An Analytical Least-Squares Solution
to the Odometer-Camera Extrinsic Calibration Problem

Chao X. Guo, Faraz M. Mirzaei, and Stergios I. Roumeliotis

Abstract— In order to fuse camera and odometer measurements, we first need to estimate their relative transformation through the so-called odometer-camera extrinsic calibration. In this paper, we present a two-step analytical least-squares solution for the extrinsic odometer-camera calibration that (i) is not iterative and finds the least-squares optimal solution without any initialization, and (ii) does not require any special hardware or the presence of known landmarks in the scene. Specifically, in the first step, we estimate a subset of the 3D relative rotation parameters by analytically minimizing a least-squares cost function. We then back-substitute these estimates in the measurement constraints, and determine the rest of the 3D transformation parameters by analytically minimizing a second least-squares cost function. Simulation and experimental results are presented that validate the efficiency and accuracy of the proposed algorithm.

I. INTRODUCTION

Cameras are widely used in robotics due to their small size, low energy expenditure, and ability to provide rich information about the robots' surroundings. Through captured images, we can observe and track landmarks (feature points) in consecutive images, from which, we may determine the camera's motion. Nevertheless, cameras have several practical limitations. First, if we cannot track a sufficient number of features in a scene during the camera motion, it may not be possible to estimate the transformation between camera poses. Second, even if we can accurately track enough points in two consecutive images, the translation between the two corresponding camera poses can only be determined up to scale. Finally, in the presence of noise, if the baseline of the camera motion is too short, the computed translation between camera poses may be very inaccurate.

Odometers are the other widely used sensors for robot localization in 2D. By integrating the encoder measurements from the odometer, we can determine the rotation and translation between robot poses. In contrast to camera observations, odometry measurements are usually accurate over short periods of time. Their open loop integration (i.e., dead-reckoning) over long periods of time, however, leads to unbounded growth in the localization uncertainty.

Combining measurements of cameras and odometers is a common approach for compensating for their individual limitations. Optimally fusing the measurements of these two sensors, however, requires accurate odometer-camera

The authors are with the Department of Computer Science & Engineering, University of Minnesota, Minneapolis, MN 55455, USA {chaguo|faraz|stergios}@cs.umn.edu

extrinsic calibration. In the absence of accurate calibration, the estimates of the robot's motion will be biased, or may even diverge. Several approaches exist in the literature for performing odometer-camera calibration. In [1], a closed-form solution is proposed for the minimal problem of the odometer-camera extrinsic calibration. However, this approach is very sensitive to noise, and it cannot handle more measurements than the minimum required. One work-around is to refine the solution of the minimal problem with iterative minimization of a nonlinear least-squares cost function. Iterative techniques, however, do not provide any guarantee of optimality, and in the absence of an accurate initial estimate, they may not find the correct solution. Analytical least-squares solutions that can handle arbitrary number of measurements are proposed in [2] and [3]. Unfortunately, these approaches either require measurements from known landmarks [2], or only obtain the three degrees of freedom (d.o.f.) of the odometer-camera transformation in 2D [3].

In this paper, we present an *analytical* least-squares solution for determining the 3D (5 d.o.f.) transformation between a camera and an odometer that are rigidly connected. The proposed method does not require any known landmarks, and it is guaranteed to find the optimal estimate in a least-squares sense, without any initialization. Specifically, we first estimate the motion of the camera from unknown landmarks that are tracked across consecutive images. By fusing these estimates with the odometry measurements, we compute the three d.o.f. of the rotation, and two d.o.f. of the translation between the odometer and the camera. Finally, we prove that the third d.o.f. of the camera-odometer translation is unobservable, and cannot be computed from the measurements.

II. RELATED WORK

A straightforward approach to solve the odometer-camera calibration problem would be to use hand-eye calibration methods [4], [5]. However, since these techniques require the sensors to rotate around at least two different axes, they cannot be directly applied to a robot moving in 2D whose rotation is constrained around a single axis (i.e., the axis perpendicular to the robot's plane of motion). To address this issue, Chang and Aggarwal present an adaptation of the hand-eye calibration that relaxes this requirement [1]. In their method, the odometer-camera calibration problem is solved in two steps. In the first step, two d.o.f. in the odometer-camera rotation are determined using images from two consecutive camera poses, as well as the odometry measurements. In the second step, the third d.o.f. of the odometer-camera rotation and two d.o.f. of the odometer-

camera translation are determined using bearing measurements of one unknown landmark from three camera poses along with the corresponding odometry measurements. The third d.o.f. of the odometer-camera translation, which is perpendicular to the plane of the robot's motion, is shown to be unobservable. While this method works perfectly in the absence of noise, in practice it may yield inaccurate estimates. Additionally, this method cannot exploit additional measurements, if more than the minimal requirement (3 camera poses and 1 landmark) exist.

In the presence of noise, the Extended Kalman Filter (EKF) is commonly used for performing odometer-camera extrinsic calibration [6], [7]. In [6], the state vector of an EKF tracking the robot's motion in 2D is augmented with the odometer-camera extrinsic transformation (three out of five d.o.f.). By processing the odometer measurements and the images of one visual landmark, the EKF estimates the extrinsic transformation between the two sensors. In [7], Martinelli *et al.* present the observability analysis for the odometer-camera calibration problem, as well as the conditions under which it is possible to accurately estimate the odometer-camera transformation. The main drawbacks of the EKF-based approaches are that they are iterative, they provide no guarantee of optimality, and their performance strongly depends on the accuracy of the initial estimate. One can initialize these iterative methods with the analytical solution of [1]. In this case, however, the EKF inherits the inaccuracy of the initial estimates, and may converge to a local minimum far from the optimal solution.

In contrast to the EKF-based methods, other approaches seek to find a closed-form least-squares solution that directly accounts for the noise in the measurements. In [2], the second step of the approach proposed in [1] is extended to exploit additional measurements in a least-squares framework that is solved analytically. Moreover, the intrinsic odometer parameters, such as the wheel radius, are estimated simultaneously. The main drawback of this approach is that it requires measurements of known landmarks. Furthermore, the first step of this method is the same as [1], which ignores the measurement noise and cannot exploit more than two images.

In [3], an analytical least-squares solution for the extrinsic calibration of a laser scanner with respect to an odometer is presented. This method employs laser-scan matching to estimate the motion of the laser scanner, and it can be adapted to cameras, if *known* landmarks are available for estimating the camera's motion. Additionally, it is assumed that the orientation of the extroceptive sensor with respect to the odometer is already known, except for the yaw component. This, in effect, limits the applicability of this calibration method to estimating only 3 d.o.f. of the unknown transformation.

