Optimal Hand-Eye Calibration

Klaus H. Strobl and Gerd Hirzinger

Institute of Robotics and Mechatronics
German Aerospace Center DLR
82230 Wessling, Germany
Klaus.Strobl@dlr.de

Abstract— This paper presents a calibration method for eye-in-hand systems in order to estimate the hand-eye and the robot-world transformations. The estimation takes place in terms of a parametrization of a stochastic model. In order to perform optimally, a metric on the group of the rigid transformations $SE(3)$ and the corresponding error model are proposed for nonlinear optimization. This novel metric works well with both common formulations $AX=XB$ and $AX=ZB$, and makes use of them in accordance with the nature of the problem. The metric also adapts itself to the system precision characteristics. The method is compared in performance to earlier approaches.

I. INTRODUCTION

This work concerns the use of visual sensors at the end-effector of robot manipulators/arms. In particular, hand-eye calibration consists of identifying the unknown position and orientation (pose) of the camera frame S_C with respect to (w.r.t.) the robot end-effector frame (also known as hand or *Tool Center Point TCP* frame) S_T, when the camera is rigidly mounted on the robot hand – see Fig. 1.

There is a strong need for an accurate hand-eye calibration. The reasons are twofold: *i)* to map sensor-centered measurements into the robot/world frame and *ii)* to allow for an accurate prediction of the pose of the sensor on the basis of the arm motion – in fact these are often complementary aspects of the same problem.

When performing hand-eye calibration on the basis of both the pose of S_T w.r.t. the robot base frame S_B $_bT^t$, and the pose of S_C w.r.t. the world/object frame S_0 $_0T^c$, there are two main approaches in order to estimate the hand-eye transformation:

1) Move the hand and observe/perceive the movement of the eye: or $AX = XB$, where A is the robot *TCP* motion $_{t_1}T^{t_2}$, B the induced camera motion $_{c_1}T^{c_2}$, and X is the hand-eye transformation $_tT^c$ to be determined. This is the classical approach. Early solutions regard the rotational part of this equation decoupled from the translational one, yielding uncomplex, fast, but error-prone formulations, since rotation estimation errors propagate to the translational part. Seminal articles are Shiu and Ahmad 1989 [1] (least squares fitting of rotation, then translation, using angle-axis representation) and Tsai and Lenz 1989 [2] (similar to [1] with closed-form solution). Zhuang and Roth 1991 [3] simplified the formulation introducing quaternions for the estimation of the rotational part, in the same way as Chou and Kamel 1991 [4], who make use of the singular value decomposition (*SVD*). Chen 1991 [5] for the first time does not decouple rotational and translational terms by using the screw theory. Wang 1992
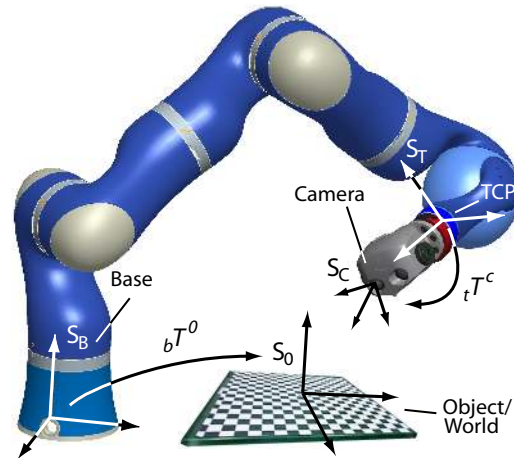

Fig. 1. Stereo camera mounted at the top of the DLR Light-Weight Robot 3.

in [6] compares [1] and [2] resulting in a slight advantage for the latter. Zhuang and Shiu 1993 [7] apply nonlinear optimization for both parts, minimizing a similar expression to Frobenius norms of homogeneous matrices of transformation errors. They additionally offer the possibility to disregard the camera orientation for the estimation. A similar approach was presented by Fassi and Legnani 2005 [8]. Park and Martin 1994 [9] perform nonlinear optimization in the same way, but again in the detached formulation. Lu and Chou 1995 [10] introduce the eight-space formulation based on quaternions, linearly optimizing both parts at the same time using the *SVD*. Horaud and Dornaika 1995 [11] nonlinearly optimize both the rotational (formulated with quaternions) and the translational parts *one-to-one*. Wei, Arbter, and Hirzinger 1998 [12] nonlinearly minimize algebraic distances performing simultaneous hand-eye and camera calibration. Daniilidis 1999 [13] introduces the dual quaternions – an algebraic representation of the screw theory to describe motions. This enables the author to find a fast *SVD*-based joint solution for rotation and translation within linear formulation. Bayro-Corrochano et al. 2000 [14] in the same way produce a *SVD*-based linear solution of the coupled problem by the use of motors within the geometric algebra framework. Andreff et al. 2001 [15] do the job properly, employing this particular formulation for *X-from-motion* applications – cf. Section II-C. They get rid of the nonlinear orthogonality constraint in $SO(3)$[1] by increasing the dimensionality of the rotational part and manage to formulate the problem as a single homogeneous linear system.

---

[1] $SO(3)$ is the *Special Orthogonal group* of 3×3 matrices or rotation group for 3D space.

2) Simultaneous estimation of the hand-eye transformation and the pose of the robot in the world: or $\boldsymbol{AX} = \boldsymbol{ZB}$, where $\boldsymbol{A}$ is the pose of the camera frame S_C w.r.t. S_0 $_0\boldsymbol{T}^c$, $\boldsymbol{B}$ the pose of the robot *TCP* S_T w.r.t. S_B $_b\boldsymbol{T}^t$, and $\boldsymbol{X}$ and $\boldsymbol{Z}$ are the eye-hand and world-base transformations to be determined, i.e. $_c\boldsymbol{T}^t$ and $_0\boldsymbol{T}^b$ respectively. To the best of our knowledge it was Wang in 1992 [6] who first submitted this formulation explicitly for hand-eye calibration. Surprisingly, none of the further approaches refer to him in this context[2]. Zhuang et al. 1994 [16] apply quaternions in order to get a simple linear solution of the rotational part by the use of the *SVD*. Rémy et al. 1997 [17] nonlinearly optimize both parts by minimizing reprojected 3D Euclidean error distances in S_0. Dornaika and Horaud 1998 [18] solve the rotational problem linearly with quaternions and also nonlinearly optimize both parts by *one-to-one* minimizing of Frobenius norms and two penalty functions – see Section III. Other approaches integrate the hand-eye calibration with the intrinsic camera calibration and minimize the **R**oot **M**ean **S**quare (*RMS*) of the image frame errors.

