MVA2002 IAPR Workshop on Machine Vision Applications, Dec. 11-13, 2002, Nara-ken New Public Hall, Nara, Japan

3—26

Plane-Based Camera Calibration: The Case of Pure Translation

Henrik Malm and Anders Heyden *
Centre for Mathematical Sciences
Lund Institute of Technology

Abstract

In this paper it is shown how to calibrate a camera from a planar calibration object with known metric structure, when the camera (or the calibration object) undergoes pure translational motion. The case of pure translation is a degenerate case in the standard formulation of plane-based camera calibration. However, if it is known that the motion is a pure translation some additional constraints can be formulated which helps us solve some interesting cases.

1 Introduction

In many applications of computer vision, as e.g. in robot vision where a camera is used as a visual sensor for a robot, calibrating the camera is an important step towards the main objective. By calibration we here mean estimating the so-called intrinsic parameters. These include e.g the focal length and the aspect ratio, i.e. the ratio of the width and the height of the imaging elements in the camera.

Traditionally calibration of a camera has been accomplished by the use of a three dimensional object with known metric structure, usually with some sort of grid pattern. See e.g. the book by Faugeras [3] for information on this kind of calibration. In recent years, so-called self-calibration methods has been common, cf. [5, 1, 2]. These methods do not need a special calibration object and rely only on the rigidity of the scene in view. However, in applications requiring high precision measurements in performing their tasks, e.g. in most industrial vision systems, the use of a carefully constructed calibration grid is often most reliable. Since very accurate knowledge of the relative 3D coordinates of points on the object is needed the construction of the grid is greatly simplified if a two dimensional planar object can be used. Zhang [7] and Sturm and Maybank [6] has independently developed principally identical algorithms for calibration from a planar object using two homogeneous linear constraints on the matrix describing the image of the so-called absolute conic. These constraints arises from the estimated homography from the object plane to the image plane at each position of the camera. By solving the linear system built up from these constraints and by a subsequent Cholesky factorization of the obtained matrix, the intrinsic parameters of the camera are obtained.

In this paper we will examine the problem of calibration from images of a planar object when the relative orientation between the object and camera does not change, i.e. the motions of the camera between obtaining the images are pure translations. This case is degenerate in the formulations of Zhang [7] and Sturm and Maybank [6]. Using only the results in these two papers a new image does not give any new constraints on the calibration parameters if the motion has been purely translational. However, using the knowledge that the motion is a pure translation we can set up some additional constraints that fits nicely with the previously developed theory.

By examining the constraints and the degrees of freedom of the calibration problem, some interesting cases can be formulated. For example, using two images where the camera has performed a pure translation between the respective positions, we can calibrate the camera when the length of the translation is unknown and the skew parameter s is set to zero. If we set $s = 0$ and the aspect ratio $\gamma = 1$, the camera can be calibrated when the direction of the translation is unknown but the length is known. One interesting case is when the translation is orthogonal to the calibration object. When calibrating a camera on a robot arm, we can then put the calibration object on the floor and let the robot arm perform a translational motion orthogonal to the floor.

In Section 2 the camera model used in the paper is explained and a brief introduction to the plane-based camera calibration of Zhang and Sturm and Maybank is given. We continue in Section 3 by developing the constraints on the calibration parameters arising from translational motion. Next, in Section 4, the results of some experiments are presented and the noise sensitivity is examined. The paper ends with some conclusions in Section 5.

2 Preliminaries

The perspective pinhole camera is used as our projection model. That is, $\lambda\vec{x} = P\vec{X}$, where $\vec{x} = (x, y, 1)$ is the 2D homogeneous coordinates for the image point, $\vec{X} = (X, Y, Z, 1)$ is the 3D homogeneous coordinates for the object point and λ is an arbitrary scale factor. P is the 3×4 projection matrix that can be decomposed as

$$P = K[\,R\,|\,t\,] = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} [\,R\,|\,t\,]\ . \qquad (1)$$

Here, β denotes the focal length, $\frac{\alpha}{\beta}$ and $\frac{\gamma}{\beta}$ the aspect ratio and the skew, respectively, and (u_0, v_0) the prin-

*Address: Box 118, S-221 00 Lund, SWEDEN. E-mail: henrik,heyden@maths.lth.se

cipal point. These are called the intrinsic parameters, and K is called the intrinsic camera matrix.