To address these drawbacks, in this work we propose a new method that, similar to [1] and [2], estimates the unknown transformation in two steps. However, in contrast to both [1] and [2], in the first step, we explicitly consider the effect of noise and *analytically* minimize a least-squares cost function to estimate two d.o.f. of the relative orientation.
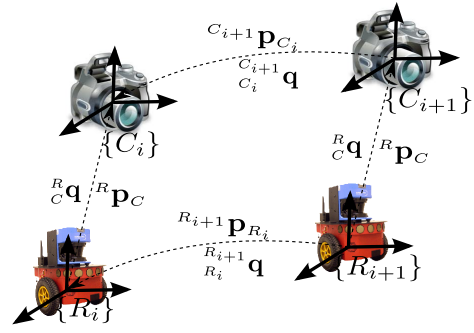

Fig. 1. The geometric relation between the camera's frame of reference, $\{C\}$, and the odometer's frame of reference, $\{R\}$, from time-step i to $i+1$.

In the second step, we *analytically* minimize another least-squares cost function to estimate the remaining d.o.f. of the relative orientation as well as two d.o.f. of the relative translation. Both these least-squares problems can handle any number of measurements, and since they are solved analytically, they need no initialization. Additionally, in contrast to [2], our proposed method does not require known landmarks, and instead, exploits unknown features to track the motion of the camera. To the best of our knowledge, our method is the first approach that, without observing known landmarks, obtains an analytical least-squares solution for all observable d.o.f. of the 3D odometer-camera transformation.

III. PROBLEM FORMULATION

Consider a robot moving in 2D equipped with an odometer whose frame of reference coincides with the robot's frame of reference, and a camera that is rigidly attached on the robot. Additionally, let us assume that the camera and the odometer have already been intrinsically calibrated. Furthermore, the camera observes landmarks whose global positions are not known, but they can be tracked across consecutive images. Our objective is to estimate the fixed, but otherwise unknown, transformation between the two sensors solely based on their measurements.

We denote the position and the orientation of the robot at time-step i relative to time-step $i+1$ by the vector $^{R_{i+1}}\mathbf{p}_{R_i}$ and the unit quaternion $^{R_{i+1}}_{R_i}\mathbf{q}$, respectively[1] (see Fig. 1), and compute them by integrating the odometry measurements. We assume the robot's motion is constrained on the xy plane, and thus $^{R_{i+1}}_{R_i}\mathbf{q}$ represents a rotation around the z axis. Similarly, the position and the orientation of the camera at time-step i relative to time-step $i+1$ are represented by $^{C_{i+1}}\mathbf{p}_{C_i} \triangleq u\,^{C_{i+1}}\bar{\mathbf{p}}_{C_i}$ and $^{C_{i+1}}_{C_i}\mathbf{q}$, respectively. Here we define the unknown scale u equal to $||^{C_1}\mathbf{p}_{C_0}||_2$ (i.e., the length of the camera's first translation), and assume that $^{C_1}\bar{\mathbf{p}}_{C_0}$ has unit norm. We then estimate u along with the odometer-camera transformation. Additionally, we assume

[1]Throughout this paper, $^A\mathbf{x}$ denotes the expression of a vector $\mathbf{x}$ with respect to frame $\{A\}$, $^A_B\mathbf{q}$ is the unit quaternion representing the orientation of frame $\{B\}$ with respect to frame $\{A\}$, and $\mathbf{R}(^A_B\mathbf{q})$ is the corresponding rotation matrix rotating vectors from frame $\{B\}$ to frame $\{A\}$. The $n \times n$ identity matrix is denoted by $\mathbf{I}_n$.

that sufficient overlap exists between each pair of consecutive images and compute the camera motion such that $^{C_{i+1}}\mathbf{p}_{C_i} = u\,^{C_{i+1}}\bar{\mathbf{p}}_{C_i}$ holds for all i. Note that in this case $^{C_{i+1}}\bar{\mathbf{p}}_{C_i}$ may not be unit-norm for $i \neq 0$. We employ the 5-point algorithm [8], [9] followed by bundle adjustment to estimate the camera's motion from the tracked visual features.

In the absence of noise, we can express the relationship between the camera's and the robot's poses as:

$$^{R_{i+1}}_{R_i}\mathbf{q} \otimes {}^R_C\mathbf{q} = {}^R_C\mathbf{q} \otimes {}^{C_{i+1}}_{C_i}\mathbf{q} \tag{1}$$

$$\left(\mathbf{R}(^{R_{i+1}}_{R_i}\mathbf{q}) - \mathbf{I}_3\right) {}^R\mathbf{p}_C = u\,\mathbf{R}(^R_C\mathbf{q})^{C_{i+1}}\bar{\mathbf{p}}_{C_i} - {}^{R_{i+1}}\mathbf{p}_{R_i} \tag{2}$$

for $i = 0, \ldots, m$. In these equations, $^R\mathbf{p}_C$ and $^R_C\mathbf{q}$ represent the 6 d.o.f. odometer-camera transformation, and $\otimes$ denotes quaternion multiplication which is defined as [10]:

$$\mathbf{q}_1 \otimes \mathbf{q}_2 \triangleq \mathcal{L}(\mathbf{q}_1)\mathbf{q}_2 = \mathcal{R}(\mathbf{q}_2)\mathbf{q}_1 \tag{3}$$

where

$$\mathcal{L}(\mathbf{q}) \triangleq \begin{bmatrix} q_4 & -q_3 & q_2 & q_1 \\ q_3 & q_4 & -q_1 & q_2 \\ -q_2 & q_1 & q_4 & q_3 \\ -q_1 & -q_2 & -q_3 & q_4 \end{bmatrix} \quad \mathcal{R}(\mathbf{q}) \triangleq \begin{bmatrix} q_4 & q_3 & -q_2 & q_1 \\ -q_3 & q_4 & q_1 & q_2 \\ q_2 & -q_1 & q_4 & q_3 \\ -q_1 & -q_2 & -q_3 & q_4 \end{bmatrix}$$

In practice, the measured robot and camera poses are perturbed by noises. We express the noisy measurements for translation vectors as $\hat{\mathbf{p}} = \mathbf{p} + \tilde{\mathbf{p}}$ where $\tilde{\mathbf{p}}$ denotes the noise, and for unit quaternions as $\hat{\mathbf{q}} = \mathbf{q} \otimes \delta\mathbf{q}^{-1}$ where $\delta\mathbf{q} \approx \left[\frac{1}{2}\delta\boldsymbol{\theta} \; 1\right]^T$ represents the *error quaternion* [10]. Substituting the noisy measurements in (1) and (2) yields:

$$^{R_{i+1}}_{R_i}\hat{\mathbf{q}} \otimes \delta\mathbf{q}_{R_i} \otimes {}^R_C\mathbf{q} = {}^R_C\mathbf{q} \otimes {}^{C_{i+1}}_{C_i}\hat{\mathbf{q}} \otimes \delta\mathbf{q}_{C_i} \tag{4}$$

$$\left(\mathbf{R}(^{R_{i+1}}_{R_i}\hat{\mathbf{q}})\mathbf{R}(\delta\mathbf{q}_{R_i}) - \mathbf{I}_3\right) {}^R\mathbf{p}_C =$$
$$u\mathbf{R}(^R_C\mathbf{q})\left({}^{C_{i+1}}\hat{\mathbf{p}}_{C_i} - {}^{C_{i+1}}\tilde{\mathbf{p}}_{C_i}\right) - \left({}^{R_{i+1}}\hat{\mathbf{p}}_{R_i} - {}^{R_{i+1}}\tilde{\mathbf{p}}_{R_i}\right) \tag{5}$$

