Multilinear Constraints in Two-dimensional Vision and Isogonal Conjugacy

Kalle Åström[*]
Dept of Mathematics, Lund University
Box 118, S-221 00 Lund, Sweden
email: kalle@maths.lth.se

Abstract

Invariance has become increasingly important in computer vision. One type of invariance is viewpoint invariance, i.e. properties of viewed objects that are invariant under camera motion. Another important concept is object invariance, i.e. properties of the image that only depend on motion and not on the viewed objects. The epipolar constraint in three dimensional vision, is an example of the latter. This paper deals with the generalisation of the epipolar constraint to multilinear constraints, for two-dimensional vision. In particular, the calibrated trilinear tensor is studied. It is shown that each such tensor is associated with two different reconstructions. The ambiguity is closely related to a concept from planar geometry, the so called *isogonal conjugacy* of two points with respect to a triangle.

1 Introduction

A central problem in three-dimensional vision is the analysis of object shape and camera motion from images, obtained by projections. The objective is to calculate the shape of the object using the shapes of the images and to calculate the camera matrices, which gives the camera movement. Traditionally this is solved by reconstructing the object and the camera motion at the same time. Recently methods have been developed which use invariance. Two main types of invariance are viewpoint invariance and object invariance. Viewpoint invariant properties are properties which are intrinsic to the viewed object and independent of camera position or motion. An example is the classical cross ratio of four collinear points. Such properties can be used to recognise or reconstruct objects without knowing the relative position of the camera to the object. Another type of invariancy is the object invariant properties, i.e. properties of images which only depend on camera motion and not on the viewed object. An example of such a property is the epipolar constraint. These properties have twofold uses. Firstly, the constraints can be used to find further image correspondences. Secondly, the constraints give information about the camera motion. This type of invariancy has become increasingly popular and is currently under intense research. Generalisations of the multilinear constraints can be found in [4]. Further generalisations to the infinitesimal case can be found in [7]. In this paper the calibrated trilinear constraint in two-dimensional vision, is studied.

2 Camera Geometry

Consider a two dimensional world consisting of a number of object points $(U^1, \dots, U^n)$. Each point U is represented in homogeneous coordinates so that $U = \begin{pmatrix} U_1 & U_2 & 1 \end{pmatrix}^T$, where U_1 and U_2 are the ordinary x and y coordinates of that point in the plane. It is useful to think of these homogeneous vectors as being defined only up to scale. Two representations of the same points are thus $U = \begin{pmatrix} U_1 & U_2 & 1 \end{pmatrix}^T \sim \begin{pmatrix} 2U_1 & 2U_2 & 2 \end{pmatrix}^T$, where we have used $\sim$ to denote equality up to scale. The image u of a point U with projection matrix P is defined as

$$\lambda u = PU \quad , \tag{1}$$

---

[*] This work has been supported by the ESPRIT project BRA EP 6448, VIVA, and the Swedish Research Council for Engineering Sciences (TFR), project 95-64-222.

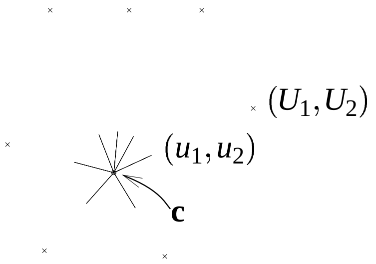

Figure 1: The image u is the direction from camera position c to the object position with homogeneous coordinates U. Think of the image of sevaral points as a set of directions.

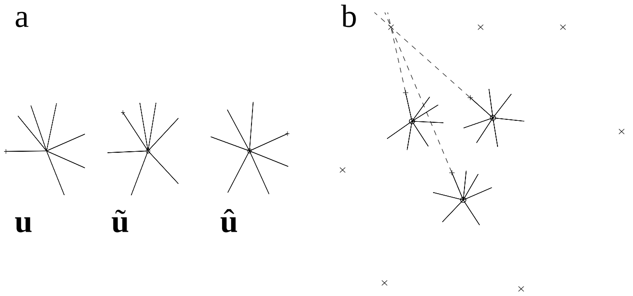

Figure 2: The structure and motion problem for two-dimensional calibrated vision. Fig. 2.A: Three images or sets of directions. Fig. 2.B: The game is to rotate and position each set of direction so that corresponding image directions meet in one point.

where P is a two by three matrix, u is a two vector $u = (u_1, u_2)^T$ and λ is a non-zero depth.

