Extrinsic Camera Calibration Using Multiple Reflections

Joel A. Hesch[1], Anastasios I. Mourikis[2], and Stergios I. Roumeliotis[1,⋆]

[1] University of Minnesota, Minneapolis MN 55455, USA
{joel,stergios}@cs.umn.edu
[2] University of California, Riverside CA 92521, USA
mourikis@ee.ucr.edu

Abstract. This paper presents a method for determining the six-degree-of-freedom (DOF) transformation between a camera and a base frame of interest, while concurrently estimating the 3D base-frame coordinates of unknown point features in the scene. The camera observes the reflections of fiducial points, whose base-frame coordinates are known, and reconstruction points, whose base-frame coordinates are unknown. In this paper, we examine the case in which, due to visibility constraints, none of the points are directly viewed by the camera, but instead are seen via reflection in multiple planar mirrors. Exploiting these measurements, we *analytically* compute the camera-to-base transformation and the 3D base-frame coordinates of the unknown reconstruction points, without *a priori* knowledge of the mirror sizes, motions, or placements with respect to the camera. Subsequently, we refine the analytical solution using a maximum-likelihood estimator (MLE), to obtain high-accuracy estimates of the camera-to-base transformation, the mirror configurations for each image, and the 3D coordinates of the reconstruction points in the base frame. We validate the accuracy and correctness of our method with simulations and real-world experiments.

1 Introduction

Extrinsic calibration – the task of computing the six-degrees-of-freedom (DOF) transformation between the camera's frame of reference and a base frame of interest – is a prerequisite for many vision-based tasks. For example, mobile robots often rely on cameras to detect and locate obstacles during their operation. When this is the case, accurate knowledge of the camera-to-body transformation is necessary for precise navigation. Estimating this transformation is often not a trivial matter: one common problem is that the robot's chassis may not lie within the camera's direct field of view (see Fig. 1), which means that one cannot apply calibration methods that rely on direct observations of known points on the robot body. This is only one example application where the camera

⋆ 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).

K. Daniilidis, P. Maragos, N. Paragios (Eds.): ECCV 2010, Part IV, LNCS 6314, pp. 311–325, 2010.

extrinsic parameters must be computed without a direct line-of-sight to any of the available fiducial points. In this work, we show that in these cases one can exploit observations of the fiducial points through reflections in multiple mirrors, to extrinsically calibrate the camera.

The objective of our work is to design an automated procedure for determining the 3D transformation between the camera frame and a *base frame* of interest, by utilizing the measurements of fiducial points, whose position in the base frame is known *a priori*. We examine the scenario in which the known points are not directly visible to the camera, but can only be observed through reflections in multiple mirrors. We maneuver the mirrors to provide the camera with multiple views of the fiducial points; 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 both treated as unknowns to be computed from the measurements. In addition to these quantities, in this paper we show how the images recorded by the camera through the mirror reflections can be used to estimate the positions of additional points in the scene, whose locations were not known *a priori*.

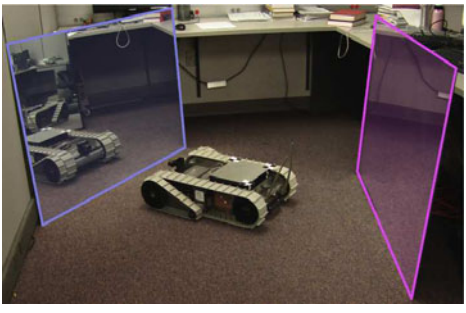

Fig. 1. A mobile robot views its reflection in two mirrors (front highlighted in blue, back highlighted in purple). The robot visually tracks point features to estimate the camera-to-base frame transformation, and a 3D point-cloud representation of its chassis.

