

Easy Calibration of Pan/Tilt Camera Heads and Online Computation of the Epipolar Correspondences

Stephan Spiess¹ and Mengxiang Li²

¹ Institute of Flexible Automation (INFA)
Vienna University of Technology
A-1040 Wien, Austria
`sp@flexaut.tuwien.ac.at`

² Computational Vision and Active Perception Laboratory (CVAP)
Department of Numerical Analysis and Computing Science
Royal Institute of Technology (KTH), S-100 44 Stockholm, Sweden

Abstract. In this article we describe a calibration procedure of a binocular camera head with two independent pan and tilt axes. For the calibration procedure itself no separate rotations of the respective axes and no fixation is required. To get reliable calibration data just a plane surface with known 2D coordinates of distinguishable target points is needed. This publication describes facts and techniques that are known to the robotic people but mostly unknown to the computer vision society.

1 Introduction

A binocular camera system usually comprises a pair of cameras mounted on a platform or a robot arm. Depending on the design of the system, the degrees of freedom (DOF) may be different. Normally, the head (neck) has two DOF: pan and tilt. Each eye (camera) may have one or two DOF: pan and tilt. In addition eyes may have the freedom of zooming, focusing and aperture control. With such a system one can manipulate its visual parameters in a controlled manner in order to extract useful information about the scene in time and space.

We will focus on the problem of calibrating the relative position and orientation of the two camera coordinate frames which depend on the commanded angular values, the geometry of the system and initially unknown errors of this description. If these errors are identified they can be included in the geometric model and together with the commanded position we obtain an exact kinematic description of the system. We can assume for simplicity that the intrinsic parameters of the cameras are known, although they may vary with different zoom or focus settings. The interdependence of these parameters can be recorded in look-up-tables [7], thus our assumption does not restrict the general case. The knowledge of the intrinsic parameters, gained by camera calibration, and the extrinsic parameters, gained by a kinematic calibration, enables the fast and

reliable computation of the essential and fundamental matrix without the need of point matching.

The presented work is related in the same way to the field of robot calibration as it is to the field of camera calibration. In both disciplines the visual system or the robot is calibrated in a world coordinate system with measurement points with accurately known position. Besides a world-base transform (the extrinsic parameters in the case of camera calibration) a least squares estimation yields the parameters of the object model such that an error function is minimized. From robot calibration it is known that the results of this estimation are only reliable if a wide range of the working space is covered in the calibration procedure [12]. For a camera with known intrinsic parameters the recovery of the pose with respect to a world coordinate system requires that at least four distinguishable non-collinear points with known 3D coordinates are in the view. The recovery gets more reliable if the points cover a larger part of the image. For the calibration of a "robot" with two axes and limited link length, like it is the case with a pan-tilt camera device, we encounter a trade-off dilemma that is described in the following. If the calibration object is very far away, then the axes of the device can move considerably and still have the object in view. But in that case we can not rely on the camera pose recovery, because the image of the object will be small. If, on the other hand, the object is very close the pose recovery gets reliable but we can only cover little of the working space of the axes of rotation, the object has to be kept in view (see figures 1, 2). A way out of this dilemma would be the use of a large calibration device with high accuracy, which is very expensive.

Li, Brady, and Wiles [5] attack the problem of binocular calibration by using point matchings at certain camera poses. At those poses the fundamental matrix is computed by an eight points algorithm and afterwards corrections are applied. Their method is only applicable to mechanisms with common elevation, that means with a common tilt axis.

Young et al. [9] describe a method where the 3D motion between different robot positions is recovered by analysing the image contents. This motion is compared to a nominal motion of separately moved axes of the robot. For a binocular camera head this approach is trapped in the described dilemma.

Davidson, Reid, and Murray [1] fixate points of known 2D coordinates to recover a plane-plane homography and calibrate their binocular head. However fixation means closed loop control of the axes of the platform and that means avoidable effort. Furthermore their approach requires that the axes of rotation are perpendicular to each other and aligned with the camera axes.

Since Zhuang, Wang, and Roth [16] use minimal models for robot and camera, parameter identification can be done simultaneously. No separate motion of axes is necessary. Every measured image point simply contributes two equations to a large system of nonlinear equations. The use of a calibration board of 0.002mm accuracy makes it difficult and expensive to confirm their results. The main drawback of their publication is the unnecessarily complicated mathematics.

