A Canonic Framework for Sequences of Images: The Uncalibrated Case

Anders Heyden, Kalle Åström
Dept of Mathematics, Lund University
Box 118, S-221 00 Lund, Sweden
email: andersp@maths.lth.se kalle@maths.lth.se

Abstract

This paper deals with the problem of analysing sequences of images taken by uncalibrated cameras. It is assumed that the correspondences between the points in the different images are known. This paper introduces a new framework for this problem. The situation of two cameras is described by the reduced fundamental matrix. Another new concept, the reduced fundamental tensor, describes the situation of three images. In the first case we get bilinear constraints on the image coordinates and in the second we get trilinear constraints. Furthermore it is shown that the trilinear constraints in general can be expressed by the bilinear and that constraints for four or more images always can be expressed by three images constraints. Finally a new canonic form for the camera matrices are given which has a lot of similarities with the calibrated case. From this description it is also possible to calculate the camera movement.

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. We will present a method where no camera calibration is needed; making it possible to reconstruct the object and the camera movement up to a projective transformation. The case with three images and six points was solved independently in [6] and [3], by different methods. The case with two images and seven points was solved in [8].

The problem of predicting new points in further images has been treated in [7], and [2]. Shashua uses affine invariants and gets two linear constraints on the coordinates of the points in the third image. Faugeras uses the fundamental matrices between the images in order to obtain two linear constraints. The problem of finding canonic description of the camera matrices has been treated in [5].

Our approach is based on the possibility to chose three corresponding points in every image as base points of an affine coordinate system. As a by-product of the reconstruction the kinetic depths, see [8], can be easily calculated. These show in some way how the camera has moved between the images. We also derive simpler forms of the bilinearities and the trilinearities, with fewer parameters and less degree of freedom than the usual. Finally we get a simpler form of the projection matrices and can calculate the camera movement.

2 Problem Formulation

We assume that the camera is described by the following standard model,

$$\begin{bmatrix} \lambda x \\ \lambda y \\ \lambda \end{bmatrix} = P \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \qquad (1)$$

where x and y are the image coordinates, P is the projection matrix, X, Y and Z are the coordinates of the object and λ is a scale factor (different for different points). Given a sequence of images from this model our aim is to reconstruct the object, to calculate mutual invariants between the images and to obtain a canonic description of this situation.

3 One Image

Since the camera matrices P are unknown we can multiply (1) from the left by an arbitrary nonsingular three by three matrix, which corresponds to chosing different affine coordinate systems in the images. A reasonable choice, to exploit this degree of freedom, is to make the first three points in the images to have affine coordinates $(1,0,0)$, $(0,1,0)$ and $(0,0,1)$, that is a standard affine basis. Furthermore we can multiply (1) from the right by an arbitrary nonsingular four by four matrix, which corresponds to chosing affine coordinates in the object. In this way we can also assume that the scale factors for the first four points are equal to 1. Assume that the affine coordinates of the fourth point in the first image are (a_1,b_1,c_1). For reasons that soon will be clear we chose the four by four matrix such that the first four points in the objects have affine coordinates $(1,0,0,0)$, $(0,1,0,0)$, $(0,0,1,0)$ and $(a_1,b_1,c_1,1)$. Putting together the equations from (1) for the first four points gives

$$\begin{bmatrix}1 & 0 & 0 & a_1\\ 0 & 1 & 0 & b_1\\ 0 & 0 & 1 & c_1\end{bmatrix} = \begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\end{bmatrix}\begin{bmatrix}1 & 0 & 0 & a_1\\ 0 & 1 & 0 & b_1\\ 0 & 0 & 1 & c_1\\ 0 & 0 & 0 & 1\end{bmatrix}, \tag{2}$$

since the only possibility for P is $P = [I\,|\,0]$. This also explains the choice of object coordinates. Since the sum of the affine coordinates of a points in the image is 1 and the sum of the first three affine coordinates in the object is 1 it can be seen that all scale factors λ in the first image are equal to 1. It also follows that an arbitrary point in the object with affine coordinates (X,Y,Z,W) projects to the image point (X,Y,Z). This means that the only unknown coordinate are the fourth and if this coordinate can be calculated we have a reconstruction.