The optimization criteria for both approaches are often sub-optimal and no attention is paid to proper parametrizations. Since the purpose of model-based[3] calibration is the accurate parametrization of the system model, *maximum accuracy optimal calibration is achieved when minimizing model fitting errors with regard to the actually erroneous elements*. Here we propose a metric on the group of rigid transformations $SE(3)$ for this purpose.

Moreover, with the exception of [17], a thorough comparison of these very different approaches is missing. This paper shows the most accurate algorithms along with the one presented here – for both approaches – and justifies their application in relation to the nature of the problem.

This paper is organized as follows: Section II states the hand-eye calibration problem and its most relevant formulations. Next Section III presents a novel metric on $SE(3)$ in order to optimally solve these problems. In addition, other representative solutions are exposed. The ideas given in Sections II and III are implemented in Sections IV and V. Section VI recapitulates and presents an outlook on future work.

II. PROBLEM DESCRIPTION

Let $_b\boldsymbol{T}^t$ be the homogeneous transformation relating the pose of the base frame S_B to the pose of the *TCP* frame S_T. $_b\boldsymbol{T}^t$ results from the calibrated forward kinematic model of the robot, the encoder readings of every joint, and possibly its control parameters. Let again $_0\boldsymbol{T}^c$ be the homogeneous transformation relating the pose of the object/world frame S_0

---

[2]Wang mainly compared the algorithms in [1] and [2], but at the same time he also produced this second family of solutions. Wang himself criticizes his *class A* calibration procedure since it yields biased results unless an error-free $_0\boldsymbol{T}^b$ is specified. He actually fails to realize the necessity of estimating $_0\boldsymbol{T}^b$ *at the same time* in order to avoid measurement innacuracies or mistakes and uses ad hoc external measurements for getting it.

[3]Approaches that do not rely on a physical model of the system may actually perform on ocassions better if they are purposefully calibrated. We therefore add the adjective *model-based* to the procedures in this work.

to the pose of the camera frame S_C (regardless of whether we use mono or stereo vision). $_0\boldsymbol{T}^c$ stems from the absolute extrinsic parameters of the camera calibration process, cf. [19].

There remain two unknown transformations $_t\boldsymbol{T}^c$ and $_0\boldsymbol{T}^b$ to be estimated. The latter does not require frequent recalibration, since manipulators are not usually shifted. On the contrary, the rigid pose of the camera frame S_C w.r.t. the *TCP* frame S_T has to be calibrated more often, since the camera(s) may be removed or rotated. These transformations should not be measured by hand since the different frames are located inside the manipulator or the sensor.

In order to uniquely determine $_t\boldsymbol{T}^c$ (and perhaps $_0\boldsymbol{T}^b$), at least $n = 3$ stations – specifically two movements with nonparallel rotation axes – are required (refer to [2], [5]).

A. Solution #1: $AX = ZB$

Next is presented the direct formulation of the *predictive parametric* model described in the last section. It enables us to *predict* values (e.g. $_b\boldsymbol{T}^t$) on the basis of a *parametric* representation of the world (e.g. $_c\boldsymbol{T}^t$). These predictions, jointly with actual measurements, make it possible to refine optimally on this parametric representation of the world – see Section III.

This first formulation directly reproduces the rigid transformations in the loop *camera-TCP-base-world-camera*:

$$_0\boldsymbol{T}^c{}_c\boldsymbol{T}^t = {}_0\boldsymbol{T}^b{}_b\boldsymbol{T}^t \quad \rightleftharpoons \quad \begin{array}{c} S_C \xrightarrow{{}_cT^t} S_T \\ {}_0T^c\uparrow \quad \nearrow \quad \uparrow {}_bT^t \\ S_0 \xrightarrow{{}_0T^b} S_B \end{array} \quad (1)$$

and evades further modeling (e.g. perspective projection or detailed joint/link information). It introduces a significant constraint in order to ensure that both $_c\boldsymbol{T}^t$ and $_0\boldsymbol{T}^b$ are consistent with the actual system. The equation is usually decomposed into its rotational and translational parts:

$$\left. \begin{array}{c} _0\boldsymbol{R}^c{}_c\boldsymbol{R}^t = {}_0\boldsymbol{R}^b{}_b\boldsymbol{R}^t \\ _0\boldsymbol{R}^c{}_c\boldsymbol{t}^t + {}_0\boldsymbol{t}^c = {}_0\boldsymbol{R}^b{}_b\boldsymbol{t}^t + {}_0\boldsymbol{t}^b \end{array} \right\} . \quad (2)$$

The solution has been calculated in different ways – refer to [6], [16]–[18].

B. Solution #2: $AX = XB$

Due to the fact that $_c\boldsymbol{T}^t$ is more often required than $_0\boldsymbol{T}^b$, most approaches eliminate the latter by writing Eq. (1) at two different instants i and j yielding the well-known hand-eye equation:

$$_{c_i}\boldsymbol{T}^{c_j}{}_c\boldsymbol{T}^t = {}_c\boldsymbol{T}^t{}_{t_i}\boldsymbol{T}^{t_j} \quad \rightleftharpoons \quad \begin{array}{c} _0\boldsymbol{T}^{c_j} \xrightarrow{{}_cT^t;\,({}_0T^b)} {}_b\boldsymbol{T}^{t_j} \\ {}_{c_i}T^{c_j}\uparrow \quad \nearrow \quad \uparrow {}_{t_i}T^{t_j} \\ _0\boldsymbol{T}^{c_i} \xrightarrow{{}_cT^t;\,({}_0T^b)} {}_b\boldsymbol{T}^{t_i} \end{array} \quad (3)$$

or

$$\left. \begin{array}{c} _{c_i}\boldsymbol{R}^{c_j}{}_c\boldsymbol{R}^t = {}_c\boldsymbol{R}^t{}_{t_i}\boldsymbol{R}^{t_j} \\ _{c_i}\boldsymbol{R}^{c_j}{}_c\boldsymbol{t}^t + {}_{c_i}\boldsymbol{t}^{c_j} = {}_c\boldsymbol{R}^t{}_{t_i}\boldsymbol{t}^{t_j} + {}_c\boldsymbol{t}^t \end{array} \right\} \quad (4)$$

first formulated in [1], [2]. This problem was geometrically analyzed recently in [8]. When writing $_{c_i}\boldsymbol{T}^{c_j} = {}_c\boldsymbol{T}^t{}_{t_i}\boldsymbol{T}^{t_j}{}_t\boldsymbol{T}^c$ it becomes clear that $_{c_i}\boldsymbol{T}^{c_j}$ and $_{t_i}\boldsymbol{T}^{t_j}$ are the same rigid transformation assessed in different frames of reference.