Using (3) in (4), we obtain:

$$\mathcal{R}(\delta\mathbf{q}_{R_i})^{R_{i+1}}_{R_i}\hat{\mathbf{q}} \otimes {}^R_C\mathbf{q} - \mathcal{R}(\delta\mathbf{q}_{C_i})^R_C\mathbf{q} \otimes {}^{C_{i+1}}_{C_i}\hat{\mathbf{q}} = \mathbf{0} \tag{6}$$

Employing the definition of the error quaternion, we express

$$\mathcal{R}(\delta\mathbf{q}) \approx \mathbf{I}_3 + \frac{1}{2}\mathbf{B}(\delta\boldsymbol{\theta}), \; \mathbf{B}(\delta\boldsymbol{\theta}) \triangleq \begin{bmatrix} 0 & \delta\theta_3 & -\delta\theta_2 & \delta\theta_1 \\ -\delta\theta_3 & 0 & \delta\theta_1 & \delta\theta_2 \\ \delta\theta_2 & -\delta\theta_1 & 0 & \delta\theta_3 \\ -\delta\theta_1 & -\delta\theta_2 & -\delta\theta_3 & 0 \end{bmatrix}$$

Subsequently, (6) can be written as:

$$^{R_{i+1}}_{R_i}\hat{\mathbf{q}} \otimes {}^R_C\mathbf{q} - {}^R_C\mathbf{q} \otimes {}^{C_{i+1}}_{C_i}\hat{\mathbf{q}} =$$
$$\frac{1}{2}\mathbf{B}(\delta\boldsymbol{\theta}_{C_i})^R_C\mathbf{q} \otimes {}^{C_{i+1}}_{C_i}\hat{\mathbf{q}} - \frac{1}{2}\mathbf{B}(\delta\boldsymbol{\theta}_{R_i})^{R_{i+1}}_{R_i}\hat{\mathbf{q}} \otimes {}^R_C\mathbf{q} \triangleq \boldsymbol{\eta}_i \tag{7}$$

Note that $\boldsymbol{\eta}_i$, $i = 0, \ldots, m$, are the residuals due to the measurement noise. To find the best estimate for the odometer-camera rotation, we have to minimize these residuals by solving the following optimization problem:

$$^R_C\mathbf{q}^* = \underset{^R_C\mathbf{q}}{\operatorname{argmin}} \sum_{i=0}^m ||\boldsymbol{\eta}_i||_2^2 \tag{8}$$

Once $^R_C\mathbf{q}^*$ is computed, we back-substitute it in (5), and formulate a least-squares problem to estimate the remaining unknowns. Also, we approximate $\mathbf{R}(\delta\mathbf{q}_{R_i}) \approx \mathbf{I}_3 - \lfloor\delta\boldsymbol{\theta}_{R_i}\times\rfloor$,

where $\lfloor\delta\boldsymbol{\theta}\times\rfloor$ is the skew-symmetric matrix, to obtain:

$$\left(\mathbf{R}(^{R_{i+1}}_{R_i}\hat{\mathbf{q}}) - \mathbf{I}_3\right) {}^R\mathbf{p}_C - u\mathbf{R}(^R_C\mathbf{q}^*)^{C_{i+1}}\hat{\mathbf{p}}_{C_i} + {}^{R_{i+1}}\hat{\mathbf{p}}_{R_i}$$
$$= \mathbf{R}(^{R_{i+1}}_{R_i}\hat{\mathbf{q}})\lfloor\delta\boldsymbol{\theta}_{R_i}\times\rfloor^R\mathbf{p}_C - u\mathbf{R}(^R_C\mathbf{q}^*)^{C_{i+1}}\tilde{\mathbf{p}}_{C_i} + {}^{R_{i+1}}\tilde{\mathbf{p}}_{R_i}$$
$$\triangleq \boldsymbol{\epsilon}_i \tag{9}$$

where $\boldsymbol{\epsilon}_i$, $i = 0, \ldots, m$, are the residuals due to the measurement noise. To estimate the remaining unknowns, we have to minimize these residuals by solving the following optimization problem:

$$^R\mathbf{p}_C^*, u^* = \underset{^R\mathbf{p}_C, u}{\operatorname{argmin}} \sum_{i=0}^m ||\boldsymbol{\epsilon}_i||_2^2 \tag{10}$$

IV. Algorithm Description

In our proposed method, we estimate the odometer-camera transformation $(^R\mathbf{p}_C, {}^R_C\mathbf{q})$ and the scale u in two steps. In the first step, we solve the algebraic least-squares problem in (8), and obtain two d.o.f. of the relative odometer-camera orientation. Moreover, we show that the third d.o.f. of the relative orientation cannot be obtained when solving the minimization problem in (8), due to the 2D-constrained motion of the robot. In the second step, we solve (10) to estimate the remaining d.o.f. of the relative orientation $^R_C\mathbf{q}$ along with the scale u, and two d.o.f. of the relative translation $^R\mathbf{p}_C$. Finally, we prove that one d.o.f. of $^R\mathbf{p}_C$ (the one perpendicular to the robot's motion plane) is unobservable, and cannot be estimated from the camera and odometer measurements.