We will say that we have **uncalibrated cameras** if nothing is known about the projection matrices P. The cameras are said to be **calibrated** if each matrix is of the type

$$P = R_\theta \begin{pmatrix} I & -\mathbf{c} \end{pmatrix} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} 1 & 0 & -\mathbf{c}_1 \\ 0 & 1 & -\mathbf{c}_2 \end{pmatrix} \quad . \tag{2}$$

The interpretation of the calibrated camera projection is that u is the direction from camera center $\mathbf{c}$ to the object U, after a rotation with angle θ. In Fig. 1 the projection of seven object points onto a pencil of seven directions is illustrated. Think of each image as such a set of directions. The shape of each image depends on both object positions (structure) and camera matrix (motion). One major goal of computer vision is to solve for structure and motion given only a set of these images.

The **structure and motion problem** is the problem of finding **structure** (object positions $(U^1, \ldots, U^n)$) and **motion** (a number of camera matrices $P(t_i)$ at time instants $(t_1, \ldots, t_m)$), using only the measured image directions $u^j(t_i) \sim P(t_i)U^j$. As in the three-dimensional case this can only be done up to an arbitrary choice of coordinate system. For example, in the calibrated camera case, where the camera matrices $P(t_i)$ are of type (2), structure and motion can only be found up to an unknown similarity transformation. A canonic reference frame can be chosen by $R_1 = I$, $\mathbf{c}(t_1) = 0$ and $|\mathbf{c}(t_m)| = 1$. This gives a unique (invariant) representation.

3 The calibrated trilinear tensor

In this section we will study the two-dimensional version of the multilinear constraints. These constraints on camera motion and image positions are interesting for several reasons.

Firstly, they can be used to calculate camera motion from known image correspondences without explicitly calculating the structure of the object points. This gives us a tool to obtain an initial estimate on motion and then structure. This initial estimate can then be used to find structure and motion with high precision using optimisation techniques, cf. [8].

Secondly, the multilinear constraints can be used to eliminate erroneous image correspondences, and may help in finding new correct ones. This is essential for robust, automatic structure and motion algorithms.

In the three-dimensional case, useful information is obtained already for two images, resulting in the calibrated bilinear form (the essential matrix) and the uncalibrated bilinear form (the fundamental matrix). However, in the two-dimensional case, no useful information is obtained until we study three images.

Theorem 3.1. *Let $u = u(t_1)$, $\tilde{u} = u(t_2)$ and $\hat{u} = u(t_3)$ be the image of the same point at three time instants t_1, t_2 and t_3, then the following trilinear constraint must hold*

$$\det \begin{pmatrix} \tilde{\mathbf{c}} & \tilde{R}u & \tilde{u} & 0 \\ \hat{\mathbf{c}} & \hat{R}u & 0 & \hat{u} \end{pmatrix} = \sum_{i,j,k} T^{i,j,k} u_i \bar{u}_j \hat{u}_k = 0 \ , \tag{3}$$

where the three camera matrices are

$$P(t_1) = \begin{pmatrix} I & 0 \end{pmatrix}, \qquad P(t_2) = \begin{pmatrix} \tilde{R} & \tilde{\mathbf{c}} \end{pmatrix}, \qquad P(t_3) = \begin{pmatrix} \hat{R} & \hat{\mathbf{c}} \end{pmatrix} \ . \tag{4}$$

Proof. See [4]. ■

Note that the above constraint involves only the **motion** parameters $(\tilde{R}, \hat{R}, \tilde{\mathbf{c}}, \hat{\mathbf{c}})$ and not the structure U. The calibrated trilinear tensor T will now be analysed in more detail. The tensor components $T^{i,j,k}$ can be expressed using rotation matrices $\tilde{R}$, $\hat{R}$ and the vectors $\tilde{\mathbf{c}}$, $\hat{\mathbf{c}}$. Introduce variables according to the following:

$$\tilde{R} = \begin{pmatrix} c_1 & -s_1 \\ s_1 & c_1 \end{pmatrix}, \quad \tilde{\mathbf{c}} = \begin{pmatrix} h_1 \\ v_1 \end{pmatrix}, \quad \hat{R} = \begin{pmatrix} c_2 & -s_2 \\ s_2 & c_2 \end{pmatrix}, \quad \hat{\mathbf{c}} = \begin{pmatrix} h_2 \\ v_2 \end{pmatrix} \ , \tag{5}$$

with $c_1 = \cos(\theta_1)$, $s_1 = \sin(\theta_1)$, $c_2 = \cos(\theta_2)$, $s_2 = \sin(\theta_2)$. Calculating the tensor components we find

$$\begin{aligned} &T^{111} = -v_1 s_2 + v_2 s_1, \quad && T^{112} = v_1 c_2 - h_2 s_1, \quad && T^{121} = h_1 s_2 - v_2 c_1, \quad && T^{122} = -h_1 c_2 + h_2 c_1, \\ &T^{211} = -v_1 c_2 + v_2 c_1, \quad && T^{212} = -v_1 s_2 - h_2 c_1, \quad && T^{221} = h_1 c_2 + v_2 s_1, \quad && T^{222} = h_1 s_2 - h_2 s_1 \ . \end{aligned} \tag{6}$$

Theorem 3.2. *A tensor $T^{i,j,k}$ is a calibrated trilinear tensor if and only if*

$$-T^{111} + T^{122} + T^{212} + T^{221} = 0 \ , \quad T^{112} + T^{121} + T^{211} - T^{222} = 0 \ . \tag{7}$$

When these conditions are fulfilled it is possible to solve (6) for $\tilde{R}$, $\hat{R}$, $\tilde{T}$, $\hat{T}$. In general there are two (possibly complex) solutions.

Proof. It is straightforward to show that these two conditions are fulfilled if the tensor components is given by (6). Conversely, assume that the tensor components $T^{i,j,k}$ fulfill (7). Introduce variables $S_1, \dots, S_8$ according to

$$\begin{aligned} &S_1 = h_2 c_1, \quad S_2 = v_2 c_1, \quad S_3 = h_2 s_1, \quad S_4 = v_2 s_1, \\ &S_5 = h_1 c_2, \quad S_6 = v_1 c_2, \quad S_7 = h_1 s_2, \quad S_8 = v_1 s_2 \ . \end{aligned} \tag{8}$$

The tensor components depend linearly on these variables. Using $\mathbf{T}$ as a vector containing the tensor components, we have the following relation:

$$\mathbf{T} = \begin{pmatrix} T^{111} \\ T^{112} \\ T^{121} \\ T^{122} \\ T^{211} \\ T^{212} \\ T^{221} \\ T^{222} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \\ S_6 \\ S_7 \\ S_8 \end{pmatrix} = \mathbf{NS} \ . \tag{9}$$

The matrix above has rank 6 so for every tensor $\mathbf{T} \in \mathbb{P}^7$, the vector $\mathbf{S} \in \mathbb{P}^7$ can be determined to lie on a projective plane of dimension 2. This plane can be parametrised using $Z = (z_1, z_2, z_3) \in \mathbb{P}^2$, as $\mathbf{S} = z_1 \mathbf{S}_1 + z_2 \mathbf{S}_2 + z_3 \mathbf{S}_3$ where $\mathbf{S}_1$ depends on the tensor components $T^{i,j,k}$ and

$$\mathbf{S}_2 = \begin{pmatrix} 1 & 0 & 0 & -1 & 1 & 0 & 0 & -1 \end{pmatrix}^T, \quad \mathbf{S}_3 = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}^T \tag{10}$$

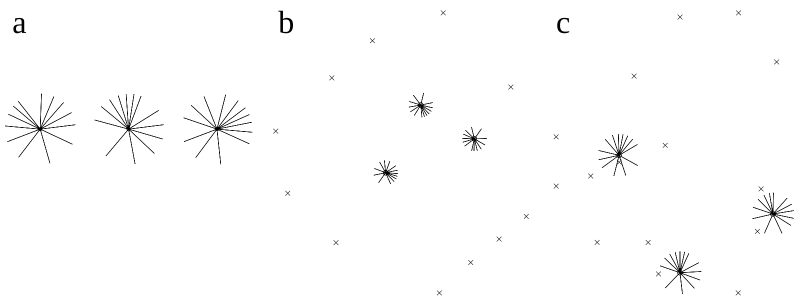

Figure 3: The figure illustrates the three images used as input in the example and the two possible solutions to the structure and motion problem as obtained from the calibrated trilinear tensor.