The solution has been calculated in different ways – refer to [1]–[15]. If necessary, the dual equation ${}_{0_i}\boldsymbol{T}^{0_j}{}_0\boldsymbol{T}^b = {}_0\boldsymbol{T}^b{}_{b_i}\boldsymbol{T}^{b_j}$ (where S_0 moves w.r.t. S_C and S_B w.r.t. S_T) also enables the estimation of ${}_0\boldsymbol{T}^b$.

C. Choice of formulation

The above equations do not hold exactly in the case of noise and $n > 3$ stations due to the erroneous rigid-body transformations ${}_b\boldsymbol{T}^{t_i}$ and ${}_0\boldsymbol{T}^{c_i}$ (or ${}_{t_i}\boldsymbol{T}^{t_j}$ and ${}_{c_i}\boldsymbol{T}^{c_j}$). The proper *optimal* way of accurately estimating ${}_t\boldsymbol{T}^c$ and ${}_0\boldsymbol{T}^b$ is thus *optimally* correcting these erroneous measurements at every station $i = 1..n$ depending on both their geometric error models and the predictions/estimations from the formulae.

The maximum likelihood method (*ML*) selects the model (e.g. ${}_t\boldsymbol{T}^c$) for which the probability of the observed data (e.g. ${}_b\boldsymbol{T}^{t_i}$) is highest or, in other words, for which its incompatibility with the model is minimized. For *Gaussian* error distributions, a typical cost function is the sum of covariance-weighted squared predictions of these particular errors. The resulting model parameters then have zero bias, lowest variance, and maximum probability if flat prior [20].

In the next section a metric for rigid-body transformation errors is presented. Experiments on the real systems *FaroArm Gold* [21] and *ARTtrack2* [22] have actually shown zero-mean Gaussian distributions[4] *if* this error metric is applied to the transformation ${}_b\boldsymbol{T}^t$ – see Section V. These errors are naturally much larger than the ones in ${}_0\boldsymbol{T}^c$ and therefore the latter shall not be considered [2]. These experiments make it clear, that for this common eye-in-hand framework it is possible to *optimally* correct ${}_b\boldsymbol{T}^{t_i}$ by minimizing a sum of covariance-weighted squared prediction errors *within* the $\boldsymbol{AX} = \boldsymbol{ZB}$ formulation in order to *optimally* estimate ${}_t\boldsymbol{T}^c$ and ${}_b\boldsymbol{T}^0$.

In the case of the $\boldsymbol{AX} = \boldsymbol{XB}$ formulation this method does *not* hold anymore, since ${}_{t_i}\boldsymbol{T}^{t_j}$ and ${}_{c_i}\boldsymbol{T}^{c_j}$ do not necessarily show Gaussian errors in this metric, but rather nonlinear functions of them. However, we propose this formulation for visual *pose-from-motion* problems (as in [15]). Here the motion ${}_{c_i}\boldsymbol{T}^{c_j}$ may actually produce Gaussian errors in this metric and moreover much larger than the ones in ${}_{t_i}\boldsymbol{T}^{t_j}$ so that the latter may be ignored. Research in this direction shall be conducted in the near future.

III. MINIMIZING ON $SE(3)$

In this section a distance metric on the Euclidean group of rigid-body motions $SE(3)$ is presented. Elements in this group are represented as a couple $(\boldsymbol{R}, \boldsymbol{t})$ where $\boldsymbol{R} \in SO(3) = \{\boldsymbol{R} \in \Re^{3 \times 3} : \boldsymbol{R}^t\boldsymbol{R} = \boldsymbol{I}, det(\boldsymbol{R}) = 1\}$ and $\boldsymbol{t} \in \Re^3$. The problem of finding a metric for $SE(3)$ can be presented in terms of a real valued objective function $\mathcal{O} : (SO(3) \times \Re^3) \to \Re$ which measures the suitability of the unknowns to fit the data.

For the hand-eye calibration problem many choices for a metric exist, but none directly referred to the actually erroneous transformations ${}_b\boldsymbol{T}^{t_i}$ and ${}_0\boldsymbol{T}^{c_i}$. Next some significant approaches are presented – they shall be assessed within the Sections IV and V:

[4]The metric in translation error actually shows χ^2 distribution in its squared form.

a) The residuals in *linear optimization methods* usually represent angles, rotation matrix elements, quaternion distances, etc. – or rather, nonlinear algebraic transformations of them.

b) A relevant metric is the one of *Horaud and Dornaika* in [11]. They optimize nonlinearly for ${}_c\boldsymbol{T}^t$ within the $\boldsymbol{AX} = \boldsymbol{XB}$ formulation as follows:

$$\{\boldsymbol{q}, \boldsymbol{t}\}^\star = \arg\min_{\boldsymbol{q}, \boldsymbol{t}} (f(\boldsymbol{q}, \boldsymbol{t}) + \lambda(1 - \boldsymbol{q} * \bar{\boldsymbol{q}})^2)$$

with

$$f(\boldsymbol{q}, \boldsymbol{t}) = \lambda_1 \sum_{i=1}^{n-1} || \boldsymbol{v}'_i - \boldsymbol{q} * \boldsymbol{v}_i * \bar{\boldsymbol{q}} ||^2 + \\ \lambda_2 \sum_{i=1}^{n-1} || \boldsymbol{q} *_{t_i}\boldsymbol{t}^{t_{i+1}} * \bar{\boldsymbol{q}} - ({}_{c_i}\boldsymbol{R}^{c_{i+1}} - \boldsymbol{I})\boldsymbol{t} -_{c_i}\boldsymbol{t}^{c_{i+1}} ||^2$$

with weights $\lambda_1 = \lambda_2 = 1$ and $\lambda = 2 \cdot 10^6$. The latter factor guarantees the consistency of the quaternion $\boldsymbol{q}$ representing ${}_c\boldsymbol{R}^t$. $\boldsymbol{v}_i$ and $\boldsymbol{v}'_i$ are the eigenvectors associated with the unitary eigenvalues of ${}_{t_i}\boldsymbol{R}^{t_{i+1}}$ and ${}_{c_i}\boldsymbol{R}^{c_{i+1}}$ respectively. This objective function $\mathcal{O}_H$ has the form of a sum of squares of nonlinear functions and can be minimized e.g. with the Levenberg-Marquardt algorithm. This method considerably improved the accuracy relative to earlier work.

