Determining the Camera to Robot-body Transformation from Planar Mirror Reflections

Joel A. Hesch[1], Anastasios I. Mourikis[1,2] and Stergios I. Roumeliotis[1]
[1] : Dept. of Computer Science and Engineering, University of Minnesota, Minneapolis, MN 55455
[2] : Dept. of Electrical Engineering, University of California, Riverside, CA 92521
email:{joel|mourikis|stergios}@cs.umn.edu

***Abstract*— This paper presents a method for estimating the six-degrees-of-freedom transformation between a camera and the body of the robot on which it is rigidly attached. The robot maneuvers in front of a planar mirror, allowing the camera to observe fiducial features on the robot from several vantage points. Exploiting these measurements, we form a maximum-likelihood estimate of the camera-to-body transformation, without assuming prior knowledge of the robot motion or of the mirror configuration. Additionally, we estimate the mirror configuration with respect to the camera for each image. We validate the accuracy and correctness of our method with simulations and real-world experiments.**

I. Introduction

Mobile robots are often used in applications that require precise motion control among static or moving obstacles (e.g., parallel parking, navigating in cluttered spaces, avoiding overhanging obstacles, etc). The main prerequisite for designing controllers for these tasks is that the robot has precise estimates of the position and orientation (pose) of its body with respect to obstacles in the environment. Obtaining these estimates is a two-step process: first the measurements of different sensors must be fused to determine the *sensor* pose, and subsequently the sensor-pose estimate must be transformed to the *robot-body* pose estimate.

The first step of this process requires that the *sensor-to-sensor* transformation be known for combining the measurements, while the second step requires the *sensor-to-robot-body* transformation. The problem of sensor-to-sensor calibration has recently received significant attention and a number of approaches exist (e.g., for odometry-camera [1], IMU-camera [2], [3], or laser scanner-camera [4], [5], [6]). However, very little attention has been devoted to determining the *sensor-to-robot-body* transformation. This is necessary in order to precisely determine the distances of points on the robot body to obstacles detected by the robot's exteroceptive sensors (e.g., cameras, laser scanners, etc).

In most cases in practice, if CAD plots of a robot are available to a user, the sensor-to-robot transformation is approximated by measuring the position of the sensor with respect to the robot's center and assuming that the sensor's main axes are aligned with those of the robot body (for orientation). If no technical drawings of the robot are available, the same process can still be applied, by using a bounding box to approximate the robot shape. In both cases, the error in the sensor-to-robot transformation must be taken into account during path planning, by introducing a safety margin. Furthermore, when additional sensors or electronics are installed on the robot that exceed the dimensions of the bounding box, the same process needs to be repeated. Any such approximate method for determining the sensor-to-robot transformation limits the space where the robot can operate in and increases the probability of collisions.

Our objective in this paper is to automate the process of determining the sensor-to-robot transformation for the case of a mobile robot equipped with a camera. Considering the most limiting situation, where no part of the robot is within the camera's field of view, we concentrate on the case that the robot can observe points on its body through its reflection on a planar mirror. In our formulation, no prior information about the camera motion or the mirror position and orientation is assumed. The configuration of the mirror and the camera-to-robot transformation are treated as unknowns to be estimated. The measurement model derived for this problem is described in Section III-A, while Section III-B presents the estimator used for determining the unknown transformation from multiple camera images. The proposed algorithm is tested both in simulation (Section IV) and experimentally (Section V) to evaluate its accuracy and investigate its sensitivity to parameters of the system. Finally, Section VI discusses the conclusions of this work and suggests interesting directions of future research.

II. Related Work

Seemingly the problem closest to the one considered here is that of hand-eye calibration, i.e., that of determining the 6 degrees-of-freedom transformation between a camera and a tool both mounted on a robot manipulator [7], [8], [9], [10], [11]. However, the hand-eye calibration problem is solved by correlating the measurements of two sensors: the camera and the actuators' encoders measuring the robot-joints' displacements. This process allows one to determine the pose of the camera with respect to the robot base. Subsequently the camera-to-tool transformation is calculated by combining the estimated camera-to-robot-base transformation with the robot-base-to-tool one (assumed known from the robot kinematics and CAD plots depicting the location of the tool with respect to the last robot joint). Unless precise technical drawings of the robot are available, hand-eye calibration methods cannot be used for determining the transformation between the camera and any point on the robot body.

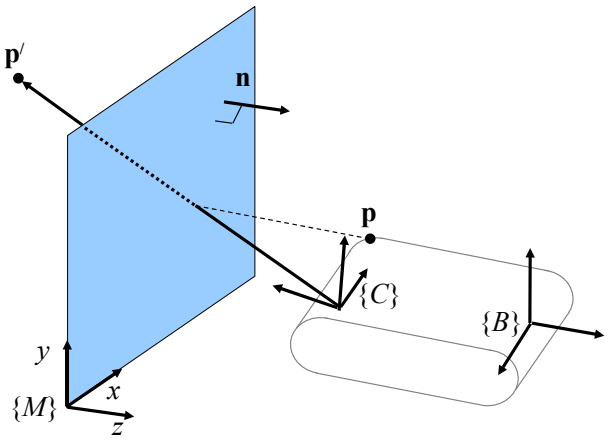

Fig. 1. An observation of a point reflection in the mirror.

At this point, we turn our attention to briefly describing cases where mirrors have been employed in robotics and computer vision applications. In [12], the authors present a Bayesian method for a humanoid robot attempting to distinguish itself (as seen in a mirror) from a human in its field of view. This is the so-called self-recognition problem, which is considered a test of intelligence in social animals.

Reflections from mirrors have also been used as "virtual cameras" [13] rigidly connected to a real camera to perform stereo vision using a single camera and planar [14], [15], [16], or conic mirrors [17]. Stereo vision is also accomplished using reflections from free-form surfaces [18], while systems with more view points are also possible [19]. Additionally, stereo systems using one moving mirror or a moving camera and two static spherical mirrors of known radii are presented in [20] and [21]. Finally, Jang *et al.* demonstrated a moving planar mirror system for 3D scene reconstruction [22]. Using a combination of fiducial points on the mirror and vanishing points in the reflections, they solve for the position of the mirror with respect to the camera. The 3D scene is determined based on synthetic stereo from multiple reflections.