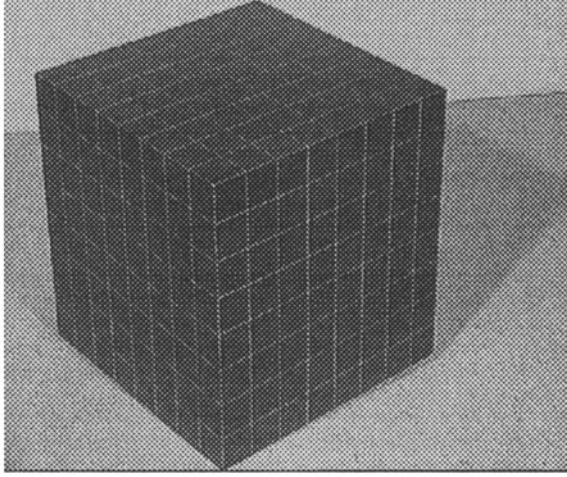


Fig. 1. A calibration object too close. A very small range of action of the cameras decreases the reliability of the kinematic calibration.

Many authors solve the problem of hand-eye calibration by solving homogeneous matrix equations of the form $AX = XB$, see [3] for a brief overview. These approaches need also the 3D recovery and suffer therefore from the mentioned dilemma. We think that solving the problem in Euclidean space is generally a bad idea, because then we have to cope with the inaccuracies introduced by pose estimation.

The subsequently presented approach is similarly to Zhuang's method capable of simultaneous calibration of camera and platform, no fixation is needed, no separate movements of axes are necessary. The mathematics used is easily understandable, the MATLAB-source code is provided [10]. Experimental results show that for the recovery of the epipolar geometry satisfactory results can be achieved without any sophisticated calibration device.

2 Problem Formulation

2.1 A Kinematic Model

Since we are interested in the calibration of the relative transformation of the two cameras, the neck does not influence our calibration result and is therefore not considered here. For the sake of simplicity we place the base coordinate system of the camera head such that its origin lies in the middle of the nominal intersections of the respective pan and tilt axes, the z axis is vertical and the y axis points towards the scene. Let A'_w denote the 6 DOF world-base transform

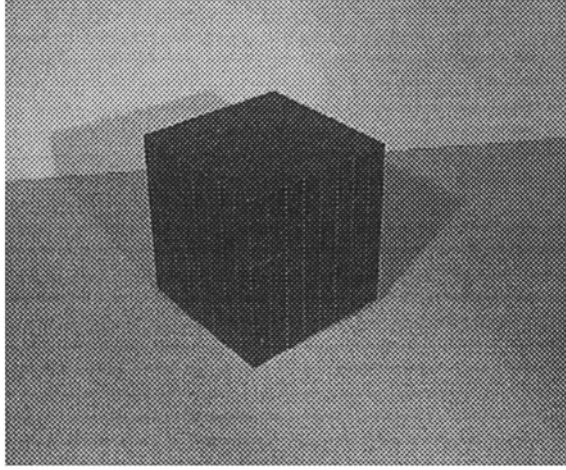


Fig. 2. A large distance enables a larger range of action but the pose recovery gets unreliable.

relating the coordinate system of a calibration plane to the base of the binocular head of the form

$$\mathbf{A}'_w = \text{Transl}(x, y, z) \text{Rot}(x, \alpha) \text{Rot}(y, \beta) \text{Rot}(z, \theta). \quad (1)$$

In the two kinematic chains first we rotate about the pan axes (vergence) then about the tilt axes (elevation). Enhancing \mathbf{A}'_w with (initially unknown) error parameters and translations to the respective pan axes yields for the left base coordinate system

$$\mathbf{A}_{w,l} = \mathbf{A}'_w \text{Transl}(e/2 + \Delta x_l, \Delta y_l, \Delta z_l) \text{Rot}(x, \Delta \alpha_l) \text{Rot}(y, \Delta \beta_l) \text{Rot}(z, \Delta \theta_l). \quad (2)$$