c) Dornaika and Horaud apply in [18] a different metric for the $\boldsymbol{AX} = \boldsymbol{ZB}$ formulation:

$$\{{}_t\boldsymbol{T}^c, {}_0\boldsymbol{T}^b\}^\star = \arg\min_{{}_t\boldsymbol{T}^c, {}_0\boldsymbol{T}^b} \Big(f({}_t\boldsymbol{T}^c, {}_0\boldsymbol{T}^b) + \\ \lambda_3 ||{}_t\boldsymbol{R}^c_c\boldsymbol{R}^t - \boldsymbol{I}||^2 + \lambda_4 ||{}_0\boldsymbol{R}^b_b\boldsymbol{R}^0 - \boldsymbol{I}||^2 \Big)$$

with

$$f({}_t\boldsymbol{T}^c, {}_0\boldsymbol{T}^b) = \lambda_1 \sum_{i=1}^{n} ||{}_0\boldsymbol{R}^c_c\boldsymbol{R}^t -_0\boldsymbol{R}^b_b\boldsymbol{R}^t ||^2 + \\ \lambda_2 \sum_{i=1}^{n} ||{}_0\boldsymbol{R}^c_c\boldsymbol{t}^t +_0\boldsymbol{t}^c -_0\boldsymbol{R}^b_b\boldsymbol{t}^t -_0\boldsymbol{t}^b ||^2$$

again with weights $\lambda_1 = \lambda_2 = 1$ and $\lambda_3 = \lambda_4 = 10^6$. This objective function $\mathcal{O}_D$ minimizes the Frobenius norm of a *residual* matrix in its rotational part.

d) Daniilidis in [13] presents a unified and fast way of formulating with dual-quaternions – algebraic counterparts of screws – for both the rotational and the translational parts. In part this method attempts to avoid the weighting problem by using the compact formulation $\check{a} = \check{q}\check{b}\bar{\check{q}}$ for $\boldsymbol{AX} = \boldsymbol{XB}$, which can be solved linearly with *SVD*.

However, it is worth noting the absence of any position/orientation weighting possibility – or at least criteria for it – in order to aim purposefully at optimal total estimators. In addition, parametrization issues have a strong influence in their usage: variable scaling (the choice of 'units' relating meaningfully to the problem structure) and, more generally, preconditioning (the choice of linear combinations of parameters to use) are crucial for convergence and unfortunately play an unexpected role in the weighting issue.

We next describe a sound objective function $\mathcal{O}$ based on an error metric on $SE(3)$. For the first time it attempts to solve the next purposes: *i)* to *optimally* reduce actual system errors, *ii)* to allow for a natural weighting of the rotational and translational components, and *iii)* the algorithm to be able to autonomously adjust the latter. In addition, it sorts well with both the $\boldsymbol{AX} = \boldsymbol{XB}$ and the $\boldsymbol{AX} = \boldsymbol{ZB}$ formulations.

A. Metric for rotation error

Any rigid body transformation can be modeled as a rotation in $SO(3)$ by an angle θ about an axis p through the origin, followed by a translation t in $\Re^3$. The rotation can thus be represented by three independent parameters – the rotation angle θ and two angles $\{\alpha, \beta\}$ defining the axis of rotation p.

Residual error rotations such as $_{\Delta t}\widetilde{R} = {}_{t}\widetilde{R}^b\,_b\widehat{R}^t$ where $_b\widetilde{R}^t$ is measured (e.g. read from joint encoders) and $_b\widehat{R}^t$ estimated by e.g. Eq. (1) present small angles θ. Experiments[5] show to a great extent arbitrary (randomly distributed) axes of rotation for these residual rotations. Based on these experiments and for general application, *the axes of rotation are considered here irrelevant*. In addition, data-driven over-special models tend to give biased results. We therefore propose

$$\mathcal{O}_i^{rot} = {}_{\Delta t}\theta_i = \pm \arccos\left(\left(\operatorname{trace}(_{\Delta t}\widetilde{R}_i) - 1\right)/2\right) = {}_{\Delta b}\theta_i \quad .$$

This geometrically-defined metric[6] for $SO(3)$ gauges the residual rotation error and proves to be frame-invariant. In [6] the similar metric $\mathcal{O}_i^{rot'} = \mathcal{N}(_b\widetilde{R}^{t_i}, {}_b\widehat{R}^{t_i}) = 2\sin(|_{\Delta t}\theta_i|/2)$ is introduced for purposes of calibration accuracy assessment.

B. Metric for translation error

To find a metric on the Euclidean space does not seem complicated, since the Euclidean distance is its natural metric. However, the metric for the translation residual error of a rotationally erroneous rigid-body motion cannot be a single Euclidean residual distance, since this measure is *not* frame-invariant, e.g. $_{t_i}\tilde{t} = {}_b\hat{t}^{t_i} - {}_b\tilde{t}^{t_i}$ differs from $_{b_i}\tilde{t} = {}_{t_i}\hat{t}^b - {}_{t_i}\tilde{t}^b$ if $\{_{\Delta t}\theta_i = {}_{\Delta b}\theta_i > 0\}$ provided that $\{||_{t_i}\tilde{t}|| > 0 \lor ||_{b_i}\tilde{t}|| > 0\}$.

In absence of further model information, we choose the equitable balance between these two symmetrical Euclidean distances as a metric for translational residual error:

$$\mathcal{O}_i^{tra} = \left(||_{t_i}\tilde{t}|| + ||_{b_i}\tilde{t}||\right)/2 \quad .$$

C. Combination of both metrics

In Section II-C it is stated that by means of the minimization of the sum of covariance-weighted squared prediction errors with zero-mean Gaussian error distributions, the *ML* method estimates the *optimal* model (e.g. $_t T^c$). Thus the total transformation error cost function results:

$$\mathcal{O}_i = \frac{(\mathcal{O}_i^{rot})^2}{{}^\star\sigma_{rot}^2} + \frac{(\mathcal{O}_i^{tra})^2}{{}^\star\sigma_{tra}^2}$$

where $^\star\sigma_{rot}^2$ and $^\star\sigma_{tra}^2$ are the 2nd moments of the *independent* Gaussian probability density functions (*pdf*s) in rotation and translation error. Eventually for the $AX = ZB$ formulation:

$$\{_t T^c, {}_b T^0\}^\star = \arg\min_{_t T^c, {}_b T^0}\left(\sum_{i=1}^n \frac{(\mathcal{O}_i^{rot})^2}{{}^\star\sigma_{rot}^2} + \frac{(\mathcal{O}_i^{tra})^2}{{}^\star\sigma_{tra}^2}\right)$$
$$= \arg\min_{_t T^c, {}_b T^0}\left(\sum_{i=1}^n (\mathcal{O}_i^{rot})^2 + \frac{(\mathcal{O}_i^{tra})^2}{({}^\star\sigma_{tra}/{}^\star\sigma_{rot})^2}\right)$$

---

[5] Experiments on pose accuracy for both the *FaroArm Gold* [21] and the *ARTtrack2* [22] systems were performed by comparing their readings with the absolute extrinsic parameters from a stereo camera calibration algorithm.

