A Canonic Framework for Sequences of Images: Similarities between Calibrated and Uncalibrated Case

Kalle Åström, Anders Heyden
Dept of Mathematics, Lund University
Box 118, S-221 00 Lund, Sweden
email: kalle@maths.lth.se andersp@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 plane that only depend on motion and not on the viewed objects. The epipolar constraint is such an example. This paper deals with the generalisation of the epipolar constraint to the problem of analysing sequences of images taken by calibrated or uncalibrated cameras. It is assumed that the correspondences between the points in the different images are known. In a previous paper it is shown how the problem of analysing sequences of images taken by an uncalibrated camera can be simplified by choosing a specific affine basis in each image plane. In this paper we show that this theory has remarkable similarities with the theory for calibrated cameras.

1 Introduction

A central problem in scene analysis is the analysis of 3D-objects from 2D-images, obtained by projections. In this paper we will concentrate on the case of a sequence of images consisting of points, with known correspondences. 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

- The constraints can be used to find further image correspondences.
- The constraints give information about the camera motion.

This type of invariancy has become increasingly popular and is currently under intense research. A generalisation of the epipolar bilinear constraint for two images to a trilinear constraint for three images is presented in [6]. There has also been a lot of discussion on generalising this further to quadrilinear constraints in four images and so on. It is shown in [3] that these multilinear forms do not contain anything new. The multilinear forms can always be obtained from the bilinear and trilinear forms.

In a previous paper it is shown how this analysis of image sequences can be simplified considerably by choosing a specific affine basis in each image plane. In this paper we show that this theory has remarkable similarities with the theory for calibrated cameras.

2 Problem Formulation

Given a sequence of images our first aim is to analyse the constraints on image points and on camera motion. These constraints will help us to determine camera motion from image correspondences without explicitly calculating the shape of the object points.

As our aim is to show how similar the uncalibrated and the calibrated case is, many of the results will be shown pairwise. In each subsequent displayed equation the calibrated case will be shown on the left hand side and its uncalibrated analogue on the right. The simplifications can be achieved after choosing a particular affine basis in each image plane as described in [3]. Thus assume that a sequence of images have been taken from different viewpoints. Let X denote an arbitrary point in space represented as a column vector in homogeneous coordinates. Denote by x_i its projection onto the i'th image plane, also in homogeneous coordinates. We assume that the camera is described by the following standard model.

$$\lambda x_i = \begin{bmatrix} \lambda x_{ix} \\ \lambda y_{ix} \\ \lambda \end{bmatrix} = P_i \begin{bmatrix} X_x \\ X_y \\ X_z \\ 1 \end{bmatrix} = P_i X, \tag{1}$$

where P_i is the 3×4 projection matrix and λ is a scale factor (different for different points). Notice that it is only the direction of x_i that is important. One should think of the 3×1 column vector as defined only up to a constant. For calibrated cameras the projection matrix P_i is of a particular type, which can be expressed using a rotation matrix R_i and a vector t_i. The situation is more complex for uncalibrated cameras. But the analysis can be simplified drastically by our particular choice of image coordinate system. Three object points are used as a partial projective basis for the object coordinate system. The image of these points are then also used as an affine basis of each image plane.

For this partial choice of coordinate system it is shown in [3] that the projection matrices P_i can be expressed using a diagonal matrix D_i and a vector t_i in a way which is analogous to the calibrated case.

$$\underbrace{P_i = [R_i | -R_i t_i]}_{\text{Calibrated case}} \qquad \underbrace{P_i = [D_i | -D_i t_i]}_{\text{Uncalibrated case}} \tag{2}$$

We will continue to write the calibrated case on the left of each displayed equation and the uncalibrated case on the right. Notice that in the uncalibrated case we still have full freedom to change the coordinate system, whereas in the uncalibrated case only partial changes are allowed. Denote by M_i the coordinate change from world coordinates to the i'th camera coordinates, cf. Fig. 1,

$$M_i = \begin{pmatrix} R_i & -R_i t_i \\ 0 & 1 \end{pmatrix} \qquad M_i = \begin{pmatrix} D_i & -D_i t_i \\ 0 & 1 \end{pmatrix} \tag{3}$$

with this notation each camera can be considered as a coordinate transformation followed by the standard camera matrix

$$P_0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{4}$$

so that $P_i = P_0 M_i$. Notice the strong similarities between the two different cases. These simplifications and similarities are important by itself. However, the interpretation of the transformations are equally important. In the calibrated case the transformation between the object coordinate system and the i'th camera coordinate system is given by the rotation matrix R_i and the translation vector t_i. In other words the focal point of the i'th camera has coordinates $\binom{t_i}{1}$ in the object coordinate system. In the uncalibrated case it can be shown that the matrix D_i contain as diagonal elements the relative depths, cf. [7], of the three basis points. In some sense D_i represent the 'orientation' of the i'th camera with respect to the basis points. Also $\binom{t_i}{1}$ is the projective coordinates of the i'th camera focal point with respect to the object coordinate system.