In contrast to the approaches above, we do not use more than one view per time instant (i.e., the points of interest are only visible through the reflection). Additionally, we do not require the dimensions or the position of the mirror with respect to the camera to be known. Instead, using images of the robot-points' reflections on consecutive camera images we determine the mirror configuration and the camera-to-robot transformation.

III. Estimating the Transformation

In this section, we describe our approach for estimating the transformation between the camera frame, $\{C\}$, and a different frame of interest, $\{B\}$, both of which are rigidly attached to a robot. Frame $\{B\}$ could be the robot-body frame itself, or a frame affixed on any other sensor of the robot (cf. Section V). Without loss of generality, we will refer to frame $\{B\}$ as the "body frame." The main contribution of this paper is a method for estimating the transformation between $\{B\}$ and $\{C\}$, using the mirror reflections of known points in multiple images.

We consider a scenario in which the robot moves in front of a planar mirror (or equivalently, in which the mirror moves and the robot remains static). During this motion, the camera records N_c images of the robot's reflection; each image contains the reflections of N_p points, whose coordinates in $\{B\}$ are known. In Fig. 1, the setup for one camera pose and one feature is shown. All the recorded measurements are utilized in a maximum-likelihood estimator for determining the transformation between the frames $\{B\}$ and $\{C\}$.

A. Measurement Model

First, we present the measurement model that describes each of the camera observations. To simplify the presentation, in this section we focus on the case of a single point, observed in a single image. The treatment of multiple points is discussed in the next section. Consider a point $\mathbf{p}$, whose position with respect to frame $\{B\}$, ${}^B\mathbf{p}$, is known[1], for instance from a CAD plot of the robot. The reflection of this point, $\mathbf{p}'$, is observed by the camera, and this observation is described by the perspective projection model:

$$\mathbf{z} = \frac{1}{p_3}\begin{bmatrix} p_1 \\ p_2 \end{bmatrix} + \boldsymbol{\eta} = \mathbf{h}({}^C\mathbf{p}') + \boldsymbol{\eta}, \qquad {}^C\mathbf{p}' = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} \tag{1}$$

where $\boldsymbol{\eta}$ is the measurement noise, assumed to be zero-mean Gaussian with covariance matrix $\sigma_\eta^2\mathbf{I}_2$, and ${}^C\mathbf{p}'$ is the position of the reflected point with respect to the camera frame (cf. Fig. 1). Our goal is to express ${}^C\mathbf{p}'$ as a function of the *known* position vector ${}^B\mathbf{p}$, the *unknown* transformation (rotation and translation) between the camera and body frame, $\{{}^C_B\mathbf{R}, {}^C\mathbf{p}_B\}$, and the *unknown* configuration of the mirror with respect to the camera. As shown in the following derivations, only three degrees of freedom of this configuration affect the measurement equation.

Without loss of generality, we assign a coordinate frame $\{M\}$ to the mirror, such that its x-y plane is the reflective surface, and the z axis is pointing towards the robot (cf. Fig. 1). Specifically, with all quantities expressed in the mirror frame, the reflection of a point ${}^M\mathbf{p}$ is a point ${}^M\mathbf{p}'$ whose z coordinate is negated:

$${}^M\mathbf{p}' = \mathbf{A}\,{}^M\mathbf{p}, \quad \text{where} \quad \mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}. \tag{2}$$

Note that matrix $\mathbf{A}$ can be written as

$$\mathbf{A} = \mathbf{I}_3 - 2\mathbf{e}_3\mathbf{e}_3^{\mathrm{T}} \tag{3}$$

where $\mathbf{I}_3$ is the 3×3 identity matrix, and $\mathbf{e}_3 = [0 \;\; 0 \;\; 1]^{\mathrm{T}}$ is the unit vector along the z axis, i.e., the unit vector normal to the mirror.

[1]Throughout this paper, ${}^X\mathbf{y}$ denotes the expression of a vector $\mathbf{y}$ with respect to frame $\{X\}$, ${}^X_W\mathbf{R}$ is the rotation matrix rotating vectors from frame $\{W\}$ to frame $\{X\}$, and ${}^X\mathbf{p}_W$ are is the position of the origin of frame $\{W\}$, expressed with respect to frame $\{X\}$.

The vector from the origin of the camera frame to the reflected point $\mathbf{p}'$, expressed in the camera frame, is:

$$\begin{aligned} {}^C\mathbf{p}' &= {}^C_M\mathbf{R}\,{}^M\mathbf{p}' + {}^C\mathbf{p}_M \\ &= {}^C_M\mathbf{R}\mathbf{A}\,{}^M\mathbf{p} + {}^C\mathbf{p}_M. \end{aligned} \quad (4)$$

Substituting the expression ${}^M\mathbf{p} = {}^C_M\mathbf{R}^{\mathrm{T}}\left({}^C\mathbf{p} - {}^C\mathbf{p}_M\right)$ and rearranging terms in the above equation yields

$${}^C\mathbf{p}' = \left({}^C_M\mathbf{R}\mathbf{A}\,{}^C_M\mathbf{R}^{\mathrm{T}}\right){}^C\mathbf{p} + \left(\mathbf{I}_3 - {}^C_M\mathbf{R}\mathbf{A}\,{}^C_M\mathbf{R}^{\mathrm{T}}\right){}^C\mathbf{p}_M \quad (5)$$