[6]The trace of a rotation matrix R is independent of the coordinate system used (as long as it is orthonormal). This implies that the trace is the sum of the eigenvalues of R, that is $1 + (\cos(\theta) + i\sin(\theta)) + (\cos(\theta) - i\sin(\theta)) = 1 + 2\cos(\theta) \Rightarrow \theta = \pm \arccos((r_{11} + r_{22} + r_{33} - 1)/2)$.

---

where $^\star\sigma_{tra}/^\star\sigma_{rot}$ is the *position/orientation precision ratio* – this is here the only required parameter for optimal estimation.

In general the weighting of translational and rotational metrics on $SE(3)$ is said to depend on the eventual task at hand. Here again this is *definitely* true, since the hand-eye calibration *task* aims primarily at the accurate estimation of the system parameters. In this way they are estimated at best.

Numerical optimization procedures are to be applied to find the solutions. This nonlinear optimization method is more stable against noise since the objective function $\mathcal{O} = \sum_{i=1}^n \mathcal{O}_i$ takes into account proper noise models.

D. Automatic optimal weighting

A further appealing virtue of this metric is the ability for it to automatically *and optimally* weight its objective function for itself in case the optimal 2nd moments of the error in $_b T^t$ are not at hand. It can be verified that
$$^\star\sigma_{rot}^2 = \sum_{i=1}^n ({}^\star\mathcal{O}_i^{rot})^2/n$$
$$^\star\sigma_{tra}^2 = \sum_{i=1}^n ({}^\star\mathcal{O}_i^{tra})^2/n$$
for $\{_t T^c, {}_b T^0\}^\star$.

Fig. 2. Weighting ratio adaption.

Fortunately, non-optimal but approximated model parameters $\{_t T^c, {}_b T^0\}$ (e.g. resulting from an optimization with arbitrary prior weightings in $\mathcal{O}_i$ with e.g. *one-to-one* weighting) yield *improved* weighting parameters, so that $\sigma_{rot}^2 = \sum_{i=1}^n (\mathcal{O}_i^{rot})^2/n \rightsquigarrow {}^\star\sigma_{rot}^2$ and $\sigma_{tra}^2 = \sum_{i=1}^n (\mathcal{O}_i^{tra})^2/n \rightsquigarrow {}^\star\sigma_{tra}^2$. Both experiments and simulations suggest that this process in general converges (see Fig. 2), so that after about 3 iterations the results become optimal. The work in Ref. [23] implements this algorithm.
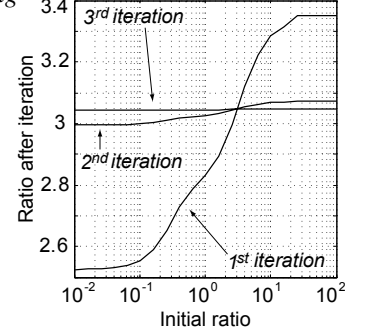
IV. SIMULATION RESULTS

Simulations were conducted to compare both solutions applying the proposed metric, as well as other representative methods. Simulations facilitate the comparison of *estimated* and *actual* values. In particular, these simulations compare nonlinear minimization on $SE(3)$ as proposed in the last section (as from now "*SE*" for $AX = ZB$ and "*se*" for $AX = XB$) with the 5 well-known methods briefly presented in Section III: linear least squares solution in rotation ("*li*"), nonlinear minimization [11] ("*nl*"), and dual-quaternion [13] ("*dq*") approaches for $AX = XB$, and linear least squares solution in rotation ("*LI*") and nonlinear minimization of Frobenius norms [18] ("*FN*") approaches for $AX = ZB$.

On the one hand, most of these methods lack of a convenient rotation/translation weighting policy. On the other hand, the novel metric proposed here is able to perform weighting automatically. We take advantage of this situation and study some of these methods in relation to their weighting parameters and, in the case of the method in [11], we introduce a novel weighting policy in order to boost performance: it aims simply at bringing both translational (Euclidean distances) and rotational (quaternion distances) errors to the the same order of magnitude.
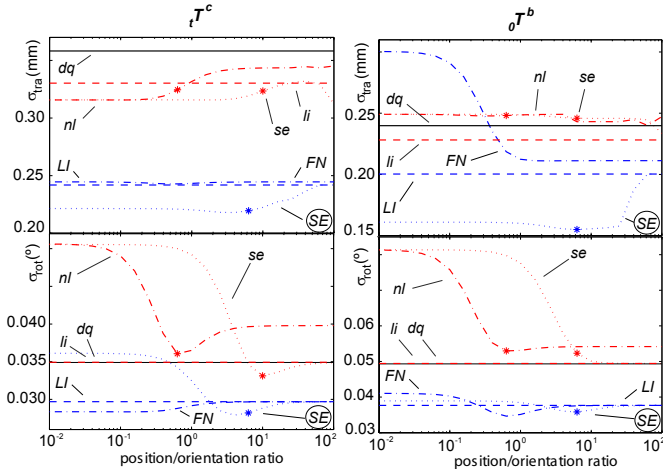
Fig. 3. Standard deviations of the parameter errors with *noise model* #1. Optimal values regarding weighting are marked with '*'.

Fig. 4. Standard deviations of the parameter errors with *noise model* #2. Optimal values regarding weighting are marked with '*'.

The simulation is conducted as follows: 9 camera stations were generated at each altitude $_0z$ of 25 and $45\,cm$ w.r.t. S_0 with both $_0x$ and $_0y \in \{-30, 0, 30\}\,cm$, i.e. $n=18$ stations altogether are used for calibration. They all focus on S_0. Since the simulation results are to be presented statistically, this series was randomly repeated 100 times in the form of a Monte-Carlo simulation. Nominal, arbitrary, but realistic values for the unknown transformations $_tT^c$ and $_bT^0$ are generated for every calibration, being the only restriction $||_tt^c||=||_bt^0||=30\,cm$. These enable the calculation of the *actual* $_bT^{t_i}$ transformations. In turn, random errors provide $_b\widetilde{T}^t$. Furthermore, the Monte-Carlo simulations for the methods *nl*, *FN*, *SE*, and *se* were repeated 21 times with different weighting factors. The averaged results show accuracy and error minimization results.

