Concurrent Extrinsic Camera Calibration and 3D Robot-body Reconstruction Using Multiple Mirrors

Joel A. Hesch, *Student Member, IEEE,* Anastasios I. Mourikis, *Member, IEEE,*
and Stergios I. Roumeliotis, *Member, IEEE*

***Abstract*—In this paper, we address the problem of concurrent extrinsic calibration of a camera to a base frame of reference and 3D robot-body reconstruction, using observations of points that can only be observed via reflection in one or more mirrors. In our scenario, the camera measures reflections of fiducial points, whose coordinates in the base frame attached to the robot body are known, and reconstruction points, whose coordinates are initially unknown and must be estimated. No prior information about the mirrors' placements or motions with respect to the camera is available *a priori*. Instead, we employ the feature measurements in order to analytically compute (i) the position and orientation of each mirror in each image, (ii) the six-degree-of-freedom (DOF) camera-to-base frame transformation, and (iii) the base-frame coordinates of the reconstruction points. We study the conditions necessary for solutions to exist and show that, in the single-mirror scenario, at least three fiducial points must be observed in three images to uniquely recover the camera-to-base frame transformation, while each reconstruction point should be viewed twice. We subsequently refine the analytic solution in a maximum-likelihood estimator (MLE), which obtains a statistically optimal estimate of the unknown quantities by accounting for the measurement noise. We study the algorithm's performance under variations of system parameters with simulations, and show the applicability of our method to real-world problems with experimental trials.**

***Index Terms*—Extrinsic camera calibration, catadioptric system, 3D reconstruction, multiple view geometry.**

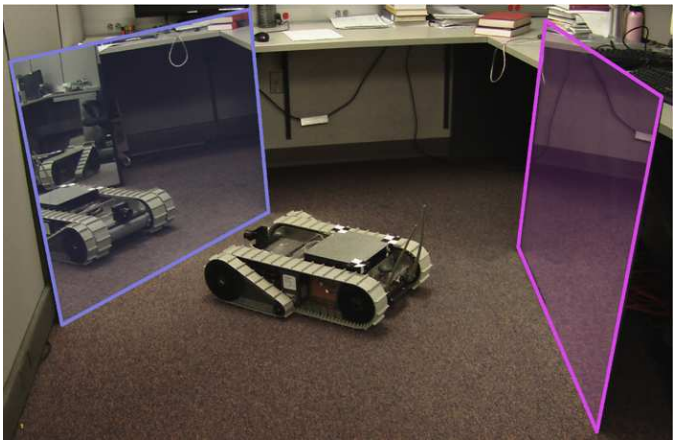


Fig. 1. A camera-equipped robot observes its reflection in two mirrors from several viewing angles. The front mirror (highlighted in blue) provides a direct reflection, whereas the double reflection comes from the back mirror (highlighted in purple). As the mirrors are moved, the robot tracks fiducial and reconstruction points with its on-board camera. These point-feature observations are employed to estimate the coordinate transformation between the camera and robot body, as well as a 3D point-cloud representation of the robot chassis.

I. Introduction

CAMERAS are utilized in a wide variety of applications ranging from surveillance and crowd monitoring to vision-based robot localization. In order to obtain meaningful geometric information from a camera, two calibration procedures must be completed. The first is intrinsic calibration, that is, determining the internal camera parameters (e.g., focal length, principal point, and distortion coefficients) that affect the image measurements [4], [5]. The second is extrinsic calibration, which is the process of computing the six-degree-of-freedom (DOF) transformation between the camera and a base frame of reference [1], [2], [6], [7], [8], [9]. In a surveillance application, the base frame may be the room or building coordinate system, whereas on a mobile robot, the base frame could be the robot-body frame (see Fig. 1). In addition to intrinsic and extrinsic camera calibration, many tasks also require determining the dimensions of a portion of the 3D scene [10], [11], [12], [13]. In an indoor surveillance scenario, knowledge of the room geometry, as well as of the locations of objects (e.g., tables and chairs) can aid in tracking and interpreting a target's actions. Similarly, in the case of a camera-equipped mobile robot, knowing the precise size and shape of the robot chassis enables more accurate path planning and obstacle avoidance.

Both extrinsic calibration and 3D reconstruction may be accomplished by hand-measuring or approximating the pertinent quantities; however, this can be time consuming and the results are often inaccurate. In particular, through manual measurements we can only determine the location of the camera housing and not the focal point itself, which is the true center of the camera's frame of reference. Furthermore, anytime the vision system changes (e.g., when the camera is replaced or moved to a different location) this laborious procedure must be repeated. For these reasons, it is imperative to develop automatic methods for estimating the camera-to-base transformation and performing 3D reconstruction.

Extrinsic camera calibration has been widely studied for the case in which known points are *directly* observed by the

This work was supported by the University of Minnesota (Digital Technology Center), the University of California Riverside (Bourns College of Engineering), and the National Science Foundation (IIS-0643680, IIS-0811946, IIS-0835637). A preliminary version of this paper was presented at the IEEE/RSJ International Conference on Intelligent Robots and Systems [1], an extension detailing the analytic solution was presented in the International Workshop on the Algorithmic Foundations of Robotics [2], and the multi-mirror case appeared at the European Conference on Computer Vision [3].

J. A. Hesch and S. I. Roumeliotis are with the Dept. of Computer Science and Engineering, University of Minnesota, Minneapolis, MN 55455, USA. email: {joel|stergios}@cs.umn.edu.

A. I. Mourikis is with the Dept. of Electrical Engineering, University of California, Riverside, CA 92521, USA. email: mourikis@ee.ucr.edu.

camera. This problem was introduced by Grunert in 1841 [14], and has since received significant attention from the computer vision community [15], [16], [17]. Unfortunately, in many realistic scenarios the known points may not lie within the camera's field of view. Figure 1 depicts an example in which a mobile robot is equipped with a forward-facing camera that cannot view any portion of the robot chassis. Instead, it must move in front of a mirror to see its body. In this work, we focus on these more limiting scenarios, in which the points of interest do not lie within the camera's field of view but can only be observed through reflection using one or more planar mirrors. This framework for extrinsic camera calibration is significantly more flexible than previous formulations which required direct observations of points, and enables extrinsic calibration and 3D reconstruction in a wider range of applications.

Specifically, the objective of our work is to design an automated procedure which accurately determines the unknown quantities by exploiting measurements of point features from different viewing angles. The point features comprise: (i) *fiducial points*, whose coordinates in the base frame (i.e., robot-body frame) are known *a priori*, and (ii) *reconstruction points*, whose coordinates are unknown and will be computed. Since the points lie outside the camera's field of view, we maneuver mirrors to provide the camera with multiple views of the robot body; however, no prior information about the mirrors' sizes or motions with respect to the camera is assumed. Instead, the configurations of the mirrors and the camera-to-base transformation are treated as unknowns to be computed from the measurements.

The main contributions of this paper are:

- We carry out an analysis of the mirror-based extrinsic calibration problem to determine the existence and number of solutions, as a function of the number of fiducial points and number of images available.
- Based on the above analysis, we identify the minimal problem formulation and propose a method for analytically computing the camera-to-base transformation, in cases where either one or more mirrors are employed to observe the fiducial points.
- We develop a method for analytically computing the 3D base-frame coordinates of each reconstruction point observed by the camera.
- We formulate a maximum-likelihood estimator (MLE) for refining the analytically computed estimates of all quantities of interest, in order to obtain a statistically optimal solution.

These contributions offer an in-depth analysis of, as well as solution methods for, the problem of mirror-based extrinsic camera calibration. Therefore, the work presented in this paper provides both a better intuitive understanding of the problem, as well as algorithmic tools that are applicable in practical scenarios.

The rest of the paper is structured as follows: In the next section, we discuss related work in the literature, and compare our work to existing approaches. Section III presents our problem formulation, as well as the analysis of the number of solutions. The analytic solution for the minimal problem is described in Section IV, while Section V details the analytic approach for computing the 3D base-frame coordinates of the reconstruction points. The MLE for refining the analytically computed quantities is formulated in Section VI. Finally, Sections VII and VIII present our evaluation of the performance of the developed algorithms, both in simulation (Section VII) and in real-world experiments with a camera-equipped robot (Section VIII).

II. Related Work

We begin with a discussion of the related approaches that address camera-to-base calibration, sensor-to-sensor calibration, hand-eye calibration, and catadioptric systems.

A. Camera-to-base calibration

The task of computing the transformation between a camera and a *generic* base frame of reference has been widely studied. Most existing methods rely on direct observations of fiducial points whose 3D base-frame coordinates are known. When three fiducial points are observed in a single image, this problem is termed the Three Point Perspective Pose Estimation Problem (P3P) [14]. Numerous algebraic and geometric studies of P3P have shown that in generic configurations there exist up to eight real solutions, four of which may lie in front of the center of perspectivity [15], [16], [17], [18]. Extensions of P3P to four observed points (P4P) and n observed points (PnP) have also been presented [19], [20]. In contrast to these approaches for camera-to-base calibration, we do not require the fiducial points to lie within the camera's field of view. Relaxing this constraint is critical for enabling calibration in a wide variety of previously unaddressed scenarios (e.g., see Fig. 1).

B. Camera-to-sensor calibration

A related problem is that of camera-to-sensor calibration, which seeks to compute the transformation between a camera and another sensor by correlating their measurements. Numerous cases have been addressed, including calibration to an odometry sensor [21], [22], an inertial measurement unit (IMU) [7], and a laser range finder [8], [23], [9]. One key difference of all these methods, compared to the mirror-based calibration approach we explore in this work, is that they require the presence of two sensors, rather than just the camera. Moreover, these methods typically require that both the camera and the additional sensor move, which enables the two sensors to estimate their motion with respect to a common frame.

C. Hand-eye calibration

Another related problem is that of hand-eye calibration, i.e., the process of determining the six-DOF transformation between a camera and a tool, which are both mounted on a robot manipulator [24], [25], [26], [27], [28]. The hand-eye problem is typically solved by correlating the measurements from the camera and the joint encoders, which measure the displacements of the manipulator joints. Specifically, using

the camera images and fiducial points, one can determine the transformation between the camera and the robot-base frames. Subsequently, the camera-to-tool transformation is calculated by combining the estimated camera-to-robot-base transformation, and the robot-base-to-tool transformation, which is assumed to be known. This necessitates the availability of precise technical drawings and limits the applicability of these methods, since they cannot determine the camera-to-base transformation for a generic base frame. Unlike the hand-eye problem, we only exploit information from a single sensor, and we do not require knowledge of the camera's motion (i.e., we do not utilize joint encoders).

D. Catadioptric systems

We turn our attention to catadioptric systems, which are camera systems that employ mirrors to perform synthetic stereo vision. These are of interest because they bear resemblance to our setup, in which one or more mirrors are utilized to compute the camera-to-base transformation and the 3D coordinates of the reconstruction points. Several different synthetic-stereo methods have been presented, utilizing a single camera and planar [11], [12], [29], or conic mirrors [30], reflections from free-form surfaces [31], or trinocular vision [13]. Additionally, stereo has been achieved with a static camera and a moving spherical mirror [32], as well as with a moving camera and two stationary spherical mirrors of known radii [33]. In contrast to the approaches above, we do not exploit *a priori* knowledge about the size or location of the mirrors. Additionally, in our method each point is only observed *once* per image, via its reflection in one or more moving planar mirrors. This requires a substantially different mathematical treatment.

Another catadioptric system that employs a moving planar mirror for catadioptric stereo vision was introduced by Jang et al. [34]. In their approach, the mirror position and orientation (pose) was determined for each image using observations of known markers on the mirror, and vanishing points in the image. After determining the pose of the mirror in all images, multiple view scene reconstruction was accomplished using all of the reflection images. In contrast to this approach, we address a broader problem which encompasses both 3D reconstruction and camera-to-base frame calibration, *without* employing known markers on the mirror. This significantly increases the difficulty of the problem. Furthermore, we address the most general case in which $N_v \geq 1$ mirrors are employed.

To the best of our knowledge, the first work addressing pose determination without a direct view of the scene was carried out by Sturm and Bonfort [35]. In particular, the authors seek to determine the pose of the camera with respect to a known object via reflections of the object in a *single* planar mirror. The analysis focuses on the two-view scenario, in which the mirror is moved in two locations, and the camera pose can be determined up to one unknown DOF. An extension to the case of more than two images recorded using the same mirror is also described. However, the issues of non-unique P3P solutions and multiple mirrors are not addressed, and no investigation of the method's accuracy is conducted. In our work, these limitations are removed.

Kumar et al. [36] presented a vision system that utilized a moving planar mirror to determine the transformations between multiple cameras with non-overlapping fields of view. A mirror was maneuvered in front of the camera rig, providing each camera with at least five views of a calibration grid, whose pose was fixed with respect to the cameras. Subsequently, the measurement constraints were transformed into a linear system of equations that was solved for the unknown extrinsic parameters of each camera. In contrast to that work, our approach requires fewer images (only three) for the single-mirror case in order to obtain a unique solution, and is also applicable to the case where more than one mirror is used.