are a basis for the nullspace of **N**. However, from the structure of the variables it follows that

$$S_1S_4 = S_2S_3, \qquad S_5S_8 = S_6S_7 \ . \tag{11}$$

Using these two quadratic constraints we get four solutions up to scale. Two of these are complex and known,

$$Z = (z_1, z_2, z_3) = \begin{pmatrix} 0 & 1 & i \end{pmatrix}^T, \quad Z = (z_1, z_2, z_3) = \begin{pmatrix} 0 & 1 & -i \end{pmatrix}^T \ . \tag{12}$$

The other two solutions are easily found through a quadratic equation. They can be complex or real depending on the tensor components. For every real solution S that fulfills (9) and (11), it is easy to solve for the motion variables,

$$\begin{aligned} &(c_1, s_1) \sim (S_1, S_3) \sim (S_2, S_4), \quad (c_2, s_2) \sim (S_5, S_7) \ \sim (S_6, S_8), \\ &(h_1, v_1) \sim (S_5, S_6) \sim (S_7, S_8), \quad (h_2, v_2) \sim (S_1, S_2) \sim (S_3, S_4) \ . \end{aligned} \tag{13}$$

■

Note that there is a two-fold ambiguity in the solution of the structure and motion problem no matter how many corresponding points we have. However, the calculations above does not take into account the sign of the directions. Thus some of the reconstructed points might be in the wrong direction (with negative depth).

We illustrate this with a simple example. In Fig. 3.A. three images $(u^1, \ldots u^n)$, $(\tilde{u}^1, \ldots \tilde{u}^n)$ and $(\hat{u}^1, \ldots \hat{u}^n)$ of the same object points are shown. A matrix **M** is constructed so that

$$\mathbf{MT} = \begin{pmatrix} u_1^1\tilde{u}_1^1\hat{u}_1^1 & u_1^1\tilde{u}_1^1\hat{u}_2^1 & u_1^1\tilde{u}_2^1\hat{u}_1^1 & u_1^1\tilde{u}_2^1\hat{u}_2^1 & u_2^1\tilde{u}_1^1\hat{u}_1^1 & u_2^1\tilde{u}_1^1\hat{u}_2^1 & u_2^1\tilde{u}_2^1\hat{u}_1^1 & u_2^1\tilde{u}_2^1\hat{u}_2^1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_1^n\tilde{u}_1^n\hat{u}_1^n & u_1^n\tilde{u}_1^n\hat{u}_2^n & u_1^n\tilde{u}_2^n\hat{u}_1^n & u_1^n\tilde{u}_2^n\hat{u}_2^n & u_2^n\tilde{u}_1^n\hat{u}_1^n & u_2^n\tilde{u}_1^n\hat{u}_2^n & u_2^n\tilde{u}_2^n\hat{u}_1^n & u_2^n\tilde{u}_2^n\hat{u}_2^n \end{pmatrix} \begin{pmatrix} T^{111} \\ T^{112} \\ T^{121} \\ T^{122} \\ T^{211} \\ T^{212} \\ T^{221} \\ T^{222} \end{pmatrix} = 0 \ . \tag{14}$$

The matrix **M** should in theory be rank deficient. In practice it has full rank due to errors in the image measurements. In our example a singular value decomposition gives the following singular values

$$\begin{pmatrix} 2.13 & 1.84 & 1.31 & 1.16 & 1.08 & 0.68 & 0.64 & 0.001 \end{pmatrix}$$

and the vector corresponding to the lowest singular value is chosen as a candidate for the tensor,

$$\mathbf{T} = \begin{pmatrix} -0.66 & -0.08 & -0.17 & -0.17 & 0.52 & -0.13 & -0.36 & 0.27 \end{pmatrix}^T$$

It can now be checked that the tensor components fulfill (7), we get

$$-T^{111} + T^{122} + T^{212} + T^{221} = 0.00040, \qquad T^{112} + T^{121} + T^{211} - T^{222} = 0.00026 \ .$$

From the tensor components the planar subspace of possible vectors S is calculated. The two solutions satisfying (11) is found. From these it is straightforward to calculate the motion parameters and then the reconstruction. The result is illustrated in Fig. 3.B and C. Note that in this particular example some reconstructed points have negative depth. In another example, Fig. 4.A and B, all points have positive depth. This makes it possible to rule out erroneous reconstructions.