As previously stated (see Section II-C), the robot arm pose measurement $_b\widetilde{T}^t$ is expected to be the main source of perturbation. Here two different noisy models are simulated: *noise model* #1 applies for general time-independent inaccurate pose data and *noise model* #2 to general time-independent inaccurate *motion* data – which is time-*dependent* inaccurate pose data since it implies growing pose innacuracy over time. It will be shown that the latter are best dealt using the $AX = XB$ formulation, whereas the former using $AX = ZB$. In particular, measurement noise was included for the *noise model* #1 in orientation[7] with $_{\Delta t}\widetilde{R}$ and $_{\Delta b}\widetilde{R}$, having rotation angles θ granted to be unbiased, with Gaussian *pdf*s with $\sigma_\theta = 0.15\,^\circ$. The axes of rotation p of these rotation matrices are uniformly distributed, i.e. $\alpha \in [-90^\circ, 90^\circ)$ with $pdf(\alpha) = 180^{-1}\,[^\circ]^{-1}$ and $\beta \in [-90^\circ, 90^\circ)$ with $pdf(\beta) \propto \arcsin(\beta/90)\,[^\circ]^{-1}$. In position, the Euclidean residual distances $_t\tilde{t}$ and $_b\tilde{t}$ also present real Gaussian *pdf*s with $\sigma_t = 0.35\,mm$, and their directions are again uniformly distributed. For *noise model* #2 we use $_{t_i}\widetilde{R}^{\Delta t_j} = {}_{t_j}\widetilde{R}^{t_i}{}_{t_i}\widehat{R}^{t_j}$ and $_{t_i}\tilde{t}^{\Delta t_j} = {}_{t_i}\hat{t}^{t_j} - {}_{t_i}\tilde{t}^{t_j}$.
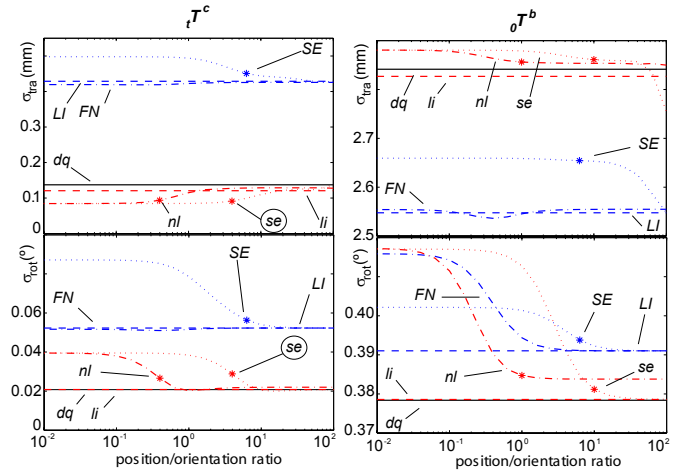
Robustness analysis in the presence of varying noise levels were performed. The conclusions are in conformity with prior works: for usual applications the superiority of an algorithm does not critically depend on the noise level. So these studies are not reported here.

A. Accuracy analysis with synthetic data

Next both accuracy and precision[8] in the estimation of $_tT^c$ and $_bT^0$ are studied in relation to the employed method, the error model, and the weighting parameters.

Primarily the simulations reflect the operation of the *ML* method as stated in Section II-C: error standard deviations are much larger than the biases – at least 10 times. In addition, under mild regularity conditions on the measurement distributions, the posterior distributions of the *ML* estimates converge asymptotically in probability to Gaussians. Therefore the accuracy analyses performed here focus on the 2nd moments of the estimation errors.

Figs. 3 and 4 show the standard deviations in position and orientation estimation for $_tT^c$ and $_0T^b$. The figures totally differ in the fact that in Fig. 3 the $AX = ZB$ approaches (upper case) show better performance, whereas in Fig. 4 ($_tT^c$) the $AX = XB$ approaches (lower case) do. In particular, for both approaches the methods developed here (*SE* and *se*) show highest precision. In addition, proper weighting proves crucial to optimal estimation – this applies with novelty to the method *nl* in [11] as well as to the ones presented here (*SE* and *se*). Note that without this weighting policy but with $\lambda_1 = \lambda_2 = 1$ (i.e. ratio $= 10^0$) the *nl* method would perform worse. Besides, results with *noise model* #2 in the right-hand side of Fig. 4 do not show relevant findings, since for this model neither of the solutions is optimal for the estimation of $_0T^b$.

Although parametrization and conditioning were taken into account, in the simulations the dual-quaternion approach does not work any better than the linear estimation methods do. This fast method may be appropriate for solving hand-eye calibration in *X-from-motion* problems.

---

[7]Note that this error model does not completely correspond to the metric proposed in Section III. Here rotational errors appear both in S_T and S_B.

[8]Accuracy refers to the agreement of estimations and actual values, whereas precision refers to the repeatability of estimations.

Fig. 5. Std. deviations of different error metrics with *noise model #1*. Optimal values regarding weighting are marked with '*'.
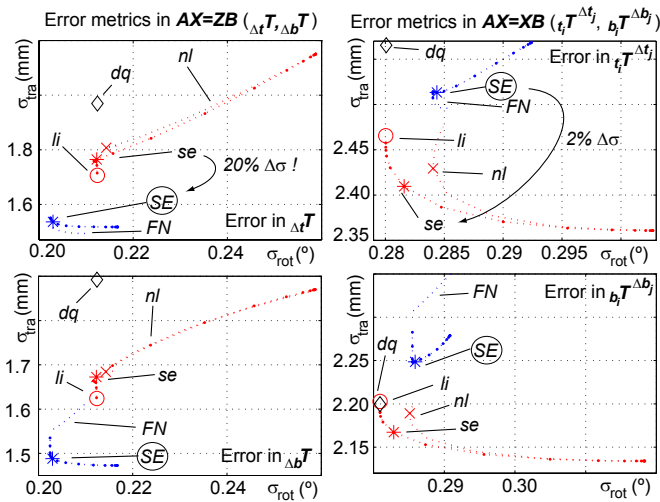

Fig. 6. Std. deviations of different error metrics with *noise model #2*. Optimal values regarding weighting are marked with '*'.

B. Model matching analysis with synthetic calibration data

Figs. 5 and 6 show the 2nd moments of the error metrics presented in Section III for the same simulations. They help to understand that: *i)* linear solutions minimizing rotational errors (*li*) represent only one potential extreme solution for the hand-eye calibration problem, *ii)* to be able to find the proper weighting is *critical* for optimal calibration, *iii)* approaches for the solution $AX = XB$ perform poorly with *noise model #1* because the error reduction they do get in $_{t_i}T^{t_j}$ becomes much smaller than the reduction the approaches for the solution $AX = ZB$ get in the error in $_bT^{t_i}$ (the opposite applies to *noise model #2*).