Our previous conference publications [1], [2], [3] have also addressed the problem of mirror-based extrinsic camera calibration. Specifically, [1] presents a maximum-likelihood estimator for the camera extrinsic parameters and mirror configurations in the single-mirror case. Under the assumption of independent Gaussian pixel noise, this estimator takes the form of a nonlinear least-squares problem, solved by iterative Gauss-Newton minimization. Despite its ease of implementation, this method has the drawback that without an accurate initial guess, the minimization may require many iterations to converge or even fail to find the correct solution. For this reason, in [2] we presented an analytic solution to the single-mirror extrinsic calibration problem, which can be employed to provide a good initial guess for the maximum-likelihood estimator. In [3], we extended the single-mirror solution to the case where more than one mirror is used, and also presented a solution for robot-body reconstruction based on observations through multiple reflections. The current paper presents a comprehensive analysis of both the single- and multi-mirror extrinsic calibration and robot-body reconstruction problems. We discuss in detail the conditions under which unique solutions are feasible, and present both an analytic algorithm as well as a maximum-likelihood solution for estimating the unknown quantities. Additionally, the paper includes extensive simulation and real-world experimental results, which demonstrate the performance of the methods presented in several setups.

III. Problem Formulation

In what follows, we first present the single- and multiple-mirror measurement models, and discuss the relationship between the mirror-based calibration problem and the n-point pose estimation problem (PnP). Subsequently, we describe the conditions under which a solution to the problem can be found.

A. Measurement Model

We start by describing the measurement equations for each of the points observed by the camera. We employ a perspective camera that is intrinsically calibrated, e.g., by a method such as the one in [37]. Only one observation of each point is available in each image: its reflection through the N_v mirrors. If we let $\mathbf{p}'$ denote the reflected point seen by the camera, then the normalized image coordinates of the measurement

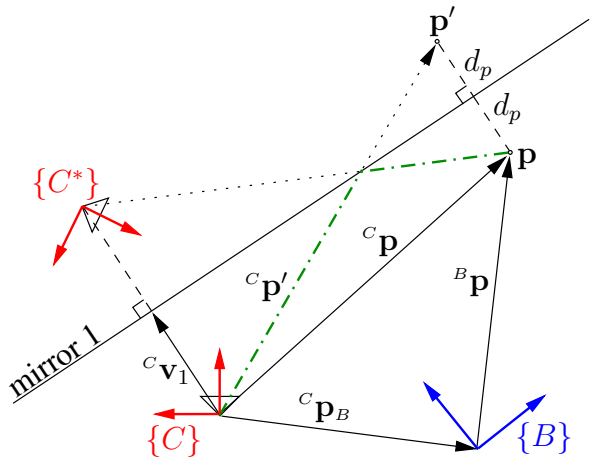


Fig. 2. In this illustration, the camera observes ${}^C\mathbf{p}'$, which is the reflection of $\mathbf{p}$ expressed in $\{C\}$. The base frame is $\{B\}$, while $\{C\}$ is the camera frame of reference, and $\{C^*\}$ is the equivalent *imaginary* camera frame which lies behind the mirror. The mirror vector, ${}^C\mathbf{v}_1 \equiv \mathbf{v}_1$, is the shortest vector from the origin of $\{C\}$ to the reflective surface. The vector ${}^C\mathbf{p}_B$ denotes the origin of $\{B\}$ with respect to $\{C\}$, while ${}^B\mathbf{p}$ and ${}^C\mathbf{p}$ denote the coordinates of $\mathbf{p}$ expressed in $\{B\}$ and $\{C\}$, respectively. The dash-dotted green line is the optical path of the reflection.

are described by the equation[1]:

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

$${}^C\mathbf{p}' = \begin{bmatrix} p_1 & p_2 & p_3 \end{bmatrix}^T, \tag{2}$$

where $\boldsymbol{\eta}$ is the noise, which is modeled as zero-mean white Gaussian with covariance matrix $\sigma_{\boldsymbol{\eta}}^2\mathbf{I}_2$. The relationship between the actual robot-body point, $\mathbf{p}$, and its reflection, $\mathbf{p}'$, depends on the mirror-reflection geometry, described next.

1) Single-mirror geometry: When the camera observes $\mathbf{p}$ via its reflection in a single mirror, from the geometry of Fig. 2 we obtain:

$${}^C\mathbf{p}' = {}^C\mathbf{p} + 2d_p\frac{{}^C\mathbf{v}_1}{\|{}^C\mathbf{v}_1\|} \tag{3}$$

$$d_p = \|{}^C\mathbf{v}_1\| - \frac{{}^C\mathbf{v}_1^T}{\|{}^C\mathbf{v}_1\|}{}^C\mathbf{p}, \tag{4}$$

where ${}^C\mathbf{p}'$ is the vector from the origin of the camera coordinate frame, $\{C\}$, to the reflected point $\mathbf{p}'$, ${}^C\mathbf{p}$ is the vector from $\{C\}$ to $\mathbf{p}$, ${}^C\mathbf{v}_1$ is the shortest vector from the origin of $\{C\}$ to the reflective surface (which we term "mirror vector"), and d_p is the distance between the mirror and the point $\mathbf{p}$. In the remainder of the paper, we will drop the reference frame from the mirror vector in order to simplify the notation (i.e., ${}^C\mathbf{v}_1 \equiv \mathbf{v}_1$).

In order to take advantage of the known coordinates of the fiducial points with respect to the base frame attached to the robot body, $\{B\}$, we make use of the frame-transformation equation between ${}^C\mathbf{p}$ and ${}^B\mathbf{p}$:

$${}^C\mathbf{p} = {}^C_B\mathbf{R}\,{}^B\mathbf{p} + {}^C\mathbf{p}_B, \tag{5}$$

[1]Throughout this paper, ${}^X\mathbf{y}$ denotes a vector $\mathbf{y}$ expressed with respect to frame $\{X\}$, ${}^X_W\mathbf{R}$ is the rotation matrix rotating vectors from frame $\{W\}$ to $\{X\}$, and ${}^X\mathbf{p}_W$ is the origin of $\{W\}$, expressed with respect to $\{X\}$. $\mathbf{I}_n$ is the $n \times n$ identity matrix, and $\mathbf{0}_{m\times n}$ is the $m \times n$ matrix of zeros.

where ${}^C_B\mathbf{R}$ is the matrix which rotates vectors from $\{B\}$ to $\{C\}$, and ${}^C\mathbf{p}_B$ is the origin of $\{B\}$ with respect to $\{C\}$. Note that the pair $\{{}^C_B\mathbf{R}, {}^C\mathbf{p}_B\}$ describes the camera-to-base transformation, which we seek to determine through the extrinsic calibration procedure. By substituting (4) and (5) into (3), after some manipulation, we obtain:

$${}^C\mathbf{p}' = \mathbf{M}_1\left({}^C_B\mathbf{R}\,{}^B\mathbf{p} + {}^C\mathbf{p}_B\right) + 2\,\mathbf{v}_1 \tag{6}$$

where $\mathbf{M}_1$ is the Householder transformation matrix corresponding to the reflection in mirror 1, given by:

$$\mathbf{M}_1 = \left(\mathbf{I}_3 - 2\frac{\mathbf{v}_1\mathbf{v}_1^T}{\mathbf{v}_1^T\mathbf{v}_1}\right). \tag{7}$$

It is useful to express (6) in homogeneous coordinates, as:

$$\begin{bmatrix} {}^C\mathbf{p}' \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{M}_1 & 2\,\mathbf{v}_1 \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix}\begin{bmatrix} {}^C_B\mathbf{R} & {}^C\mathbf{p}_B \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix}\begin{bmatrix} {}^B\mathbf{p} \\ 1 \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{A}_1 & \mathbf{b}_1 \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix}\begin{bmatrix} {}^B\mathbf{p} \\ 1 \end{bmatrix}. \tag{8}$$

In the last equation, the pair $\mathbf{A}_1 = \mathbf{M}_1{}^C_B\mathbf{R}$ and $\mathbf{b}_1 = \mathbf{M}_1{}^C\mathbf{p}_B + 2\mathbf{v}_1$, defines a composite Euclidean/reflection transformation, which converts ${}^B\mathbf{p}$ into ${}^C\mathbf{p}'$. Intuitively, we can consider the pair $\{\mathbf{A}_1, \mathbf{b}_1\}$ as a transformation from $\{B\}$ to an imaginary camera, $\{C^*\}$, with a *left-handed* reference frame, which lies behind the mirror and observes the true points (see Fig. 2).

Equations (1) and (8) define the measurement model in the single-mirror case. The observed image coordinates, $\mathbf{z}$, are a function of the known vector ${}^B\mathbf{p}$, the unknown camera-to-base transformation $\{{}^C_B\mathbf{R}, {}^C\mathbf{p}_B\}$, and the unknown configuration of the mirror with respect to the camera, $\mathbf{v}_1$. Note that, while in general six DOF are needed to fully determine the placement of the mirror, only three DOF appear in the measurement equation, via the 3×1 vector $\mathbf{v}_1$. The remaining three DOF, which correspond to rotations of the mirror about $\mathbf{v}_1$ and translations of the mirror within the mirror plane, do not affect the measurements.

2) Multiple-mirror geometry: We next present the geometric relationships arising in the multiple-mirror case, starting for simplicity with the two-mirror case, which is shown in Fig. 3. When two mirror reflections are required to view the robot-body, the optical ray from the point to the camera is first reflected by mirror 1, to yield the point $\mathbf{p}''$, and then reflected by mirror 2, to yield the point $\mathbf{p}'$, which is the one observed by the camera [see (1)]. From Fig. 3, by analogy to the previous case of a single mirror, we can write the following expression for $\mathbf{p}''$

$${}^C\mathbf{p}'' = \left(\mathbf{I}_3 - 2\frac{\mathbf{v}_1\mathbf{v}_1^T}{\mathbf{v}_1^T\mathbf{v}_1}\right){}^C\mathbf{p} + 2\,\mathbf{v}_1 = \mathbf{M}_1\,{}^C\mathbf{p} + 2\,\mathbf{v}_1. \tag{9}$$

While the point $\mathbf{p}'$ is expressed as a function of $\mathbf{p}''$ as:

$${}^C\mathbf{p}' = \left(\mathbf{I}_3 - 2\frac{\mathbf{v}_2\mathbf{v}_2^T}{\mathbf{v}_2^T\mathbf{v}_2}\right){}^C\mathbf{p}'' + 2\,\mathbf{v}_2 = \mathbf{M}_2{}^C\mathbf{p}'' + 2\,\mathbf{v}_2, \tag{10}$$

where $\mathbf{M}_2$ is the reflection matrix corresponding to mirror 2. We substitute (10) into (9) to obtain:

$${}^C\mathbf{p}' = \mathbf{M}_2\left(\mathbf{M}_1{}^C\mathbf{p} + 2\,\mathbf{v}_1\right) + 2\,\mathbf{v}_2. \tag{11}$$

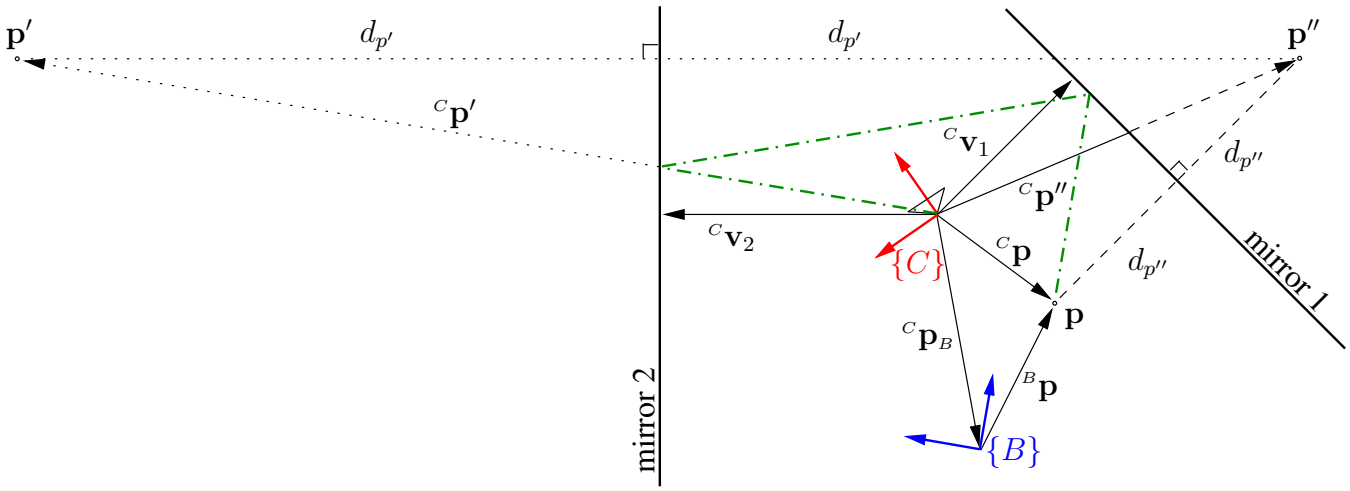


Fig. 3. Point $\mathbf{p}$ is visible to the camera $\{C\}$ via its double reflection, ${}^C\mathbf{p}'$, which is generated by first reflecting $\mathbf{p}$ in mirror 1, to obtain $\mathbf{p}''$, and then reflecting $\mathbf{p}''$ in mirror 2 to obtain $\mathbf{p}'$. The mirrors are described by the mirror vectors, ${}^C\mathbf{v}_1$ and ${}^C\mathbf{v}_2$, which are the shortest vectors from the origin of $\{C\}$ to the respective reflective surfaces. The points $\mathbf{p}''$ and $\mathbf{p}$ are at distance $d_{p''}$ from mirror 1, while $\mathbf{p}''$ and $\mathbf{p}'$ are at distance $d_{p'}$ from mirror 2. The dash-dotted green line denotes the optical path of the reflection.