Where e denotes the eye to eye distance of the camera head. For the right base coordinate system replace l by r and e by $-e$. Now the z axis of $\mathbf{A}_{w,l}$ is aligned with the left pan axis and similarly the z axis of $\mathbf{A}_{w,r}$ is aligned with the right pan axis. We follow the two different kinematic chains to the respective cameras. Using the Denavit-Hartenberg model [8,14] we obtain

$$\mathbf{A}_{pan,l} = \text{Rot}(z, \phi_l + \pi/2) \text{Transl}(\Delta a_l, 0, 0) \text{Rot}(x, \Delta \alpha_l - \pi/2) \quad (3)$$

as the transformation relating the pan and tilt axis. A similar equation is obtained for the right kinematic chain. Here ϕ_l denotes the variable pan angle. In the model we assume that zero pan angles result in a straight forward look with approximately parallel optical axes. This is no restriction to the general case, because adding a appropriate constant offsets to the angular values will always

lead to such a system. Generally the Denavit-Hartenberg model has four (error) parameters, but here only two, the normal distance Δa_l of the two axes and the twist angle $\Delta \alpha_l$ appear. That is because the unused parameters are redundant with Δz_l and $\Delta \theta_l$.

The rotations about the tilt axes and the transformations into the camera frames have again 6 DOF. Again left and right transformations are the same.

$$\begin{aligned} A_{\text{tilt},l} = & \text{Rot}(z, \psi_l + \pi) \text{Rot}(y, \pi/2) \text{Rot}(z, \Delta \theta_l) \text{Rot}(y, \Delta \beta_l) \\ & \text{Rot}(x, \Delta \alpha_l) \text{Transl}(\Delta x, \Delta y, \Delta y) \end{aligned} \quad (4)$$

With ψ_l we denoted the left tilt angle.

All together we have 28 error parameters, which corresponds to the rule that a serial manipulator with rotational axes consisting of rigid links has $4N + 6$ parameters [11,16], where N is the number of degrees of freedom and our head is considered as two 2 DOF robots.

2.2 Camera Model

The left camera frame is described by the homogeneous transformation

$$\begin{aligned} C_l(\phi_l, \psi_l, \text{param}) = & A_{w,l} A_{\text{pan},l} A_{\text{tilt},l} \\ C_l = & \begin{pmatrix} r_{11} & r_{12} & r_{13} & X_0 \\ r_{21} & r_{22} & r_{23} & Y_0 \\ r_{31} & r_{32} & r_{33} & Z_0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{aligned} \quad (5)$$

In our case the positive z axis of the camera frame directs away from the scene. In this case we derive the left image coordinates $(x_{\text{mod}}, y_{\text{mod}})^T$ as the projection of a 3D point $(X, Y, Z)^T$ by equations 6, 7 [4].

$$x_{\text{mod}} = x_0 + (x - x_0)d_r - f_x \frac{r_{11}\hat{X} + r_{21}\hat{Y} + r_{31}\hat{Z}}{r_{13}\hat{X} + r_{23}\hat{Y} + r_{33}\hat{Z}} \quad (6)$$

$$y_{\text{mod}} = y_0 + (y - y_0)d_r - f_y \frac{r_{12}\hat{X} + r_{22}\hat{Y} + r_{32}\hat{Z}}{r_{13}\hat{X} + r_{23}\hat{Y} + r_{33}\hat{Z}} \quad (7)$$

where $f_x = f/s_x$, $f_y = f/s_y$, f is the focal length and s_x, s_y are the pixel lengths in x and y directions. We use the abbreviations $\hat{X} = X - X_0$, $\hat{Y} = Y - Y_0$, $\hat{Z} = Z - Z_0$. The coordinates of the principal point are (x_0, y_0) and d_r denotes the radial distortion of the lens with

$$d_r = a_1 \left(\frac{r^2}{R_0^2} - 1 \right) + a_2 \left(\frac{r^4}{R_0^4} - 1 \right) \quad (8)$$

$$r = \sqrt{(x - x_0)^2 + (y - y_0)^2}. \quad (9)$$

The constant R_0 is set to half of the image border length. This computation of the distortion exhibits better numerical stability than a conventional model [6,15]. The intrinsic parameters $f_x, f_y, x_0, y_0, a_1, a_2$ are not accurately known and therefore for each camera the number of parameters is now $14+6=20$.

