IEEE TRANSACTIONS ON

ROBOTICS AND AUTOMATION

A PUBLICATION OF THE IEEE ROBOTICS AND AUTOMATION SOCIETY

FEBRUARY 1998 VOLUME 14 NUMBER 1 IRAUEZ (ISSN 1042–296X)

PAPERS

SHORT PAPERS

CALLS FOR PAPERS

IEEE TRANSACTIONS ON ROBOTICS AND AUTOMATION

IEEE TRANSACTIONS ON ROBOTICS AND AUTOMATION is published by the IEEE Robotics and Automation Society. All members of the IEEE are eligible for membership in the Society and will receive this TRANSACTIONS upon payment of the annual Society membership fee of $19.00 plus an annual subscription fee of $6.00. For information on joining, write to the IEEE Service Center at the address below. *Member copies of Transactions/Journals are for personal use only.*

(Continued on inside back cover)
 (Continued on inside back cover)

IEEE TRANSACTIONS ON ROBOTICS AND AUTOMATION (ISSN 1042–296X) is published bimonthly by The Institute of Electrical and Electronics Engineers, Inc. Responsibility for the contents rests upon the authors and not upon the IEEE, the Society/Council, or its members. **IEEE Corporate Office:** 345 East 47 Street, New York, NY 10017-2394. **IEEE Operations Center:** 445 Hoes Lane, P.O. Box 1331, Piscataway, NJ 08855-1331. **NJ Telephone:** 732-981-0060. **Price/Publication Information:** Individual copies: IEEE Members $10.00 (first copy only), nonmembers $20.00 per copy. (Note: Add $4.00 postage and handling charge to any order from $1.00 to $50.00, including prepaid orders.) Member and nonmember subscription prices available upon request. Available in microfiche and microfilm. Periodicals Postage Paid at New York, NY and at additional mailing offices. **Postmaster:** Send address changes to IEEE TRANSACTIONS ON ROBOTICS AND AUTOMATION, IEEE, 445 Hoes Lane, P.O. Box 1331, Piscataway, NJ 08855-1331. GST Registration No. 125634188. Printed in U.S.A.

[17] G. S. Young, T. H. Hong, M. Herman, and J. C. S. Yang, "Kinematic calibration of an active camera system," in *Proc. IEEE Conf. Comput. Vision Pattern Recognition*, 1992, pp. 748–751.

[18] H. Zhuang and Z. S. Roth, "A linear solution to the kinematic parameter identification of robot manipulators," *IEEE Trans. Robot. Automat.*, vol. 9, pp. 174–185, Apr. 1993.

Active Self-Calibration of Robotic Eyes and Hand–Eye Relationships with Model Identification

Guo-Qing Wei, Klaus Arbter, and Gerd Hirzinger

***Abstract*— In this short paper, we first review research results of camera self-calibration achieved in photogrammetry, robotics and computer vision. Then we propose a method for self-calibration of robotic hand cameras by means of active motion. Through tracking a set of world points of unknown coordinates during robot motion, the internal parameters of the cameras (including distortions), the mounting parameters as well as the coordinates of the world points are estimated. The approach is fully autonomous, in that no initial guesses of the unknown parameters are to be provided from the outside by humans for the solution of a set of nonlinear equations. Sufficient conditions for a unique solution are derived in terms of controlled motion sequences. Methods to improve accuracy and robustness are proposed by means of best model identification and motion planning. Experimental results in both a simulated and a real environments are reported.**

***Index Terms*— Active motion, hand-cameras, hand–eye calibration, model identification, motion planning, self-calibration, unique solution.**

I. Introduction

In order to use cameras for estimating robot motion for object manipulation, it is usually necessary to do the following three calibrations: camera calibration, hand–eye calibration, and robot calibration. In this paper, we address the first two problems, assuming the third, i.e., robot calibration, has been done.