Exploiting the relationship ${}^C\mathbf{p} = {}^C_B\mathbf{R}\,{}^B\mathbf{p} + {}^C\mathbf{p}_B$, and rewriting (11) in homogeneous coordinates, we obtain

$$\begin{bmatrix} {}^C\mathbf{p}' \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{M}_2 & 2\,\mathbf{v}_2 \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{M}_1 & 2\,\mathbf{v}_1 \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \begin{bmatrix} {}^C_B\mathbf{R} & {}^C\mathbf{p}_B \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \begin{bmatrix} {}^B\mathbf{p} \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_2 & \mathbf{b}_2 \\ \mathbf{0}_{3\times 1} & 1 \end{bmatrix} \begin{bmatrix} {}^B\mathbf{p} \\ 1 \end{bmatrix}, \tag{12}$$

where

$$\mathbf{A}_2 = \mathbf{M}_2\,\mathbf{M}_1\,{}^C_B\mathbf{R} \tag{13}$$

$$\mathbf{b}_2 = \mathbf{M}_2\,\mathbf{M}_1\,{}^C\mathbf{p}_B + 2\,\mathbf{M}_2\mathbf{v}_1 + 2\,\mathbf{v}_2. \tag{14}$$

By comparing (12) to (8), we observe that the use of a second mirror resulted in the introduction of a second reflection transformation in the expression for ${}^C\mathbf{p}'$. Extending this property to the general case of N_v mirrors, we obtain the following expression for ${}^C\mathbf{p}'$:

$$\begin{bmatrix} {}^C\mathbf{p}' \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{M}_{N_v} & 2\,\mathbf{v}_{N_v} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \cdots \begin{bmatrix} \mathbf{M}_1 & 2\,\mathbf{v}_1 \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \begin{bmatrix} {}^C_B\mathbf{R} & {}^C\mathbf{p}_B \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \begin{bmatrix} {}^B\mathbf{p} \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{N_v} & \mathbf{b}_{N_v} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \begin{bmatrix} {}^B\mathbf{p} \\ 1 \end{bmatrix}, \tag{15}$$

where $\mathbf{M}_\ell$, $\ell = 1, \ldots, N_v$, denotes the Householder reflection matrix corresponding to the vector $\mathbf{v}_\ell$:

$$\mathbf{M}_\ell = \left(\mathbf{I}_3 - 2\frac{\mathbf{v}_\ell\mathbf{v}_\ell^T}{\mathbf{v}_\ell^T\mathbf{v}_\ell}\right). \tag{16}$$

Together, (15) and (1) describe the measurement model when N_v mirrors are used to observe each point. A number of interesting remarks can be made about the structure of these equations. First, we see that the pair $\{\mathbf{A}_{N_v}, \mathbf{b}_{N_v}\}$ describes a combined Euclidean/reflection transformation, which depends on the camera-to-base transformation and the configuration of all mirrors. Moreover, these quantities (which are the unknowns in the extrinsic calibration problem) affect the camera observations *only* through the pair $\{\mathbf{A}_{N_v}, \mathbf{b}_{N_v}\}$. This means that computing this pair using the point observations in an image provides us with all the information available in the image for determining the unknowns.

The special significance of the pair $\{\mathbf{A}_{N_v}, \mathbf{b}_{N_v}\}$ warrants further investigation of its properties. First, we note that its components can be expressed recursively in terms of $\{\mathbf{A}_{N_v-1}, \mathbf{b}_{Nv-1}\}$. Specifically, based on (15), we can write:

$$\begin{aligned} \mathbf{A}_{N_v} &= \mathbf{M}_{N_v}\cdots\mathbf{M}_1\,{}^C_B\mathbf{R} \\ &= \mathbf{M}_{N_v}\mathbf{A}_{N_v-1} \end{aligned} \tag{17}$$

$$\begin{aligned} \mathbf{b}_{N_v} &= \mathbf{M}_{N_v}\cdots\mathbf{M}_1\,{}^C\mathbf{p}_B + 2\,\mathbf{M}_{N_v}\cdots\mathbf{M}_2\mathbf{v}_1 + \cdots \\ &\quad + 2\,\mathbf{M}_{N_v}\mathbf{M}_{N_v-1}\mathbf{v}_{N_v-2} + 2\,\mathbf{M}_{N_v}\mathbf{v}_{N_v-1} + 2\,\mathbf{v}_{N_v} \\ &= \mathbf{M}_{N_v}\mathbf{b}_{N_v-1} + 2\,\mathbf{v}_{N_v}. \end{aligned} \tag{18}$$

In fact, this recursive structure can also be extended to include the camera-to-base transformation, by defining:

$$\mathbf{A}_0 = {}^C_B\mathbf{R} \tag{19}$$

$$\mathbf{b}_0 = {}^C\mathbf{p}_B. \tag{20}$$

This fact will be useful in the presentation of our analytic solution in Section IV.

The matrix $\mathbf{A}_{N_v}$ has additional useful properties. Specifically, from (17) we observe that this matrix is the product of N_v Householder reflection matrices, $\mathbf{M}_\ell$, $\ell = 1, \ldots, N_v$, and one rotation matrix, ${}^C_B\mathbf{R}$. Both Householder and rotation matrices are unitary matrices; however, Householder matrices have determinant equal to -1, while rotation matrices' determinants equal +1. Using these properties, we can show that $\mathbf{A}_{N_v}$ is always unitary, and when N_v is even, $\mathbf{A}_{N_v}$ is itself a rotation matrix [i.e., $\det(\mathbf{A}_{N_v}) = +1$], while when N_v is odd, $\det(\mathbf{A}_{N_v}) = -1$. The usefulness of these properties will become evident in Section IV-A.

B. Relationship to PnP

At this point, it is interesting to examine the relationship between the measurement model described above and the

TABLE I
DEFINITION OF SYMBOLS

N_f	number of fiducial points
N_r	number of reconstruction points
N_c	number of images
N_v	number of mirrors
i	index for points
j	index for mirror configurations / images
k	index for the MLE iteration
ℓ	index for mirrors
$\{C\}$	camera frame
$\{B\}$	base frame
${}^C_B\mathbf{R}$	camera-to-base rotation matrix
${}^C\mathbf{p}_B$	camera-to-base translation vector
${}^C\mathbf{p}$	point of interest in the camera frame
${}^C\mathbf{p}'$	reflection of $\mathbf{p}$, observed by the camera
${}^C\mathbf{v}_{\ell(j)}$	mirror vector for the j-th configuration of mirror ℓ
$\mathbf{M}_{\ell(j)}$	reflection matrix corresponding to ${}^C\mathbf{v}_{\ell(j)}$
$\mathbf{A}_{N_v}$	effective rotation/reflection matrix in the N_v-mirror case
$\mathbf{b}_{N_v}$	effective translation vector in the N_v-mirror case
$\mathbf{z}_{ij}$	normalized image coordinates of point i observed in image j

measurement model associated with the "traditional" perspective n-point camera pose determination problem (PnP). The goal of the latter is to determine the six-DOF transformation $\{{}^C_B\mathbf{R}, {}^C\mathbf{p}_B\}$ between the camera frame $\{C\}$ and a base frame $\{B\}$, based on observations of N_f fiducial points whose base-frame coordinates, ${}^B\mathbf{p}_i$, $i = 1, \ldots, N_f$, are known. The observations of each of these points are described by the following equations:

$$\mathbf{z}_i = \frac{1}{p_{3i}} \begin{bmatrix} p_{1i} \\ p_{2i} \end{bmatrix}, \quad {}^C\mathbf{p}'_i = \begin{bmatrix} p_{1i} & p_{2i} & p_{3i} \end{bmatrix}^T \tag{21}$$

$$\begin{bmatrix} {}^C\mathbf{p}'_i \\ 1 \end{bmatrix} = \begin{bmatrix} {}^C_B\mathbf{R} & {}^C\mathbf{p}_B \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \begin{bmatrix} {}^B\mathbf{p}_i \\ 1 \end{bmatrix}. \tag{22}$$

By comparison of the above two equations to (1) and (15), it becomes clear that the structure of the two sets of equations is similar. The difference is that in the PnP problem the unknowns are described by the transformation pair $\{{}^C_B\mathbf{R}, {}^C\mathbf{p}_B\}$, while in the mirror-based extrinsic calibration problem all the unknowns are included in the pair $\{\mathbf{A}_{N_v}, \mathbf{b}_{N_v}\}$, which depends on $\{{}^C_B\mathbf{R}, {}^C\mathbf{p}_B\}$, as well as on the mirror vectors $\mathbf{v}_\ell$, $\ell = 1, \ldots, N_v$. Thus, the number of unknowns in the mirror-based problem is higher. Despite this difference, the similarity of the two problems allows us to build upon existing solutions for the PnP problem, and utilize them in the analytic method we propose in Section IV.

C. Problem statement and conditions for computing a solution

We now formally state the problem we solve in this paper. Our goal is to determine (i) the six-DOF transformation, $\{{}^C_B\mathbf{R}, {}^C\mathbf{p}_B\}$, between the camera frame $\{C\}$ and the base frame $\{B\}$, and (ii) the base-frame coordinates of N_r "reconstruction points," ${}^B\mathbf{p}_i$, $i = 1, \ldots, N_r$. To accomplish these two objectives, we use the observations in N_c images of the N_r reconstruction points, as well as of N_f fiducial points, whose coordinates are known in $\{B\}$. The observation of the ith point in the jth image is denoted by $\mathbf{z}_{ij}$, with $i = 1, \ldots, (N_f + N_r)$ and $j = 1, \ldots, N_c$. Each of these observations is the result of reflection in N_v mirrors, and is described by the measurement

TABLE II
CONDITIONS FOR COMPUTING A SOLUTION WITH $N_v = 1$: NUMBER OF IMAGES VS. NUMBER OF POINTS

	$N_c = 1, 2$	$N_c \geq 3$
$N_f < 3$	Not enough information to determine $\mathbf{A}$ and $\mathbf{b}$	
$N_f = 3$	Discrete solutions for $\mathbf{A}_{1(j)}$ and $\mathbf{b}_{1(j)}$, infinite solutions for ${}^C_B\mathbf{R}$, ${}^C\mathbf{p}_B$, and $\mathbf{v}_{1(j)}$	Discrete solutions for $\mathbf{A}_{1(j)}$ and $\mathbf{b}_{1(j)}$, unique solution for ${}^C_B\mathbf{R}$, ${}^C\mathbf{p}_B$, and $\mathbf{v}_{1(j)}$ **[$N_c = 3$: Minimal problem with a unique solution]**
$N_f > 3$	Unique solutions for $\mathbf{A}_{1(j)}$ and $\mathbf{b}_{1(j)}$, infinite solutions for ${}^C_B\mathbf{R}$, ${}^C\mathbf{p}_B$, and $\mathbf{v}_{1(j)}$	Unique solutions for $\mathbf{A}_{1(j)}$ and $\mathbf{b}_{1(j)}$, unique solution for ${}^C_B\mathbf{R}$, ${}^C\mathbf{p}_B$, and $\mathbf{v}_{1(j)}$

model presented in Section III-A. Since the configuration of the mirrors is not known *a priori*, the mirror vectors corresponding to each image, $\mathbf{v}_{\ell(j)}$, with $\ell = 1, \ldots, N_v$ and $j = 1, \ldots, N_c$, are additional unknowns in the problem. Note that, in general, each of the mirrors may move to a new location for each image. However, we will show that for a unique solution to exist, some restrictions need to be imposed on the motion of the mirrors. For easy reference, Table I summarizes the notation used throughout the paper.

Our solution approach is to first compute an initial estimate for all unknowns using an analytic algorithm (Sections IV and V), and to subsequently refine the closed-form solution using an MLE (Section VI). For both steps, the existence and number of solutions is an important consideration. We now examine what are the requirements on the number of fiducial points, N_f, the number of images, N_c, and the motion of the mirrors, to ensure that a unique solution can be computed. In this section, we focus on the existence of solutions for the first objective stated above (computing the camera-to-base transformation), while the requirements for computing a solution to the second objective are discussed in Section V.

As explained in Section III-A2, the pair $\{\mathbf{A}_{N_v}, \mathbf{b}_{N_v}\}$ fully describes the reflection geometry for each image, and encapsulates all the relevant unknowns. Thus, computing this pair for each of the available images is the first step to an analytic solution. To examine the conditions under which a solution for $\{\mathbf{A}_{N_v}, \mathbf{b}_{N_v}\}$ can be computed, we leverage the analogy between our problem and PnP, discussed in Section III-B. Specifically, due to the similarity of the problem structure, the conditions under which the pair $\{\mathbf{A}_{N_v}, \mathbf{b}_{N_v}\}$ can be determined in the mirror-based calibration problem are the same as the conditions for computing the camera-to-base transformation in the PnP. There exist three cases of interest:

- $N_f < 3$**:** When less than three fiducial points are available in PnP, there is not enough information to determine the transformation between the camera and the base frame. This is because at least three points are needed to define the transformation between two frames of reference: if, for example, only two points are available, then any rotation of the two frames about the line which passes through the two points would not affect the measurements. Analogously, when $N_f < 3$, the mirror-based calibration problem cannot be solved.
- $N_f = 3$**:** When exactly three non-collinear fiducial points

are available the problem at hand is analogous to a P3P problem. In this case, we can employ existing P3P algorithms to compute a discrete number of solutions for the pair $\{\mathbf{A}_{N_v}, \mathbf{b}_{N_v}\}$ for each of the available images (see Section III-B). Several approaches have been proposed for the P3P since its first introduction [14], and it is well known that there may exist up to four pairs of solutions for the problem, where for each pair, one solution lies in front of the center of perspectivity and one behind it [17].