Thus, the problem we address is that of jointly estimating the camera's configuration, mirrors' configurations, and scene structure, using observations of points visible only though a number of reflections. The main contribution of this work is an algorithm for *analytically* computing all the unknown quantities, given the available measurements. The analytically computed estimates are subsequently refined by a maximum likelihood estimator (MLE), implemented by an iterative nonlinear minimization process, to obtain the estimates with the highest precision possible while accounting for measurement noise. In addition to the theoretical importance of an analytical solution, its practical utility is demonstrated by both our simulation results and our real-world experiments. These tests demonstrate that using the analytical solution to seed the MLE results in accurate estimates, which can be computed in a small number of iterations.

2 Related Work

Extrinsic camera calibration has been widely studied for the case in which known points are *directly observed* by the camera [1,2,3]. Unfortunately, in many realistic scenarios, the known points may not lie within the camera's field of view

(see Fig. 1). This motivates studying the more limiting scenarios, in which the points of interest can only be observed through reflection using one or more planar mirrors. The literature in this field is substantially sparser. We note that catadioptric systems in which one or more mirrors are employed to reconstruct a scene, such as those presented in [4,5,6,7] are not directly relevant here. First, in these methods the location of the mirrors is assumed to be known in advance. Second, in these systems each point is observed multiple times in each image (directly, as well as through reflections). In our method each point is only observed *once* per image, via its reflection in the moving planar mirrors.

A system which employs a moving planar mirror for 3D scene reconstruction was introduced by Jang et al. [8]. By exploiting a combination of known markers on a moving mirror and vanishing points in the reflections, they first solved for the position of the mirror with respect to the camera, and subsequently determined the 3D scene based on synthetic stereo from multiple reflections. In contrast to this approach, we do not utilize known mirror markers, since doing so would introduce constraints on the mirror motions (i.e., the markers must always be visible to the camera). This enhances the flexibility of our method, but it renders our problem more challenging, since multiple images are required to compute the mirror configurations.

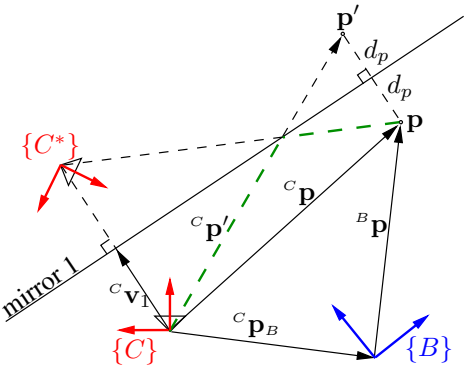

Fig. 2. The camera observes ${}^C\mathbf{p}'$, which is the reflection of $\mathbf{p}$. The base frame is $\{B\}$, while $\{C\}$ is the camera frame, and $\{C^*\}$ is the equivalent *imaginary* camera frame which lies behind the mirror. ${}^C\mathbf{v}_1$ is the shortest vector from $\{C\}$ to the mirror. The distance d_p is measured from $\mathbf{p}$ to the mirror, along $\mathbf{v}_1$. The vector ${}^C\mathbf{p}_B$ is 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 dashed green line is the path of the reflection.

Kumar et al. [9] presented a vision system that utilized a moving planar mirror to determine the transformations between multiple cameras with non-overlapping fields of view. Each camera in turn viewed the reflection of a calibration grid, whose position with respect to the cameras was fixed. To solve the problem, each camera was required to view the calibration pattern from five vantage points. Subsequently, the measurement constraints were transformed into a set of linear equations which were solved for the unknown transformations. In contrast to [9], the method presented here requires observations of only three fiducial points, and is also applicable in cases where reflection in a single mirror is not sufficient to make the points visible to the camera.

Finally, in our previous work we addressed the problem of extrinsic camera calibration using a *single* mirror, presenting both an analytical solution [10], and

an MLE to obtain estimates of the camera-to-base transformation [11]. We now extend this single-mirror extrinsic calibration method and address the multi-mirror case, as described in the following sections.