For camera calibration, the basic theory has been developed in the field of photogrammetry [17]. Calibration approaches used nowadays can be generally categorized into two classes: test-field calibration and self-calibration. Test-field calibration determines the camera internal and external parameters from images of a set of control points whose three-dimensional (3-D) coordinates are known in a world coordinate system [5], [17], [22], [26], [29]. Since it is difficult to fabricate and also maintain for a long period a highly accurate control field, the self-calibration methods were developed, which estimate not only the camera parameters but also the coordinates of the control points, based on multiple images of the same control field acquired at different camera stations. The collinearity constraint [11] and the coplanarity constraint [12], [17, pp. 259–260] were the most popular equations used in self-calibration, the later of which eliminates the world point coordinates. A major difficulty with the photogrammetric self-calibration approach is that good initial guesses of the unknown parameters have to be provided from the outside by humans for an iterative procedure to converge to the correct solution. In Computer Vision, Faugeras *et al.* [6] recently proposed a camera self-calibration approach which involves only camera internal parameters. "The method, however, was found to be noise sensitive and also computationally intensive, in spite of some improvements [14] in the formulation" (Luong and Faugeras [15]) The above self-calibration methods are passive in the sense that no knowledge about the camera motion is assumed. Active methods employ knowledge about camera motion, in terms of either a movable mechanical devices or a robot [1], [3], [16]. Among the self-calibration methods, either passive or active, developed in computer vision, almost all of them do not consider lens distortions (and some of the formulations are only applicable to the no-distortion case, e.g., [6], [16]). Besides, most of the above methods do not address the problem of how to automate the process of getting initial values of the unknown parameters in nonlinear iteration.

For hand–eye calibration, two basic hand–eye configurations were used in the robotics literature: static cameras and on-hand cameras. In the case of stationary cameras, hand–eye calibration was performed by moving the hand and tracking in the image, a single point (e.g., a light emitting diode, LED) on the gripper [2], [10], [13]. When the coordinates of the LED with respect to the hand coordinate system are known, hand–eye calibration is equivalent to camera calibration [10], [13], where the control points are generated by hand movements. These approaches, however, cannot deal with multiple points with unknown relative positions. The capability of utilizing multiple points is important in improving robustness and reducing the number of robot motions required. For the camera-on-hand configuration, earlier work on hand–eye calibration assume that the cameras have been calibrated in advance [21], [24]. By moving the robot hand to at least three stations, the hand–eye calibration problem was shown to be equivalent to solving equations of the form $A_iX = XB_i$ [21], [24], [25], [30]. In this approach, the robot motion matrix A_i is calculated from the known robotic kinematics; while the camera motion matrix B_i is determined by camera extrinsic calibration in terms of a known control field [24]. Recently, Zhuang *et al.* [31] calibrated a hand-camera, the hand–eye transformation, and the robot together by using a known control field, assuming that the image center and scale factor are known *a priori*. It was suggested in [31] that gauging devices be used to manually measure some parameters as the initial values in the solution of a set of nonlinear equations.

In this paper, we propose a complete autonomous approach for self-calibration of hand-cameras and hand–eye relationships.

The proposed approach has the following features

1) No metric control points are used. The coordinates of the calibration points are determined by the calibration itself. Besides, the number of object points can be as few as one *and* as many as one wants.
2) The initial values of all the unknowns for starting the iteration in the solution of a system of nonlinear equations for the calibration are found automatically, all in closed forms, by the calibration method itself.
3) Lens distortions are considered.
4) The method does not rely on any system knowledge or any pre-calibration or partial calibration of the camera's internal or external parameters.

As another salient feature, the method identifies the best lens distortion model and plans the robot motion such that both robustness and accuracy can be improved.

Manuscript received November 22, 1996. This paper was recommended for publication by Editor A. Goldenberg upon evaluation of the reviewers' comments.

The authors are with the Institute of Robotics and System Dynamics, German Aerospace Research Establishment, Oberpfaffenhofen 82234, Germany (e-mail: Guo-Qing.Wei@dlr.de, Klaus.Arbter@dlr.de, and Gerd.Hirzinger@dlr.de).