- $N_f > 3$**:** When more than three non-collinear fiducial points are visible, the problem is analogous to the PnP, with $n > 3$. By analogy to the PnP, in this case we can determine a unique solution for the pair $\{\mathbf{A}_{N_v}, \mathbf{b}_{N_v}\}$, barring degenerate configurations [17].

Based on the above discussion, we see that when at least three fiducial points are available (i.e., $N_f \geq 3$), we can compute one or more discrete solutions for $\{\mathbf{A}_{N_v(j)}, \mathbf{b}_{N_v(j)}\}$, for $j = 1, \ldots, N_c$. Our next task is to use these solutions in order to determine the camera-to-base transformation, as well as all the mirror vectors. In order to establish conditions under which this is possible, in what follows we examine the number of unknowns in the system, and compare it with the number of available constraint equations.

Recall that, as shown in Section III-A2, $\mathbf{A}_{N_v}$ is a unitary matrix, and consequently it only has three DOF. This means that determining the pairs $\{\mathbf{A}_{N_v(j)}, \mathbf{b}_{N_v(j)}\}$ for $j = 1, \ldots, N_c$ only provides 6 independent constraint equations per image [3 from the unitary matrix $\mathbf{A}_{N_v}$, and 3 from the vector $\mathbf{b}_{N_v}$, see (17)-(20)]. Thus, from N_c images, we can obtain $6N_c$ constraint equations. On the other hand, the six-DOF camera-to-base transformation corresponds to 6 unknowns and, if N_v mirrors are used, each of the mirrors introduces 3 unknowns per image (the three elements of the vector $\mathbf{v}_{\ell(j)}$). Therefore, if in each of the N_c images each mirror moves to a new location, the total number of unknowns is equal to $3N_vN_c + 6$. At this point we distinguish two cases of interest:

1) Single-mirror-based calibration: When $N_v = 1$, then the total number of unknowns is $3N_c + 6$. Comparing this number with $6N_c$, the number of available constraints, we identify four cases:

- $N_c = 1$: When one image is available, there are 6 constraint equations arising from the image measurements with 9 unknowns (i.e., an underdetermined system), and thus we cannot compute a discrete set of solutions.
- $N_c = 2$: When two images are available, there are 12 constraint equations with 12 unknowns (i.e., a square system). However, we show in [38] that in this case there exists an unobservable mode in the system, which corresponds to rotations about the vector $\mathbf{v}_{1(1)} \times \mathbf{v}_{1(2)}$. Thus, even though the number of unknowns equals the number of available constraints, no discrete solutions can be computed.[2]
- $N_c = 3$: When three images are available, the nonlinear system is overdetermined with 15 unknowns and 18 constraints. If the mirror vectors $\mathbf{v}_{1(j)}$, $j = 1, 2, 3$ are linearly independent, a unique solution for ${}^C_B\mathbf{R}$, ${}^C\mathbf{p}_B$, and $\mathbf{v}_{1(j)}$, $j = 1, 2, 3$, can be computed, by application of the algorithm presented in Section IV.
- $N_c \geq 3$: Finally, if more than three images are available, we can also compute a unique solution provided that at least three mirror vectors are linearly independent.

From the above discussion we conclude that in the single-mirror case the minimal problem, for which a unique solution can be computed, is one with $N_f = 3$ and $N_c = 3$, under the constraint that the three mirror vectors are linearly independent. For easy reference, Table II summarizes the results for the single-mirror case

2) Multiple-mirror-based calibration: When $N_v > 1$ and each mirror moves to a new location for each image, then the number of unknowns, $3N_vN_c + 6$, is always larger than the number of constraints available, $6N_v$. Therefore, unless some restrictions are imposed on the motion of the mirrors, no discrete set of solutions can be computed. We propose to move the mirrors in a specific order that reduces the number of unknowns and allows for computing a unique solution. Denoting the j-th configuration of mirror ℓ as $\mathbf{v}_{\ell(j)}$, our strategy requires that for each configuration of the first ℓ mirrors ($\mathbf{v}_{1(j)}$ to $\mathbf{v}_{\ell(j)}$), the $(\ell+1)$-th mirror is moved to at least three different configurations ($\mathbf{v}_{\ell+1(j)}$, $\mathbf{v}_{\ell+1(j')}$, $\mathbf{v}_{\ell+1(j'')}$) with an image recorded in each of them.

This approach is illustrated in Fig. 4 for the simple case of two mirrors. Each "leaf node" in the ternary tree corresponds to one image, and tracing the path from the leaf node to the root of the tree specifies the mirror configurations for that image. The index j also corresponds to the image number at the leaf nodes, since there is one image recorded per configuration of mirror N_v. We employ this fact to simplify the notation, and we index both the mirror configurations and the images with j.

In the two-mirror case, the total number of mirror-1 configurations is three, and the total number of mirror-2 configurations is nine. In the general case of N_v mirrors, this strategy results in 3^{N_v} images with 6×3^{N_v} constraints, and hence $6 + 3(3 + 3^2 + \ldots + 3^{N_v}) = 6 + \frac{9}{2}(3^{N_v} - 1)$ unknowns. Since $6 \times 3^{N_v} > 6 + \frac{9}{2}(3^{N_v} - 1)$, the resulting problem is an over-determined one, and a unique solution can be computed, as shown in the next section. The sequence of mirror motions described above is motivated by the recursive structure of the measurement equations, described in Section III-A. Using this strategy, it is possible to first determine all the mirror vectors for mirror N_v, then for mirror $N_v - 1$, and to continue recursively until the camera-to-base transformation is found.

IV. Analytic Solution for the Camera-to-Base Transformation

In this section, we present the analytic solution to the mirror-based extrinsic calibration problem. The main steps of the proposed approach are:

- In the first step, we compute the pair $\{\mathbf{A}_{N_v(j)}, \mathbf{b}_{N_v(j)}\}$, for each image $j = 1, \ldots, N_c$ (Section IV-A).
- In the second step, the solutions obtained above are used to compute the mirror vectors for the N_v-th mirror, $\mathbf{v}_{N_v(j)}$, $j = 1, \ldots, N_c$ (Section IV-B).

[2]Sturm and Bonfort [35] also report this case, providing an alternative algebraic proof for the unobservable direction.

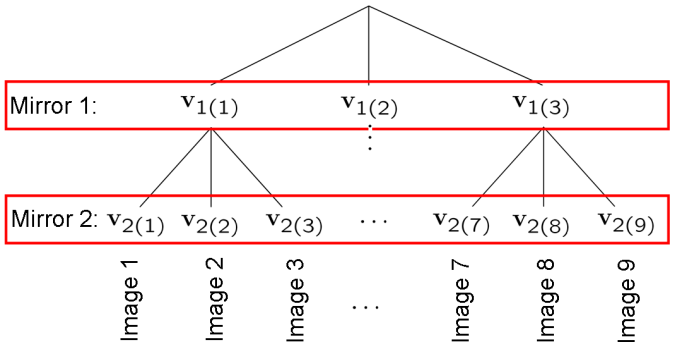


Fig. 4. The mirror configurations for the two-mirror case are depicted as a full ternary tree.

Algorithm 1 Extrinsic Calibration and 3D Reconstruction

Input: Observations of N_f fiducial (known) points and N_r reconstruction (unknown) points tracked in N_c images

Output: Camera-to-base transformation $\{{}^{C}_{B}\mathbf{R}, {}^{C}\mathbf{p}_{B}\}$ and 3D base-frame coordinates of reconstruction points ${}^{B}\mathbf{p}_i$, $i = 1, \ldots, N_r$

for each image $j = 1, \ldots, N_c$ **do**
  If N_v is odd, convert to N_f-point pose estimation problem (23)
  Solve PnP to obtain $\{\mathbf{A}_{N_v(j)}, \mathbf{b}_{N_v(j)}\}$ (24)
end for
for each mirror $\ell = 1, \ldots, N_v$ **do**
  Compute mirror configurations from (27)-(30)
  Eliminate ℓ-th mirror from the problem (31)-(32)
end for
Compute camera-to-base rotation ${}^{C}_{B}\mathbf{R}$ from (19)
Compute camera-to-base translation ${}^{C}\mathbf{p}_{B}$ from (20)
for each reconstruction point ${}^{B}\mathbf{p}_i$, $i = 1, \ldots, N_r$ **do**
  Compute 3D position using the observations from all images (35)
end for
Refine the solution using the maximum-likelihood estimator (see Section VI)

- Given the pairs $\{\mathbf{A}_{N_v(j)}, \mathbf{b}_{N_v(j)}\}$ and the vectors $\mathbf{v}_{N_v(j)}$, $j = 1, \ldots, N_c$, we compute the pairs $\{\mathbf{A}_{N_v-1(j)}, \mathbf{b}_{N_v-1(j)}\}$, and use them to solve for all the $\mathbf{v}_{N_v-1(j)}$, $j = 1, \ldots, \frac{N_c}{3}$ mirror vectors. This process is repeated recursively for $\ell = N_v - 1, N_v - 2, \ldots, 0$, to compute the configurations of all the mirrors and finally, the camera-to-base transformation (Section IV-C).

Once these steps are complete, we proceed to compute a solution analytically for the position of the reconstruction points, as described in Section V. For reference, the steps of the solution process are described in Algorithm 1.

A. Solving for $\{\mathbf{A}_{N_v(j)}, \mathbf{b}_{N_v(j)}\}$

As explained in Section III-B, the mathematical formulation of the mirror-based calibration problem is similar to the formulation of the PnP problem, with the key difference being that in the mirror-based problem the pair $\{\mathbf{A}_{N_v(j)}, \mathbf{b}_{N_v(j)}\}$ replaces the pair $\{{}^{C}_{B}\mathbf{R}, {}^{C}\mathbf{p}_{B}\}$ in the measurement equations. We now show that this similarity can be leveraged in order to compute $\{\mathbf{A}_{N_v(j)}, \mathbf{b}_{N_v(j)}\}$ for each image j. Specifically, we take advantage of the fact that, as discussed in Section III-A2, when N_v is even $\mathbf{A}_{N_v(j)}$ is a rotation matrix. Therefore, for N_v even, the pair $\{\mathbf{A}_{N_v(j)}, \mathbf{b}_{N_v(j)}\}$ is "equivalent" to a Euclidean transformation, described by a rotation matrix $\mathbf{A}_{N_v(j)}$ and a translation vector $\mathbf{b}_{N_v(j)}$. In turn, this means that we can directly apply any PnP solution method to obtain $\mathbf{A}_{N_v(j)}$ and $\mathbf{b}_{N_v(j)}$. In the experiments presented in this paper, $N_f = 3$, and we solve the corresponding P3P problem using the solution presented by Fischler and Bolles [16].

The situation is slightly different when N_v is odd, since in this case $\mathbf{A}_{N_v(j)}$ is no longer a rotation matrix (it is unitary but its determinant is -1), and we cannot directly apply a PnP solution. Although it is possible to formulate and solve a modified PnP problem for this case, there is a simpler method which allows us to use existing PnP algorithms. In particular, we transform the problem to convert $\mathbf{A}_{N_v(j)}$ into a rotation matrix, by applying an additional *known* reflection to the measurements. In our implementation, we accomplish this by negating all of the y-coordinates of the observed points, which corresponds to reflecting all the points across the xz-plane of the camera frame. Thus, the measurement equation in this case becomes $\mathbf{z}_{ij} = \mathbf{h}({}^{\breve{C}}\mathbf{p}_{ij}) + \boldsymbol{\eta}$, where

$$\begin{aligned}\begin{bmatrix} {}^{\breve{C}}\mathbf{p}_{ij} \\ 1 \end{bmatrix} &= \begin{bmatrix} (\mathbf{I}_3 - 2\mathbf{e}_2\mathbf{e}_2^T) & \mathbf{0}_{3\times 1} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \begin{bmatrix} {}^{C}\mathbf{p}'_{ij} \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} (\mathbf{I}_3 - 2\mathbf{e}_2\mathbf{e}_2^T) & \mathbf{0}_{3\times 1} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{A}_{N_v(j)} & \mathbf{b}_{N_v(j)} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \begin{bmatrix} {}^{B}\mathbf{p}_i \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} {}^{\breve{C}}_{B}\mathbf{R}_{(j)} & {}^{\breve{C}}\mathbf{p}_{B(j)} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \begin{bmatrix} {}^{B}\mathbf{p}_i \\ 1 \end{bmatrix}, \end{aligned} \tag{23}$$

where $\mathbf{e}_2 = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T$. Note that the matrix ${}^{\breve{C}}_{B}\mathbf{R}_{(j)}$ is a rotation matrix, not a reflection. Thus, after negating the y-coordinates of all the points in the image, we can solve the PnP to obtain solution(s) for the unknown transformation $\{{}^{\breve{C}}_{B}\mathbf{R}_{(j)}, {}^{\breve{C}}\mathbf{p}_{B(j)}\}$. Subsequently, given the PnP solution, we recover $\mathbf{A}_{N_v(j)}$ and $\mathbf{b}_{N_v(j)}$ through the following relationship which follows directly from (23)

$$\begin{bmatrix} \mathbf{A}_{N_v(j)} & \mathbf{b}_{N_v(j)} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} = \begin{bmatrix} (\mathbf{I}_3 - 2\mathbf{e}_2\mathbf{e}_2^T) & \mathbf{0}_{3\times 1} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \begin{bmatrix} {}^{\breve{C}}_{B}\mathbf{R}_{(j)} & {}^{\breve{C}}\mathbf{p}_{B(j)} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \tag{24}$$

for $j = 1, \ldots, N_c$.