3 Problem Formulation

Our main goal in this work is to simultaneously determine: (i) the six-DOF transformation between the camera frame, $\{C\}$, and a base frame of interest, $\{B\}$, and (ii) the 3D base-frame coordinates of N_r "reconstruction points." To this end, we assume that N_f fiducial points, whose coordinates in the base frame are known *a priori*, are observed in N_c images. We address the most limiting scenario, in which the points do not lie in the camera's direct field of view, but are only visible via reflection in N_v mirrors (each point is reflected N_v times, and each point is only observed once in each image). Since the placement of the mirrors in each image is unknown, in addition to the camera configuration and the positions of the reconstruction points, it is necessary to jointly estimate the configurations of the mirrors in all the images. In what follows, we start by presenting the model describing the camera measurements.

3.1 Measurement Model

The camera observes each point, $\mathbf{p}$, via its reflection $\mathbf{p}'$, as shown in Fig. 2. The measurement model which describes this observation is divided in two components: (i) the camera projection model and (ii) the expression which describes the geometric relationship between $\mathbf{p}'$ and $\mathbf{p}$ as a function of the mirror and camera configurations.

Single-mirror constraint: In the single-mirror scenario, we obtain two equations from geometry (see Fig. 2):

$$^{C}\mathbf{p}' = {}^{C}\mathbf{p} + 2d_p \frac{^{C}\mathbf{v}_1}{\|^{C}\mathbf{v}_1\|}, \quad d_p = \|^{C}\mathbf{v}_1\| - \frac{^{C}\mathbf{v}_1^T}{\|^{C}\mathbf{v}_1\|} {}^{C}\mathbf{p}, \tag{1}$$

where $^{C}\mathbf{p}'$ is the vector from the origin of $\{C\}$ to the reflected point $\mathbf{p}'$, $^{C}\mathbf{p}$ is vector from $\{C\}$ to $\mathbf{p}$, $^{C}\mathbf{v}_1$ is the mirror vector, which is the shortest vector form the origin of $\{C\}$ to the reflective surface, and d_p is the distance between the mirror and the point $\mathbf{p}$ measured along the direction of $^{C}\mathbf{v}_1$. In order to simplify the notation, we refer to $^{C}\mathbf{v}_1$ as $\mathbf{v}_1$ in the remainder of the paper. In addition to the two geometric constraints derived from Fig. 2, we also exploit the coordinate transformation between $^{C}\mathbf{p}$ and $^{B}\mathbf{p}$, i.e.,

$$^{C}\mathbf{p} = {}^{C}_{B}\mathbf{R}\, {}^{B}\mathbf{p} + {}^{C}\mathbf{p}_{B}, \tag{2}$$

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 $\{C\}$. We substitute (2) into (1), and rearrange the terms to obtain

$$^{C}\mathbf{p}' = \left(\mathbf{I}_3 - 2\frac{\mathbf{v}_1\mathbf{v}_1^T}{\mathbf{v}_1^T\mathbf{v}_1}\right) {}^{C}\mathbf{p} + 2\,\mathbf{v}_1 = \mathbf{M}_1\,({}^{C}_{B}\mathbf{R}\, {}^{B}\mathbf{p} + {}^{C}\mathbf{p}_{B}) + 2\,\mathbf{v}_1, \tag{3}$$

where $\mathbf{M}_1 = \left(\mathbf{I}_3 - 2\left(\mathbf{v}_1\mathbf{v}_1^T/\mathbf{v}_1^T\mathbf{v}_1\right)\right)$ is the Householder transformation matrix corresponding to the mirror reflection. Equation (3) is equivalently expressed in homogeneous coordinates as

$$\begin{bmatrix} {}^C\mathbf{p}' \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{M}_1 & 2\,\mathbf{v}_1 \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \begin{bmatrix} {}^C_B\mathbf{R} & {}^C\mathbf{p}_B \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \begin{bmatrix} {}^B\mathbf{p} \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{b}_1 \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \begin{bmatrix} {}^B\mathbf{p} \\ 1 \end{bmatrix}, \tag{4}$$

where 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 homogeneous/reflection transformation, which converts ${}^B\mathbf{p}$ into ${}^C\mathbf{p}'$.