A. Solving for the 2 d.o.f. of Orientation

We start by decomposing the unknown unit quaternion $^R\mathbf{q}_C$ into three unit quaternions, corresponding to Z-Y-Z Euler angles α, β, γ as:

$$^R_C\mathbf{q} = \mathbf{q}_Z(\alpha) \otimes \mathbf{q}_Y(\beta) \otimes \mathbf{q}_Z(\gamma) \tag{11}$$

where $\mathbf{q}_Z(\alpha) = \begin{bmatrix} 0 & 0 & s(\alpha/2) & c(\alpha/2) \end{bmatrix}^T$, $\mathbf{q}_Y(\beta) = \begin{bmatrix} 0 & s(\beta/2) & 0 & c(\beta/2) \end{bmatrix}^T$, s stands for sin, and c stands for cos. The quaternion $\mathbf{q}_Z(\gamma)$ is defined similar to $\mathbf{q}_Z(\alpha)$. Then the term $\boldsymbol{\eta}_i$ in (8) can be expressed as:

$$\boldsymbol{\eta}_i = {}^{R_{i+1}}_{R_i}\hat{\mathbf{q}} \otimes \mathbf{q}_Z(\alpha) \otimes \mathbf{q}_Y(\beta) \otimes \mathbf{q}_Z(\gamma)$$
$$- \mathbf{q}_Z(\alpha) \otimes \mathbf{q}_Y(\beta) \otimes \mathbf{q}_Z(\gamma) \otimes {}^{C_{i+1}}_{C_i}\hat{\mathbf{q}} \tag{12}$$

Since both $^{R_{i+1}}_{R_i}\mathbf{q}$ and $\mathbf{q}_Z(\alpha)$ represent rotations around the z axis, they can commute to yield:

$$\boldsymbol{\eta}_i = \mathbf{q}_Z(\alpha) \otimes {}^{R_{i+1}}_{R_i}\hat{\mathbf{q}} \otimes \mathbf{q}_Y(\beta) \otimes \mathbf{q}_Z(\gamma)$$
$$- \mathbf{q}_Z(\alpha) \otimes \mathbf{q}_Y(\beta) \otimes \mathbf{q}_Z(\gamma) \otimes {}^{C_{i+1}}_{C_i}\hat{\mathbf{q}} \tag{13}$$

Employing (3), (13) can be written as:

$$\boldsymbol{\eta}_i = \mathcal{L}(\mathbf{q}_Z(\alpha)) \left({}^{R_{i+1}}_{R_i}\hat{\mathbf{q}} \otimes \mathbf{q}_{YZ} - \mathbf{q}_{YZ} \otimes {}^{C_{i+1}}_{C_i}\hat{\mathbf{q}}\right) \tag{14}$$

where

$$\mathbf{q}_{YZ} \triangleq \mathbf{q}_Y(\beta) \otimes \mathbf{q}_Z(\gamma) = \begin{bmatrix} s(\beta/2)s(\gamma/2) \\ s(\beta/2)c(\gamma/2) \\ c(\beta/2)s(\gamma/2) \\ c(\beta/2)c(\gamma/2) \end{bmatrix} \tag{15}$$

Since $\mathcal{L}(\mathbf{q}_Z(\alpha))$ is always an orthonormal matrix for any α, it can be factored out from (14). Therefore, the optimization problem in (8) can be expressed as:

$$_C^R\mathbf{q}^* = \underset{_C^R\mathbf{q}}{\operatorname{argmin}} \sum_{i=0}^{m} ||\boldsymbol{\eta}_i'||_2^2 \tag{16}$$

$$\boldsymbol{\eta}_i' \triangleq {}_{R_i}^{R_{i+1}}\widehat{\mathbf{q}} \otimes \mathbf{q}_{YZ} - \mathbf{q}_{YZ} \otimes {}_{C_i}^{C_{i+1}}\widehat{\mathbf{q}}$$

The absence of α in the above optimization problem implies that, in contrast to the general hand-eye calibration problem, not all d.o.f. of the unknown rotation $_C^R\mathbf{q}$ can be obtained solely based on (16). This is not an unexpected result, since according to the observability analysis of the hand-eye calibration [11], rotations around at least two non-parallel axes are required in order to estimate all three d.o.f. of $_C^R\mathbf{q}$. In the current problem, however, the robot (and the camera) are restricted to move on a 2D plane, and hence they can rotate around only one axis.

By employing (3), we can write the optimization problem in (16) as:

$$_C^R\mathbf{q}^* = \underset{_C^R\mathbf{q}}{\operatorname{argmin}} \sum_{i=0}^{m} \left|\left|\left(\mathcal{L}({}_{R_i}^{R_{i+1}}\widehat{\mathbf{q}}) - \mathcal{R}({}_{C_i}^{C_{i+1}}\widehat{\mathbf{q}})\right)\mathbf{q}_{YZ}\right|\right|_2^2 \tag{17}$$

Moreover, we can express (17) in a matrix form as:

$$\mathbf{q}_{YZ}^* = \underset{\mathbf{q}_{YZ}}{\operatorname{argmin}} ||\mathbf{M}\mathbf{q}_{YZ}||^2, \quad \mathbf{M} \triangleq \begin{bmatrix} \mathcal{L}({}_{R_m}^{R_{m+1}}\widehat{\mathbf{q}}) - \mathcal{R}({}_{C_m}^{C_{m+1}}\widehat{\mathbf{q}}) \\ \vdots \\ \mathcal{L}({}_{R_0}^{R_1}\widehat{\mathbf{q}}) - \mathcal{R}({}_{C_0}^{C_1}\widehat{\mathbf{q}}) \end{bmatrix} \tag{18}$$

In order to solve this optimization problem, let us first investigate the rank of each block row of $\mathbf{M}$, which can be expressed as:

$$\mathcal{L}({}_{R_i}^{R_{i+1}}\widehat{\mathbf{q}}) - \mathcal{R}({}_{C_i}^{C_{i+1}}\widehat{\mathbf{q}}) =$$

$$\begin{bmatrix} c(\frac{\phi_i}{2})-c(\frac{\phi_i'}{2}) & -s(\frac{\phi_i}{2})-s(\frac{\phi_i'}{2})k_Z & s(\frac{\phi_i'}{2})k_Y & -s(\frac{\phi_i'}{2})k_X \\ s(\frac{\phi_i'}{2})k_Z+s(\frac{\phi_i}{2}) & c(\frac{\phi_i}{2})-c(\frac{\phi_i'}{2}) & -s(\frac{\phi_i'}{2})k_X & -s(\frac{\phi_i'}{2})k_Y \\ -s(\frac{\phi_i'}{2})k_Y & s(\frac{\phi_i'}{2})k_X & c(\frac{\phi_i}{2})-c(\frac{\phi_i'}{2}) & s(\frac{\phi_i}{2})-s(\frac{\phi_i'}{2})k_Z \\ s(\frac{\phi_i'}{2})k_X & s(\frac{\phi_i'}{2})k_Y & s(\frac{\phi_i'}{2})k_Z-s(\frac{\phi_i}{2}) & c(\frac{\phi_i}{2})-c(\frac{\phi_i'}{2}) \end{bmatrix}$$

where ϕ_i denotes the angle of rotation between consecutive robot poses (corresponding to ${}_{R_i}^{R_{i+1}}\mathbf{q}$), and $(\phi_i', \begin{bmatrix} k_X, k_Y, k_Z \end{bmatrix}^T)$ denote the angle and axis of rotation between consecutive camera poses (corresponding to ${}_{C_i}^{C_{i+1}}\mathbf{q}$). Note that since the camera and the robot are rigidly connected, in the absence of noise, ϕ_i and ϕ_i' are the same, and $\begin{bmatrix} k_X, k_Y, k_Z \end{bmatrix}^T$ is fixed over time. In this case we have:

$$\mathcal{L}({}_{R_i}^{R_{i+1}}\widehat{\mathbf{q}}) - \mathcal{R}({}_{C_i}^{C_{i+1}}\widehat{\mathbf{q}}) = s(\frac{\phi_i}{2}) \begin{bmatrix} 0 & -1-k_Z & k_Y & -k_X \\ k_Z+1 & 0 & -k_X & -k_Y \\ -k_Y & k_X & 0 & 1-k_Z \\ k_X & k_Y & -1+k_Z & 0 \end{bmatrix}$$