Publisher Item Identifier S 1042-296X(98)01432-3.

The paper is organized as follows. In Section II, we present the method of active camera calibration. In Section III, we address the problem of how to identify the best distortion model and to design robot motion. In Section IV, the method is tested and compared with a modified Tsai's algorithm. Conclusions are given in Section V.

II. Self-Calibration of a Hand–Eye System

Suppose a camera is rigidly mounted on a robot gripper. We denote the camera coordinate system by $\langle c \rangle$: $X_c - Y_c - Z_c$, the gripper coordinate system by $\langle g \rangle$: $X_g - Y_g - Z_g$. The transformation from the camera coordinate system to the hand (gripper) coordinate system is represented by the rotation R_{cg} and translation $\mathbf{t}_{cg}$ as

$$\begin{pmatrix} X_g \\ Y_g \\ Z_g \end{pmatrix} = R_{cg} \begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix} + \mathbf{t}_{cg} \tag{1}$$

or in homogeneous form as

$$\begin{pmatrix} X_g \\ Y_g \\ Z_g \\ 1 \end{pmatrix} = H_{cg} \begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix} \tag{2}$$

where

$$H_{cg} = \left[\begin{array}{ccc|c} & R_{cg} & & \mathbf{t}_{cg} \\ - & --- & - & -- \\ 0 & 0 & 0 & 1 \end{array} \right] \triangleq H(R_{cg}, \mathbf{t}_{cg}). \tag{3}$$

Assume there is a set of world points $\{P_i\}$, whose coordinates in an initial camera coordinate system $\langle c_0 \rangle$ are represented by (X_{i0}, Y_{i0}, Z_{i0}), $i = 1, 2, \cdots, N$, where N is the number of points. If we move the robot hand to M different stations $\langle g_j \rangle$, $j = 0, 1, 2, \cdots, M$, where $\langle g_0 \rangle$ stands for initial hand station, then the coordinates (X_{ij}, Y_{ij}, Z_{ij}) of the ith point at the jth camera station $\langle c_j \rangle$ are

$$\begin{pmatrix} X_{ij} \\ Y_{ij} \\ Z_{ij} \\ 1 \end{pmatrix} = H_{c0j} \begin{pmatrix} X_{i0} \\ Y_{i0} \\ Z_{i0} \\ 1 \end{pmatrix} \tag{4}$$

where H_{c0j} is the homogeneous transformation matrix from $\langle c_0 \rangle$ to $\langle c_j \rangle$. If we use $H_{g0j} = H(R_{g0j}, \mathbf{t}_{g0j})$ to denote the robot motion matrix from $\langle g_0 \rangle$ to $\langle g_j \rangle$, with R_{g0j} and $\mathbf{t}_{g0j}$ being the rotation and translation components, respectively, then H_{c0j} can be computed as

$$H_{c0j} = H_{cg}^{-1} H_{g0j} H_{cg}. \tag{5}$$

It can be easily shown that the rotation and translation components of H_{c0j} are

$$R_{c0j} = R_{cg}^T R_{g0j} R_{cg} \tag{6}$$

$$\mathbf{t}_{c0j} = R_{cg}^T (R_{cg} - I)\mathbf{t}_{cg} + R_{cg}^T \mathbf{t}_{g0j}. \tag{7}$$

Fig. 1 illustrates the chain of transformations.