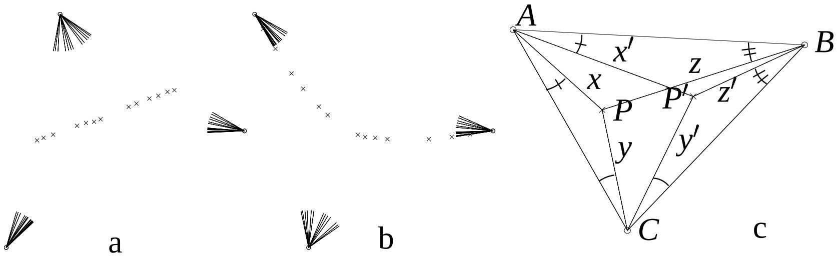

Figure 4: Fig. A and B illustrate two reconstructions for the same three images. Notice that the positions of the cameras are mirrored, and that the position of the object points are completely different. Fig.C illustrate a point P and its isogonal conjugate point P' with respect to triangle ABC

4 Understanding the ambiguity

In the numerical examples (Fig. 3.B-C and Fig. 4.A-B) one can almost see that the position of the camera centers in the two reconstructions are reversed. One suspects that the two-fold ambiguity has something to do with planar geometry. It turns out that the two-fold ambiguity is a consequence of the following theorem, which is illustrated in Fig. 4.C, cf. [1, p. 113]. Further reading can be found in [2, p. 97] and [3, p. 273].

Theorem 4.1. *Let ABC be a triangle. Let x, y and z be lines respectively through A, B and C that intersects one point, say P. Let the line x' be the reflection of x in the bisector of angle A and similarly for y' and z', then the three lines x' y' and z' intersect at one point, say P'.*

The points P and P' are called **isogonal conjugate points** with respect to the triangle ABC. One interesting property of these points is that they are focal points for a conic inscribed in the triangle, i.e. a conic which are tangent to all three sides of the triangle.

Thus for each possible solution of the structure and motion problem for three cameras, a conjugate solution can be found by first taking the isogonal conjugate points and then invert the whole situation. This can be seen in Fig. 4.A and B.

5 Conclusions

In this paper we have studied the calibrated trilinear tensor in two-dimensional vision. It has been shown that each trilinear tensor corresponds to two different camera motions. This ambiguity is closely linked to the concept of isogonal conjugacy. It is also shown how the trilinear tensor can be used to solve the structure and motion problem using simple linear methods.

References

[1] Baker, H. F., An Introduction to Plane Geometry, *Cambridge University Press*, 1943.

[2] Coxeter, H. S. M., The Real Projective Plane, Third Edition *Springer-Verlag, New York, NY, USA*, 1993.

[3] Altshiller-Court, N., College Geometry, *Barnes and Noble, New York*, 1952.

[4] Heyden, A. and Åström, K., A Canonical Framework for Sequences of Images, *Proc. IEEE workshop on Representation of Visual Scenes, MIT, Cambridge, MA, USA*, 1995.

[5] Shashua, A., Trilinearity in Visual Recognition by Alignment, *ECCV'94, Lecture notes in Computer Science, Vol 800. Ed. Jan-Olof Eklund, Springer-Verlag* 1994, pp. 479-484.

[6] Sparr, G., A common framework for kinetic depth, reconstruction and motion for deformable objects, *ECCV'94, Lecture notes in Computer Science, Ed. Jan-Olof Eklund, Springer-Verlag* 1994, pp. 471-481.

[7] Åström, K. and Heyden, A., Multilinear forms in the velocity case, *to be presented at CVPR'96*, 1996.

[8] Åström, K., Automatic Mapmaking *Selected Papers from the 1st IFAC International Workshop on Intelligent Autonomous Vehicles, Southampton, UK*, 1993.