11} & a_{11}b_{24} - a_{14}b_{21} & 0 \\ a_{21}b_{14} - a_{24}b_{11} & a_{21}b_{24} - a_{24}b_{21} & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{9}$$

$$\mathbf{T}_2^{\cdot\cdot} = \begin{bmatrix} a_{12}b_{14} - a_{14}b_{12} & a_{12}b_{24} - a_{14}b_{22} & 0 \\ a_{22}b_{14} - a_{24}b_{12} & a_{22}b_{24} - a_{24}b_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{10}$$

$$\mathbf{T}_3^{\cdot\cdot} = \begin{bmatrix} a_{13}b_{14} - a_{14}b_{13} & a_{13}b_{24} - a_{14}b_{23} & -a_{14}b_{33} \\ a_{23}b_{14} - a_{24}b_{13} & a_{23}b_{24} - a_{24}b_{23} & -a_{24}b_{33} \\ a_{33}b_{14} & a_{33}b_{24} & 0 \end{bmatrix}. \tag{11}$$

There are only 16 free parameters in the affine trifocal tensor, and the condition $\det(\mathbf{T}_i^{\cdot\cdot}) = 0$, for $i = 1, 2, 3$, is promptly satisfied for $i = 1, 2$, and since the one-dimensional null spaces of $\mathbf{T}_1^{\cdot\cdot}$ and $\mathbf{T}_2^{\cdot\cdot}$ are the same, the basis of the one-dimensional null spaces of $\mathbf{T}_i^{\cdot\cdot}, i = 1, 2, 3$, must have a common perpendicular, as imposed in [8] for the general trifocal tensor.

4 Linear Computation of the Affine Trifocal Tensor

The main drawback of the linear methods for the computation of the trifocal tensor is the need of a large amount of correct matches for a reliable computation, which is usually

difficult to obtain. The linear computation with a minimum number of points fails because the appropriate constraints are not used, and the extra degrees of freedom reduce the error for the given matches at the expenses of an increase in error for the other points. Since the affine trifocal tensor is much less unconstrained, it is expected that its linear computation give a much more reliable result, what will be confirmed by the experiments in Section 5.

Let $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \lambda_3]^{\mathrm{T}}$ and $\mathbf{u} = [u^1, u^2, u^3]^{\mathrm{T}}$ be a line and a point in the line, respectively, in the homogeneous image space spanned by matrix $\tilde{\mathbf{P}}_1$. The superscripts $'$ and $''$ are appended to features (points or lines) to indicate they are in the homogeneous image space spanned by camera matrices $\tilde{\mathbf{P}}_2$ and $\tilde{\mathbf{P}}_3$, respectively.

The linear method for the computation of the trifocal tensor [9, 10] relies on the equation

$$u^i \lambda'_j \lambda''_k T_i^{jk} = 0, \tag{12}$$

that can be used for correspondent triplets of points or lines indifferently. For a triplet of points $\mathbf{u} \leftrightarrow \mathbf{u}' \leftrightarrow \mathbf{u}''$, the point $\mathbf{u}$ is selected and two different lines passing through each one of the points $\mathbf{u}'$ and $\mathbf{u}''$ are arbitrarily chosen. Each one of the four pairs of lines so generated, say $\boldsymbol{\lambda}'$ and $\boldsymbol{\lambda}''$, yields an independent equation with the form of (12). If a triplet $\boldsymbol{\lambda} \leftrightarrow \boldsymbol{\lambda}' \leftrightarrow \boldsymbol{\lambda}''$ of lines is given, two different points $\mathbf{u}$ on the line $\boldsymbol{\lambda}$ are randomly chosen, resulting in two independent equations. Thus, for a general trifocal tensor, one needs at most *l* line correspondences and *p* point correspondences, where $2l + 4p \geq 26$, to solve for the entries of the tensor. Remember that the overall scale of the tensor is not important. In the affine case, there are only 16 non-zero entries in the tensor, and thus *l* line and *p* point correspondences satisfying $2l + 4p \geq 15$ are enough, in accordance with [17].

In real situations, (12) will not be satisfied exactly, and the trifocal tensor may be estimated by a linear least-squares method. In the affine approximation the solution of the least-squares problem is simplified in comparison to the general case, since it involves the computation of the eigenvectors of a 16×16 matrix, instead of a 27×27 one.

An important issue is the minimum number of points that must be used in the computation of the trifocal tensor *if the constraints are taken into account*. In the projective case it is shown in [18] that although 5 points provide 20 equations and the constrained tensor holds only 18 dof, 6 points are necessary for its computation. This restriction appears in [18] as a characteristic of the algorithm, but the argument below, based on the theory of invariants [19, 11, chapter 1], shows that this result does not depend on the algorithm used.