Recalling the definition of $\mathbf{A}$ in (3), we write:

$$\begin{aligned} {}^C_M\mathbf{R}\mathbf{A}\,{}^C_M\mathbf{R}^{\mathrm{T}} &= {}^C_M\mathbf{R}\left(\mathbf{I}_3 - 2\mathbf{e}_3\mathbf{e}_3^{\mathrm{T}}\right){}^C_M\mathbf{R}^{\mathrm{T}} \\ &= \mathbf{I}_3 - 2\,{}^C\mathbf{n}\,{}^C\mathbf{n}^{\mathrm{T}}, \end{aligned} \quad (6)$$

where ${}^C\mathbf{n} = {}^C_M\mathbf{R}\mathbf{e}_3$ is the mirror's normal vector expressed in frame $\{C\}$. Substitution in (5) yields:

$$\begin{aligned} {}^C\mathbf{p}' &= \left(\mathbf{I}_3 - 2\,{}^C\mathbf{n}\,{}^C\mathbf{n}^{\mathrm{T}}\right){}^C\mathbf{p} + 2\left({}^C\mathbf{n}^{\mathrm{T}}\,{}^C\mathbf{p}_M\right){}^C\mathbf{n} \\ &= \left(\mathbf{I}_3 - 2\,{}^C\mathbf{n}\,{}^C\mathbf{n}^{\mathrm{T}}\right){}^C\mathbf{p} + 2d\,{}^C\mathbf{n} \end{aligned} \quad (7)$$

where we have defined $d = {}^C\mathbf{n}^{\mathrm{T}}\,{}^C\mathbf{p}_M$, which is the camera-to-mirror distance. At this point, we employ the following expression for the position of the point $\mathbf{p}$ in the camera frame:

$${}^C\mathbf{p} = {}^C_B\mathbf{R}\,{}^B\mathbf{p} + {}^C\mathbf{p}_B \quad (8)$$

Using this expression in (7), we obtain:

$${}^C\mathbf{p}' = \left(\mathbf{I}_3 - 2\,{}^C\mathbf{n}\,{}^C\mathbf{n}^{\mathrm{T}}\right)\left({}^C_B\mathbf{R}\,{}^B\mathbf{p} + {}^C\mathbf{p}_B\right) + 2d\,{}^C\mathbf{n} \quad (9)$$

Finally, by letting $\mathbf{v} = d\,{}^C\mathbf{n}$ denote the vector from the origin of the camera frame to the mirror surface, and rearranging terms, the expression becomes

$${}^C\mathbf{p}' = \left(\mathbf{I}_3 - 2\frac{\mathbf{v}\mathbf{v}^{\mathrm{T}}}{\mathbf{v}^{\mathrm{T}}\mathbf{v}}\right){}^C_B\mathbf{R}\,{}^B\mathbf{p} + \left(\mathbf{I}_3 - 2\frac{\mathbf{v}\mathbf{v}^{\mathrm{T}}}{\mathbf{v}^{\mathrm{T}}\mathbf{v}}\right){}^C\mathbf{p}_B + 2\mathbf{v}.$$

This equation, along with (1), defines the measurement model for observing the reflection of the point $\mathbf{p}$. It involves the known position, ${}^B\mathbf{p}$, of the point with respect to frame $\{B\}$, the unknown transformation, $\{{}^C_B\mathbf{R}, {}^C\mathbf{p}_B\}$, between the camera and the body frame, and the configuration of the mirror with respect to the camera. We point out that, even though the transformation between the mirror and camera frame has six degrees of freedom, only three of these degrees of freedom appear in the measurement equation. These are expressed by the vector $\mathbf{v} = d\,{}^C\mathbf{n}$, which has two degrees of freedom from the mirror normal, ${}^C\mathbf{n}$, and one degree of freedom from the camera-to-mirror distance, d. The remaining three degrees of freedom, which correspond to rotation about $\mathbf{v}$ and to translations of the origin of the mirror frame in the mirror plane, do not affect the measurements, and are unobservable.

B. Maximum Likelihood Estimation of the Transformation

We now proceed with the description of a maximum likelihood estimator (MLE) for determining the unknown transformation between the camera and body frames. We consider the case where N_p points in the body frame, denoted as ${}^B\mathbf{p}_i$, $i = 1 \ldots N_p$, are observed in N_c images of the camera. The observation of the ith point in the jth image ($j = 1 \ldots N_c$) is given by the equation (cf. (1)):

$$\mathbf{z}_{ij} = \mathbf{h}({}^{C_j}\mathbf{p}'_i) + \boldsymbol{\eta}_{ij}, \qquad \text{where}$$

$${}^{C_j}\mathbf{p}'_i = \left(\mathbf{I}_3 - 2\frac{\mathbf{v}_j\mathbf{v}_j^{\mathrm{T}}}{\mathbf{v}_j^{\mathrm{T}}\mathbf{v}_j}\right){}^C_B\mathbf{R}\,{}^B\mathbf{p}_i + \left(\mathbf{I}_3 - 2\frac{\mathbf{v}_j\mathbf{v}_j^{\mathrm{T}}}{\mathbf{v}_j^{\mathrm{T}}\mathbf{v}_j}\right){}^C\mathbf{p}_B + 2\mathbf{v}_j$$

The vector $\mathbf{v}_j$ describes the mirror configuration when the jth image is recorded. In the following, we use $\mathcal{Z}$ to denote the set of all available measurements.

It is interesting to examine the number of unknown parameters that exist in our problem. The unknown camera-to-body transformation $\{{}^C_B\mathbf{R}, {}^C\mathbf{p}_B\}$ introduces six degrees of freedom, while each new image introduces an additional three unknowns, corresponding to the mirror configuration $\mathbf{v}_j$. Thus, the total number of unknown parameters is $6+3N_c$. On the other hand, each point observation provides two independent scalar measurements (the image coordinates of the projection), and thus we obtain a total of $2N_pN_c$ measurements, which we employ for estimating all unknown parameters. When the number of measurements equals or exceeds the number of unknowns, we expect to be able to compute a solution for all unknown parameters. It should be noted that a minimum of $N_p = 3$ non-collinear points observed in $N_c = 3$ images from different viewing angles are required for obtaining a solution. The 3 non-collinear points are necessary to define a coordinate frame. If only points on a single line are used, the frame's rotation about the line cannot be determined. Moreover, using only 2 camera views (or more which differ by rotations about a single axis) is not sufficient to uniquely determine the transformation between the camera and robot-body frames [23].