V. EXPERIMENTAL RESULTS

In this section the performance of the different algorithms in real systems is presented. The pose of S_T w.r.t. S_B stems from readings in the *ARTtrack2* system [22]. $_0T^c$ is in turn provided by a stereo camera calibraton algorithm derived from the monocular one in [19], refer to [23]. Experiments aim at verifying correct operation of the methods in real systems. As in Section IV, the estimation accuracy *is* evidence of correct operation. Unfortunately, since ground-truth information is usually missing in experiments, calibration accuracy may not be *directly* assessed. It is possible to *indirectly* evaluate calibration performance: this should be done by verifying the system model matching capability in relation to the calibration results, since the results that best fit the model stem from an optimal calibration process [20]. An example for this case is testing the ability to predict $_bT^t$ on the basis of $_0T^c$ and the solutions $_t^{\star}\widehat{T}^c$ and $_b^{\star}\widehat{T}^0$ in several verification stations – see Section V-B. Alternatively, it is possible to *indirectly* evaluate calibration performance by using some task-dependent metrics[9].

<hr>

[9]For instance you may evaluate the ability to predict camera poses or image projections using only measurement data. These practices are undesirable since external assessments incorporate task-dependent requirements. For instance, for a very long link $_tT^c$ the camera pose estimation accuracy may be extremely influenced by the *TCP* orientation error. In this case orientation error minimization algorithms would be preferred – but they certainly *do not* parameterize the system model properly – recall Section III-C.
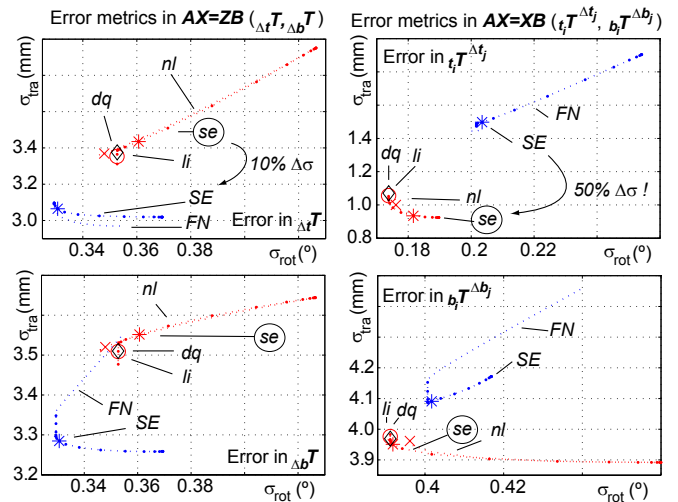
A. Model matching analysis with real calibration data

Following on from the last section, we next observe the model matching capability of different calibration algorithms on real experiments in order to relate them to their virtual counterparts in Section IV. Again, the 2nd moments of the metrics presented in Section III are presented for an experiment consisting of 10 stations. Results in Fig. 7 resemble to a great extent Fig. 5 in a slightly different order of magnitude[10]. This suggests that the system presents *noise model #1*. In this case Section IV points to *SE* (minimization on $SE(3)$ within $AX = ZB$) in order to optimally estimate $_tT^c$ and $_bT^0$.
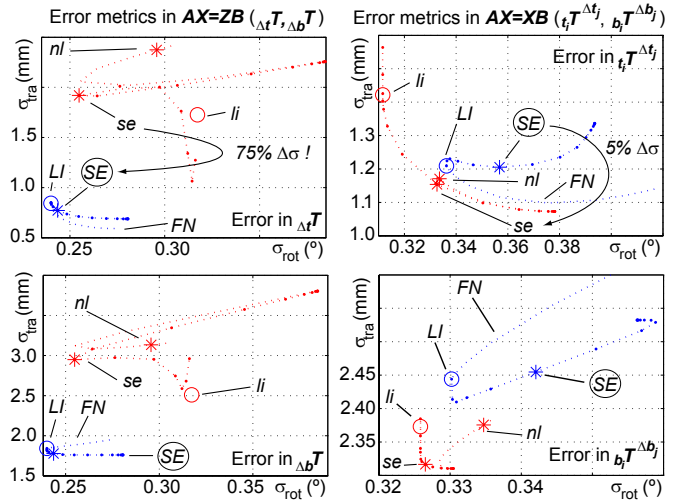

Fig. 7. Std. dev. of different error metrics in a real hand-eye calibration. Optimal values regarding weighting are marked with '*'.

B. Model matching analysis with real verification data

The last results do not suffice to indirectly evaluate model-based calibration performance, since the calibration process purposefully forced this particular data to comply with the parameterized model. To verify general predictive capability of

<hr>

[10]An important attribute of the automatic weighting method exposed in Section III-D is its independency of the order of magnitude – the only required parameter being a *ratio* of the precisions of the system.

TABLE I

$_t\widehat{\boldsymbol{T}}^c$ and $_b\widehat{\boldsymbol{T}}^0$ from a previous calibration with 10 stations							$_t\widehat{\boldsymbol{T}}^c$ and $_b\widehat{\boldsymbol{T}}^0$ from a comprehensive calibration with 37 stations							
$\boldsymbol{AX} = \boldsymbol{ZB}$			$\boldsymbol{AX} = \boldsymbol{XB}$				$\boldsymbol{AX} = \boldsymbol{ZB}$			$\boldsymbol{AX} = \boldsymbol{XB}$				
SE	FN	LI	se	nl	dq	li	SE-t	FN-t	LI-t	se-t	nl-t	dq-t	li-t	
σ_θ	0.210	0.213	0.213	0.227	0.239	0.516	0.270	0.199	0.200	0.200	0.215	0.216	0.341	0.241	[°]
σ_t	1.343	1.388	1.386	2.401	2.597	3.735	1.969	1.171	1.217	1.216	1.463	1.390	2.424	1.549	[mm]

Eq. (1), data external to the calibration process was acquired in the form of 27 additional *TCP* stations and their corresponding camera pose estimations. Tab. I presents the 2nd moments of the metrics in Section III for 27 verification stations with erroneous $_b\boldsymbol{T}^t$. In the left-hand side the calibration results $_t\widehat{\boldsymbol{T}}^c$ and $_b\widehat{\boldsymbol{T}}^0$ from the last section were used. The *verification* stations still present low discrepancies between measurements and predictions. In the right-hand side the results (*-t*) of an ex**t**ensive hand-eye calibration with 37 stations were used. The latter show slightly better results as expected. This is due to the few stations used for the former calibration[11] as well as to remaining modeling errors (e.g. camera pose estimation). The proposed *SE-t* approach is the least affected in this concern. Apart from that, all results confirm superior performance for the *SE* estimation approach.