4 Two Images

We now consider a second image. Let the scale factors for the first four points be p_1, p_2, p_3 and p_4 and the affine coordinates for the fourth point in the second image be (a_2,b_2,c_2). Then the camera matrix is given from

$$\begin{bmatrix}1 & 0 & 0 & a_2\\ 0 & 1 & 0 & b_2\\ 0 & 0 & 1 & c_2\end{bmatrix} = \begin{bmatrix}p_1 & 0 & 0 & f_1\\ 0 & p_2 & 0 & f_2\\ 0 & 0 & p_3 & f_3\end{bmatrix}\begin{bmatrix}1 & 0 & 0 & a_1\\ 0 & 1 & 0 & b_1\\ 0 & 0 & 1 & c_1\\ 0 & 0 & 0 & 1\end{bmatrix}, \tag{3}$$

where $(f_1,f_2,f_3) = (p_4a_2 - p_1a_1, p_4b_2 - p_2b_1, p_4c_2 - p_3c_1)$. That is $P = [D_p\,|\,f]$, where D_p is the diagonal matrix obtained from (p_1,p_2,p_3) and $f = (f_1,f_2,f_3)$.

Consider now an arbitrary point (X,Y,Z,W) in the object and its projection (x_1,y_1,z_1) and (x_2,y_2,z_2) in the first and second image. It follows from (2) and (3) that

$$(x_1,y_1,z_1) = (X,Y,Z), \qquad (px_2,py_2,pz_2) = (p_1X + f_1W, p_2Y + f_2W, p_3Z + f_3W), \tag{4}$$

where p are the scale factor for the point. From these equation we can eliminate p and (X,Y,Z,W) which gives

$$\det\begin{bmatrix}p_1x_1 & f_1 & x_2\\ p_2y_1 & f_2 & y_2\\ p_3z_1 & f_3 & z_2\end{bmatrix} = 0. \tag{5}$$

Rewriting this determinant gives

$$\begin{bmatrix}x_1 & y_1 & z_1\end{bmatrix}\begin{bmatrix}0 & p_1f_3 & -p_1f_2\\ -p_2f_3 & 0 & p_2f_1\\ p_3f_2 & -p_3f_1 & 0\end{bmatrix}\begin{bmatrix}x_2\\ y_2\\ z_2\end{bmatrix} = 0, \tag{6}$$

or $x^TF\bar{x} = 0$, where x and $\bar{x}$ are the affine coordinate vectors in image one and two and F are the fundamental matrix for this particular choice of coordinates. We call F the *reduced fundamental matrix*. It can be seen from (6) that F can be factorised as

$$F = \begin{bmatrix}p_1 & 0 & 0\\ 0 & p_2 & 0\\ 0 & 0 & p_3\end{bmatrix}\begin{bmatrix}0 & f_3 & -f_2\\ -f_3 & 0 & f_1\\ f_2 & -f_1 & 0\end{bmatrix} = p_1p_2p_3\begin{bmatrix}0 & f_3/p_3 & -f_2/p_2\\ -f_3/p_3 & 0 & f_1/p_1\\ f_2/p_2 & -f_1/p_1 & 0\end{bmatrix}\begin{bmatrix}p_1^{-1} & 0 & 0\\ 0 & p_2^{-1} & 0\\ 0 & 0 & p_3^{-1}\end{bmatrix}, \tag{7}$$

that is as a product of a diagonal and a skew matrix. The diagonal matrix contains the kinetic depths, see [8] and [4], on the diagonal and the skew matrix is formed from the vector f. From (6) and (7) it is obvious that $f = \mu e_{21}$ where e_{21} are the affine coordinates of the epipole in the second image and $D_p^{-1} f = \nu e_{21}$ where e_{12} are the affine coordinates of the epipole in the first image, and μ and ν are scale factors needed to make the sum of the affine coordinates equal to 1. The components of the reduced fundamental matrix can be calculated linearly from eight point matches and then the kinetic depths and the epipoles can be recovered linearly. Finally, the eliminated W-coordinates can be calculated which gives the reconstruction.