Let the vector of all unknown parameters be denoted by $\mathbf{x}$. This vector comprises the unknown transformation, as well as the vectors $\mathbf{v}_j$, $j = 1 \ldots N_c$, that describe the mirror configuration. In our implementation, we adopt the unit-quaternion representation of rotation [24], and thus $\mathbf{x}$ is:

$$\mathbf{x} = \begin{bmatrix} {}^C\mathbf{p}_B^{\mathrm{T}} & {}^C\bar{q}_B^{\mathrm{T}} & \mathbf{v}_1^{\mathrm{T}} & \ldots & \mathbf{v}_{N_c}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \quad (10)$$

where ${}^C\bar{q}_B$ is the unit quaternion representation of the rotation between frames $\{B\}$ and $\{C\}$. The likelihood of the measurements is given by:

$$\begin{aligned} L(\mathcal{Z};\mathbf{x}) &= \prod_{i=1}^{N_p}\prod_{j=1}^{N_c} p(\mathbf{z}_{ij};\mathbf{x}) \\ &= \prod_{i=1}^{N_p}\prod_{j=1}^{N_c} \frac{1}{2\pi\sigma_\eta^2}\exp\left[-\frac{\left(\mathbf{z}_{ij}-\mathbf{h}\left({}^{C_j}\mathbf{p}'_i\right)\right)^{\mathrm{T}}\left(\mathbf{z}_{ij}-\mathbf{h}\left({}^{C_j}\mathbf{p}'_i\right)\right)}{2\sigma_\eta^2}\right] \\ &= \prod_{i=1}^{N_p}\prod_{j=1}^{N_c} \frac{1}{2\pi\sigma_\eta^2}\exp\left[-\frac{(\mathbf{z}_{ij}-\mathbf{h}_{ij}(\mathbf{x}))^{\mathrm{T}}(\mathbf{z}_{ij}-\mathbf{h}_{ij}(\mathbf{x}))}{2\sigma_\eta^2}\right] \end{aligned}$$

where the dependence on $\mathbf{x}$ is explicitly shown. Maximizing the likelihood is equivalent to maximizing its logarithm,

which in turn is equivalent to minimizing the quantity:

$$J(\mathbf{x}) = \sum_{i=1}^{N_p}\sum_{j=1}^{N_c}(\mathbf{z}_{ij} - \mathbf{h}_{ij}(\mathbf{x}))^{\mathrm{T}}(\mathbf{z}_{ij} - \mathbf{h}_{ij}(\mathbf{x})) \qquad (11)$$

The minimization of this cost function is a nonlinear least-squares problem, and thus we employ the Gauss-Newton iterative minimization algorithm for estimating $\mathbf{x}$. During each iteration k of the algorithm, the estimate is changed by:

$$\delta\mathbf{x}^{(k)} = \left(\sum_{i,j}\mathbf{H}_{ij}^{(k)\mathrm{T}}\mathbf{H}_{ij}^{(k)}\right)^{-1}\left(\sum_{i,j}\mathbf{H}_{ij}^{(k)\mathrm{T}}\left(\mathbf{z}_{ij} - \mathbf{h}_{ij}(\mathbf{x}^{(k)})\right)\right)$$

where $\mathbf{H}_{ij}^{(k)}$ is the Jacobian of the measurement $\mathbf{z}_{ij}$ with respect to $\mathbf{x}$, evaluated at the current iterate, $\mathbf{x}^{(k)}$. The exact structure of this matrix is given in the Appendix. It is worth mentioning that because this structure is sparse, the matrix to be inverted in the above equation is a sparse one. Thus, $\delta\mathbf{x}^{(k)}$ can be evaluated very efficiently (the computational cost of the operation can be shown to be linear in the number of images).

The parameter correction, $\delta\mathbf{x}^{(k)}$, has the following structure:

$$\delta\mathbf{x}^{(k)} = \begin{bmatrix} \delta{}^{C}\mathbf{p}_{B}^{(k)} \\ \delta\boldsymbol{\theta}^{(k)} \\ \delta\mathbf{v}_1^{(k)} \\ \vdots \\ \delta\mathbf{v}_{N_c}^{(k)} \end{bmatrix} \qquad (12)$$

where all vectors on the right-hand side are 3×1 vectors. With this notation, the updates for the iterates of the parameters ${}^{C}\mathbf{p}_B$ and $\mathbf{v}_j$ are written as:

$$\begin{aligned} {}^{C}\mathbf{p}_{B}^{(k+1)} &= {}^{C}\mathbf{p}_{B}^{(k)} + \delta{}^{C}\mathbf{p}_{B}^{(k)} \\ \mathbf{v}_j^{(k+1)} &= \mathbf{v}_j^{(k)} + \delta\mathbf{v}_j^{(k)}, \quad j = 1\ldots N_c \end{aligned}$$

To ensure that the unit-length quaternion constraint is properly accounted for, a multiplicative error model is used for the quaternion iterates [24]:

$$\begin{aligned} {}^{C}\bar{q}_{B}^{(k+1)} &= \delta\bar{q}^{(k)} \otimes {}^{C}\bar{q}_{B}^{(k)}, \quad \text{with} \\ \delta\bar{q}^{(k)} &= \begin{bmatrix} \frac{1}{2}\delta\boldsymbol{\theta}^{(k)} \\ \sqrt{1 - \frac{1}{4}\delta\boldsymbol{\theta}^{(k)\mathrm{T}}\delta\boldsymbol{\theta}^{(k)}} \end{bmatrix}. \end{aligned} \qquad (13)$$

where $\otimes$ denotes quaternion multiplication. Employing this formulation for the quaternion updates enables us to have minimal error parametrization, since $\delta\boldsymbol{\theta}^{(k)}$ is a 3×1 vector.

