

Euclidean Reconstruction from Constant Intrinsic Parameters*

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

Abstract

In this paper a new method for Euclidean reconstruction from sequences of images taken by uncalibrated cameras, with constant intrinsic parameters, is described. Our approach leads to a variant of the so called Kruppa equations. It is shown that it is possible to calculate the intrinsic parameters as well as the Euclidean reconstruction from at least three images.

The novelty of our approach is that we build our calculation on a projective reconstruction, obtained without the assumption on constant intrinsic parameters. This assumption simplifies the analysis, because a projective reconstruction is already obtained and we need ‘only’ to find the correct Euclidean reconstruction among all possible projective reconstructions.

1. Introduction

There exist several different methods to make Euclidean reconstruction, that is to reconstruct the object up to Euclidean transformations. However, in reconstruction without any knowledge about the scene, the scale ambiguity is always present, because it is impossible to distinguish between a large object far away and a small object close to the camera. This means that it is only possible to reconstruct the object up to similarity transformations, that is Euclidean transformation plus a uniform change of scale. In the sequel the term **Euclidean reconstruction** will always mean reconstruction up to similarity transformations.

It is well known that it is possible to reconstruct an object up to similarity transformations, given images from calibrated cameras, see [11]. The drawback of this calibrated reconstruction is that the cameras have to be calibrated before the reconstruction can be made.

When making projective reconstruction, the cameras need not be calibrated and reconstruction can be made directly. One drawback with projective reconstructions is that the object is only reconstructed up to projective transformations. Thus the resulting reconstruction might have severe projective distortions. Sometimes this reconstruction is sufficient, but in other cases it would be desirable to make a more precise reconstruction, for example up to similarity transformations. The intrinsic parameters are allowed to change arbitrarily between the different views in projective reconstruction. However, it is unlikely that all the intrinsic parameters can change drastically between the different exposure times if the images are obtained from a sequence with a short time interval between subsequent images. For example, the skew and aspect ratio are often constant when the same camera is used, because they depend primarily on the geometry of the light sensitive elements in the camera.

In some applications it is even possible to assume that the intrinsic parameters are constant for all images in the sequence. In particular, if the images are taken with exactly the same camera, without change of focus or zoom, this is the case. This problem, reconstruction from constant intrinsic parameters, have been considered in [3] and [6], but we will give a different formulation here. The difference between our approach and that of [3], is that the latter only uses the epipolar transformations between image 1 and image i , while we are using a complete representation, using the camera matrices as a starting-point, to impose the Euclidean constraints. In [6] a large minimisation problem is solved using the camera matrices and the reconstruction as parameters. A number of solutions to smaller minimisation problems is used to give a starting-point to the final minimisation. The special case where the camera is known only to rotate between subsequent images have been studied in [5].

The motivation for our approach, starting with a projective reconstruction, is that there exists a lot of good methods to obtain a projective reconstruction from image sequences

*This work has been supported by the Swedish Research Council for Engineering Sciences (TFR), project 95-64-222.

taken by uncalibrated cameras. Once such a reconstruction is obtained, it is often desirable to present it in some Euclidean space. If it is known that the intrinsic parameters have been unchanged during the sequence such a Euclidean reconstruction can be obtained afterwards, without using the image data; only the camera matrices and the reconstruction is used. If accuracy is essential, it is possible to use this result as a starting-point of a nonlinear global optimisation procedure, using all image data, as in the last step of the method outlined in [6]. A similar method (starting with a projective reconstruction) has recently been outlined in [13]. They proceed in two steps, first making affine reconstruction and then Euclidean and need at least 4 images. We proceed directly towards a Euclidean reconstruction and need only 3 images.

2. Projective reconstruction

When some a priori information about the cameras is available, e.g. constant intrinsic parameters, one way to use this information is to make reconstruction in two steps.

1. Make projective reconstruction from uncalibrated cameras, that is without using the extra information.
2. Use the extra information to pick out a class of Euclidean equivalent configurations among all projective equivalent configurations obtained from the uncalibrated reconstruction.

Given a sequence of images $(\mathcal{Y}_i)_{i=1}^m$, where the image \mathcal{Y}_i is represented by the image coordinates of matched points in the sequence. Assume that the image coordinates (x_i, y_i) in image i have been obtained from the corresponding object coordinates (X, Y, Z) and the i :th 3×4 camera matrix P_i using the standard pinhole camera model:

$$\lambda \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = P_i \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \quad (1)$$

where λ_i is the inverse relative depth of the point, see [9]. The goal of a projective reconstruction technique is to calculate both the camera matrices, P_i , and the object coordinates from the image coordinates.