The situation is illustrated in Fig. 1. In this figure we have also included the change of coordinate system between different camera coordinates M_{ij} which we define as $M_{ij} = M_i M_j^{-1}$. This block matrix can be seen to be

$$M_{ij} = \begin{pmatrix} R_i R_j^{-1} & R_i(t_j - t_i) \\ 0 & 1 \end{pmatrix} \qquad M_{ij} = \begin{pmatrix} D_i D_j^{-1} & D_i(t_j - t_i) \\ 0 & 1 \end{pmatrix} \tag{5}$$

and we define

$$R_{ij} = R_i R_j^{-1} \qquad D_{ij} = D_i D_j^{-1} \tag{6}$$

$$t_{ij} = R_i(t_j - t_i) \qquad t_{ij} = D_i(t_j - t_i) \tag{7}$$

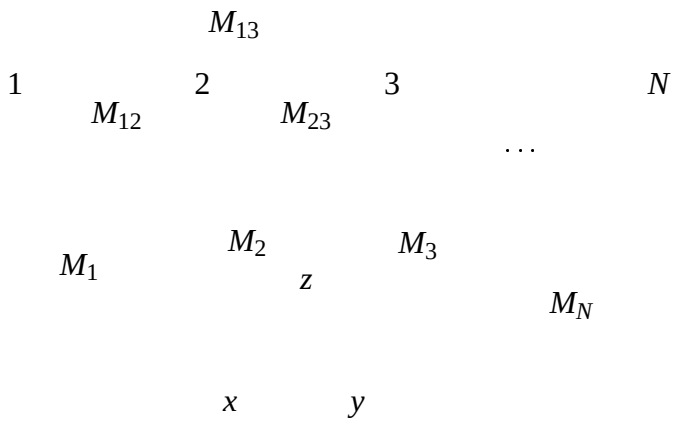

Figure 1: A sequence of images and coordinate transformations between different camera coordinate systems and the object coordinate system.

Thus R_{ij} is the change of orientation from the j'th camera coordinate system to the i'th camera coordinate system. Similarly D_{ij} can be shown to be the kinetic depths for to the three basis points. Through the multilinear forms, which can be calculated from the images, we obtain partial information about the relative coordinate transformations M_{ij}. This partial information is often difficult to use and analyse. It is therefore natural to try to estimate the full motion of the cameras instead, i.e. the coordinate transformations M_i, $i = 1, \ldots, n$.

3 Two images

The epipolar constraint between two cameras can be written as

$$x_i^T E_{ij} x_j = 0 \qquad\qquad x_i^T F_{ij} x_j = 0 \tag{8}$$

where x_i is a point in image i and x_j is a corresponding point in image j. In the calibrated case the essential matrix E_{ij} is a matrix with singular values λ, λ and 0. In the uncalibrated case the fundamental matrix F_{ij} is simply a rank two matrix. In [3] it is shown that with the specific choice of basis points the essential matrix in the calibrated case and the fundamental matrix in the uncalibrated case have similar decompositions,

$$E_{ij} = T_{t_{ij}} R_{ij} \qquad\qquad F_{ij} = T_{t_{ij}} D_{ij} \tag{9}$$

where T_t is the matrix representing cross product with t. The bilinear (epipolar) constraint can also be expressed as

$$\operatorname{rank} \begin{bmatrix} x_i & t_{ij} & R_{ij} x_j \end{bmatrix} = 2 \qquad\qquad \operatorname{rank} \begin{bmatrix} x_i & t_{ij} & D_{ij} x_j \end{bmatrix} = 2 \tag{10}$$

This can be interpreted that the direction x_j after a transformation to the i'th camera coordinate system is coplanar with x_i and the direction between the two cameras. This formulation will be generalised to the multiple image case.

The bilinear constraints can be calculated from point and/or curve correspondences in the two images. Thus (R_{ij}, t_{ij}) or (D_{ij}, t_{ij}) can be calculated, but only up to an unknown scale factor. This information can then be used both to find further image correspondences but also as we shall see to calculate the full camera motion, up to an unknown scale factor. In the two image case we do as follows. First we choose object coordinate system e.g. R_1 as the identity and t_1 as the origin. Then (R_2, t_2) can easily be chosen in accordance with Eqs. (6) and (7).