After the Gauss-Newton algorithm converges to a minimum (convergence is determined by a threshold on the norm of $\delta\mathbf{x}^{(k)}$), the covariance of the resulting parameter estimates can be determined by the expression:

$$\mathbf{P} = \sigma_\eta^2\left(\sum_{i,j}\mathbf{H}_{ij}^{(k)\mathrm{T}}\mathbf{H}_{ij}^{(k)}\right)^{-1} \qquad (14)$$

IV. Simulations

In this section, we present simulation results that demonstrate the feasibility of computing the camera-to-body transformation using the proposed approach. Moreover, the effects of several parameters on the estimates' accuracy are studied.

A. MLE convergence

First, we describe the results of Monte-Carlo simulations, which verify the correctness of the estimates computed by the MLE. In Figs. 2(a) and 2(b), we plot the estimation errors (solid blue lines) and corresponding 3σ values (dashed red lines) for the camera position and attitude estimates, respectively. For these plots, 50 Monte Carlo trials were carried out, in each of which four points were tracked in 250 images. In our simulation setup the robot remains static, while the mirror moves to generate different views of the robot. Specifically, the mirror was placed at a distance of 0.5 m in front of the camera, and rotated by a range of 25° about both the vertical and the horizontal directions. The plots of Figs. 2(a) and 2(b) show that the estimate errors are commensurate with the computed covariance, which indicates that the MLE correctly converges to the global minimum. Regarding the accuracy of the estimates, we note that the estimate for the translation along the camera optical axis (z axis) is the least accurate. This makes sense intuitively, since the estimation of depth using perspective cameras is typically less accurate than the estimation of bearing.

The initial estimates in each Monte Carlo trial are selected as follows: for the body-to-camera transformation, the initial estimates are corrupted by randomly generated errors, with standard deviation equal to 2 cm for position, and 5° for attitude. In a real-world setting, this level of accuracy can easily be accomplished by manual measurement. For the mirror normal vector we use as initial guess a vector parallel to the camera optical axis, and finally for the mirror distance, a value corrupted by error with standard deviation equal to 5% is used. It should be pointed out that in all our simulation tests, the MLE converged to the correct minimum (after an average of 15 iterations), even though the described initialization for the mirror configuration is crude. This is very important for practical purposes, since obtaining an accurate initial guess for the mirror configuration can be challenging in an experimental setup.

B. Parameters affecting estimation accuracy

We next turn our attention to studying the accuracy of the estimated camera-to-body transformation. In particular, we explore the effects of the following parameters: (i) number of images, (ii) camera-to-mirror distance, (iii) range of the mirror's angular motion, and (iv) number of points tracked on the robot. We consider a "base" case, in which four points are observed in 200 images, while a mirror placed at a distance of 0.5 m is rotated by 25° in two directions. Then, we vary each of the aforementioned parameters individually, in order to examine its effects on the estimation accuracy. The plots of Fig. 3 show the standard deviation for the position and

attitude estimates. In each subplot, the standard deviation for the least accurate among the three axes is reported. Some key observations are the following:

- As expected, increasing the number of images results in better estimation accuracy. However, the improvement follows a "law of diminishing return," i.e., when a large number of images is already available, the impact of recording more observations is smaller.
- Changing the distance of the mirror to the camera has a very significant effect on the accuracy of the estimates. For instance, for the particular simulation setup keeping the mirror at a distance of 0.8 m results in standard deviation for the camera position equal to approximately 1 cm. This level of accuracy can potentially be achieved with manual measurements alone, and thus in this case the use of the mirror may not be justified. Hence, it becomes clear that the mirror distance should be kept as small as possible.
- As we would expect, increasing the range of the mirror's angular motion results in improved estimation accuracy. The effect on the attained accuracy is significant, and thus every effort should be made to move the mirror in the widest range of motion allowed by the camera's field of view.
- Increasing the number of observed points increases the estimation accuracy. We point out that for generating Fig. 3(d), points were randomly placed in a cube with side length equal to 20 cm. For each number of points, 10 Monte Carlo trials were run, and the reported accuracy is the average of these trials.

V. Experiments

The method described in the preceding sections was employed for computing the transformation between a camera and the frame attached on a SICK laser scanner. For this goal, three fiducial points were placed in known positions on the laser range finder (cf. Fig. 4). The origin of frame $\{B\}$ coincides with the top-left fiducial point, while both frames $\{B\}$ and $\{C\}$ are right-hand systems with the axes of $\{B\}$ approximately aligned with those of $\{C\}$. The three fiducial points were tracked using the KLT algorithm [25], [26] in 1000 images, recorded by a Firewire camera with resolution of 1024×768 pixels. In this particular configuration, the camera is placed facing forward, while the SICK is facing upward[2].

A planar mirror was maneuvered in different spacial configurations, and in distances varying between 30 and 50 cm from the camera, in order to generate a wide range of views. All the measurements were processed in the MLE described in Section III-B, to obtain an estimate for the transformation between the two frames of interest. The Gauss-Newton minimization converged after 17 iterations, to the following solution for the transformation: ${}^{C}\mathbf{p}_{B} = \begin{bmatrix} -14.80 & -15.96 & -14.95 \end{bmatrix}^{\mathrm{T}}$ cm, and

[2]This non-standard mounting of the laser scanner is chosen to allow building 3D models of the environment, by having the robot move while the SICK records vertical "slices" of its surroundings.