In this section we will concentrate on describing the constraints appearing when calibrating a single camera from a planar object. The orientation and origin of the world coordinate system can be chosen so that the plane of the calibration object has $Z = 0$. We then get

$$\lambda \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = K[\, r_1\, r_2\, r_3\, t\,] \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} = K[\, r_1\, r_2\, t\,] \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \ , \quad (2)$$

where r_i is the i:th column of R. In this way the object point is related to the corresponding image point by a homography H:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = H \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}, \quad H = \frac{1}{\lambda} K[\, r_1\, r_2\, t\,] \ . \quad (3)$$

The homography H can be estimated for each image, cf. [7] for details. Let h_i be the i:th column in H. We then have

$$[\, h_1\, h_2\, h_3\,] = \frac{1}{\lambda} K[\, r_1\, r_2\, t\,] \quad (4)$$

and

$$r_1 = \lambda K^{-1} h_1 \quad (5)$$
$$r_2 = \lambda K^{-1} h_2 \ . \quad (6)$$

Introduce $\omega = K^{-T}K^{-1}$. Since r_1 and r_2 are orthonormal the following constraints, involving h_1, h_2 and ω, can be derived from (5) and (6).

$$r_1^T r_1 = \lambda^2 {h_1}^T K^{-T} K^{-1} h_1 = \lambda^2 {h_1}^T \omega h_1 = 1 \quad (7)$$
$$r_2^T r_2 = \lambda^2 {h_2}^T K^{-T} K^{-1} h_2 = \lambda^2 {h_2}^T \omega h_2 = 1 \quad (8)$$
$$r_1^T r_2 = \lambda^2 {h_1}^T K^{-T} K^{-1} h_2 = \lambda^2 {h_1}^T \omega h_2 = 0 \quad (9)$$

These equations could of course be simplified so that the unknown scale factor λ is excluded:

$${h_1}^T \omega h_2 = 0 \quad (10)$$
$${h_1}^T \omega h_1 = {h_2}^T \omega h_2 \quad (11)$$

The matrix ω describes the image of the absolute conic, cf. [4], and we now have two linear constraints on this symmetric matrix from each different image of the plane. By using three different views of the plane we have enough constraints to solve for ω. The intrinsic camera matrix K can then be obtained by Cholesky factorization and matrix inversion.

3 New Constraints

In this paper we concentrate on the case where the motion of the camera is pure translational between the images obtained. Let H' be the estimated homography from the object to the image after the translation. Then

$$H' = \frac{1}{\lambda'} K[\, r_1\, r_2\, t + R\bar{t}\,] = \quad (12)$$
$$= \frac{1}{\lambda'} K[\, r_1\, r_2\, t + \bar{t}_1 r_1 + \bar{t}_2 r_2 + \bar{t}_3 r_3\,] \ , \quad (13)$$

where $\bar{t}$ is the translation vector expressed in the coordinate system of the calibration object, i.e. in the world coordinate system. Note K and R are unchanged from before the translation. Let us have a look at the third columns in the matrices H and H', denoted by h_3 and h'_3, respectively.

$$h_3 = \frac{1}{\lambda} K t \quad (14)$$
$$h'_3 = \frac{1}{\lambda'} K(t + \bar{t}_1 r_1 + \bar{t}_2 r_2 + \bar{t}_3 r_3) \ . \quad (15)$$

Let

$$\hat{h}_3 = \frac{\lambda'}{\lambda} h'_3 - h_3 \ . \quad (16)$$

Then, using (4),

$$\hat{h}_3 = \frac{1}{\lambda} K(\bar{t}_1 r_1 + \bar{t}_2 r_2 + \bar{t}_3 r_3) = \bar{t}_1 h_1 + \bar{t}_2 h_2 + \frac{1}{\lambda} K \bar{t}_3 r_3 \quad (17)$$

and subsequently

$$r_3 = \frac{\lambda}{\bar{t}_3} K^{-1}(\hat{h}_3 - \bar{t}_1 h_1 - \bar{t}_2 h_2) \ . \quad (18)$$