VI. CONCLUSION AND FUTURE WORK

This paper presented a calibration method for eye-in-hand systems. These systems are the most common approach used to provide autonomy to robot manipulators.

Starting out, this work distinguishes between the two common solutions of the hand-eye calibration problem. A *novel physically-based metric on* $SE(3)$ is presented in order to optimally estimate within these model formulations. This metric in turn makes it possible to select the appropriate formulation in relation to the actual system characteristics. The metric, together with its *experimentally validated error model*, are the starting point in order to perform *optimal estimation* in the form of the Maximum Likelihood method – translational and rotational errors are weighted with the particular precision characteristics of the manipulator. Furthermore, *the weights are adapted automatically* by the algorithm.

Simulations and experiments illustrate the theoretical analysis. Comparison with other methods shows improved performance. Specific versions of some representative approaches were used and in particular a *novel weighting policy* for the well-known approach in [11] is presented.

Future work will include further accuracy experiments as well as the implementation in *X-from-motion* systems.

ACKNOWLEDGMENT

---

[11]Optimal estimation provides unbiased results only for numerous erroneous data. Here the common number of 10 stations is used.

REFERENCES

[1] Y. C. Shiu and S. Ahmad, "Calibration of Wrist–Mounted Robotic Sensors by Solving Homogeneous Tranform Equations of the Form $\boldsymbol{AX} = \boldsymbol{XB}$," *IEEE Trans. on Robs. and Aut.*, vol. 5, no. 1, pp. 16–29, February 1989.

[2] R. Y. Tsai and R. K. Lenz, "A New Technique for Fully Autonomous and Efficient 3D Robotics Hand/eye Calibration," *IEEE Trans. on Robs. and Aut.*, vol. 5, no. 3, pp. 345–358, June 1989.

[3] H. Zhuang and Z. S. Roth, "Comments on 'Calibration of Wrist–Mounted Robotic Sensors by Solving Homogenous Transform Equations of the Form $\boldsymbol{AX} = \boldsymbol{XB}$'," *IEEE Trans. on Robs. and Aut.*, vol. 7, no. 6, pp. 877–878, December 1991.

[4] J. Chou and M. Kamel, "Finding the Position and Orientation of a Sensor on a Robot Manipulator using Quaternions," *Int. Journal of Robs. Research*, vol. 10, no. 3, pp. 240–254, June 1991.

[5] H. H. Chen, "A Screw Motion Approach to Uniqueness Analysis of Head–Eye Geometry," in *Proc. of the IEEE Conf. on Comp. Vision and Pattern Recognition (CVPR)*, Maui, Hawaii, June 1991, pp. 145–151.

[6] C.-C. Wang, "Extrinsic Calibration of a Vision Sensor Mounted on a Robot," *IEEE Trans. on Robs. and Aut.*, vol. 8, no. 2, 161–175, 1992.

[7] H. Zhuang and Y. C. Shiu, "A Noise–Tolerant Algorithm for Robotic Hand–Eye Calibration with or without Sensor Orientation Measurement," *IEEE Trans. on Systems, Man and Cybernetics*, vol. 23, no. 4, pp. 1168–1175, July 1993.

[8] I. Fassi and G. Legnani, "Hand to Sensor Calibration: A Geometrical Interpretation of the Matrix Equation $\boldsymbol{AX} = \boldsymbol{XB}$," *Journal of Robotic Systems*, vol. 22, no. 9, pp. 497–506, 2005.

[9] F. C. Park and B. J. Martin, "Robot Sensor Calibration – Solving $\boldsymbol{AX} = \boldsymbol{XB}$ on the Euclidean Group," *IEEE Trans. on Robs. and Aut.*, vol. 10, no. 5, pp. 717–721, October 1994.

[10] Y.-C. Lu and J. C. Chou, "Eight–Space Quaternion Approach for Robotic Hand–Eye Calibration," in *Proc. of the IEEE Int. Conf. on Systems, Man and Cybernetics*, Vancouver, BC, Canada, Oct. 1995, pp. 3316–3321.

[11] R. Horaud and F. Dornaika, "Hand–Eye Calibration," *Int. Journal of Robs. Research*, vol. 14, no. 3, pp. 195–210, June 1995.

[12] G.-Q. Wei, K. Arbter, and G. Hirzinger, "Active Self–Calibration of Robotic Eyes and Hand–Eye Relationships with Model Identification," *IEEE Trans. on Robs. and Aut.*, vol. 14, no. 1, pp. 158–166, Feb. 1998.

[13] K. Daniilidis, "Hand–Eye Calibration Using Dual Quaternions," *Int. Journal of Robs. Research*, vol. 18, no. 3, pp. 286–298, June 1999.

[14] E. Bayro–Corrochano, K. Daniilidis, and G. Sommer, "Motor–Algebra for 3D Kinematics: The Case of Hand–Eye Calibration," *Journal for Mathem. Imaging and Vision*, vol. 13, no. 2, pp. 79–100, Oct. 2000.

[15] N. Andreff, R. Horaud, and B. Espiau, "Robot Hand–Eye Calibration Using Structure–from–Motion," *Int. Journal of Robs. Research*, vol. 20, no. 3, pp. 228–248, March 2001.

[16] H. Zhuang, Z. S. Roth, and R. Sudhakar, "Simultaneous Robot/World and Tool/Flange Calibration by Solving Homogeneous Transformation Equations of the Form $\boldsymbol{AX} = \boldsymbol{YB}$," *IEEE Trans. on Robs. and Aut.*, vol. 10, no. 4, pp. 549–554, August 1994.

[17] S. Rémy, M. Dhome, J. M. Lavest, and N. Daucher, "Hand–Eye Calibration," in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems IROS*, Grenoble, France, September 1997, pp. 1057–1065.

[18] F. Dornaika and R. Horaud, "Simultaneous Robot–World and Hand–Eye Calibration," *IEEE Trans. on Robs. and Aut.*, vol. 14, no. 4, pp. 617–622, August 1998.

[19] Z. Zhang, "A Flexible new Technique for Camera Calibration," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, November 2000.

[20] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

[21] FARO Technologies Inc. [Online]. Available: http://www.faro.com/

[22] A.R.T. GmbH. [Online]. Available: http://www.ar-tracking.de/

[23] CalLab 2005 and CalDe. Inst. of Robotics and Mechatronics, German Aerospace Center. [Online]. Available: http://www.robotic.dlr.de/callab/