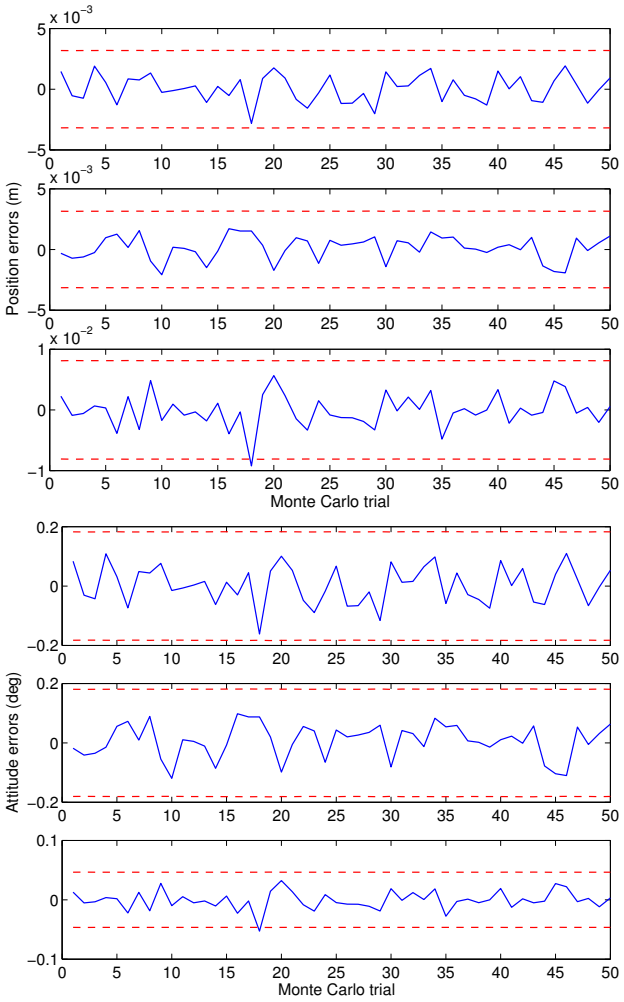

Fig. 2. The estimation errors and corresponding 3σ bounds for the camera-to-body transformation, in 50 Monte-Carlo simulation trials. The 3σ values are computed as three times the square root of the corresponding diagonal elements of the covariance matrix.

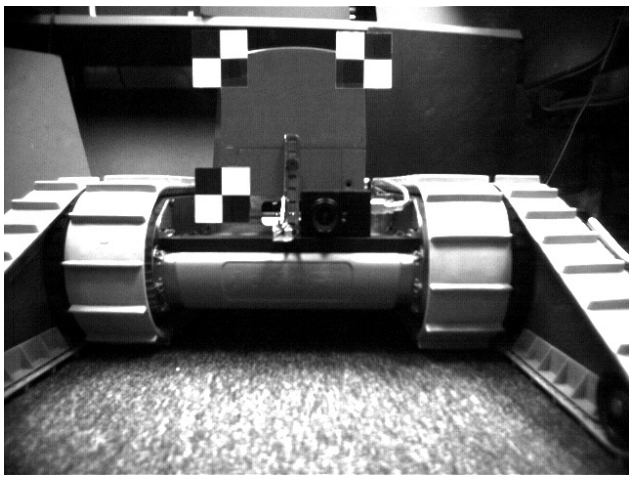

Fig. 4. An example image recorded in the experiment. The three fiducial points are seen in the corners of the upward-facing SICK laser scanner.

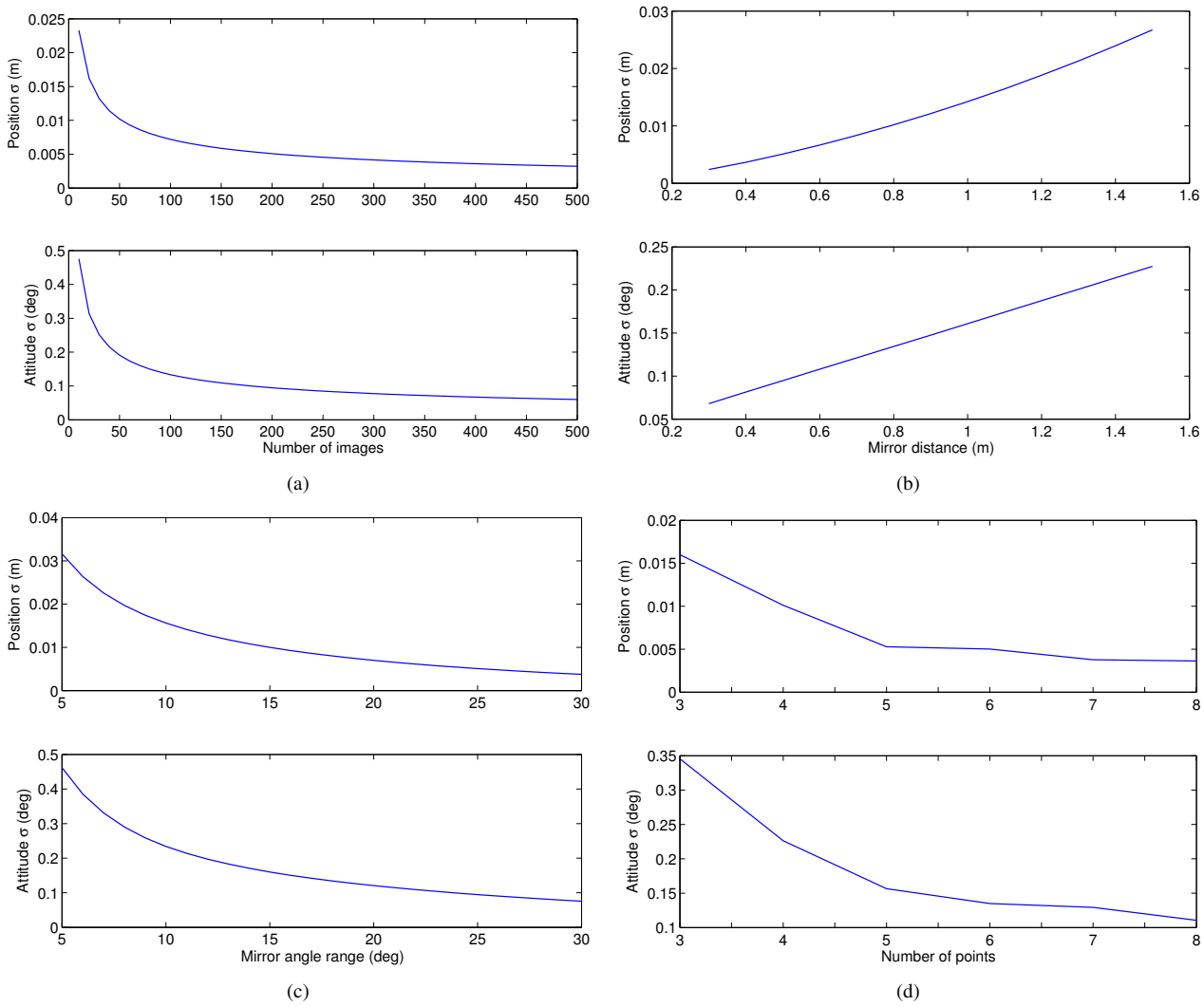

Fig. 3. The standard deviation for the position and attitude, as a function of various parameters affecting the estimation accuracy: (a) Number of recorded images. (b) Distance between the mirror and the camera. (c) Range of the mirror's angular motion. (d) Number of points tracked on the robot body.