B. Determining the configuration of the N_v-th mirror

Once the pair $\{\mathbf{A}_{N_v(j)}, \mathbf{b}_{N_v(j)}\}$, has been determined for all images ($j = 1, \ldots, N_c$), we then employ these matrices in order to compute the vectors of mirror N_v, $\mathbf{v}_{N_v(j)}$. For simplicity, we first present the solution for the minimal case, in which exactly three images ($N_c = 3$) are taken for each configuration of mirrors 1 through $N_v - 1$, and we momentarily ignore the possibility of multiple PnP solutions for the pair $\{\mathbf{A}_{N_v(j)}, \mathbf{b}_{N_v(j)}\}$. These issues are discussed at the end of this subsection. To demonstrate the procedure, we focus on the first three images recorded by the camera. Let us assume that for these images the N_v-th mirror is moved to three different locations, described by the vectors $\mathbf{v}_{N_v(j)}$, $j = 1, 2, 3$, while

all the other mirrors in the system remain in place (as in the situation depicted in Fig. 4). In this case, we can write the equations [see (17)]:

$$\mathbf{A}_{N_v(j)} = \mathbf{M}_{N_v(j)}\mathbf{A}_{N_v-1(1)}, \quad j = 1, 2, 3. \tag{25}$$

Let us now consider the unit vector $\mathbf{r}_{jj'}$ that is perpendicular to both $\mathbf{v}_{N_v(j)}$ and $\mathbf{v}_{N_v(j')}$ $(j, j' \in \{1, 2, 3\},\ j \neq j')$. This vector satisfies $\mathbf{v}_{N_v(j)}^T \mathbf{r}_{jj'} = \mathbf{v}_{N_v(j')}^T \mathbf{r}_{jj'} = 0$, and using these properties, we can show that [see (25) and (16)]:

$$\begin{aligned}\mathbf{A}_{N_v(j)}\mathbf{A}_{N_v(j')}^T\mathbf{r}_{jj'} &= \mathbf{M}_{N_v(j)}\mathbf{A}_{N_v-1(1)} \\ &\quad \times \mathbf{A}_{N_v-1(1)}^T\mathbf{M}_{N_v(j')}^T\mathbf{r}_{jj'} \\ &= \mathbf{M}_{N_v(j)}\mathbf{M}_{N_v(j')}^T\mathbf{r}_{jj'} \\ &= \mathbf{r}_{jj'},\end{aligned} \tag{26}$$

where we also used the fact that $\mathbf{A}_{N_v-1(1)}$ is unitary, $\mathbf{A}_{N_v-1(1)}\mathbf{A}_{N_v-1(1)}^T = \mathbf{I}_3$. Equation (26) states that the unit vector $\mathbf{r}_{jj'}$ is the eigenvector of $\mathbf{A}_{N_v(j)}\mathbf{A}_{N_v(j')}^T$ corresponding to the unit eigenvalue. Since the matrices $\mathbf{A}_{N_v(j)}$, $j = 1, 2, 3$ have already been computed using the PnP in the previous step, we are able to determine the vectors $\mathbf{r}_{jj'}$ up to sign, via eigen-decomposition of $\mathbf{A}_{N_v(j)}\mathbf{A}_{N_v(j')}^T$. Once this step is complete, we can determine the vectors $\mathbf{v}_{N_v(j)}$, $j = 1, 2, 3$ up to scale via the following relationships:

$$\mathbf{v}_{N_v(1)} = c_{N_v(1)}\,\mathbf{r}_{13} \times \mathbf{r}_{12} \tag{27}$$

$$\mathbf{v}_{N_v(2)} = c_{N_v(2)}\,\mathbf{r}_{21} \times \mathbf{r}_{23} \tag{28}$$

$$\mathbf{v}_{N_v(3)} = c_{N_v(3)}\,\mathbf{r}_{13} \times \mathbf{r}_{23}, \tag{29}$$

where $c_{N_v(j)}$, $j = 1, 2, 3$, are unknown scalars that can be determined using (18). Specifically, by substituting the above three equations in (18) for each of the first three images, we formulate the following overdetermined linear system:

$$\begin{bmatrix} \mathbf{M}_{N_v(1)} & 2\,\mathbf{r}_{13}\times\mathbf{r}_{12} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} \\ \mathbf{M}_{N_v(2)} & \mathbf{0}_{3\times1} & 2\,\mathbf{r}_{21}\times\mathbf{r}_{23} & \mathbf{0}_{3\times1} \\ \mathbf{M}_{N_v(3)} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & 2\,\mathbf{r}_{13}\times\mathbf{r}_{23} \end{bmatrix} \begin{bmatrix} \mathbf{b}_{N_v-1(1)} \\ c_{N_v(1)} \\ c_{N_v(2)} \\ c_{N_v(3)} \end{bmatrix} = \begin{bmatrix} \mathbf{b}_{N_v(1)} \\ \mathbf{b}_{N_v(2)} \\ \mathbf{b}_{N_v(3)} \end{bmatrix}. \tag{30}$$

Solving this linear system provides a unique solution for the vector $[\mathbf{b}_{N_v-1(1)}^T \quad c_{N_v(1)} \quad c_{N_v(2)} \quad c_{N_v(3)}]^T$, which in turn, allows us to compute $\mathbf{v}_{N_v(j)}$, $j = 1, 2, 3$, using (27)-(29). In the above system, the vectors $\mathbf{b}_{N_v(j)}$ appearing on the right-hand side are available from the PnP solution for the pairs $\{\mathbf{A}_{N_v(j)}, \mathbf{b}_{N_v(j)}\}$, $j = 1, 2, 3$, the vectors $\mathbf{r}_{jj'}$ are computed by the eigen-decomposition of the matrices $\mathbf{A}_{N_v(j)}\mathbf{A}_{N_v(j')}^T$, and finally, the matrices $\mathbf{M}_{N_v(j)}$ are computed using the vectors in (27)-(29). Note that the definition of the Householder matrix $\mathbf{M}_\ell$ in (16) is invariant to the length or sign of the vector $\mathbf{v}_\ell$, thus the matrices appearing on the left-hand side in the above equation can be evaluated even before the scale factors $c_{N_v(j)}$ are known.

1) Dealing with multiple solutions: Up to this point, we have assumed that each of the PnP solutions was unique, in order to simplify the presentation of the analytic solution. However, in the general case multiple PnP solutions may exist (e.g., up to 4 admissible ones when $N_f = 3$). In that case, we compute an analytic solution by following the above procedure for each of the PnP solutions for each of the images, and select the solution which yields the minimum residual in (30). For instance, if four solutions exist for each of the pairs $\{\mathbf{A}_{N_v(j)}, \mathbf{b}_{N_v(j)}\}$, $j = 1, 2, 3$, then we will apply the above procedure $4 \times 4 \times 4 = 64$ times, and keep the best solution. If the measurements were noise-free, we would expect only one solution to have zero error, since the nonlinear system we are solving is over-constrained. In practice, when noise is present, we have found that choosing the solution with the minimum error is a suitable way of rejecting the invalid solutions.

2) Dealing with multiple images: In our mirror-motion strategy, we require that mirror N_v is moved to at least three locations for each configuration of mirrors 1 through $N_v - 1$. However, to increase the accuracy of the estimates, we may choose to record more than three images. This corresponds to a measurement tree (see Fig. 4) with a branching factor larger than three. In this case, we select triplets of images to analytically compute the mirror vectors for all of the N_v-th mirror locations. Amongst all triplets we select the one that yields the minimum residual in (30) to use in the recursive process described in the following section.

C. Solving for the remaining mirror vectors, and the camera-to-base transformation

Following the procedure described in the previous section, we can determine the mirror vectors $\mathbf{v}_{N_v(j)}$ for the last mirror in all the N_c images. Once this process is complete, we can eliminate all the variables related to mirror N_v, i.e., the vectors $\mathbf{v}_{N_v(j)}$, $j = 1, \ldots, N_c$. This is accomplished by the following equations, which result from (17) and (18):

$$\mathbf{A}_{N_v-1(j')} = \mathbf{M}_{N_v(j)}^{-1}\mathbf{A}_{N_v(j)} \tag{31}$$

$$\mathbf{b}_{N_v-1(j')} = \mathbf{M}_{N_v(j)}^{-1}\left(\mathbf{b}_{N_v(j)} - 2\,\mathbf{v}_{N_v(j)}\right), \tag{32}$$

for $j = 1, \ldots, 3^{N_v} = N_c$, $j' = \lceil j/3 \rceil$ (where $\lceil\cdot\rceil$ denotes the round-up operation). Note that since at least three different images (three different values of j) correspond to the same index j', we obtain three estimates of the pair $\{\mathbf{A}_{N_v-1(j')}\mathbf{b}_{N_v-1(j')}\}$. Due to the presence of noise, these estimates will generally be slightly different. From these estimates we compute a single, "average" estimate for $\mathbf{A}_{N_v-1(j')}$ and $\mathbf{b}_{N_v-1(j')}$, using the least-squares method presented in [2], which accounts for the fact that $\mathbf{A}_{N_v-1(j')}$ is a unitary matrix.

Once the above procedure is complete, we have effectively converted the original N_v-mirror problem into an instance of an (N_v-1)-mirror problem. Hence, we can follow the process outlined in Section IV-B to compute all the mirror vectors for mirror $N_v - 1$. Subsequently, we can eliminate all the variables associated with mirror $N_v - 1$, and create an instance of a $(N_v - 2)$-mirror problem. Proceeding recursively in this manner, we can compute all the mirror vectors in the system. Most importantly, this recursive process also provides us with the camera-to-base transformation, $\{{}_B^C\mathbf{R}, {}^C\mathbf{p}_B\}$, which is the main goal of the calibration process [see (19) and (20)].

V. Analytic Solution for Robot-body Reconstruction

In the previous section, we described how to analytically determine the camera-to-base transformation and the mirror

configurations using the observations of the fiducial points. We now turn our attention to computing the 3D base-frame coordinates of the reconstruction points, whose positions are *not known a priori*. We present an analytic method for this problem, which employs the solution for the camera-to-base transformation and the mirror configurations computed in the previous step.

Consider a single reconstruction point, ${}^B\mathbf{p}_i$, observed via reflection through N_v mirrors in N_c images. If we denote the measured unit-vector direction towards the reflected point as

$$ {}^C\hat{\mathbf{p}}'_{ij} = \frac{1}{\sqrt{1+\mathbf{z}_{ij}^T\mathbf{z}_{ij}}} \begin{bmatrix} \mathbf{z}_{ij} \\ 1 \end{bmatrix}, \tag{33} $$

then we can write the expression:

$$ s_{ij}\,{}^C\hat{\mathbf{p}}'_{ij} = \mathbf{A}_{N_v(j)}\,{}^B\mathbf{p}_i + \mathbf{b}_{N_v(j)}, \quad j = 1,\ldots,N_c, \tag{34} $$

where the scalar s_{ij} is the *unknown* distance to the reflected point in image j. The measurement model in (34) is equivalent to the perspective projection model defined in Section III-A, but has the property that it is linear in the unknowns s_{ij} and ${}^B\mathbf{p}_i$. When ${}^B\mathbf{p}_i$ is observed in at least two images (i.e., $N_c \geq 2$), we can form an overdetermined set of linear equations:

$$ \begin{bmatrix} \mathbf{A}_{N_v(1)} & -{}^C\hat{\mathbf{p}}_{i1} & \ldots & \mathbf{0}_{3\times 1} \\ \mathbf{A}_{N_v(2)} & \mathbf{0}_{3\times 1} & \ldots & \mathbf{0}_{3\times 1} \\ \vdots & & \ddots & \vdots \\ \mathbf{A}_{N_v(N_c)} & \mathbf{0}_{3\times 1} & \ldots & -{}^C\hat{\mathbf{p}}_{iN_c} \end{bmatrix} \begin{bmatrix} {}^B\mathbf{p}_i \\ s_{i1} \\ \vdots \\ s_{iN_c} \end{bmatrix} = \begin{bmatrix} -\mathbf{b}_{N_v(1)} \\ -\mathbf{b}_{N_v(2)} \\ \vdots \\ -\mathbf{b}_{N_v(N_c)} \end{bmatrix}. $$

The matrices $\mathbf{A}_{N_v(j)}$ and $\mathbf{b}_{N_v(j)}$, $j = 1,\ldots,N_c$, appearing above are available from the analytic solution presented in Section IV, while the vectors ${}^C\hat{\mathbf{p}}_{ij}$ are computed using the image measurements, as shown in (33). Thus the only unknowns in the system are the scale factors s_{ij} and the feature position, ${}^B\mathbf{p}_i$. This system has $3N_c$ linear equations in N_c+3 unknowns, and by solving it we obtain the point's coordinates in the base frame.