5 Three Images

We now consider three images concurrently. Let the scale factors for the first four points in the third image be q_1, q_2, q_3 and q_4 and the affine coordinates for the fourth point in the third image be (a_3, b_3, c_3). Then the camera matrix for camera three is given by

$$P_3 = \begin{bmatrix} q_1 & 0 & 0 & \bar{f}_1 \\ 0 & q_2 & 0 & \bar{f}_2 \\ 0 & 0 & q_3 & \bar{f}_3 \end{bmatrix} \tag{8}$$

where $(\bar{f}_1, \bar{f}_2, \bar{f}_3) = (q_4 a_3 - q_1 a_1, q_4 b_3 - q_2 b_1, q_4 c_3 - q_3 c_1)$. It follows that $P = [D_q \,|\, \bar{f}]$, where D_q is the diagonal matrix obtained from (q_1, q_2, q_3) and $\bar{f} = (\bar{f}_1, \bar{f}_2, \bar{f}_3)$. Consider now an arbitrary point (X, Y, Z, W) in the object and its projection (x_3, y_3, z_3) in the third image. It follows from (3) that

$$(qx_3, qy_3, qz_3) = (q_1 X + \bar{f}_1 W, q_2 Y + \bar{f}_2 W, q_3 Z + \bar{f}_3 W), \tag{9}$$

where q is the scale factor for the point. From (4) and (9) we can eliminate p, q and (X, Y, Z, W) which gives

$$\text{rank} \begin{bmatrix} p_1 x_1 & f_1 & x_2 & 0 \\ p_2 y_1 & f_2 & y_2 & 0 \\ p_3 z_1 & f_3 & z_2 & 0 \\ q_1 x_1 & \bar{f}_1 & 0 & x_3 \\ q_2 y_1 & \bar{f}_2 & 0 & y_3 \\ q_3 z_1 & \bar{f}_3 & 0 & z_3 \end{bmatrix} = 3. \tag{10}$$

From this equation it follows that there are three algebraically independent equations in the image coordinates. Two of them can be chosen as the bilinear constraints from the reduced fundamental matrix and one is a combined expression involving coordinates of points in all three images. It can also be shown that there are four linearly independent trilinear expressions in the image coordinates. It also follows from (10) that it is possible to recover p, q, f, and $\bar{f}$ up to scale factors. If we fix the scale for p and q then f and $\bar{f}$ can be recovered without unknown scale factors. This means that the camera matrices can be completely recovered from this information. We remark that this can be done linearly from at least seven point matches in three images by the reduced fundamental tensor.

Considering the pair of images two and three we can derive the reduced fundamental matrix from (10) by eliminating x_1, x_2 and x_3 which gives

$$\det \begin{bmatrix} q_1 x_2 & p_1 x_3 & q_1 f_1 - p_1 \bar{f}_1 \\ q_2 y_2 & p_2 y_3 & q_2 f_2 - p_2 \bar{f}_2 \\ q_3 z_2 & p_3 z_3 & q_3 f_3 - p_3 \bar{f}_3 \end{bmatrix} = p_1 p_2 p_3 \det \begin{bmatrix} r_1 x_2 & x_3 & r_1 f_1 - \bar{f}_1 \\ r_2 y_2 & y_3 & r_2 f_2 - \bar{f}_2 \\ r_3 z_2 & z_3 & r_3 f_3 - \bar{f}_3 \end{bmatrix} = 0, \tag{11}$$

where $r_i = q_i / p_i$ are the kinetic depths between image two and three. The epipole of camera 2 in image 3 is a multiple of $\hat{f} = (r_1 f_1 - \bar{f}_1, r_2 f_2 - \bar{f}_2, r_3 f_3 - \bar{f}_3)$. This means that if f, $\bar{f}$ and $\hat{f} = D_r f - \bar{f}$ can be determined up to unknown scale factors from reduced fundamental matrices between the three different pairs of images, then it is possible to recover the camera matrices if f, $\bar{f}$ and $\hat{f}$ are not collinear. Geometrically this can be viewed as vector addition of f and $\hat{f}$ with appropriate lengths giving $\bar{f}$ also with appropriate length.