In search for new calibration constraints containing $\hat{h}_3$ and $\bar{t}$, scalar products including r_3, r_1 and r_2 are written down. Taking the scalar product of the orthonormal vectors r_1 and r_3 and using (7) and (9) gives

$$\begin{aligned} r_1^T r_3 &= \frac{\lambda^2}{\bar{t}_3} h_1^T K^{-T} K^{-1}(\hat{h}_3 - \bar{t}_1 h_1 - \bar{t}_2 h_2) \\ &= \frac{\lambda^2}{\bar{t}_3}(h_1^T \omega \hat{h}_3 - \bar{t}_1 h_1^T \omega h_1 - \bar{t}_2 h_1^T \omega h_2) \\ &= \frac{\lambda^2}{\bar{t}_3}(h_1^T \omega \hat{h}_3 - \frac{\bar{t}_1}{\lambda^2}) = 0 \ . \end{aligned}$$

So

$$h_1^T \omega \hat{h}_3 = \frac{\bar{t}_1}{\lambda^2} \ . \quad (19)$$

Similarly, the scalar product of r_2 and r_3 gives

$$h_2^T \omega \hat{h}_3 = \frac{\bar{t}_2}{\lambda^2} \ . \quad (20)$$

It remains to examine the scalar product of r_3 with itself which should be equal to $|r_3| = 1$,

$$\begin{aligned} r_3^T r_3 &= \frac{\lambda^2}{\bar{t}_3^2}(\hat{h}_3^T - \bar{t}_1 h_1^T - \bar{t}_2 h_2^T)\omega(\hat{h}_3 - \bar{t}_1 h_1 - \bar{t}_2 h_2) \\ &= \frac{\lambda^2}{\bar{t}_3^2}(\hat{h}_3^T \omega \hat{h}_3 - \frac{\bar{t}_1^2}{\lambda^2} - \frac{\bar{t}_2^2}{\lambda^2}) = 1 \ . \end{aligned}$$

This gives

$$\hat{h}_3^T \omega \hat{h}_3 = \frac{|\bar{t}|^2}{\lambda^2} \ . \quad (21)$$

By letting $\bar{\omega} = \lambda^2 \omega$, the complete set of constraints arising from two images of a plane when the camera undergoes translation $\bar{t}$ looks like

$$h_1^T \bar{\omega} h_2 = 0 \quad (22)$$
$$h_1^T \bar{\omega} h_1 = 1 \quad (23)$$
$$h_2^T \bar{\omega} h_2 = 1 \quad (24)$$
$$h_1^T \bar{\omega} \hat{h}_3 = \bar{t}_1 |\bar{t}| \quad (25)$$
$$h_2^T \bar{\omega} \hat{h}_3 = \bar{t}_2 |\bar{t}| \quad (26)$$
$$\hat{h}_3^T \bar{\omega} \hat{h}_3 = |\bar{t}|^2 \quad (27)$$

Since the first and the second columns, h_1 and h_2, in the homography H are parallel to the first and the second columns, h'_1 and h'_2, in H' respectively and the scale factors λ and λ' are unknown individually, we get 11 known distinct elements from the two estimated homographies. There are 6 degrees of freedom for the pose of the camera in the first image. Therefore there are $11 - 6 = 5$ degrees of freedom left for the intrinsic parameters and the translational motion. That is, if we want to calculate all the 5 intrinsic parameters we need to now the translation $\bar{t}$ completely. If we e.g. set the skew $s = 0$, the camera can be calibrated when the length of the translation is unknown. If set $s = 0$ and the aspect ratio $\gamma = 1$, we can calibrate the camera when the direction of the translation is unknown but the length is known.

After solving the system (22)-(27) of equations in the unknowns of $\bar{\omega}$, $\bar{t_1}$, $\bar{t_2}$ and $|\bar{t}|$ we perform a Cholesky factorization on $\bar{\omega}$. Inverting and scaling the resulting matrix gives us the intrinsic calibration matrix K. The scale factor λ is easily found since K should have a 1 in the bottom right position.

4 Experiments

Calibration has been performed on computer generated data to get a measurement on how sensitive the calculations are to errors and noise in the image data. Projections of a grid with 9 times 6 points was calculated. The simulated camera was rotated 6° degrees around the x-axis, 30° degrees around the y-axis and −12° degrees around the z-axis in this order in relation to the calibration grid. The distance between the grid point both horizontally and vertically was 5 length units (l.u.) which is to be compared initial position of the camera which was 100 l.u. from the plane in the z direction and 10 l.u. in the y direction. We looked at three different cases:

Case 1 The length of the translation $\bar{t}$ is known, but the direction is unknown. The aspect ratio α is set equal to 1 and the skew s is set to 0.