It is interesting to note that the above presented solution only requires knowledge of the pair $\{\mathbf{A}_{N_v(j)}, \mathbf{b}_{N_v(j)}\}$ for each of the images, but *not necessarily* knowledge of the camera-to-base transformation or of the mirror vectors. Therefore, there are fewer requirements for computing the 3D position of the reconstruction points than for determining the camera pose in the base frame. Specifically, if at least two images are available, for which the pair $\{\mathbf{A}_{N_v(j)}, \mathbf{b}_{N_v(j)}\}$ can be computed (i.e., at least two images with $N_f \geq 3$), a solution for the reconstruction points' coordinates can be found. This is the case *regardless* of the number of mirrors, an observation which may have useful practical implications.

VI. MLE Refinement of Analytic Solutions

In the preceding sections, we showed how to compute the analytic solutions for the camera-to-base transformation, the mirror configurations, and the reconstruction points. However, the accuracy of these solutions may not be the best possible, due to the fact that the statistical properties of the noise have not been properly accounted for. For this reason, we refine the analytic solution with an MLE for all the unknowns.

Let the vector of all unknown parameters be denoted by $\mathbf{x}$. This vector comprises the unknown transformation $\{{}^C_B\mathbf{R}, {}^C\mathbf{p}_B\}$, the vectors $\mathbf{v}_{\ell(j)}$, $\ell = 1,\ldots,N_v$, $j = 1,\ldots,3^\ell$, that describe the mirror configurations, and the 3D base-frame coordinates of the reconstruction points, ${}^B\mathbf{p}_i$, $i = 1,\ldots,N_r$. In our implementation, we adopt the unit-quaternion representation of rotation [39], and thus $\mathbf{x}$ is

$$ \mathbf{x} = \begin{bmatrix} {}^C\mathbf{p}_B^T & {}^C\bar{q}_B^T & {}^C\mathbf{v}_{1(1)}^T & \ldots & {}^C\mathbf{v}_{N_v(3^{N_v})}^T & {}^B\mathbf{p}_1^T & \ldots & {}^B\mathbf{p}_{N_r}^T \end{bmatrix}^T, \tag{35} $$

where ${}^C\bar{q}_B$ is the unit quaternion representation of the rotation between frames $\{B\}$ and $\{C\}$. We use $\mathcal{Z}$ to denote the set of all available measurements, and the likelihood of the measurements is given by

$$ \begin{aligned} L(\mathcal{Z};\mathbf{x}) &= \prod_{i=1}^{N_f+N_r}\prod_{j=1}^{N_c} p(\mathbf{z}_{ij};\mathbf{x}) \\ &= \prod_{i=1}^{N_f+N_r}\prod_{j=1}^{N_c} \frac{1}{2\pi\sigma_\eta^2}\exp\left[-\frac{\left(\mathbf{z}_{ij}-\mathbf{h}\left({}^C\mathbf{p}'_{ij}\right)\right)^T\left(\mathbf{z}_{ij}-\mathbf{h}\left({}^C\mathbf{p}'_{ij}\right)\right)}{2\sigma_\eta^2}\right] \\ &= \prod_{i=1}^{N_f+N_r}\prod_{j=1}^{N_c} \frac{1}{2\pi\sigma_\eta^2}\exp\left[-\frac{(\mathbf{z}_{ij}-\mathbf{h}_{ij}(\mathbf{x}))^T(\mathbf{z}_{ij}-\mathbf{h}_{ij}(\mathbf{x}))}{2\sigma_\eta^2}\right], \end{aligned} \tag{36} $$

where $\mathbf{h}\left({}^C\mathbf{p}'_{ij}\right)$ is the measurement function defined in (1), and the dependence of the measurements on $\mathbf{x}$ is explicitly shown. Maximizing the likelihood is equivalent to maximizing its logarithm, which in turn is equivalent to minimizing the following cost function:

$$ J(\mathbf{x}) = \sum_{i,j}(\mathbf{z}_{ij}-\mathbf{h}_{ij}(\mathbf{x}))^T(\mathbf{z}_{ij}-\mathbf{h}_{ij}(\mathbf{x})). \tag{37} $$

The minimization of this cost function is a nonlinear least-squares problem, and thus we employ the Levenberg-Marquardt (LM) iterative minimization algorithm for estimating the parameter vector $\mathbf{x}$ [10]. During iteration k of the algorithm, the estimate is changed by $\delta\mathbf{x}^{(k)}$, where

$$ \left(\sum_{i,j}\mathbf{H}_{ij}^{(k)T}\mathbf{H}_{ij}^{(k)} + \lambda\mathbf{I}_{6+3N_c+3N_r}\right)\delta\mathbf{x}^{(k)} = \left(\sum_{i,j}\mathbf{H}_{ij}^{(k)T}\left(\mathbf{z}_{ij}-\mathbf{h}_{ij}(\mathbf{x}^{(k)})\right)\right), \tag{38} $$

where λ is the LM damping coefficient, and $\mathbf{H}_{ij}^{(k)}$ is the Jacobian of the measurement function $\mathbf{h}_{ij}(\mathbf{x})$ with respect to $\mathbf{x}$, evaluated at the current iterate, $\mathbf{x}^{(k)}$. The structure of this matrix in the N_v-mirror case is shown in Appendix A. It is important to note that because the structure of $\mathbf{H}_{ij}^{(k)}$ is sparse, the above system is also sparse. Thus, $\delta\mathbf{x}^{(k)}$ can be evaluated very efficiently.

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

$$ \delta\mathbf{x}^{(k)} = \begin{bmatrix} \delta{}^C\mathbf{p}_B^{(k)T} & \delta\boldsymbol{\theta}^{(k)T} & \delta{}^C\mathbf{v}_{1(1)}^{(k)T} & \ldots & \delta{}^C\mathbf{v}_{N_v(3^{N_v})}^{(k)T}, \\ \delta{}^B\mathbf{p}_1^{(k)T} & \ldots & \delta{}^B\mathbf{p}_{N_r}^{(k)T} \end{bmatrix}^T, \tag{39} $$

where all vectors on the right-hand side are 3×1 error vectors. With this notation, the updates for the iterates of the parameters ${}^{C}\mathbf{p}_{B}$, ${}^{C}\bar{q}_{B}$, ${}^{C}\mathbf{v}_{\ell(j)}$, and ${}^{B}\mathbf{p}_{i}$ are written as:

$$ {}^{C}\mathbf{p}_{B}^{(k+1)} = {}^{C}\mathbf{p}_{B}^{(k)} + \delta {}^{C}\mathbf{p}_{B}^{(k)} \tag{40} $$

$$ {}^{C}\bar{q}_{B}^{(k+1)} = \delta\bar{q}^{(k)} \otimes {}^{C}\bar{q}_{B}^{(k)}, \quad \delta\bar{q}^{(k)} = \begin{bmatrix} \frac{1}{2}\delta\boldsymbol{\theta}^{(k)} \\ \sqrt{1 - \frac{1}{4}\delta\boldsymbol{\theta}^{(k)T}\delta\boldsymbol{\theta}^{(k)}} \end{bmatrix} \tag{41} $$

$$ {}^{C}\mathbf{v}_{\ell(j)}^{(k+1)} = {}^{C}\mathbf{v}_{\ell(j)}^{(k)} + \delta {}^{C}\mathbf{v}_{\ell(j)}^{(k)}, \qquad \ell = 1, \ldots, N_v, \tag{42} $$

$$ {}^{B}\mathbf{p}_{i}^{(k+1)} = {}^{B}\mathbf{p}_{i}^{(k)} + \delta {}^{B}\mathbf{p}_{i}^{(k)}, \qquad i = 1, \ldots, N_r, \tag{43} $$

where $\otimes$ denotes quaternion multiplication (for more details on the quaternion convention, please refer to [39]).

Once the LM iterates converge (determined by a threshold on the norm of $\delta\mathbf{x}^{(k)}$), the covariance of the resulting parameter estimates is computed with the expression

$$ \mathbf{P} = \sigma_{\eta}^{2} \left(\sum_{i,j} \mathbf{H}_{ij}^{(k)T} \mathbf{H}_{ij}^{(k)} \right)^{-1}. \tag{44} $$

VII. Simulations

We hereafter present simulation results for evaluating the performance of the proposed single- and multi-mirror approaches when computing the camera-to-base transformation and the 3D base-frame coordinates of the reconstruction points.

A. Single mirror: analytic solution and MLE accuracy

We first examine the accuracy of the analytically computed camera-to-base transformation (see Section IV) and reconstruction points (see Section V), as well as the uncertainty in the MLE estimates (see Section VI) in the single-mirror case. In particular, we investigate how the estimation performance is affected by the following parameters: (i) pixel noise, (ii) camera-to-mirror distance, and (iii) range of the mirror's angular motion. We consider a *base* case, in which three fiducial points and a single reconstruction point, placed at the corners of a square with sides measuring 20 cm, are observed in three images. The points are seen via their reflections in a planar mirror, which is placed at a distance of 0.3 m from the camera and rotated by 25° in two directions. We vary each of the aforementioned parameters individually to examine their effects on the solution accuracy.

In Fig. 5, we plot the position and attitude accuracy of the analytic camera-to-base transformation and the analytically computed 3D base-frame coordinates of the reconstruction point, along with the accuracy of the MLE estimates. Specifically, to evaluate the analytic solution's accuracy we depict the RMS error for the least accurate axis, averaged over 100 Monte Carlo runs. The MLE accuracy is shown by the standard deviation (1σ) for the least certain axis, computed from the covariance estimate [see (44)]. Some key observations based on the results shown in Fig. 5 are the following:

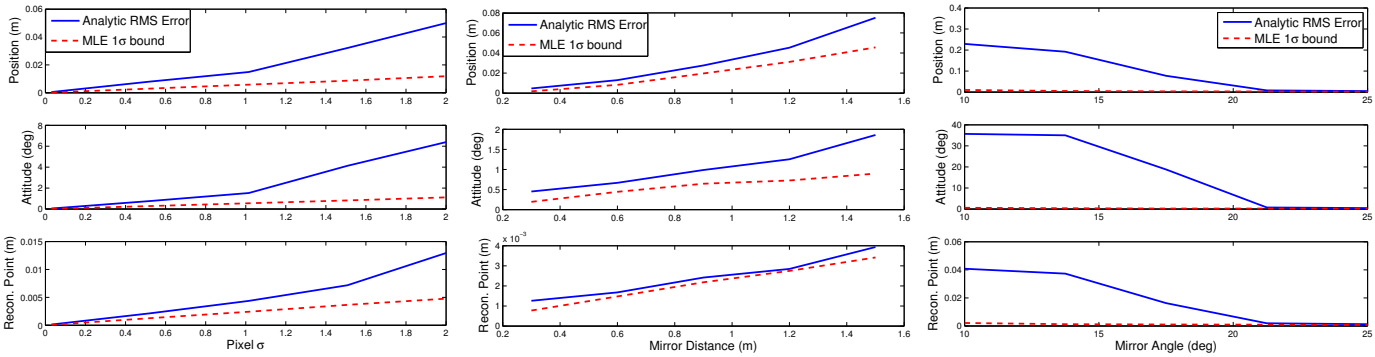
- Increasing the camera's pixel noise decreases the accuracy of the computed solution and the MLE estimates [see Fig. 5(a)]. When the camera measurements become substantially noisy, e.g., $\sigma = 2$ pixels, the average RMS error of the analytic solution is 5 cm in position and 6.4° in attitude. After refinement the 1σ uncertainty bound from the MLE is 1.2 cm in position and 1.1° is attitude. At this noise level, the analytic error and MLE 1σ bound for the reconstruction point are 1.3 cm and 4.7 mm, respectively. Using such trade-off curves, a suitable sensor can be selected based on the calibration accuracy required for a specific task.
- Changing the distance from the mirror to the camera has a significant effect on the accuracy of the estimate of the camera position with respect to the base frame [see Fig. 5(b)]. In the extreme case, when the mirror is at a distance of 1.5 m, the average RMS error for the analytic solution for position is approximately 7.5 cm and drops to 4.5 cm after application of the MLE. The magnitude of this error suggests that the mirror distance should be kept small. Additionally, it highlights the need to refine our analytically-computed transformation with an MLE.
- Increasing the range of the mirror's angular motion results in improved accuracy both for the analytic solution and the MLE estimates [see Fig. 5(c)]. The effect is significant and every effort should be made to move the mirror in the widest range of directions allowed by the camera's field of view.

B. Two mirrors: analytic solution and MLE accuracy

We next examine the two-mirror case. As in the previous section, we evaluate the accuracy of the analytically-computed camera-to-base transformation (see Section IV), and of the reconstruction point's 3D base-frame coordinates (see Section V), as well as the uncertainty in the MLE estimates (see Section VI). The base case we employ is identical to the single mirror case, with the addition of a second mirror behind the camera, at a distance equal that of the camera from the first mirror. The three fiducial points and the single reconstruction point are viewed via their double reflections, in mirrors 1 and 2 (see Fig. 3).