There exists several techniques to make projective reconstruction, see [1], [7], [12], [8]. Any of these methods is applicable for the forthcoming analysis.

When projective reconstruction from a sequence, $(\mathcal{Y}_i)_{i=1}^m$, of images has been made, this reconstruction is usually represented by the coordinates of the object in projective normal form. Another equivalent way to represent this information is to give a representation of the camera matrices, $(P_i)_{i=1}^m$, in the sequence. Thus the camera matrices

obey

$$\mathbf{x}_i \sim P_i \mathbf{X}, \quad i = 1, \dots, m, \quad (2)$$

where \mathbf{x}_i denotes extended image coordinates in image i , \mathbf{X} denotes extended object coordinates in one reconstruction and \sim means equality up to scale. In our approach to Euclidean reconstruction we will use the camera matrices, P_i , as a representation of the projective reconstruction.

Since all projective transformations of the object are possible reconstructions, we can replace \mathbf{X} by $\tilde{\mathbf{X}}$, where $\mathbf{X} \sim A\tilde{\mathbf{X}}$ in (2), for any nonsingular 4×4 matrix A . In this way it is always possible to arrange that the first camera matrix, P_1 , is

$$P_1 = [I | 0].$$

The second camera matrix, P_2 , can then be chosen uniquely up to a choice of the plane at infinity and the global scale, that is

$$P_2 = [Q_2 | t_2],$$

where Q_2 belongs to a 3-parameter family of 3×3 matrices and t_2 is defined up to a scale factor. This can be seen from the fact that when P_1 has been chosen as above, then a coordinate transformation

$$\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \sim A \begin{bmatrix} \tilde{X} \\ \tilde{Y} \\ \tilde{Z} \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ a & b & c & d \end{bmatrix} \begin{bmatrix} \tilde{X} \\ \tilde{Y} \\ \tilde{Z} \\ 1 \end{bmatrix} \quad (3)$$

can be made without changing P_1 . This denotes a 4-parameter family of transformations. The parameter d represents a uniform change of scale and in the sequel we may assume that $d = 1$, which gives a 3-parameter family of transformations. The three remaining parameters (a, b, c) determines the plane at infinity. The following definition is convenient.

Definition 2.1. By a **solution to the uncalibrated reconstruction problem** for a sequence, $(\mathcal{Y}_i)_{i=1}^m$, is meant a sequence of camera matrices $(P_i)_{i=1}^m$, with $P_1 = [I | 0]$, such that (2) is valid. ■

Observe that the P_i :s are unique up to a multiplication by a 4×4 matrix of the same type as A in (3), since the inverse of such a matrix has the same form.

3. Problem formulation

In this section we assume that the camera is modelled by a pinhole camera

$$\lambda \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha_x & s & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} [R_i | -R_i t_i] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \quad (4)$$

with the same α_x , α_y , s , x_0 and y_0 for each camera. Here α_x and α_y represents magnifications in the x and y direction in the optical system, that is the focal length and the aspect ratio, s describes the skew that model nonrectangular light sensitive elements in the CCD-array, and (x_0, y_0) denotes the principal point, that is the orthogonal projection of the focal point onto the image plane.

Suppose that a solution to the uncalibrated reconstruction problem is known, that is a sequence $(P_i)_{i=1}^m$ of camera matrices, obeying (2).

Problem 3.1. *Given a sequence of images taken with a camera, modeled in (4), with constant intrinsic parameters and a solution to the uncalibrated reconstruction problem. What kind of reconstruction can be obtained and how many images are needed?*

According to (4) the camera matrices P_i , obtained from the solution to the uncalibrated reconstruction problem, can be written

$$P_i A \sim K [R_i | -R_i t_i] , \quad (5)$$

where A stands for an appropriate projective change of coordinates in the object and

$$K = \begin{bmatrix} \alpha_x & s & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} . \quad (6)$$

We may assume that $R_1 = I$ and $t_1 = \bar{0}$ in (5), since the origin and the orientation of the first camera can be chosen arbitrarily. This observation means that

$$A = \begin{bmatrix} \alpha_x & s & x_0 & 0 \\ 0 & \alpha_y & y_0 & 0 \\ 0 & 0 & 1 & 0 \\ a & b & c & 1 \end{bmatrix} = \begin{bmatrix} K & 0 \\ [a \ b \ c] & 1 \end{bmatrix} , \quad (7)$$

that is an 8-parameter family of projective transformations. We now reformulate Problem 3.1.