Case 2 The length $\bar{t}$ is unknown but the direction of the translation is known. The skew s is set to 0.

Case 3 Both the length and the direction of the translation $\bar{t}$ is known. All intrinsic parameters are calculated.

Noise with a standard deviation of 0.5 pixels was added to the projected points. This is to be compared two the artificial image size which was approximately 320 times 240 pixels. The calibration was simulated 100 times with and the mean and the standard deviation of the different parameters were calculated. Two camera positions were used with the length of the translation chosen as $\bar{t} = 15$ and the direction parallel to the vector $(5, 3, 10)$. The exact calibration matrix was

$$K = \begin{bmatrix} 650 & 0 & 160 \\ 0 & 650 & 120 \\ 0 & 0 & 1 \end{bmatrix} . \qquad (28)$$

The resulting mean and the standard deviations for simulations and calculations according to the three different cases was

$$K_1 = \begin{bmatrix} 650.7 \pm 14.6 & 0 & 158.9 \pm 8.4 \\ 0 & 650.7 \pm 14.6 & 121.1 \pm 11.1 \\ 0 & 0 & 1 \end{bmatrix}$$

$$K_2 = \begin{bmatrix} 650.5 \pm 24.4 & 0 & 159.4 \pm 10.5 \\ 0 & 650.3 \pm 23.5 & 121.3 \pm 11.1 \\ 0 & 0 & 1 \end{bmatrix}$$

$$K_3 = \begin{bmatrix} 651.1 \pm 15.7 & -0.03 \pm 2.2 & 159.3 \pm 6.4 \\ 0 & 650.8 \pm 14.6 & 121.1 \pm 9.1 \\ 0 & 0 & 1 \end{bmatrix}$$

The bias in the estimations seems almost negligible. The standard deviations are rather high but considering that only two images were used for the calculations the results are pretty reasonable. The use of more images could probably make the calculations more precise. Of course the sensitivities in the calibration technique varies a lot with the pose of the camera in relation to the image plane. As for general plane-based calibration there exists critical configurations, as discussed in [6], also in this translation based calibration which makes the calculations impossible. We have, however, in the presented experiments tried to simulate a setup which seems to be reasonable for a real case.

5 Conclusions

An extension of plane-based camera calibration to deal with image sequences were the orientation of the camera relative to the calibration plane does not change, has been presented in this paper. The standard constraints on the calibration parameters, obtained from the estimated homographies from the calibration plane to the image plane, has been extended with constraints also containing parameters of the pure translational motion. These constraints form together a homogeneous collection of equations that can that can be used to solve some interesting cases, e.g. the situation when the calibration object is placed on the floor and a robot translates the camera down towards the floor.

Through experiments on computer generated data we have tried to analyze the sensitivities of the current technique. More experiments are definitely needed to evaluate the usefulness of the method. These experiments should include the usage of more images and extensive tests on real data. It should also be possible to apply the calibration technique to a rigid stereo head, where the translational motion parameters then would be the same for both cameras in the head.

6 Acknowledgments*

This work was sponsored by the Swedish Research Council for Engineering Sciences (TFR), project 2000-530, and SSF Junior Individual Grant, project 99-241. Thanks to Björn Johansson for some helpful discussions.

References

[1] O. Faugeras, T. Luong, S. Maybank, "Camera self-calibration: theory and experiments", *Proc. 3rd European Conf. Computer Vision*, Stockholm, Sweden, 471-478, May 1994.

[2] R. Hartley, "An algorithm for self calibration from several views", *Proc. Conf. Computer Vision and Pattern Recognition,* Seattle, Washington, USA, 908-912, June 1994.

[3] O. Faugeras, "Three-Dimensional Computer Vision: A Geometric Viewpoint", MIT Press, 1993.

[4] R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision", Cambridge University Press, 2000.

[5] S. J. Maybank and O. D. Faugeras, "A theory of self-calibration of a moving camera", *Int. J. Computer Vision,* Vol. 8(2), 123-152, August 1992.

[6] P. Sturm, S. Maybank, "On plane-based camera calibration: A general algorithm, singularities, applications", *Proc. Conf. Computer Vision and Patterns Recognition,* Vol 1., 432-437, 1999.

[7] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations", *Proc. Int. Conf. Computer Vision*, Vol 1., 666-679 September 1999.