4 A Sequence of Images

The bilinear constraint between two images contains all information about the relative camera motion between the two images. If many images are taken the bilinear constraint can be calculated from each image pair. This information is however, cumbersome and contains to many degrees of freedom. For more than two images it is better to consider the projection matrices themselves, instead of the multilinear constraints or the bilinear constraints between each image pair.

A collection of bilinear constraints are only consistent if

$$R_{ij} = R_{ik}R_{kj} \qquad D_{ij} = D_{ik}D_{kj} \tag{11}$$

$$t_{ij}, t_{ik}, R_{ik}t_{kj} \text{ are coplanar} \qquad t_{ij}, t_{ik}, D_{ik}t_{kj} \text{ are coplanar} \tag{12}$$

For general camera positions, consistent bilinear constraints determines the motion uniquely up to similarity transformations in the calibrated case and uniquely up to projective transformations in the uncalibrated case. For special camera positions, in particular if the camera move along a straight line, the bilinear constraints are not enough. They do not determine the motion uniquely. In this case the trilinear constraint can be used to determine the camera motion. The trilinear constraint can be formulated as follows.

$$\text{rank}\begin{bmatrix} x_1 & t_{12} & R_{12}x_2 & 0 \\ x_1 & \lambda t_{13} & 0 & R_{13}x_3 \end{bmatrix} = 3 \qquad \text{rank}\begin{bmatrix} x_1 & t_{12} & D_{12}x_2 & 0 \\ x_1 & \lambda t_{13} & 0 & D_{13}x_3 \end{bmatrix} = 3 \tag{13}$$

where x_1, x_2 and x_3 are corresponding points in images 1, 2 and 3 respectively. Notice the scale factor λ in the trilinear constraint. The interpretation is that not only is the direction from camera 1 to camera 2, t_{12}, and the direction from camera 1 to camera 3, t_{13}, determined from the trilinear constraint, but also the relative depths of camera 2 with respect to camera 3. Thus even for degenerate camera motion the trilinear constraints determine the motion (R_i, t_i), $i = 1, \ldots, n$ or (D_i, t_i), $i = 1, \ldots, n$.

Generalisations of the bilinear (epipolar) constraint and the trilinear constraint can be found. These N-linear constraints can be formulated as

$$\text{rank}\begin{bmatrix} x_1 & t_{12} & R_{12}x_2 & \ldots & 0 \\ \vdots & \vdots & 0 & \ddots & 0 \\ x_1 & \lambda_N t_{1N} & 0 & \ldots & R_{1N}x_N \end{bmatrix} = N \qquad \text{rank}\begin{bmatrix} x_1 & t_{12} & D_{12}x_2 & \ldots & 0 \\ \vdots & \vdots & 0 & \ddots & 0 \\ x_1 & \lambda_N t_{1N} & 0 & \ldots & D_{1N}x_N \end{bmatrix} = N \tag{14}$$

However these N-linear constraints do not add much to the understanding of image sequences since each constraint can be expressed in the bilinear and trilinear constraints as explained in [3]. The full camera motion is a better parametrisation for this information.

5 Conclusions

In this paper we have demonstrated some of the similarities between the geometry of calibrated and uncalibrated image sequences. The question of epipolar constraint versus motion is easier to understand in the calibrated case. The strong similarities between the calibrated and the uncalibrated case makes it somewhat easier to understand the interplay between epipolar constraints and motion in uncalibrated image sequences.

References

[1] Faugeras, O., D., What can be seen in three dimensions with an uncalibrated stereo rig?, *ECCV'92, Lecture notes in Computer Science, Vol 588. Ed. G. Sandini, Springer-Verlag* 1992, pp. 563-578.

[2] Heyden, A., Reconstruction from Three Images of Six Point Objects, *Proc. Symposium on Image Analysis, SSAB, Halmstad, Sweden, 1994.*

[3] Heyden, A., Åström, K., A Canonic Framework for Sequences of Images: The Uncalibrated Case, *Proc. Symposium on Image Analysis, SSAB, Linköping, Sweden, 1995.*

[4] Luong, Q., Vieville, T., Canonic Representations for the Geometries of Multiple Projective Views, *ECCV'94, Lecture notes in Computer Science, Vol 800. Ed. Jan-Olof Eklund, Springer-Verlag* 1994, pp. 589-599.

[5] Quan, L., Invariants of 6 points from 3 uncalibrated images, *ECCV'94, Lecture notes in Computer Science, Vol 801. Ed. Jan-Olof Eklund, Springer-Verlag* 1994, pp. 459-470.

[6] Shahsua, 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.

[7] Sparr, G., A common framework for kinetic depth, reconstruction and motion for deformable objects, *ECCV'94, Lecture notes in Computer Science, Vol 800. Ed. Jan-Olof Eklund, Springer-Verlag* 1994, pp. 471-481.