Problem 3.1. *Given a sequence of camera matrices, $(P_i)_{i=1}^m$, with $P_1 = [I | 0]$, find a , b , c and K such that for all i ,*

$$P_i \begin{bmatrix} K & 0 \\ [a \ b \ c] & 1 \end{bmatrix} \sim K [R_i | -R_i t_i] ,$$

where R_i denotes arbitrary rotation matrices and t_i denotes arbitrary vectors.

Observe that given one such solution, $(P_i)_{i=1}^m$, all other possible camera matrices giving the same Euclidean reconstruction can be obtained from right multiplication with

$$\begin{bmatrix} R & -Rt \\ 0 & 1 \end{bmatrix} , \quad (8)$$

where R denotes a 3×3 orthogonal matrix and t a 3×1 vector, that is a 6-parameter family of reconstructions representing rotations, translations and change of global scale.

4. Solution

Looking at the first three columns in (5) gives

$$P_i \tilde{A} \sim K R_i , \quad (9)$$

where \tilde{A} denotes the first three columns of A . It follows that

$$P_i \tilde{A} \tilde{A}^T P_i^T \sim K R_i R_i^T K^T = K K^T , \quad (10)$$

that is

$$P_i \begin{bmatrix} K K^T & [a \ b \ c] K \\ K^T [a \ b \ c]^T & a^2 + b^2 + c^2 \end{bmatrix} P_i^T \sim K K^T . \quad (11)$$

The equations in (11) are related to the so called Kruppa equations, found in [10], [4] and [3].

Definition 4.1. For a given P_i , the constraints on K and $[a \ b \ c]$ in (11) are called the **Kruppa constraints**. ■

We have shown the following theorem.

Theorem 4.1. *Given a sequence, $(\mathcal{Y}_i)_{i=1}^m$, of images taken by cameras with constant intrinsic parameters. Let $(P_i)_{i=1}^m$ be a solution to the uncalibrated reconstruction problem. Then it is possible to make reconstruction up to similarity transformations if and only if the Kruppa constraints are fulfilled, when $m \geq 3$.*

Proof. The Kruppa constraints in (11) contain 5 equations, because the matrices of both members are symmetric, and the homogeneity reduces the number of equations with 1. Furthermore, it can be shown that these equations, in general, are independent, in the sense that they define transversal intersections of hyperplanes. The first camera matrix $P_1 = [I | 0]$ automatically fulfils the Kruppa constraints. Then each camera matrix, P_i , apart from the first one, gives 5 equations in the unknowns, α_x , α_y , s , x_0 , y_0 , a , b and c . This shows that it is in general possible to obtain a unique solution when 3 images are available. ■

5. Example

Consider the three images in Figure 1 of a scene consisting of simple blocks. We have extracted 32 corresponding points (corners) in the images. This was done manually in order to avoid false matches and to illustrate the applicability of our approach. The chosen coordinate system in each image has its origin in the centre, the x -axis in horizontal direction and the y -axis in vertical direction. Finally all coordinates were divided by 400 in order to get coordinates between -1 and 1 , which increases the accuracy of the results.

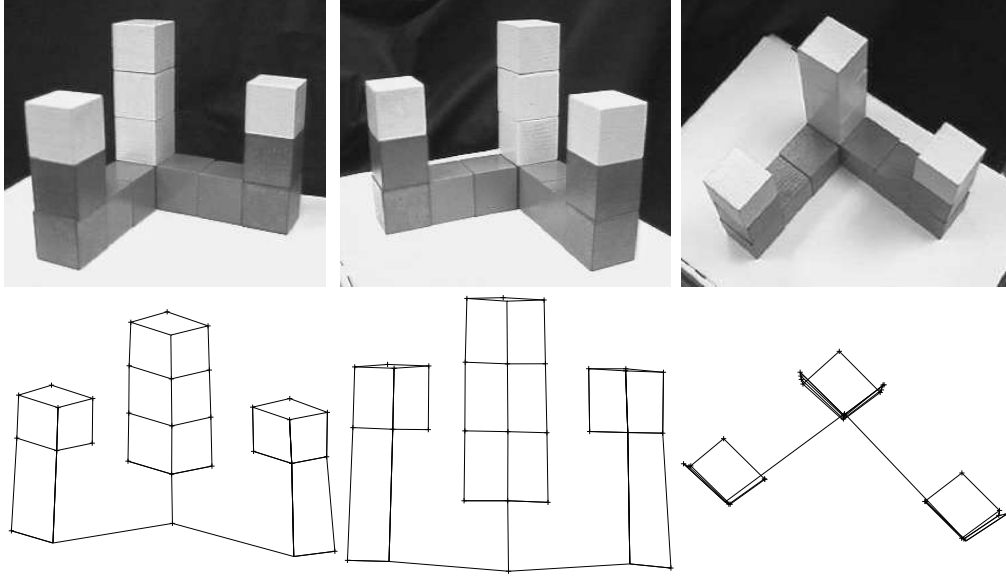


Figure 1. Above: Three images of a simple scene consisting of blocks, taken by the same camera (constant intrinsic parameters). Below: Euclidean reconstruction from three different views.