Using Gaussian elimination, it is easy to show that this matrix has rank two [2]. Moreover, the rank of matrix $\mathbf{M}$ is also two since block rows of $\mathbf{M}$ differ only by $s(\frac{\phi_i}{2})$. Therefore, the solution of (18) lies in the two-dimensional null space of the matrix $\mathcal{L}({}_{R_i}^{R_{i+1}}\widehat{\mathbf{q}}) - \mathcal{R}({}_{C_i}^{C_{i+1}}\widehat{\mathbf{q}})$. A unique solution for

[2] The two vectors spanning the null space of $\mathbf{M}$ are $\begin{bmatrix} k_Z-1 & 0 & -k_X & -k_Y \end{bmatrix}^T$ and $\begin{bmatrix} k_X & k_Y & 1+k_Z & 0 \end{bmatrix}^T$.

the unknown unit quaternion (up to a sign ambiguity) can be obtained by enforcing the following two constrains:

$$q_{YZ1}q_{YZ4} = q_{YZ2}q_{YZ3} \quad , \quad ||\mathbf{q}_{YZ}||_2 = 1 \tag{19}$$

where q_{YZj} denotes the jth element of $\mathbf{q}_{YZ}$ and the former constraint is implied by the definition of $\mathbf{q}_{YZ}$ (see (15)).

In the presence of noise, $\mathbf{M}$ may not be rank deficient. In this situation, we compute the minimizer of (18) by enforcing (19) on a linear combination of the two eigenvectors of $\mathbf{M}^T\mathbf{M}$ that correspond to its two smallest eigenvalues. Specifically, if we assume that $\mathbf{t}_1$ and $\mathbf{t}_2$ are the eigenvectors corresponding to the smallest eigenvalues of $\mathbf{M}^T\mathbf{M}$, we can choose $\mathbf{q}_{YZ}^* = a\mathbf{t}_1 + b\mathbf{t}_2$. Then, the solution to (18) can be obtained by computing a and b using the constraints in (19).

B. Solving for the Remaining Unknowns

Once $\mathbf{q}_{YZ}^*$ is computed, we backsubstitute it in (9) to obtain:

$$\boldsymbol{\epsilon}_i = \mathbf{N}_i{}^R\mathbf{p}_C - u\mathbf{R}_Z(\alpha)\mathbf{R}(\mathbf{q}_{YZ}^*){}^{C_{i+1}}\widehat{\mathbf{p}}_{C_i} + {}^{R_{i+1}}\widehat{\mathbf{p}}_{R_i} \tag{20}$$

where

$$\mathbf{N}_i \triangleq \mathbf{R}({}_{R_i}^{R_{i+1}}\mathbf{q}) - \mathbf{I}_3 = \begin{bmatrix} \cos\phi_i - 1 & -\sin\phi_i & 0 \\ \sin\phi_i & \cos\phi_i - 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

In the absence of noise, all the terms in the third row of (20) are zero for any time-step i. This is not, however, the case in practise. If we treat $\mathbf{q}_{YZ}^*$ as a known parameter, we have implicitly assumed that the third nonzero element in (20) is only due to the measurement noise in ${}^{C_{i+1}}\widehat{\mathbf{p}}_{C_i}$. We can eliminate the component of noise causing this effect by projecting ${}^{C_{i+1}}\widehat{\mathbf{p}}_{C_i}$ to the plane that is perpendicular to the vector defined by the third row of matrix $\mathbf{R}(\mathbf{q}_{YZ}^*)$. This projection guarantees that the third row of (20) is always equal to zero. Clearly, this technique is an approximation, since the uncertainty in the estimated $\mathbf{q}_{YZ}^*$ also contributes to the perturbation in (20). Our simulation results, however, have shown that this measurement projection improves the accuracy of the estimates for both α and ${}^R\mathbf{p}_C$.

Note that based on (20), the third element of ${}^R\mathbf{p}_C$ is multiplied by zeros. Therefore, it does not affect the cost function in (20), and it cannot be estimated from the given measurements. This is not unexpected because if we change the height of the 2D plane of the robot motion while keeping the camera motion plane fixed, all the measurements received by the odometer and the camera (i.e., ${}^{C_{i+1}}\widehat{\mathbf{p}}_{C_i}$, ${}^{C_{i+1}}_{C_i}\widehat{\mathbf{q}}$, ${}^{R_{i+1}}\widehat{\mathbf{p}}_{R_i}$, and ${}^{R_{i+1}}_{R_i}\widehat{\mathbf{q}}$) remain unaffected. Thus, the translation between the odometer and the camera in the z direction is unobservable, and we can remove the third row in (20) to obtain:

$$\boldsymbol{\epsilon}_i' = \begin{bmatrix} \cos\phi_i - 1 & -\sin\phi_i \\ \sin\phi_i & \cos\phi_i - 1 \end{bmatrix} \begin{bmatrix} p_X \\ p_Y \end{bmatrix}$$
$$- u\begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} p_{i_1} \\ p_{i_2} \end{bmatrix} + {}^{R_{i+1}}\widehat{\mathbf{p}}_{R_i}' \tag{21}$$

In (21), $\begin{bmatrix} p_X & p_Y \end{bmatrix}^T$ denotes the first two elements of ${}^R\mathbf{p}_C$ corresponding to the translation between the camera and the

odometer in the xy plane, $\begin{bmatrix} p_{i1} & p_{i2} \end{bmatrix}^T$ denotes the first two elements of the vector $\mathbf{R}(\mathbf{q}_{YZ}^*)^{C_{i+1}}\widehat{\mathbf{p}}_{C_i}$, and $^{R_{i+1}}\widehat{\mathbf{p}}'_{R_i}$ denotes the first two elements of the vector $^{R_{i+1}}\widehat{\mathbf{p}}_{R_i}$. It is easy to show that (21) can be written as:

$$
\boldsymbol{\epsilon}'_i = \underbrace{\begin{bmatrix} \cos\phi_i - 1 & -\sin\phi_i \\ \sin\phi_i & \cos\phi_i - 1 \end{bmatrix}}_{\mathbf{J_i}} \begin{bmatrix} p_X \\ p_Y \end{bmatrix}
$$
$$
- u \underbrace{\begin{bmatrix} p_{i_1} & -p_{i_2} \\ p_{i_2} & p_{i_1} \end{bmatrix}}_{\mathbf{K_i}} \begin{bmatrix} \cos\alpha \\ \sin\alpha \end{bmatrix} + \underbrace{^{R_{i+1}}\mathbf{p}_{R_i}'}_{\mathbf{t_i}} \quad (22)
$$