Suppose (u_{ij}, v_{ij}) are the measured image coordinates of the ith world point P_i at the jth camera station $\langle c_j \rangle$. Then the following perspective equations can be obtained as the measurement equations:

$$\overline{u}_{ij} = f_x \frac{X_{ij}}{Z_{ij}},\ \overline{v}_{ij} = f_y \frac{Y_{ij}}{Z_{ij}}; \qquad i = 1, \cdots, N; \quad j = 0, 1, \cdots, M \tag{8}$$

where f_x and f_y are the effective focal lengths in the x and y directions of the image plane, respectively; and $(\overline{u}_{ij}, \overline{v}_{ij})$ are the distortion-compensated frame buffer coordinates from the measured ones (u_{ij}, v_{ij}), according to a distortion model containing both radial and tangential components[1] [4]:

$$\overline{u}_i = u_i - u_0 + p_1(3x^2 + y^2/k^2) + 2p_2 y^2 + x(1 + k_1 r^2) \tag{9}$$

$$\overline{v}_i = v_i - v_0 + 2p_1 xy + p_2(x^2k^2 + 3y^2) + y(1 + k_1 r^2) \tag{10}$$

with

$$x = u_i - u_0;\ y = v_i - v_0 \tag{11}$$

and

$$r^2 = k^2x^2 + y^2 \tag{12}$$

where we have, for simplicity, omitted the subscript j; the parameters (u_0, v_0) are the coordinates of the image center; p_1, p_2, and k_1 are the tangential and radial distortion coefficients, respectively; $k = f_y/f_x$ is the image plane scale factor, which is the ratio of the between-pixel distances d_x and d_y in the x- and y- directions of the camera CCD plane.

By inserting (3)–(5) into (8), we can now state the camera self-calibration problem as: determine from the measurement equations (8), the world coordinates (X_{i0}, Y_{i0}, X_{i0}), $i = 1, \cdots, N$, the camera internal parameters $u_0, v_0, f_x, f_y, p_1, p_2, k_1$, and the hand–eye configuration parameters (external parameters) R_{cg} and $\mathbf{t}_{cg}$, based on the robot motion parameters $\{H_{g0j}\}$ and the image coordinate measurements $\{(u_{ij}, v_{ij})\}$. Here, we use the roll–pitch–yaw angles (α, β, γ) to parameterize the hand–eye rotation matrix R_{cg}.

In the above formulation, assumptions are made of the known robot motions $\{H_{g0j}\}$. This is an assumption adopted in most previous work on hand–eye calibration, e.g., Tsai and Lenz [24], Shiu and Ahmad [21], to name a few. This assumption is not so hard to meet for most industrial robots, since we do not require a high *absolute* positioning precision. Rather, it is the relative motion that is used in the calibration. Often, it is helpful to use the *measured* amounts of relative motions (from joint angles) rather than the *commanded* ones in the calibration equations, because of error accumulation in the execution of a specified motion sequence.

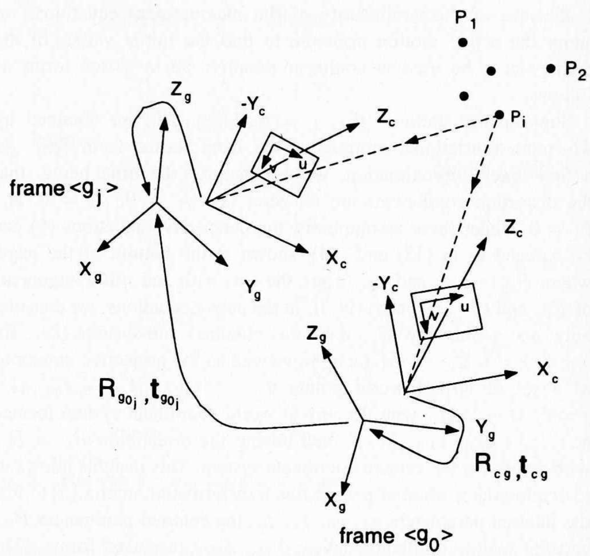

Fig. 1. The hand and eye coordinate systems and hand motion.

[1] Notice that the distortion model is expressed directly in the frame buffer instead of in the CCD plane, so that a scale factor k is involved, as compared with that in [4].

Because of the nonlinearity of the measurement equations, we adopt the active motion principle to find the initial values of the unknowns to be used in nonlinear iteration, all in closed forms as follows.