First a projective reconstruction was obtained using the reduced trilinear tensor, first introduced in [14], see also [8]. The 15 components of the reduced trilinear tensor was calculated linearly directly from the reduced affine coordinates in the images. From the tensor components the kinetic depths and translational vectors were calculated, also linearly, see [8] again. Finally, the camera matrices and the projective reconstruction were obtained from a simple change of coordinates from reduced affine coordinates to original image coordinates. The following camera matrices were obtained:

$$\begin{aligned}
 P1 &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \\
 P2 &= \begin{bmatrix} 0.9866 & -0.3261 & -0.0521 & 0.2769 \\ 0.2207 & 0.9868 & 0.0019 & 0.0216 \\ -0.2658 & 0.0256 & 1.0012 & -0.0308 \end{bmatrix} \\
 P3 &= \begin{bmatrix} 0.8084 & -0.0744 & 0.0020 & 0.0720 \\ -0.0096 & 0.6470 & 0.0414 & -0.2543 \\ -0.0502 & 0.2563 & 0.9547 & -0.0793 \end{bmatrix}
 \end{aligned} \tag{12}$$

This technique is simple to use and all calculations made are linear, i.e. using singular value decomposition of matrices. This means that the noise is not modelled accurately, but we get a fast solution. Since there are one step further to go in order to get a Euclidean reconstruction it is not worth the time to use a more sophisticated method at this stage.

In order to recover the intrinsic parameters and the plane at infinity (11) was used in the form

$$\begin{aligned}
 0 &= f_i(K, a, b, c, \lambda_i) := \\
 &= K K^T - \lambda_i P_i \begin{bmatrix} K K^T & [a \ b \ c] K \\ K^T [a \ b \ c]^T & a^2 + b^2 + c^2 \end{bmatrix} P_i^T .
 \end{aligned} \tag{13}$$

The following goal function was used to formulate an optimisation problem:

$$\begin{aligned}
 F(K, a, b, c, \lambda_1, \lambda_2) &= \\
 &= \|f_1(K, a, b, c, \lambda_1)\|_F + \|f_2(K, a, b, c, \lambda_2)\|_F ,
 \end{aligned} \tag{14}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. As a starting-point the following was used

$$\begin{aligned}
 K_0 &= \begin{bmatrix} 3 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 1 \end{bmatrix} , \\
 (a_0, b_0, c_0) &= (-1, -1, 4), \quad (\lambda_1, \lambda_2) = (1, 1) ,
 \end{aligned} \tag{15}$$

where the minus sign before element (2, 2) in K_0 is needed because the coordinate systems in the images were negatively oriented. The initial values for (a_0, b_0, c_0) and (λ_1, λ_2) were obtained from (13) by inserting K_0 and noticing that (13) is linear in (a_0, b_0, c_0) and (λ_1, λ_2) . F was then

minimised using Nelder-Meads simplex method, chosen because it is a robust method, giving the following result:

$$K = \begin{bmatrix} 2.2394 & 0.0779 & -0.0021 \\ 0 & -2.0699 & 0.0048 \\ 0 & 0 & 1.0000 \end{bmatrix}$$

$$A = \begin{bmatrix} 2.2394 & 0.0779 & -0.0021 & 0 \\ 0 & -2.0699 & 0.0048 & 0 \\ 0 & 0 & 1.0000 & 0 \\ -1.1603 & -0.5113 & 4.3219 & 1 \end{bmatrix} \quad (16)$$

Using A in (16) the camera matrices P_i in (12) can be transformed to new camera matrices, \tilde{P}_i , using $\tilde{P}_i = P_i A$, giving the following camera matrices:

$$\tilde{P}_1 = \begin{bmatrix} 2.2394 & 0.0779 & -0.0021 \\ 0 & -2.0699 & 0.0048 \\ 0 & -0.0000 & 1.0000 \end{bmatrix} \quad (17)$$

$$\tilde{P}_2 = \begin{bmatrix} 1.8881 & 0.6102 & 1.1408 \\ 0.4692 & -2.0365 & 0.0994 \\ -0.5594 & -0.0578 & 0.8686 \end{bmatrix} \quad (18)$$

$$\tilde{P}_3 = \begin{bmatrix} 1.7268 & 0.1800 & 0.3111 \\ 0.2735 & -1.2100 & -1.0545 \\ -0.0203 & -0.4940 & 0.6133 \end{bmatrix} \quad (19)$$