where $\mathbf{J}_i$, $\mathbf{K}_i$ and $\mathbf{t}_i$ are all known quantities. Then, we can express the optimization problem (10) in the following compact form:

$$
\mathbf{m}^* = \underset{\mathbf{m}}{\operatorname{argmin}} ||\mathbf{Gm} + \mathbf{w}||_2^2 \quad (23)
$$

where

$$
\mathbf{G} \triangleq \begin{bmatrix} \mathbf{J}_m & \mathbf{K}_m \\ \mathbf{J}_{m-1} & \mathbf{K}_{m-1} \\ \vdots & \vdots \\ \mathbf{J}_0 & \mathbf{K}_0 \end{bmatrix}, \quad \mathbf{m} \triangleq \begin{bmatrix} p_X \\ p_Y \\ u\cos\alpha \\ u\sin\alpha \end{bmatrix}, \quad \mathbf{w} \triangleq \begin{bmatrix} \mathbf{t}_m \\ \mathbf{t}_{m-1} \\ \vdots \\ \mathbf{t}_0 \end{bmatrix}
$$

The solution $\mathbf{m}^*$ to the optimization problem in (23) can be obtained using linear least squares, from which the estimates for p_X, p_Y, u, and α can be easily extracted.

C. Iterative Refinement

Once the five d.o.f. of the odometer-camera transformation and the scale are computed, we may refine the estimates by iteratively minimizing the following weighted batch least-squares cost function:

$$
^R\mathbf{p}_C^*, {}^R\mathbf{q}_C^*, u^* = \operatorname{argmin}_{^R\mathbf{p}_C, {}^R_C\mathbf{q}, u} \sum_{i=1}^m (\mathcal{C}_i + \mathcal{D}_i) \quad (24)
$$
$$
\mathcal{C}_i \triangleq \left|\left| (\mathbf{R}(^{R_{i+1}}_{R_i}\widehat{\mathbf{q}}) - \mathbf{I})^R\mathbf{p}_C - u\mathbf{R}(^R_C\mathbf{q})^{C_{i+1}}\widehat{\mathbf{p}}_{C_i} + {}^{R_{i+1}}\widehat{\mathbf{p}}_{R_i} \right|\right|^2_{\mathbf{W}_{i1}}
$$
$$
\mathcal{D}_i \triangleq \left|\left| {}^{R_{i+1}}_{R_i}\widehat{\mathbf{q}} \otimes^R_C \mathbf{q} - {}^R_C\mathbf{q} \otimes^{C_{i+1}}_{C_i} \widehat{\mathbf{q}} \right|\right|^2_{\mathbf{W}_{i2}}
$$

where $||\mathbf{x}||_{\mathbf{W}}^2 \triangleq \mathbf{x}^T\mathbf{W}^{-1}\mathbf{x}$, and $\mathbf{W}_{i1}$ and $\mathbf{W}_{i2}$ are the covariance matrices of the corresponding measurement residuals. For detailed information on how to minimize this cost function, and a discussion on how to address the gauge freedom in $\mathcal{D}_i$, we refer the interested reader to [12].

V. SIMULATION RESULTS

In order to validate the proposed algorithm, we have performed a number of simulations. In each simulation, 1000 Monte Carlo trials are carried out, with the robot's initial pose set to coincide with the global frame. The robot's consecutive translations $^{R_{i+1}}\mathbf{p}_{R_i}$, $i = 1 \ldots N$, are generated randomly from a normal distribution with zero mean and 0.2 m standard deviation along the x and y axes of the global frame. The angles ϕ_i of the consecutive robot's rotations, $\mathbf{R}(^{R_{i+1}}_{R_i}\mathbf{q})$, are generated from a uniform random distribution $\mathcal{U}[-\pi/2, \pi/2]$. The noise in the robot's translation measurements is generated using a normal distribution with standard deviation proportional to the length of the robot's motion,

TABLE I
PERCENTAGE OF DIVERGENCE IN SIMULATIONS

	GT-B	ALS-B	ALSP-B	CM-B
Divergence	1.09%	1.09%	1.09%	1.54%

while the rotation measurements' standard deviations vary between 0 to 0.4 radian are used.

We randomly generated each element of the relative odometer-camera translation, $^R\mathbf{p}_C$, using the uniform distribution $\mathcal{U}[-0.1, 0.1]$ m. The angle of the odometer-camera rotation is generated randomly using a uniform distribution in the interval $[-\pi, \pi]$, while each element of the rotation axis is generated randomly using a normal distribution with zero mean and 0.1 m standard deviation. Based on the generated camera-odometer transformation, $^R_C\mathbf{q}$ and $^R\mathbf{p}_C$, as well as the robot motions, $^{R_{i+1}}_{R_i}\mathbf{q}$ and $^{R_{i+1}}\mathbf{p}_{R_i}$, we employ (1) and (2) to compute the true consecutive camera transformations $^{C_{i+1}}_{C_i}\mathbf{q}$ and $^{C_{i+1}}\mathbf{p}_{C_i}$. We then randomly generate 12 3D landmarks at each time-step and project them on the image plane of the cameras at two consecutive time-steps using the true camera configuration. In order to simulate the noise of a real camera, we add a zero-mean Gaussian noise with standard deviations between $0-1$ mm (equivalent to $0-2$ pixels on a camera with focal length of 400 pixels) to the projection of the landmarks on the image plane. In our simulations, we use the 8-point algorithm [13] followed by bundle adjustment to compute the transformation between consecutive camera poses[3].

We hereafter denote the results for each of the implemented methods using the following abbreviations:

- ALSP: The proposed algorithm of Section IV.
- ALS: The proposed algorithm of Section IV without projection of $^{C_{i+1}}\widehat{\mathbf{p}}_{C_i}$ (see Section IV-B).
- CM: The closed-form solution to the minimal problem, as proposed by Chang *et al.* [1].
- X-B: The solution to the weighted batch least-squares problem (see Section IV-C) initialized by the results of method X, where X is any of the above-mentioned approaches.
- GT-B: The solution to the weighted batch least-squares problem initialized by the ground truth values.

Note that the method of Chang *et al.* (CM), does not use all the available measurements. Specifically, in the first step of CM, we only need one robot motion, along with the camera's rotation estimation. In the second step of CM, measurements of one landmark and two robot motions are required. Consequently, we have several choices for the subset of the measurements that can be used to estimate the unknowns. Hence, we employ the RANSAC algorithm to select the best subset of the measurements for running CM.