The trifocal tensor, in the projective case, can be used to reconstruct the scene apart from a 3D homography. Except for particular configurations (e. g. coplanar or collinear points), there are no projective invariants for less then six points in space [12], and thus 5 points do not provide the necessary information about the projective structure of the triplet of cameras. A more remarkable result is that the computation of the affine trifocal tensor cannot be done with less then 4 point correspondences, and thus the linear algorithm, in the affine case, makes the best possible use of the available information. To state that it is enough to observe that there is no affine invariant for less then 4 generic points in space, although 4 points define a tetrahedron whose volume, under the group of affine transformations, is a relative invariant of weight 1. This is a direct consequence of the identity $\mathbf{A}\mathbf{a} \cdot (\mathbf{A}\mathbf{b} \times \mathbf{A}\mathbf{c}) = \det(\mathbf{A})\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$.

5 Preliminary Results

5.1 Experiments with Synthetic Data

To evaluate the validity of the proposed model for the affine tensor, some experiments were done. Firstly, 10 points X_i, $i = 1, 2, \ldots, 10$, were randomly placed with an uniform distribution inside a square box with side 2 and centred at the point (0,0,21). Then, the points were projected through cameras $\tilde{\mathbf{P}}_1$, $\tilde{\mathbf{P}}_2$ and $\tilde{\mathbf{P}}_3$, where the centre C_i and the unitary normal vector to the image plane of camera $\tilde{\mathbf{P}}_i$, $\mathbf{n}_i$, are

$$\begin{array}{rclrcl} C_1 &=& [\ 0.0 \quad 0.0 \quad 0.0\ ]^{\mathrm{T}}, & \mathbf{n}_1 &=& [\ 0.0000 \quad 0.0000 \quad 1.0000\ ]^{\mathrm{T}}; \\ C_2 &=& [\ -1.0 \quad 0.5 \quad 2.0\ ]^{\mathrm{T}}, & \mathbf{n}_2 &=& [\ 0.0393 \quad -0.0174 \quad 0.9991\ ]^{\mathrm{T}}; \\ C_3 &=& [\ 0.5 \quad 1.0 \quad 1.0\ ]^{\mathrm{T}}, & \mathbf{n}_3 &=& [\ -0.0349 \quad -0.0262 \quad 0.9990\ ]^{\mathrm{T}}. \end{array}$$

The centroid of the points X_i, $i = 1, 2, \ldots, 10$ is $X_0 = [\ -0.1238 \quad 0.0257 \quad 21.0217\ ]^{\mathrm{T}}$ and thus $\max_i \|C_i - X_0\| = 21.0471$, for $i = 1$. This is in agreement with the rule of thumb that for a good affine approximation the average depth of the scene must be around 10 times less then the viewing distance. The points X_i and the configuration of the cameras are shown in fig. 2.

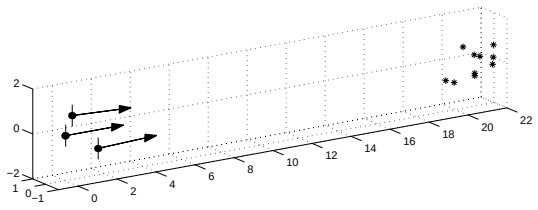

Figure 2: Configuration of the cameras and points. The arrows are the normal vectors to the image plane of each camera, starting from the respective camera centres. The open dot is the origin of the image coordinate system, and the stars are the 3D points.

Fig. 3 shows the result of the transfer of lines and points using the affine trifocal tensor. The tensor was computed with only four point correspondences, and the transfer error is hardly perceived by the eye.

The effects of noise in image data are shown in fig. 4. In this experiment the position of the points on each image was perturbed with zero mean Gaussian noise, with variance in the interval (0,1), and the trifocal tensor was calculated in the projective and affine cases, both with the minimum number of points. It is evident that the linear algorithm has an extremely bad performance in the projective case, but gives a reasonable accurate result if the affine approximation is used.

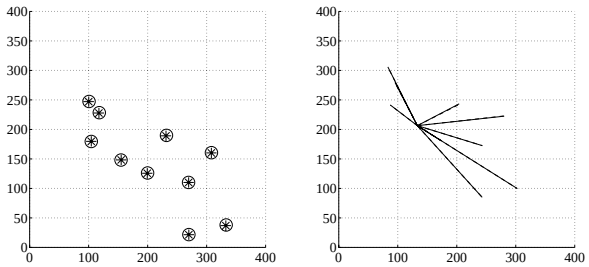

Figure 3: Transfer of points and lines. a) The stars are the correct point images, and the circles are the transfered points. b) The continuous lines are the correct line images, and the dashes show the transfered lines.

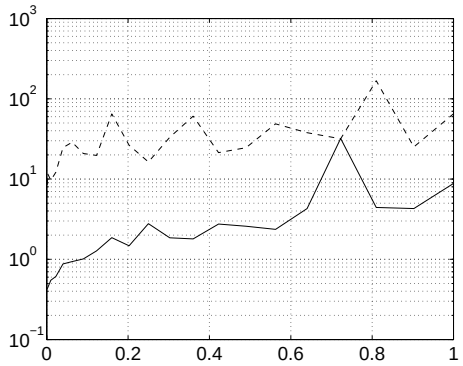

Figure 4: Stability of the linear computation of the trifocal tensor with minimum number of points. The dashed line is the projective case (7 points) and the solid line is the affine case (4 points). The x axis is the noise variance, and the y axis is the mean square error (in pixels) of the transfered points. As pointed in [5], the linear algorithm fails completely for the projective case when the minimum number of points is used, but the affine approximation performs well for realistic amounts of noise.