Suppose the stations $\langle g_j \rangle$, $j = 1, 2, \cdots, M_t$ are obtained by M_t pure translational motions of the hand started from $\langle g_0 \rangle$. As a first order approximation, we assume, for the time being, that the distortion coefficients are all zero, i.e., $p_1 = 0$, $p_2 = 0$, and $k_1 = 0$. Under these assumptions, the perspective equations (8) can be reduced as in (13) and (14), shown at the bottom of the page, where $R_{cg}(m, n)$ and $t_{g0j,n}$ are the (m, n)th and nth components of R_{cg} and $\mathbf{t}_{g0j}$, respectively. If, in the above equations, we consider only one point, say P_1, then the obtained sub-system (i.e., for $i = 1, j = 1, 2, \cdots, M_t$) can be viewed as the projective equations of a set of *virtual* world points $\mathbf{v}_j = (t_{g0j,1}, t_{g0j,2}, t_{g0j,3})^T$, $j = 0, 1, \cdots, M_t$, with the virtual world coordinate system located at $\mathbf{t}_v = (X_{10}, Y_{10}, Z_{10})^T$ and having the orientation $R_v = R_{cg}^T$ with respect to the camera coordinate system. This insights allows us to employ the method of perspective transformation matrix [5] to find the internal parameters u_0, v_0, f_x, f_y, the external parameters R_{cg}, and the world coordinates (X_{10}, Y_{10}, Z_{10}) in closed forms. [The perspective transformation matrix here refers to the 3×4 matrix which transforms the 3-D virtual points $\{\mathbf{v}_j\}$ to the two-dimensional (2-D) frame buffer coordinates $\{(u_{1j}, v_{1j})\}$.] The obtained values for u_0, v_0, f_x, f_y and R_{cg} are then substituted into (13) and (14), for $i = 2, 3, \cdots, N$. The resulting equations can be rearranged in a linear form on (X_{i0}, Y_{i0}, Z_{i0}) for $i = 2, 3, \cdots, N$, by multiplying both sides with the denominator. From the obtained linear system, the world coordinates can be easily solved for in closed forms. So far, we have obtained $u_0, v_0, f_x, f_y, R_{cg}$, and the world coordinates $(X_{i0}, Y_{i0}, Z_{i0})^T$. Suppose then the motion stations $\langle g_j \rangle$, $j = M_t + 1, \cdots, M$ contain nonzero rotational components, which we call *compound motions*. By substituting all the obtained parameters into (4)–(8) for $j = M_t + 1, \cdots, M$, the measurement (8) can be similarly rearranged in a linear form on $\mathbf{t}_{cg}$ and can be easily solved for the hand–eye translation parameters. The complete procedure of estimating the unknown parameters above also leads to the following sufficient conditions for a unique solution.

Lemma 1: If a robot undergoes: a) 5 translational motions ($M_t = 5$), among which no more than 3 of the translation vectors with respect to the initial hand system are coplanar and b) 2 compound motions whose axes of rotation (with respect to the initial hand system) do not coincide with $\mathbf{t}_{cg}$, then the solutions for all the unknown parameters are unique.

Proof: See [27].

If the image data are free of noise and there exist no lens distortions, the values of the unknown parameters estimated above are exact, since we have introduced no approximations in the estimation. In general cases, however, when noises and distortions are both present, the obtained values can only be regarded as good initial guesses because the estimation errors accumulate in a sequential way. To refine the estimates, the Newton–Raphson method is used to adjust all the parameters (including the distortion coefficients) simultaneously, by local linearization of the original measurement (8) about the initial values. It can be easily demonstrated that the normal equations of the locally-linearized measurement equations take the following form:

$$\begin{pmatrix} A_1 & \mathbf{0} & \cdots & \mathbf{0} & B_1 \\ \mathbf{0} & A_2 & \cdots & \mathbf{0} & B_2 \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \mathbf{0} & \mathbf{0} & \cdots & A_N & B_N \\ B_1^T & B_2^T & \cdots & B_N^T & B_0 \end{pmatrix} \begin{pmatrix} \delta\mathbf{x}_1 \\ \delta\mathbf{x}_2 \\ \cdot \\ \cdot \\ \cdot \\ \delta\mathbf{x}_N \\ \delta\mathbf{x}_0 \end{pmatrix} = \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{e}_N \\ \mathbf{e}_0 \end{pmatrix} \quad (15)$$

where $\delta\mathbf{x}_i$ is a 3×1 increment vector for the world coordinates, $i = 1, \cdots, N$, $\delta\mathbf{x}_0$ a $m \times 1$ increment vector for the internal and external parameters, with m being the number of internal and external parameters; $\mathbf{e}_i$'s are the residual errors; A_i, B_i, and B_0 are matrices of size 3×3, $3 \times m$ and $m \times m$, respectively; The bordered block-diagonal structure of the normal matrix in (15) allows us to use a partitioning scheme proposed by Brown [17] to solve the normal equations. This avoids the inversion of a very large matrix when the number of world points is large. The computational burden in solving the normal equation by partitioning is made only linearly proportional to the number of world points. It should be noted that the Levenberg–Marquardt algorithm [19], which has been widely used in nonlinear optimizations, takes no account of the structure of the normal matrix. Its computational cost increases quadratically with the dimensions of the problem, which can be very large in problems like uncalibrated 3-D reconstruction [18] or self-calibration. We give, in the Appendix, details of the matrix reduction scheme and derive the Gauss–Markov theorem [20] for the estimation of parameter variances in the reduction case.

III. Model Identification and Motion Planning

In this section, we shall deal with two issues related to improvement of robustness and accuracy of the proposed method.

A. Model Identification

Model identification is concerned with the choice of the best model in describing a problem. Since the calibration equations derived in the last section are based on physical processes, the number of parameters employed should be minimum, except in the distortion model, which may be camera dependent. We suppose here that the distortion models of (9) and (10) cover the possible distortions, so that there is only a possible over-parameterization in the model. Overparameterization may cause the variances of some of the estimated parameters to increase, especially when there are few measurements. Although the computed parameters *as a whole* in the case of over-parameterization may still be useful in performing the correct transformation from the sensor space to the world space, the individual parameters when used alone for other purposes are not as reliable, because of the correlations in the estimates. To cope with this problem, we use the statistic inference method to deduce whether some specific (or all) distortion components should be excluded from the final calibration procedure, thus increasing the reliability of the estimated parameters.

The student-distribution (t-distribution) [8], [20] could be used to test whether a certain variable takes on a presumed value under a selected significance level $(1 - \alpha)$, based on the estimated variance of the variable [7]. A problem with the t-test is that it provides

$$u_{ij} - u_0 = f_x \frac{R_{cg}(1, 1)t_{g0j,1} + R_{cg}(2, 1)t_{g0j,2} + R_{cg}(3, 1)t_{g0j,3} + X_{i0}}{R_{cg}(1, 3)t_{g0j,1} + R_{cg}(2, 3)t_{g0j,2} + R_{cg}(3, 3)t_{g0j,3} + Z_{i0}} \quad (13)$$

$$u_{ij} - u_0 = f_x \frac{R_{cg}(1, 2)t_{g0j,1} + R_{cg}(2, 2)t_{g0j,2} R_{cg}(3, 2)t_{g0j,3} + Y_{i0}}{R_{cg}(1, 3)t_{g0j,1} + R_{cg}(2, 3)t_{g0j,2} + R_{cg}(3, 3)t_{g0j,3} + Z_{i0}}, \quad i = 1, \cdots, N;\ j = 0, 1, \cdots, M_t \quad (14)$$

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | **sales@docketalarm.com** | **1-866-77-FASTCASE**

fastcase®
Smarter legal research.