---

[3]Note that in practical situations when outliers in the landmark matches may exist, the 5-point algorithm followed by RANSAC should be the method of choice. However, in our controlled simulation, we do not have outliers, and thus using the 8-point algorithm, which has better numerical stability (since it does not require solving polynomial equations), is preferred.
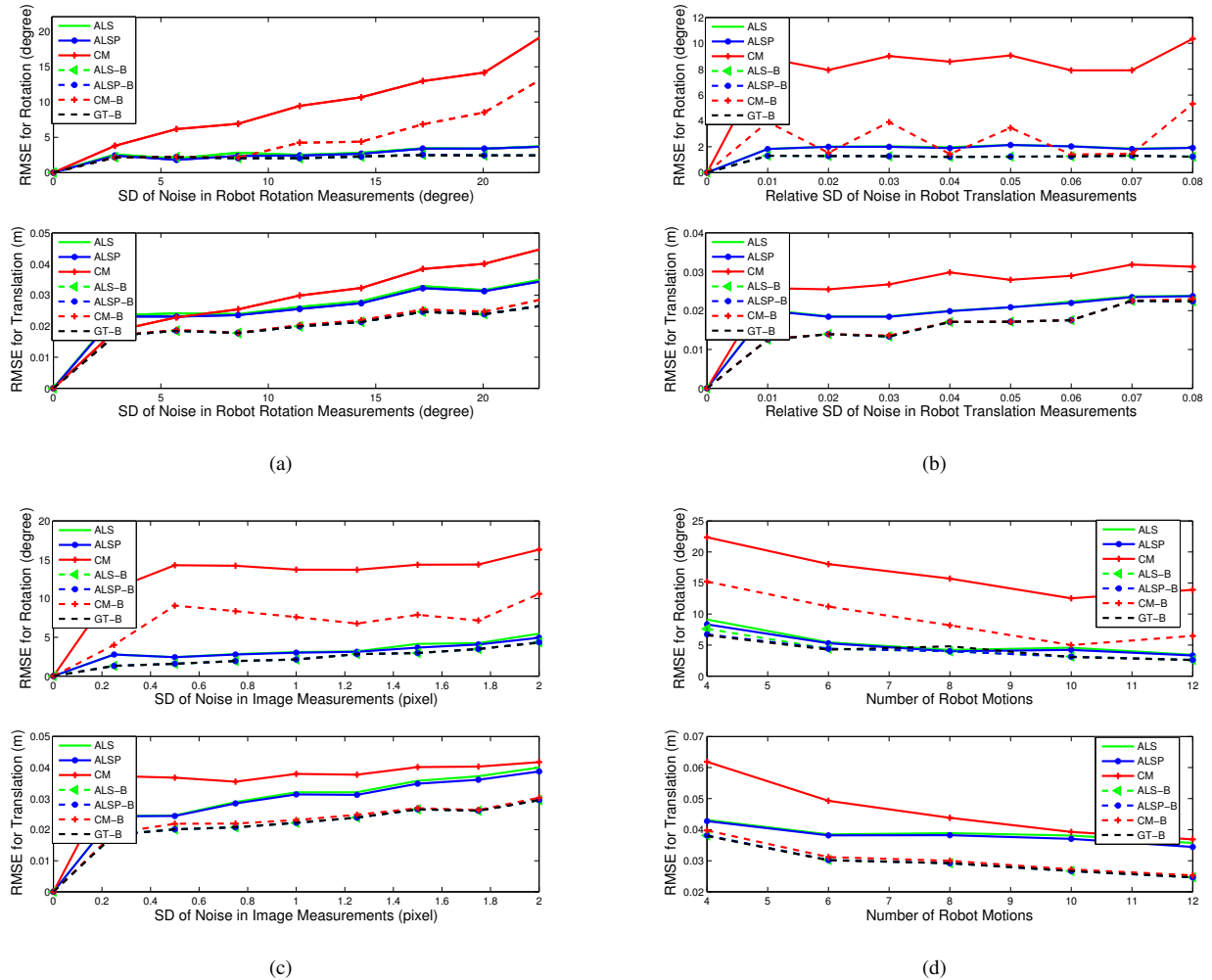
Fig. 2. The RMSE of the estimated odometer-camera transformation versus: (a) The standard deviation (SD) of the noise in the robot's rotation measurement; (b) The SD of the noise in the robot's translation measurement relative to the length of its motion; (c) The SD of the noise in the image; (d) The number of robot motions.

The Root Mean Square Error (RMSE) of the odometer-camera transformation computed by the aforementioned methods using various noise levels and number of motions is shown in Fig. 2. To compute the error in ${}_{C}^{R}\hat{\mathbf{q}}$, we first determine the error quaternion, $\delta\mathbf{q} = {}_{C}^{R}\hat{\mathbf{q}}^{-1} \otimes {}_{C}^{R}\mathbf{q}$, and then compute the tilt angle error $\delta\boldsymbol{\theta}$ from $\delta\mathbf{q} \approx [\frac{1}{2}\delta\boldsymbol{\theta} \ 1]^{T}$.

The key findings from the simulation are the following: First, the ALSP and the ALS yield significantly more accurate estimates than CM. This is not unexpected, since the CM only uses a subset of the measurements (selected through RANSAC), while the ALSP and the ALS exploit all the available measurements. The more interesting observation is that while the ALSP-B and the ALS-B result in accuracies that are almost indistinguishable from that of the GT-B, the CM-B consistently results in inferior accuracy, due to cases where the iterative estimator becomes trapped in local minima. Finally, note the slightly higher accuracy of the ALSP compared to the ALS, particularly in higher noise scenarios, which justifies its application in practice.

Throughout our simulations, we observed that the true calibration parameters are usually located within the 3σ bounds of the computed estimates. This, however, was not the case in some instances of the simulations with high measurement noise levels. The percentages of these *divergences* are reported in Table I. As evident from this table, even if we initialize the batch least-squares algorithm with the ground truth, the final estimate may not converge to the correct solution. This can be explained by noting that in high measurement noise regimes, the true calibration parameters may not be located in the convergence basin of the weighted batch least-squares cost function. It is clear from these results that the ALSP-B and the ALS-B diverge as often as the GT-B, while the CM-B leads to almost 50% more divergences.

VI. EXPERIMENTAL RESULTS

In order to demonstrate the validity of our algorithm in practice, we have conducted several experiments using a Pioneer III robot equipped with a camera and an odometer, both of which have already been intrinsically calibrated. The robot and the camera are rigidly connected and the transformation between them does not change during the experiment. The resolution of the captured images is 752×480 pixels and the images are received at a rate of 5Hz.