5.2 Experiments with Real Data

A further experiment, this time with real data, was also done. From three images of a calibration grid, lines were fitted to edges found using Canny's edge detector [2]. Squares where automatically extracted, and four point matches in the three images were selected by hand. With this triplets, the affine trifocal tensor was linearly computed and then used to transfer squares from images 5a and 5b to image 5c. The mean distance from correspondent vertices of transfered squares to extracted squares is 1.2633 pixels, enough to produce a correct matching for all squares in the grid.

a) b) c)

Figure 5: The dots show the points selected for the computation of affine trifocal tensor. The squares in images a) and b) were used to predict the squares in image c).

6 Conclusions

The trifocal tensor plays for trinocular rigs the same role as the fundamental matrix plays for stereo rigs, allowing recovery of structure apart from a 3D homography [9]. It is an effective tool for solving problems in several major topics of Computer Vision [1, 14, 17, 6]. The major contribution of this paper is the introduction of an *affine trifocal tensor*. This entity is connected to the general trifocal tensor in the same way that the affine fundamental matrix [16] and the affine camera [11] are connected to the general fundamental matrix and the projective camera. The affine trifocal tensor has several advantages over the general trifocal tensor: less point and line correspondences are needed for its computation, that can be reliably and robustly done by a linear algorithm, and it naturally satisfies most of the constraints of a generic trifocal tensor. The linear algorithm for the computation of the trifocal tensor [9] was adapted to the affine case, and experiments with synthetic and real data confirmed the usefulness of the ideas here developed.

Acknowledgements

P. R. S. Mendonça would like to acknowledge CAPES for the grant BEX 1165/96-8 that partially funded this research.

References

[1] G. Barrett, E. amd Gheen and P. Payton. Representation of three-dimensional object structure as cross-rations of determinants of stereo image points. In J. L. Mundy, A. Zisserman, and D. Forsyth, editors, *Applications of Invariance in Computer Vision*, volume 825 of *Lecture Notes in Computer Science*, pages 47–66. Springer-Verlag, 1994.

[2] J.F. Canny. A computational approach to edge detection. *IEEE Trans. Pattern Analysis and Machine Intell.*, 8:679–698, 1986.

[3] J. A. Eisele and R. M. Mason. *Applied Tensor and Matrix Analysis*. John Wiley & Sons, 1970.

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

[5] O. Faugeras and T. Papadopoulo. A nonlinear method for estimating the projective geometry of 3 views. In *Proc. 6th Int. Conf. on Computer Vision*, pages 477–484, 1998.

[6] O. Faugeras, L. Quan, and P. Sturm. Self-calibration of a 1d projective camera and its application to the self-calibration of a 2d projective camera. In *Proc. 5th European Conf. on Computer Vision*, 1998.

[7] R. Hartley. Projective reconstruction and invariants from multiple images. *IEEE Trans. Pattern Analysis and Machine Intell.*, 16(10):1036–1041, 1994.

[8] R. Hartley. A linear method for the reconstruction of lines and points. In *Proc. 5th Int. Conf. on Computer Vision*, pages 882–887, 1995.

[9] R. Hartley. Lines and points in three views and the trifocal tensor. *Int. Journal of Computer Vision*, 22(2):125–140, 1997.

[10] R. Hartley. Minimising algebraic error in geometric estimation problem. In *Proc. 6th Int. Conf. on Computer Vision*, pages 469–476, 1998.

[11] J.L. Mundy and A.Zissermann editors. *Geometric Invariance in Computer Vision*. MIT Press, 1992.

[12] L. Quan. Invariants of six points and projective reconstruction from three uncalibrated images. *IEEE Trans. Pattern Analysis and Machine Intell.*, 17(1):34–46, 1995.

[13] L. Quan. Self-calibration of an affine camera from multiple views. *Int. Journal of Computer Vision*, 19(1):93–105, 1996.

[14] L. Quan and T. Kanade. Affine structure from line correspondences with uncalibrated affine cameras. *IEEE Trans. Pattern Analysis and Machine Intell.*, 19(8):834–845, 1997.

[15] L. Quan and Y. Ohta. A new linear method for euclidean motion/structure from three calibrated affine cameras. In *Proc. Conf. Computer Vision and Pattern Recognition*, 1998.

[16] L. S. Shapiro, A. Zisserman, and M. Brady. 3d motion recovery via affine epipolar geometry. *Int. Journal of Computer Vision*, 16:147–182, 1995.

[17] P. Torr. *Motion Segmentation and Outlier Detection*. PhD thesis, Department of Engineering Science, University of Oxford, 1995.

[18] P. H. S. Torr and A. Zisserman. Robust parametrization and computation of the trifocal tensor. *Image and Vision Computing*, 15(8):591–605, 1997.

[19] I. Weiss. Geometric invariants and object recognition. *Int. Journal of Computer Vision*, 10(3):207–231, 1993.