${}^{C}\mathbf{q}_{B} = \begin{bmatrix}0.0045 & 0.0774 & 0.0389 & 0.9962\end{bmatrix}^{\mathrm{T}}$. The corresponding 3σ accuracy bounds are $\begin{bmatrix}1.1 & 1.6 & 5.0\end{bmatrix}$ mm for the position, and $\begin{bmatrix}0.24 & 0.23 & 0.06\end{bmatrix}$ degrees for the orientation estimates. These results agree with the values computed by manual measurement. Most importantly, we point out that the accuracy attained by the MLE is substantially higher than that attained by manual measurement, which demonstrates the usefulness of the proposed method.

VI. Conclusions and Future Work

In this paper, we propose a method for estimating the 6 degrees-of-freedom transformation between the camera and robot-body frames. To this end, the camera observes the reflections in a planar mirror of known robot points, while the robot moves in front of the mirror (or equivalently, while the mirror moves in front of the camera). The measurements collected are processed in a maximum-likelihood estimator, which produces estimates for the camera-to-robot-body transformation, as well as for the mirror configuration in each of the images. The approach was validated both in simulation and using real-world experimentation. One of the key advantages of the proposed method is its ease of use, as the only information needed are the coordinates of a number of points on the robot, which can be obtained from CAD plots. The mirror configuration is concurrently estimated, and need not be known in advance.

In our future work, we plan to extend this method to the case where the coordinates of the points on the robot-body frame are not known *a priori*, but are estimated along with the body-to-camera transformation and the mirror configurations. It can be shown that in this case the scale is unobservable, and therefore the distance between at least two of the points should be manually measured. However, this is not difficult to accomplish, and does not require availability of CAD plots for the robot. Moreover, we plan to investigate the observability properties of the camera-to-robot transformation. Through a nonlinear observability analysis [27], we intend to identify singular cases of motion, for which the transformation cannot be determined. This is important for practical purposes, as it would enable us to avoid such cases,

and ensure estimates of sufficient accuracy. Finally, we will seek to compute an initial guess for the sought parameters in closed form. If such a guess were available, it would enable a better initialization of the nonlinear optimization algorithm, which would, in turn, speed up convergence.

APPENDIX A

The Jacobian of the measurement $\mathbf{z}_{ij}$ with respect to $\mathbf{x}$, $\mathbf{H}_{ij}$, is given by:

$$\mathbf{H}_{ij} = \mathbf{H}_{c_{ij}} \begin{bmatrix} \mathbf{H}_{p_{ij}} & \mathbf{H}_{q_{ij}} & 0 & \dots & \underbrace{\mathbf{H}_{v_{ij}}}_{j-\text{th image}} & \dots & 0 \end{bmatrix}$$

where $\mathbf{H}_{c_{ij}}$ is the Jacobian of the perspective projection model with respect to ${}^{C_j}\mathbf{p}'_i$:

$$\mathbf{H}_{c_{ij}} = \frac{1}{p_3} \begin{bmatrix} 1 & 0 & -\frac{p_1}{p_3} \\ 0 & 1 & -\frac{p_2}{p_3} \end{bmatrix}$$

and $\mathbf{H}_{p_{ij}}$, $\mathbf{H}_{q_{ij}}$, and $\mathbf{H}_{v_{ij}}$, are the Jacobians of ${}^{C_j}\mathbf{p}'_i$ with respect to the position, rotation, and mirror configuration, respectively:

$$\mathbf{H}_{p_{ij}} = \mathbf{I}_3 - 2\frac{\mathbf{v}_j\mathbf{v}_j^{\mathrm{T}}}{\mathbf{v}_j^{\mathrm{T}}\mathbf{v}_j}$$

$$\mathbf{H}_{q_{ij}} = \left(\mathbf{I}_3 - 2\frac{\mathbf{v}_j\mathbf{v}_j^{\mathrm{T}}}{\mathbf{v}_j^{\mathrm{T}}\mathbf{v}_j}\right) \lfloor {}^{C}_{B}\mathbf{R}\,{}^{R}\mathbf{p}_i \times \rfloor$$

$$\mathbf{H}_{v_{ij}} = 2\left(1 - \frac{\mathbf{v}_j^{\mathrm{T}}{}^{C}\mathbf{p}_i}{\mathbf{v}_j^{\mathrm{T}}\mathbf{v}_j}\right)\mathbf{I}_3 - 2\frac{\mathbf{v}_j\,{}^{C}\mathbf{p}_i^{\mathrm{T}}}{\mathbf{v}_j^{\mathrm{T}}\mathbf{v}_j} + 4\mathbf{v}_j\mathbf{v}_j^{\mathrm{T}}\frac{\mathbf{v}_j^{\mathrm{T}}{}^{C}\mathbf{p}_i}{\left(\mathbf{v}_j^{\mathrm{T}}\mathbf{v}_j\right)^2}$$

ACKNOWLEDGMENTS

This work was supported by the University of Minnesota (DTC), and the National Science Foundation (EIA-0324864, IIS-0643680). Anastasios Mourikis was supported by the UMN Doctoral Dissertation Fellowship. Joel Hesch was supported by the NIH Neuro-physical-computational Sciences Fellowship.