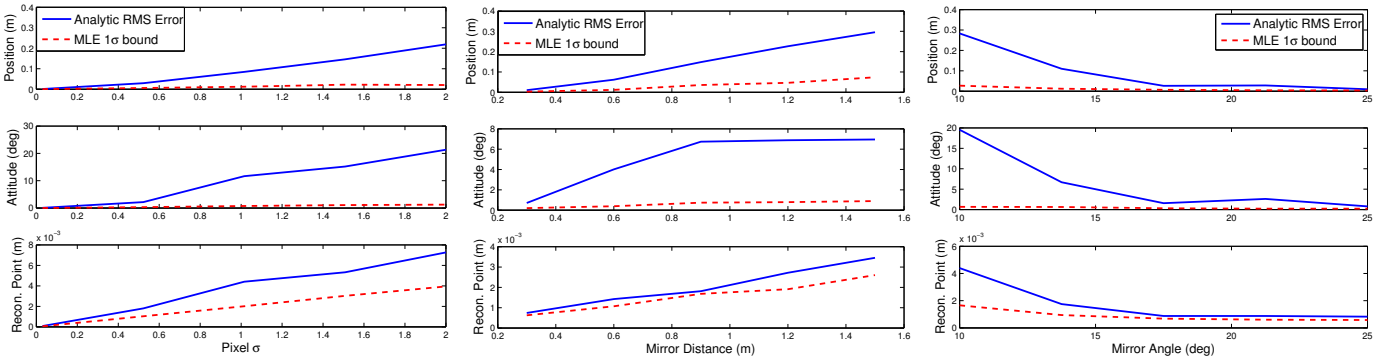
In Fig. 6, we plot the position and attitude accuracy of the analytic solution for the camera-to-base transformation and the accuracy of the reconstruction point, along with the accuracy of the MLE estimates. The performance trends follow the single-mirror scenario, in particular:

- As the pixel noise increases, the accuracy of the analytic solution, as well as of the MLE decreases [see Fig. 6(a)]. As expected, the use of two mirrors introduces additional unknowns compared to the single mirror case, and this degrades the accuracy of the camera-to-base transformation estimation. However, it is interesting to observe that this is *not* the case for the reconstruction point. This is due to the fact that, as discussed in Section V, computing the 3D point's coordinates does not require knowledge of the camera-to-base transformation or the mirror vectors.
- As the distance between the camera and the mirrors increases, the accuracy also decreases [see Fig. 6(b)]. We note that the effect is more pronounced than in the



(a) (b) (c)

Fig. 5. Single mirror: analytic solution and MLE accuracy for the camera-to-base transformation and reconstruction point plotted versus: (a) pixel noise, (b) mirror distance, and (c) range of mirror rotation. The analytic solution's accuracy is computed as the average RMS error over 100 trials, while the MLE accuracy is plotted as the standard deviation computed from the estimated covariance.



(a) (b) (c)

Fig. 6. Two mirrors: analytic solution and MLE accuracy for the camera-to-base transformation and reconstruction point plotted versus: (a) pixel noise, (b) mirror distance, and (c) range of mirror rotation. The analytic solution's accuracy is computed as the average RMS error over 100 trials, while the MLE accuracy is plotted as the standard deviation computed from the estimated covariance.

single-mirror case since in the two-mirror simulation *both* mirrors are moving farther away from the camera, and the effective depth to the points is twice as large.

- Increasing the range of the mirrors' angular motions improves the overall system performance [see Fig. 6(c)]. In the two-mirror case, the effective baseline created by moving the mirrors is larger than in the single-mirror scenario, and the performance improves more rapidly with the increase in the mirror-rotation range.

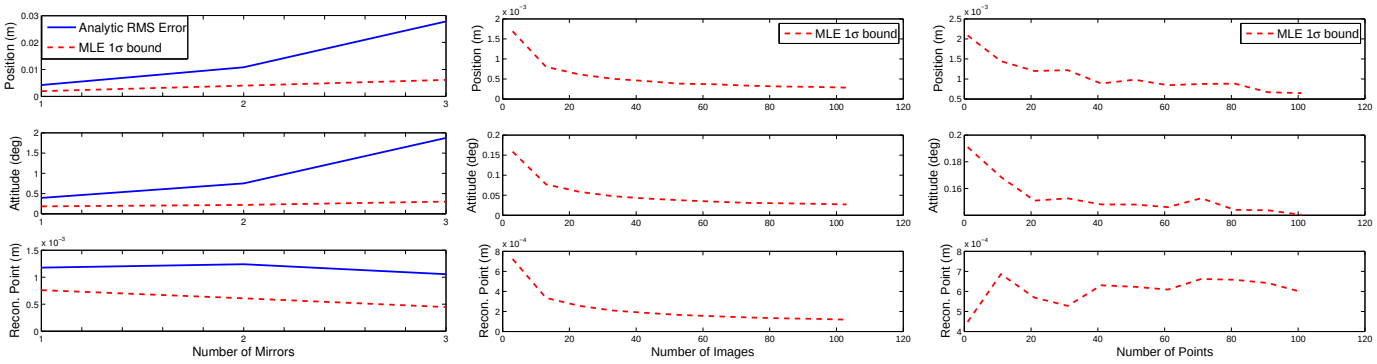
Note that, for both the single- and multi-mirror cases, using the analytic solution as an initial guess for the MLE enables the latter to converge to the correct minimum 100% of the time for non-singular measurement configurations. On average, fewer iterations were required (approx. 4) when compared to using a naïve initial guess (approx. 18). This shows that a precise analytic solution improves the speed and robustness of the overall estimation process.

C. Analysis of Additional Parameters

There are several additional parameters which affect the solution accuracy: (i) number of mirrors in the system, (ii) number of images recorded, and (iii) number of reconstruction points (see Fig. 7).

- Using the same base case as in Sections VII-A and VII-B, we examine the effect of adding a third mirror in the system. Figure 7(a) depicts the results which demonstrate that as the number of mirrors increases, the solution accuracy decreases. This is primarily due to the dramatic increase in effective depth to the robot body with each additional mirror (it typically doubles with each mirror). Moreover, we note that for a large number of mirrors (four or more), it becomes difficult to even generate realistic scenarios in which the reflected points fall within the camera's field of view. Hence, while these instances provide validation of the method, and additional intuition for the problem, they may be of little practical importance.
- Increasing the number of images[3] results in higher accuracy for the MLE, as shown in Fig. 7(b). However, the improvement follows a "law of diminishing returns," i.e., when a large number of images is already available, the impact of recording more measurements is small.
- The final parameter which we study is the number of reconstruction points. Figure 7(c) depicts the MLE 1σ

[3]Since the analytic solution does not benefit from additional images (i.e., we solve the base case, and we do not exploit additional measurements at the analytic level), we have not plotted the analytic RMS values on this graph.



(a) (b) (c)

Fig. 7. Effect of additional system parameters on the accuracy of the camera-to-base transformation and the reconstruction point: (a) number of mirrors, (b) number of images, and (c) number of reconstruction points.

bound for the camera-to-base transformation and the least accurately estimated reconstruction point when using 3 to 100 reconstruction points[4]. It is quite interesting to note that as more reconstruction points become available, the camera-to-base transformation becomes increasingly more accurate. This is due to the fact that the observations of unknown points provide us with information about the motion of the camera between images (analogous to the epipolar geometry).

VIII. Experiments

The proposed methods were also evaluated in real-world experiments to assess their performance and effectiveness in practice. We first present results for performing single-mirror camera-to-base calibration and 3D robot-body reconstruction (see Section VIII-A). Subsequently, we show the results for extrinsic calibration with two mirrors (see Section VIII-B).

A. Single mirror: camera-to-base transformation and 3D reconstruction

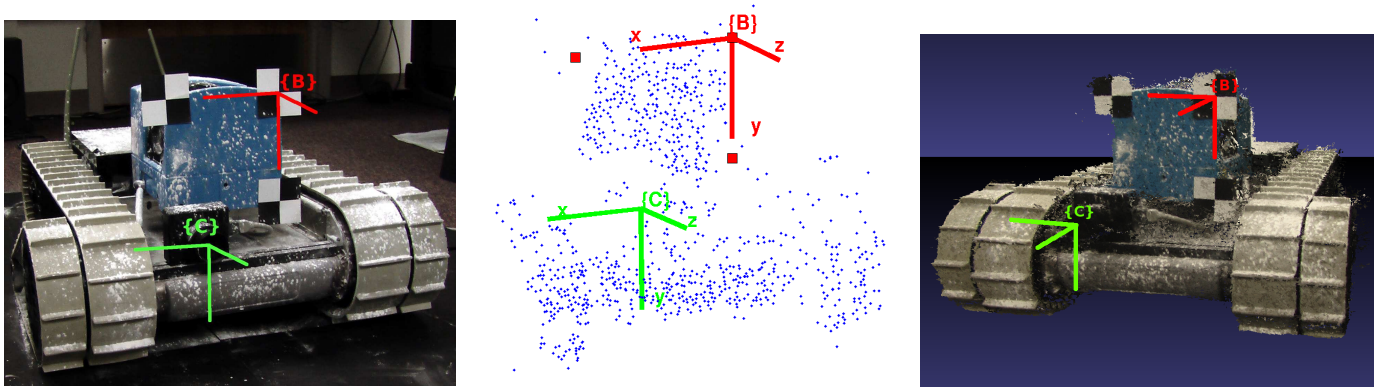
The methods described in Sections IV, V, and VI were employed for computing the transformation between a camera and the body frame of the robot on which it is attached, and for determining the point-cloud representation of the robot-body (i.e., the 3D coordinates of the reconstruction points). Specifically, the camera-to-base transformation is computed from observations of three fiducial points, placed in known positions on the robot as shown in Fig. 8(a). The origin of $\{B\}$ coincides with the top-left fiducial point; both $\{B\}$ and $\{C\}$ are right-handed systems with the axes of $\{B\}$ approximately aligned with those of $\{C\}$. These points were tracked using the Kanade-Lucas-Tomasi (KLT) feature tracker [43] in 1000 images, recorded by a Firewire camera with resolution of 1024×768 pixels.

A planar mirror was maneuvered in different spatial configurations (rotating about two axes), and at distances varying between 30 and 50 cm from the camera, in order to generate a wide range of views. All the measurements were processed to compute the transformation analytically: ${}^{C}\mathbf{p}_{B} = \begin{bmatrix}-14.13 & -10.25 & -13.89\end{bmatrix}^{T}$ cm, and ${}^{C}\bar{q}_{B} = \begin{bmatrix}-0.0401 & -0.0017 & -0.0145 & 0.9991\end{bmatrix}^{T}$. This initial solution was refined using the MLE described in Section VI, to obtain a better estimate for the transformation between the two frames of interest. The Levenberg-Marquardt iterative minimization with pseudo-Huber robust cost function [10] converged after 10 iterations, to the following solution for the transformation: ${}^{C}\mathbf{p}_{B} = \begin{bmatrix}-11.88 & -13.81 & -8.72\end{bmatrix}^{T}$ cm, and ${}^{C}\bar{q}_{B} = \begin{bmatrix}0.0063 & 0.0298 & 0.0009 & 0.9995\end{bmatrix}^{T}$. The corresponding 3σ uncertainty bounds are $\begin{bmatrix}1.516 & 1.762 & 8.862\end{bmatrix}$ mm for the position, and $\begin{bmatrix}0.1176 & 0.2782 & 0.0483\end{bmatrix}$ deg for the orientation estimates. We point out that the estimates agree with our best guess from manual measurement. We believe that the attained accuracy (given by the 3σ bounds from the MLE) is sufficiently high for most practical applications. However, if higher accuracy is required this can be achieved by increasing the number of images and points, as well as decreasing the distance between the camera and mirror (see Fig. 5).

In addition to computing the camera-to-base transformation, we computed a 3D point cloud representation of the front of the robot chassis (i.e., the 3D coordinates of the reconstruction points). To this end, approximately 1000 SIFT features were matched between two views and utilized as reconstruction points [44]. After analytically computing the 3D coordinates of each point, we employed the MLE (Section VI) to estimate their base-frame coordinates. Figures 8(a) and 8(b) depict the robot from an angled view, along with the 3D point cloud representation of the robot chassis.

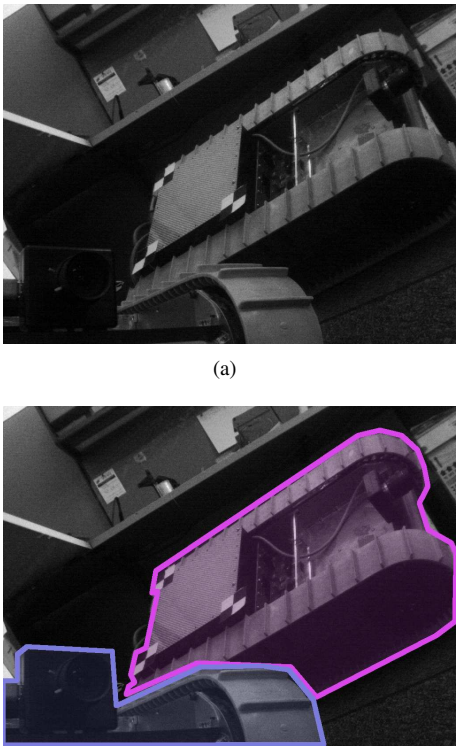
In the single-mirror approach, the reconstruction points only cover the front of the robot body due to limitations in the camera's field of view. In order to capture the whole chassis, we augmented our point-wise reconstruction with a dense (per pixel) reconstruction method using a hand-held camera [42] [see Fig. 8(c)]. The hand-held camera does not have limitations in placement, hence the entire robot chassis can be reconstructed with respect to it, up to scale. Using the locations of the fiducial points in the previous (calibration-based) 3D reconstruction expressed in the base frame, we obtain the scale, and location of each point on the chassis with respect to the robot's camera. Precise knowledge of the

[4]Since the number of reconstruction points does not affect the accuracy of the analytical solution, we omit the analytic RMS values from the plot.