In our experiments, the robot moves around in a

TABLE II

ESTIMATION OF ODOMETER-CAMERA ROTATION FROM FOUR INDEPENDENT EXPERIMENTS (EULER ANGLE PARAMETRIZATION IN DEGREES)

	# of Motions	ALSP	ALSP-B	CM	CM-B
1	20	$[1.7; -87.1; 90.0]$	$[1.7; -86.5; 90.0] \pm [1.7; 3.4; 1.7]$	$[1.1; -87.7; 90.0]$	$[1.1; -86.5; 93.4] \pm [1.7; 3.4; 1.7]$
2	40	$[2.9; -87.7; 89.4]$	$[1.1; -87.7; 99.7] \pm [1.1; 2.3; 1.1]$	$[3.4; -89.4; 90.5]$	$[1.1; -87.7; 99.7] \pm [1.1; 2.3; 1.1]$
3	26	$[0.0; -87.1; 90.0]$	$[0.0; -87.1; 83.7] \pm [2.9; 5.2; 2.9]$	$[0.0; -87.7; 90.0]$	$[0.0; -87.1; 83.7] \pm [2.9; 5.2; 2.9]$
4	23	$[0.0; -86.5; 90.5]$	$[3.4; -88.8; 87.1] \pm [1.7; 2.3; 1.7]$	$[3.4; -85.9; 90.5]$	$[2.3; -94.5; 20.6] \pm [1.7; 2.9; 1.7]$

TABLE III

ESTIMATION OF ODOMETER-CAMERA RELATIVE TRANSLATION FROM FOUR INDEPENDENT EXPERIMENTS (XY PLANE COMPONENTS IN cm)

	# of Motions	ALSP	ALSP-B	CM	CM-B
1	20	$[10.89; 1.98]$	$[9.64; 1.96] \pm [1.63; 1.71]$	$[9.65; 0.05]$	$[9.63; 1.95] \pm [1.63; 1.71]$
2	40	$[10.58; 0.99]$	$[8.98; 1.87] \pm [1.89; 1.91]$	$[8.11; 0.01]$	$[8.98; 1.87] \pm [1.89; 1.91]$
3	26	$[8.21; 0.39]$	$[8.99; 0.38] \pm [1.76; 1.78]$	$[8.25; -0.08]$	$[9.00; 0.38] \pm [1.76; 1.78]$
4	23	$[9.22; 0.91]$	$[8.81; 3.31] \pm [2.26; 2.32]$	$[7.34; 3.20]$	$[23.31; 5.02] \pm [2.46; 2.48]$

10m×10m room. Once images are captured, we first extract SIFT features, which are then used in the 5-point algorithm along with RANSAC to compute an initial estimate for the transformation between the camera poses. We then employ a batch least-squares method to refine these estimates. The transformation between the robot poses are computed by integrating the odometry measurements. The camera and odometer data are captured by the same computer with time stamps, and synchronized according to the system time.

The estimates of the relative odometer-camera transformation by ALSP, ALSP-B, CM, CM-B, as well as the 3σ bounds from four independent experiments are provided in Tables II and III. Since we do not have the ground truth of the relative odometer-camera transformation, we are unable to verify the calibration result accurately. However, the estimated calibration parameters, are consistent with the manually measured rotation of $[0° \ -90° \ 90°]^T$, and the translation of $[9 \ 1.5]^T$ cm. In the first three experiments, although the estimates of our proposed method and Chang's method are slightly different, they converge to similar values after applying batch least-squares refinement. However, in the fourth experiment the batch least-squares algorithm initialized by Chang's method converges to a value far from all other estimates, while the batch least-squares algorithm initialized by our method obtains a result consistent with the first three experiments.

VII. CONCLUSION

In this paper, we presented a new analytical method for the extrinsic calibration of odometer-camera pairs that does not require any special hardware or known landmarks. In particular, we first formulate a least-squares problem to estimate a subset of the odometer-camera rotation parameters. Once computed, these parameters are used to formulate a second least-squares problem for estimating the remaining unknown parameters of the odometer-camera transformation. We presented methods to solve both least-squares problems analytically without requiring any iterations or initialization. Finally, we provided extensive simulation and experimental results that confirm the validity of the proposed approach.

Currently, we are investigating the possibility of analytically estimating the intrinsic parameters of the odometer as well as the odometer-camera extrinsic calibration in a least-squares framework, without observations of known landmarks.

REFERENCES

[1] Y. Chang, X. Lebegue, and J. Aggarwal, "Calibrating a mobile camera's parameters," *Pattern Recognition*, vol. 26, no. 1, pp. 75–88, 1993.

[2] G. Antonelli, F. Caccaale, F. Grossi, and A. Marino, "Simultaneous calibration of odometry and camera for a differential drive mobile robot," in *Proc. of the IEEE International Conference on Robotics and Automation*, Anchorage, AK, May 3–7 2010, pp. 5417–5422.

[3] A. Censi, L. Marchionni, and G. Oriolo, "Simultaneous maximum-likelihood calibration of odometry and sensor parameters," in *Proc. of the IEEE International Conference on Robotics and Automation*, Pasadena, CA, May 19–23 2008, pp. 2098–2103.

[4] R. Tsai and R. Lenz, "A new technique for fully autonomous and efficient 3d robotics hand/eye calibration," vol. 5, no. 3, pp. 345–358, June 1989.

[5] K. H. Strobl and G. Hirzinger, "Optimal hand-eye calibration," in *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Beijing, China, Oct. 9–15 2006, pp. 4647–4653.

[6] A. Martinelli, D. Scaramuza, and R. Siegwart, "Automatic self-calibration of a vision system during robot motion," in *Proc. of the IEEE International Conference on Robotics and Automation*, Orlando, FL, May 15–19 2006, pp. 43–48.

[7] A. Martinelli, "Local decomposition and observability properties for automatic calibration in mobile robotics," in *Proc. of the IEEE International Conference on Robotics and Automation*, Kobe, May 12–17 2009, pp. 4182–4188.

[8] D. Nister, "An efficient solution to the five-point relative pose problem," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 26, no. 6, pp. 756–770, June 2004.

[9] D. Nister, O. Naroditsky, and J. Bergen, "Visual odometry for ground vehicle applications," vol. 23, no. 1, pp. 3–20, Jan. 2006.

[10] N. Trawny and S. I. Roumeliotis, "Indirect kalman filter for 3d attitude estimation," University of Minnesota, Dept. of Comp. Sci. & Eng., Tech. Rep., March 2005.

[11] K. Daniilidis and E. Bayro-Corrochano, "The dual quaternion approach to hand-eye calibration," in *Proc. of the IEEE International Conference on Pattern Recognition*, Aug. 25–29 1996, pp. 318–322.

[12] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon, "Bundle adjustment - a modern synthesis," in *Vision Algorithms: Theory and Practice*, B. Triggs, A. Zisserman, and R. Szeliski, Eds. Springer–Verlag, 2000.

[13] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, March 2004.