REFERENCES

[1] A. Martinelli and R. Siegwart, "Observability properties and optimal trajectories for on-line odometry self-calibration," in *Proceedings of the IEEE Conference on Decision and Control*, San Diego, CA, Dec. 13–15, 2006, pp. 3065–3070.

[2] F. M. Mirzaei and S. I. Roumeliotis, "A Kalman filter-based algorithm for IMU-camera callibration," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, San Diego, CA, Oct. 29–Nov. 2, 2007, pp. 2427–2434.

[3] ——, "A Kalman filter-based algorithm for IMU-camera calibration: Observability analysis and performance evaluation," *IEEE Transactions on Robotics*, 2008, (to appear).

[4] X. Brun and F. Goulette, "Modeling and calibration of coupled fish-eye CCD camera and laser range scanner for outdoor environment reconstruction," in *Proceedings of the International Conference on 3D Digital Imaging and Modeling*, Montreal, QC, Canada, Aug. 21–23, 2007, pp. 320–327.

[5] Q. Zhang and R. Pless, "Extrinsic calibration of a camera and laser range finder (improves camera calibration)," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sendai, Japan, Sept. 28–Oct. 2, 2004, pp. 2301–2306.

[6] S. Wasielewski and O. Strauss, "Calibration of a multi-sensor system laser rangefinder/camera," in *Proceedings of the Intelligent Vehicles Symposium*, Detroit, MI, Sept. 25–26, 1995, pp. 472–477.

[7] R. Y. Tsai and R. K. Lenz, "A new technique for fully autonomous and efficient 3D robotics hand/eye calibration," *IEEE Transactions on Robotics and Automation*, vol. 5, no. 3, pp. 345–358, June 1989.

[8] K. Daniilidis, "Hand-eye calibration using dual quaternions," *International Journal of Robotics Research*, vol. 18, no. 3, pp. 286–298, 1999.

[9] N. Andreff, R. Horaud, and B. Espiau, "Robot hand-eye calibration using structure-from-motion," *International Journal of Robotics Research*, vol. 20, no. 30, pp. 228–248, 2001.

[10] Y. Shiu and S. Ahmad, "Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form AX = XB," *IEEE Transactions on Robotics and Automation*, vol. 5, no. 1, pp. 16–29, 1989.

[11] J. Chou and M. Kamel, "Finding the position and orientation of a sensor on a robot manipulator using quaternions," *International Journal of Robotics Research*, vol. 10, no. 3, pp. 240–254, 1991.

[12] K. Gold and B. Scassellati, "A Bayesian robot that distinguishes "self" from "other"," in *Proceedings of the 29th Annual Meeting of the Cognitive Science Society*, Nashville, TN, Aug. 1–4, 2007.

[13] S. Baker and S. K. Nayar, "A theory of single-viewpoint catadioptric image formation," *International Journal of Computer Vision*, vol. 35, no. 2, pp. 175–196, Nov. 1999.

[14] J. Gluckman and S. K. Nayar, "Planar catadioptric stereo: Geometry and calibration," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Ft. Collins, CO, June 23–25, 1999, pp. 22–28.

[15] A. A. Goshtasby and W. A. Gruver, "Design of a single-lens stereo camera system," *Pattern Recognition*, vol. 26, no. 6, pp. 923–937, June 1993.

[16] M. Inaba, T. Hara, and H. Inoue, "A stereo viewer based on a single camera with view-control mechanisms," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Yokohama, Japan, July 26–30, 1993, pp. 1857–1865.

[17] G. Jang, S. Kim, and I. Kweon, "Single camera catadioptic stereo system," in *Proceedings of the 6th Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras (OMNIVIS/ICCV)*, Beijing, China, Oct. 21, 2005.

[18] A. Würz-Wessel and F. K. Stein, "Calibration of a free-form surface mirror in a stereo vision system," in *Proceedings of the IEEE Intelligent Vehicle Symposium*, vol. 2, Versailles, France, June 17–21, 2002, pp. 471–476.

[19] B. K. Ramsgaard, I. Balslev, and J. Arnspang, "Mirror-based trinocular systems in robot-vision," in *Proceedings of the IEEE International Conference on Pattern Recognition*, Barcelona, Spain, Sept. 3–7, 2000, pp. 499–502.

[20] M. Kanbara, N. Ukita, M. Kidode, and N. Yokoya, "3D scene reconstruction from reflection images in a spherical mirror," in *Proceedings of the IEEE 18th International Conference on Pattern Recognition*, Hong Kong, China, Aug. 20–24, 2006, pp. 874–897.

[21] S. K. Nayar, "Sphereo: Determining depth using two specular spheres and a single camera," in *Proceedings of the SPIE conference on Optics, Illumination, and Image Sensing for Machine Vision*, vol. 1005, Nov. 1988, pp. 245–254.

[22] K. H. Jang, D. H. Lee, and S. K. Jung, "A moving planar mirror based approach for cultural reconstruction," *Computer Animation and Virtual Worlds*, vol. 15, no. 3–4, pp. 415–423, July 2004.

[23] J. A. Hesch, A. I. Mourikis, and S. I. Roumeliotis, "Camera to robot-body calibration using planar mirror reflections," University of Minnesota, Dept. of Comp. Sci. & Eng., Tech. Rep. 2008-001, July 2008.

[24] N. Trawny and S. I. Roumeliotis, "Indirect Kalman filter for 3D attitude estimation," University of Minnesota, Dept. of Comp. Sci. & Eng., Tech. Rep. 2005-002, Mar. 2005.

[25] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proceedings of the International Joint Conference on Artificial Intelligence*, Vancouver, B.C., Canada, Aug. 24–28, 1981, pp. 674–679.

[26] J. Shi and C. Tomasi, "Good features to track," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Washington, DC, June 27–July 2, 1994, pp. 593–600.

[27] R. Hermann and A. Krener, "Nonlinear controllability and observability," *IEEE Transactions on Automatic Control*, vol. 22, no. 5, pp. 728–740, Oct. 1977.