(a) (b) (c)

Fig. 8. Single-mirror experiment: (a) A view of the robot from the side. (b) The estimated 3D point cloud of reconstruction points seen from a similar angle. (c) A dense reconstruction obtained using Arc3D [42]. A layer of talc powder was applied in order to increase the texture on the robot body. Using talc is reasonable since it does not change the geometry of the chassis and it can be easily wiped off after the experiment.



(a)

(b)

Fig. 9. Two-mirror experiment: (a) Image recorded during experimentation. Two reflections of the robot are visible which provide different viewpoints of the chassis. (b) The same image with the single mirror and two-mirror reflections highlighted in blue and purple, respectively (see Fig. 1). The fiducial points are only visible through the two-mirror reflection.

robot's chassis in the camera frame can be used to increase the robustness and efficiency of obstacle-avoidance algorithms when navigating through tight spaces.

B. Two mirrors: camera-to-base transformation

We also evaluated the proposed method in a two-mirror scenario with a camera-equipped mobile robot to compute the transformation between the camera frame of reference and the robot-body frame (see Fig. 1). Frame $\{B\}$ is right-handed with its x-axis pointing towards the front of the robot and its z-axis pointing upwards. The rear-right fiducial marker coincides with the origin of $\{B\}$, while the other two markers lie on its x- and y-axis, respectively. Due to the relative placement of the camera and the fiducials, they cannot be observed directly by the camera nor can they be seen in the reflection in the front mirror [see Figs. 9(a) and 9(b)]. Instead, the markers were only visible via their double reflections, first in the rear mirror, then in the front mirror.

As in the single-mirror experiment, the camera was connected to the robot's on-board computer via Firewire, and recorded 900 gray-scale images at 10 Hz with resolution 1024×768 pixels. During the experiment the markers were tracked using KLT [43]. The front mirror was moved continuously through the image sequence, while the rear mirror was moved in three configurations (300 images per configuration). We computed the analytic solution for both mirror vectors in every image, as well as the transformation between the camera-to-base translation, ${}^{C}\mathbf{p}_{B} = \begin{bmatrix}11.26 & -4.48 & -52.48\end{bmatrix}^{T}$ cm, and orientation, ${}^{C}\mathbf{q}_{B} = \begin{bmatrix}-0.5005 & 0.5063 & -0.4931 & 0.4998\end{bmatrix}^{T}$.

We initialized the MLE with the analytically computed quantities, and the Levenberg-Marquardt minimization converged after three iterations. The final estimates for translation and orientation were ${}^{C}\mathbf{p}_{B} = \begin{bmatrix}10.54 & -4.42 & -53.01\end{bmatrix}^{T}$ cm, and ${}^{C}\mathbf{q}_{B} = \begin{bmatrix}-0.5026 & 0.5040 & -0.4931 & 0.5000\end{bmatrix}^{T}$, respectively. The corresponding 3σ bounds computed from the diagonal components of the MLE estimated covariance were $\begin{bmatrix}9.75 & 6.92 & 6.44\end{bmatrix}$ mm in position and $\begin{bmatrix}0.445 & 0.684 & 0.356\end{bmatrix}$ deg in orientation. The obtained accuracy is suitable for use in many real-world applications, however, we note that the accuracy decreases as the number of mirrors increases. This also agrees with the observations from our simulation trials (see Figs. 5 and 6).

IX. Conclusions and Future Work

In this paper, we addressed the problem of concurrently estimating the camera-to-base frame transformation and per-

forming 3D robot-body reconstruction, using observations of points that lie outside the camera's natural field of view. To extend the camera's visual coverage, we employ a sequence of mirrors to observe the points via reflections. We require that at least three of the features (fiducial points) be known in the base frame in advance, while all the remaining features' positions are estimated (reconstruction points). We do not exploit any prior knowledge about the mirrors' motions or placements with respect to the camera, but instead, we treat the mirror poses as unknown parameters to be determined along with the camera-to-base transformation and the reconstruction points' coordinates. Using the geometry of the camera-mirror observations, we formulate this problem as a system of non-linear equations which we solve analytically, using PnP as an intermediate step. Subsequently, we present an MLE which allows us to compute high-accuracy estimates of all system parameters, by properly accounting for the measurement noise statistics.

We plan to extend this method to the case where no fiducial points are available (i.e., none of the points' coordinates are known *a priori*), but all points' positions are estimated with respect to the camera frame, along with the mirror configurations. In this case, the scale is unobservable, and thus the distance between at least two of the points should be measured manually.

APPENDIX A

The Jacobian of the measurement function $\mathbf{z}_{ij}$ (point i measured in image j) with respect to the parameter vector $\mathbf{x}$ is given by:

$$\mathbf{H}_{ij} = \mathbf{H}_{c_{ij}} \begin{bmatrix} \mathbf{H}_{p_{ij}} & \mathbf{H}_{q_{ij}} & \dots & \mathbf{H}_{v_{ij}} & \dots & \mathbf{H}_{{}^B p_i} & \dots \end{bmatrix}, \tag{45}$$

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

$$\mathbf{H}_{c_{ij}} = \frac{1}{p_3} \begin{bmatrix} 1 & 0 & -\frac{p_1}{p_3} \\ 0 & 1 & -\frac{p_2}{p_3} \end{bmatrix} \tag{46}$$

and $\mathbf{H}_{p_{ij}}$, $\mathbf{H}_{q_{ij}}$, $\mathbf{H}_{v_{ij}}$, and $\mathbf{H}_{{}^B p_i}$, are the Jacobians of ${}^C\mathbf{p}'_{ij}$ with respect to the camera position, camera rotation, the mirror vector $\mathbf{v}_{ij}$, and the feature position, respectively:

$$\mathbf{H}_{p_{ij}} = \mathbf{A}_{N_v(j)} {}^C_B\mathbf{R}^T \tag{47}$$

$$\mathbf{H}_{q_{ij}} = \mathbf{A}_{N_v(j)} \lfloor {}^B\mathbf{p}_i \times \rfloor {}^C_B\mathbf{R}^T \tag{48}$$

$$\mathbf{H}_{v_{ij}} = 2\mathbf{A}_{N_v(j)} \mathbf{A}^{-1}_{\ell(j)} \left(\left(1 - \frac{\mathbf{v}^T_{\ell(j)} \mathbf{g}_{\ell-1(j)}}{\mathbf{v}^T_{\ell(j)} \mathbf{v}_{\ell(j)}} \right) \mathbf{I}_3 - \frac{\mathbf{v}_{\ell(j)} \mathbf{g}^T_{\ell-1(j)}}{\mathbf{v}^T_{\ell(j)} \mathbf{v}_{\ell(j)}} + 2\mathbf{v}_{\ell(j)} \mathbf{v}^T_{\ell(j)} \frac{\mathbf{v}^T_{\ell(j)} \mathbf{g}_{\ell-1(j)}}{\left(\mathbf{v}^T_{\ell(j)} \mathbf{v}_{\ell(j)} \right)^2} \right) \tag{49}$$

$$\mathbf{H}_{{}^B p_i} = \mathbf{A}_{N_v(j)} \tag{50}$$

where $\mathbf{g}_{\ell-1(j)} = \mathbf{A}_{\ell-1(j)} {}^B\mathbf{p}_i + \mathbf{b}_{\ell-1(j)}$. If a fiducial (known) point is measured, then $\mathbf{H}_{{}^B p_i}$ does not appear in the corresponding measurement Jacobian, since the coordinates of the point are not in the parameter vector.

REFERENCES

[1] J. A. Hesch, A. I. Mourikis, and S. I. Roumeliotis, "Determining the camera to robot-body transformation from planar mirror reflections," in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, Nice, France, Sep. 22–26, 2008, pp. 3865–3871.

[2] ——, "Mirror-based extrinsic camera calibration," in *Proc. of the Int. Workshop on the Algorithmic Foundations of Robotics*, Guanajuato, Mexico, Dec. 7–9, 2008, pp. 285–299.

[3] ——, "Extrinsic camera calibration using multiple reflections," in *Proc. of the European Conference on Computer Vision*, Crete, Greece, Sep. 5–11, 2010, pp. 311–325.

[4] J. Heikkila and O. Silven, "A four-step camera calibration procedure with implicit image correction," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, Jun. 17–19, 1997, pp. 1106–1112.

[5] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.

[6] W. Bourgeous, L. Ma, P. Chen, Z. Song, and Y. Chen, "Simple and efficient extrinsic camera calibration based on a rational model," in *Proc. of the IEEE Int. Conf. on Mechatronics and Automation*, Luoyang, China, Jun. 25–28, 2006, pp. 177–182.

[7] F. M. Mirzaei and S. I. Roumeliotis, "A Kalman filter-based algorithm for IMU-camera calibration: Observability analysis and performance evaluation," *IEEE Trans. on Robotics*, vol. 24, no. 5, pp. 1143–1156, Oct. 2008.

[8] X. Brun and F. Goulette, "Modeling and calibration of coupled fish-eye CCD camera and laser range scanner for outdoor environment reconstruction," in *Proc. of the Int. Conf. on 3D Digital Imaging and Modeling*, Montréal, Canada, Aug. 21–23, 2007, pp. 320–327.

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

[10] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521623049, 2000.

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

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

[13] B. K. Ramsgaard, I. Balslev, and J. Arnspang, "Mirror-based trinocular systems in robot-vision," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Barcelona, Spain, Sep. 3–7, 2000, pp. 499–502.

[14] J. A. Grunert, "Das pothenotische problem in erweiterter gestalt nebst bber seine anwendugen in der geodäsie," *Grunerts Archiv für Mathematik und Physik*, pp. 238–248, 1841, Band 1.

[15] E. L. Merritt, "Explicitly three-point resection in space," *Photogrammetric Engineering*, vol. XV, no. 4, pp. 649–655, 1949.

[16] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applicatlons to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.

[17] R. M. Haralick, C.-N. Lee, K. Ottenberg, and M. Nölle, "Review and analysis of solutions of the three point perspective pose estimation problem," *Int. Journal of Computer Vision*, vol. 13, no. 3, pp. 331–356, Dec. 1994.

[18] F. J. Müller, "Direkte (exakte) losung des einfachen rückawärtseinschneidens im raume," *Allgemeine Vermessungs-Nachrichten*, 1925.

[19] R. Horaud, B. Conio, O. Leboulleux, and B. Lacolle, "An analytic solution for the perspective 4-point problem," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, San Diego, CA, Jun. 4–8, 1989, pp. 500–507.

[20] Y. Wu and Z. Hu, "PnP problem revisited," *Journal of Mathematical Imaging and Vision*, vol. 24, no. 1, pp. 131–141, Jan. 2006.

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

[22] A. Martinelli, "Using the distribution theory to simultaneously calibrate the sensors of a mobile robot," in *Proc. of Robotics: Science and Systems*, Seattle, WA, Jun. 25–29, 2009.

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

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

[25] N. Andreff, R. Horaud, and B. Espiau, "Robot hand-eye calibration using structure-from-motion," *Int. Journal of Robotics Research*, vol. 20, no. 3, pp. 228–248, Mar. 2001.

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

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

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

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

[30] G. Jang, S. Kim, and I. Kweon, "Single camera catadioptic stereo system," in *Proc. of the Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras*, Beijing, China, Oct. 21, 2005.

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

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

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

[34] 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, Jul. 2004.

[35] P. Sturm and T. Bonfort, "How to compute the pose of an object without a direct view?" in *Proc. of the Asian Conf. on Computer Vision*, Hyderabad, India, Jan. 13–16, 2006, pp. 21–31.

[36] R. K. Kumar, A. Ilie, J.-M. Frahm, and M. Pollefeys, "Simple calibration of non-overlapping cameras with a mirror," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Anchorage, AK, Jun. 24–26, 2008.

[37] J.-Y. Bouguet, "Camera calibration toolbox for matlab," 2006. [Online]. Available: http://www.vision.caltech.edu/bouguetj/calibdoc/

[38] 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., MARS Lab, Tech. Rep. 2008-001, Jul. 2008. [Online]. Available: http://www-users.cs.umn.edu/∼joel/

[39] N. Trawny and S. I. Roumeliotis, "Indirect Kalman filter for 3D attitude estimation," University of Minnesota, Dept. of Comp. Sci. & Eng., MARS Lab, Tech. Rep. 2005-002, Mar. 2005. [Online]. Available: http://www-users.cs.umn.edu/ trawny/

[40] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon, "Bundle adjustment – a modern synthesis," in *Vision Algorithms: Theory and Practice*, vol. 1883. Springer-Verlag, 2000, pp. 298–372.

[41] D. Nistér, "An efficient solution to the five-point relative pose problem," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Madison, WI, Jun. 16–22, 2003, pp. 195–202.

[42] M. Vergauwen and L. V. Gool, "Web-based 3D reconstruction service," *Machine Vision and Applications*, vol. 17, no. 6, pp. 411–426, Oct. 2006.

[43] J. Shi and C. Tomasi, "Good features to track," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Washington, DC, Jun. 27–Jul. 2, 1994, pp. 593–600.

[44] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.