A similar equation holds for the right image coordinates.

2.3 Cost Function

Every measured target point with known 3D coordinates and viewed with the left camera produces an image point $(x_{meas}, y_{meas})^T$. Together with the currently commanded angular variables ϕ_l, ψ_l this point contributes two equations

$$x_{mod} = x_{meas} \quad (10)$$

$$y_{mod} = y_{meas} \quad (11)$$

to a system of nonlinear equations. The equations depend on the unknown parameters and the known joint variables and image coordinates. We can solve for the unknown parameters by a least squares estimation if we have at least 10 different points recorded with different commanded poses.

In practice we have to use an overdetermined system to gain reliable results. Due to measurement noise and limits of the model, equations 10, 11 will never be fulfilled accurately. Thus the cost function for m measurement points can be written as

$$\sum_{i=1}^m ((x_{mod,i} - x_{meas,i})^2 + (y_{mod,i} - y_{meas,i})^2). \quad (12)$$

The minimization can be done with the Gauss-Newton method which converges very fast in the neighbourhood of a solution. This method is the iterated application of a locally linearised balanced adjustment [4]. The iterations are stopped when the amount of adjustment on each of the parameters is below a prescribed threshold. The Jacobian can be approximated using finite differences. A start estimation for the intrinsic camera parameters can be obtained from the camera specifications. The initial error parameters of the pan/tilt unit can be set to zero.

The world-to-base transformation (eq. 1) has to be measured manually with measurement tape to obtain a feasible start value for the parameter estimation. Since the accuracy of this measurement is mostly far worse than the initial estimation of the other parameters, an estimation procedure for the world-to-base parameters (eq. 2) with the other parameters fixed to their default values should be done before doing the adjustment with all parameters. After obtaining the final parameters the parameters of the world-to-base transform are worthless for the epipolar problem.

Although like [16] the described technique is capable of simultaneous calibration of all parameters we recommend in the case of imperfect calibration data the splitting into two stages. Stage one being the camera calibration, thus the identification of the intrinsic parameters, and stage two being the kinematic calibration with fixed intrinsic parameters.

3 Epipolar Geometry

Let $D = C_l^{-1} C_r$ with elements d_{ij} be the transformation from the left camera frame to the right. Then the *essential matrix* [13] is given by

$$E = TD_{3 \times 3} \quad (13)$$

where $D_{3 \times 3}$ is the upper left 3×3 submatrix of D and

$$T = \begin{pmatrix} 0 & d_{34} & -d_{24} \\ -d_{34} & 0 & d_{14} \\ d_{24} & -d_{14} & 0 \end{pmatrix}. \quad (14)$$

If we neglect distortion the transformation of a point $(x, y, 1)^T$ given in homogeneous pixel coordinates into a metric representation is given by a multiplication with a matrix [2,7]

$$P^{-1} = \begin{pmatrix} -\frac{1}{f_x} & 0 & \frac{x_0}{f_x} \\ 0 & -\frac{1}{f_y} & \frac{y_0}{f_y} \\ 0 & 0 & 1 \end{pmatrix}. \quad (15)$$

Now the *fundamental matrix* can be derived as

$$F = (P_l^{-1})^T E P_r^{-1}. \quad (16)$$

For any point $(x_l, y_l, 1)^T$ in the left image given in homogeneous pixel coordinates F yields the epipolar line $Ax + By + C = 0$ in the right image as follows

$$(A, B, C) = (x_l, y_l, 1)F. \quad (17)$$

Note that x and y are given in pixel coordinates. For the reverse direction we have to use F^T . Further note that this nice property is only valid for a distortion-free lens. That means in practice we have to correct the distortion of a measured image point $(x, y)^T$ by an appropriate shift about $(-d_r(x - x_0), -d_r(y - y_0))^T$.

4 Experimental Results

In all simulations and experiments we first calibrated the cameras with the calibration cube shown in figure 1 and fixed the intrinsic parameters for the kinematic calibration. This was done because the accuracy of the used calibration object for the camera calibration was several times higher than the accuracy of the 3D coordinates of the target points used for kinematic calibration. The calibration worked well, despite of the fact that we measured the 2D points of measurement targets manually, details can be found in [17] and can be extracted from [10].