Making rq -decomposition of \tilde{P}_i gives the following r -matrices:

$$\begin{bmatrix} 2.2394 & 0.0779 & -0.0021 \\ 0 & -2.0699 & 0.0048 \\ 0 & 0 & 1.0000 \end{bmatrix}$$

$$\begin{bmatrix} 2.2070 & 0.1149 & -0.0941 \\ 0 & -2.0213 & -0.0545 \\ 0 & 0 & 1.0000 \end{bmatrix} \quad (20)$$

$$\begin{bmatrix} 2.2359 & 0.0529 & 0.1077 \\ 0 & -2.0649 & -0.0879 \\ 0 & 0 & 1.0000 \end{bmatrix}$$

These matrices are close to K in (16). Finally the Euclidean reconstruction is obtained from the projective one using the projective transformation defined by A in (16). The resulting Euclidean reconstruction is shown from three different viewpoints in Figure 1. The result seems to be close to the real solution.

We remark that if one wants as good accuracy as possible, this solution can be taken as a starting-point in a big minimisation problem, where all object points, all rotation matrices, translation vectors and intrinsic parameters are used as parameters and the distance between projected points and detected points in the images are minimised as described in [6]

6. Conclusions

In this paper we have described a situation, where Euclidean reconstruction is possible, when some a priori information of the camera is available.

When the camera has constant intrinsic parameters, it is possible to make Euclidean reconstruction from the Kruppa constraints. This can, in general, be done uniquely when at least 3 images are available.

A property of this reconstruction technique is that the constraints are based on a projective reconstruction, obtained as if the cameras were uncalibrated with different intrinsic parameters. This approach divides the problem into two steps. The first step is projective reconstruction, which is well known. The second step is imposing the constraints that gives Euclidean reconstruction. The assumptions about the cameras are easier to handle using this two-step procedure. Similar ideas of a stratified approach can be found in [2].

References

- [1] O. D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? *ECCV'92, Lecture notes in Computer Science*, 588:563–578, 1992.
- [2] O. D. Faugeras. Stratification of three-dimensional projective, affine and metric representations. *Journal of the Optical Society of America A.*, 12(3):465–484, March 1995.
- [3] O. D. Faugeras, Q.-T. Luong, and S. J. Maybank. Camera Self-Calibration: Theory and Experiments. *ECCV'92, Lecture notes in Computer Science*, 588:321–334, 1992.
- [4] O. D. Faugeras and S. J. Maybank. Motion from Point Matches: Multiplicity of Solutions. *Technical Report*, (1157), 1990.
- [5] R. I. Hartley. Self-Calibration from Multiple Views with a Rotating Camera. *ECCV'94, Lecture notes in Computer Science*, 800:471–478, 1992.
- [6] R. I. Hartley. Euclidean Reconstruction from Uncalibrated Views. *Applications of Invariance in Computer Vision, Lecture notes in Computer Science*, 825:237–256, 1994.
- [7] R. I. Hartley. Projective Reconstruction and Invariants from Multiple Images. *IEEE Trans. Pattern Anal. Machine Intell.*, 16(10):1036–1041, 1994.
- [8] A. Heyden. Reconstruction from Image Sequences by means of Relative Depths. *Proc. ICCV'95*, pages 1058–1063, 1995.
- [9] A. Heyden and K. Åström. A Canonical Framework for Sequences of Images. *Proc. IEEE Workshop on Representation of Visual Scenes*, pages 45–52, 1995.
- [10] E. Kruppa. Zur Ermittlung eines Objektes Zwei Perspektiven mit innerer Orientierung. *Sitz-Ber. Akad. Wiss., Wien, math. naturw. Kl. Abt. IIa*, 122, 1939-1948.
- [11] S. J. Maybank. *Theory of Reconstruction from Image Motion*. Springer-Verlag, Berlin, Heidelberg, New York, 1993.
- [12] R. Mohr and E. Arbogast. It can be done without camera calibration. *Pattern Recognition Letters*, 12(1):39–43, 1991.
- [13] M. Pollefeys, L. Van Gool, and A. Oosterlinck. The modulus constraint: a new constraint for self-calibration. *Proc. ICPR'96*.
- [14] M. E. Spetsakis and J. Aloimonos. A Unified Theory of Structure from Motion. *Proc. DARPA IU Workshop*, pages 271–283, 1990.