6 Four Images

In the same way as above it can be shown that considering four images, elimination of object coordinates and kinetic depths gives

$$
\text{rank}\begin{bmatrix}
p_1x_1 & f_1 & x_2 & 0 & 0 \\
p_2y_1 & f_2 & y_2 & 0 & 0 \\
p_3z_1 & f_3 & z_2 & 0 & 0 \\
q_1x_1 & \bar{f}_1 & 0 & x_3 & 0 \\
q_2y_1 & \bar{f}_2 & 0 & y_3 & 0 \\
q_3z_1 & \bar{f}_3 & 0 & z_3 & 0 \\
s_1x_1 & \tilde{f}_1 & 0 & 0 & x_4 \\
s_2y_1 & \tilde{f}_2 & 0 & 0 & y_4 \\
s_3z_1 & \tilde{f}_3 & 0 & 0 & z_4
\end{bmatrix} = 4, \tag{12}
$$

where s_i are the kinetic depths between image 1 and 4, $\tilde{f}_i$ are a multiple of the epipole in image 4 from camera 1 and x_4, y_4 and z_4 are affine coordinates in the fourth image. By considering all five by five subdeterminants it can be seen that we get nothing new, just bilinearities from pairs of images and trilinearities from triples of images.

7 Sequence of Images

It has been shown above that the camera matrices can be written as

$$
P_1 = [I\,|\,0], \qquad P_i = [D_i\,|\,-f_i], \quad i = 2\ldots n, \tag{13}
$$

where D_i are diagonal matrices formed from the kinetic depths from the three basis points, f_i are multiples of the affine coordinates for the epipoles of camera 1 in image i and $n+1$ is the number of images. Equivalently this can be written as

$$
P_1 = [I\,|\,0], \qquad P_i = [D_i\,|\,-D_i\bar{f}_i], \quad i = 2\ldots n, \tag{14}
$$

where $\bar{f}_i$ is a multiple of the affine coordinates of the epipole in image 1 from camera i. Notice the similarities with the calibrated case, see also [1]. Now it is possible to calculate the positions of the center, Z_i, of camera i as the nullspace of P_i, which is $(\bar{f}_i, 1) = (\bar{f}_{i1}, \bar{f}_{i2}, \bar{f}_{i3}, 1)$, where $\bar{f}_i$ are a multiple of the affine coordinates of the epipole in image 1 of camera i. Then all possible camera locations can be calculated from this representation by a projective transformation.

8 Conclusions

In this paper we have given a new canonic representation of the camera matrices in a sequence of images from uncalibrated cameras. A new form of the fundamental matrix, called the reduced fundamental matrix, has been given. It has been shown that it can be factorised as a product of a diagonal and a skew matrix. We have furthermore derived the bilinear and trilinear relations and shown that in general the trilinearities follows from the bilinearities and that there is no need to consider m-linearities, with $m \geq 4$. Finally we have shown how the camera movement can be calculated from this representation.

References

[1] Åström. K., Heyden, A., Canonic Framework for Sequences of Images: Similarities between Calibrated and Uncalibrated Case, *Proc. Symposium on Image Analysis, SSAB, Linköping, Sweden, 1995.*

[2] 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.

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

[4] Heyden, A., Reconstruction from Image Sequences by means of Relative Depths, *submitted to ICCV-95.*

[5] 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.

[6] 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.

[7] 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. 479-4484.

[8] Sparr, G., A Common Framework for Kinetic Depth, Reconstruction and Motion for Deformable Objects, *ECCV'94, Lecture notes in Computer Science, Vol 801. Ed. J-O. Eklund, Springer-Verlag* 1994, pp. 471-482.