5 Summary

We presented a minimal, but complete offset model for the kinematics of a binocular head with separated eye pan and tilt and a, for the calibration procedure, fixed neck. We demonstrated that the online calculation of the essential and fundamental matrix can be done to a, for the sake of computer vision, satisfactory precision with easy mathematics and without expensive calibration mechanisms. The accuracy of the obtained results is in the range of that of competing methods but accomplished with very little effort.

References

1. A. Davidson, I. Reid, D. Murray: "The active camera as a projective pointing device." *Proc. of the British Machine Vision Conf. '95*, Birmingham 1995, pp. 453 – 462.
2. O. Faugeras: *Three-Dimensional Computer Vision. A Geometric Viewpoint*. Cambridge, MA, MIT Press, 1993.
3. R. Horaud, F. Dornaika: "Hand-Eye Calibration." *The Int. Journal of Robotics Research*, Vol. 14, No. 3, June 1995, pp. 195 – 210.
4. K. Kraus: *Photogrammetry*. Dümmler Verlag Bonn, 1993.
5. F. Li, M. Brady, C. Wiles: "Fast Computation of the Fundamental Matrix for an Active Stereo Vision System." *Computer Vision – ECCV'96*. Lecture Notes in Computer Science, Vol. 1064, ed. by B. Buxton and R. Cipolla, Springer-Verlag, 1996.
6. M. Li: "Camera Calibration of the KTH Head-Eye System." Technical Report ISRN KTH/NA/P-9407-SE, Dep. of Numerical Analysis and Computing Science, Royal Institute of Technology, Stockholm 1996.
7. M. Li: "Camera Calibration of a Head-Eye System for Active Vision." *Computer Vision – ECCV'94*, Lecture Notes in Computer Science, Vol. 800, ed. by J.-O. Eklundh, Springer-Verlag 1994, pp. 543 – 554.
8. R. Paul: *Robot Manipulators: Mathematics, Programming, and Control*. Cambridge, MA, MIT Press, 1981.
9. G.-S. Young, T.-H. Hong, M. Herman, J. Yang: "Kinematic Calibration of an Active Camera System." *Proc. of IEEE Conf. CV '92*, pp. 748 – 751.
10. MATLAB source code for the calibration; anonymous ftp to [antigone.infa.tuwien.ac.at](ftp://antigone.infa.tuwien.ac.at/File/pub/BinoCal.tar.Z). File /pub/BinoCal.tar.Z
11. B. Mooring, Z. Roth, M. Driels: *Fundamentals of Manipulator Calibration*. New York, Wiley & Sons, 1991.
12. K. Schröder: "Theory of kinematic modelling and numerical procedures for robot calibration." *Robot Calibration*, edited by R. Bernhardt and S. Albright, Chapman & Hall, London 1993.
13. R. Tsai, T. Huang: "Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Surfaces." *IEEE PAMI*, Vol. PAMI-6, No. 1, Jan. 1984.
14. M. Vincze, K.M. Filz, H. Gander, J.P. Prenninger, G. Zeichen: "A Systematic Approach to Model Arbitrary Non Geometric Kinematic Errors." *Advances in Robot Kinematics and Computational Geometry*, edited by J. Lenarcic and B. Ravani, Kluwer Academic Publishers, 1994, pp. 129 – 138.
15. J. Weng, P. Cohen, M. Herniou: "Camera Calibration with Distortion Models and Accuracy Evaluation." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14, No. 10, Oct. 1992.
16. H. Zhuang, K. Wang, Z. Roth: "Simultaneous Calibration of a Robot and a Hand-Mounted Camera". *IEEE Transactions on Robotics and Automation*, Vol. 11, No. 5, Oct. 1995, pp. 649 – 660.
17. S. Spiess, M. Li: "Kinematic Calibration of an Active Binocular Head for Online Computation of the Epipolar Geometry", *Technical Report of the Computer Vision and Active Perception Laboratory*, Royal Institute of Technology, Stockholm, Sweden. See www.bion.kth.se/abstracts/cvap205.html