N_v-mirror constraint: The single-mirror case is readily extended to the N_v-mirror case, by noting that each additional mirror in the system adds a reflection transformation parameterized by the corresponding mirror vector. Hence, the geometric relationship for a base-frame point observed through N_v mirrors is

$$\begin{bmatrix} {}^C\mathbf{p}' \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{M}_{N_v} & 2\,\mathbf{v}_{N_v} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \cdots \begin{bmatrix} \mathbf{M}_1 & 2\,\mathbf{v}_1 \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \begin{bmatrix} {}^C_B\mathbf{R} & {}^C\mathbf{p}_B \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \begin{bmatrix} {}^B\mathbf{p} \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{N_v} & \mathbf{b}_{N_v} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \begin{bmatrix} {}^B\mathbf{p} \\ 1 \end{bmatrix}, \tag{5}$$

where $\{\mathbf{A}_{N_v}, \mathbf{b}_{N_v}\}$ is a homogeneous transformation comprising the N_v mirror vectors and the camera-to-base transformation. Their structure can be defined recursively by expanding (5), i.e.,

$$\begin{aligned} \mathbf{A}_{N_v} &= \mathbf{M}_{N_v} \cdots \mathbf{M}_1{}^C_B\mathbf{R} = \mathbf{M}_{N_v}\mathbf{A}_{N_v-1} && (6) \\ \mathbf{b}_{N_v} &= \mathbf{M}_{N_v} \cdots \mathbf{M}_1{}^C\mathbf{p}_B + 2\,\mathbf{M}_{N_v} \cdots \mathbf{M}_2\mathbf{v}_1 + \cdots \\ &\quad + 2\,\mathbf{M}_{N_v}\mathbf{M}_{N_v-1}\mathbf{v}_{N_v-2} + 2\,\mathbf{M}_{N_v}\mathbf{v}_{N_v-1} + 2\,\mathbf{v}_{N_v} \\ &= \mathbf{M}_{N_v}\mathbf{b}_{N_v-1} + 2\,\mathbf{v}_{N_v}. && (7) \end{aligned}$$

We extend this recursive structure to include the camera-to-base transformation:

$$\mathbf{A}_0 = {}^C_B\mathbf{R}, \quad \mathbf{b}_0 = {}^C\mathbf{p}_B, \tag{8}$$

which will simplify the discussion of our analytical solution (see Sect. 4.2).

Perspective projection model: The reflected point, $\mathbf{p}'$, is observed by the camera whose intrinsic camera parameters are assumed to be known [12]. The normalized image coordinates of the measurement are described by the perspective projection model:

$$\mathbf{z} = \frac{1}{p_3}\begin{bmatrix} p_1 \\ p_2 \end{bmatrix} + \boldsymbol{\eta} = \mathbf{h}({}^C\mathbf{p}') + \boldsymbol{\eta}, \quad {}^C\mathbf{p}' = \begin{bmatrix} p_1 & p_2 & p_3 \end{bmatrix}^T, \tag{9}$$

where $\boldsymbol{\eta}$ is the pixel noise, which is modeled as a zero-mean white Gaussian process with covariance matrix $\sigma_\eta^2\mathbf{I}_2$. Equations (5) and (9) define the measurement model that expresses the point's observed image coordinates $\mathbf{z}$, as a function of the vector ${}^B\mathbf{p}$, the *unknown* camera-to-base transformation $\{{}^C_B\mathbf{R}, {}^C\mathbf{p}_B\}$, and the *unknown* configurations of the mirrors with respect to the camera, $\mathbf{v}_1, \ldots, \mathbf{v}_{N_v}$.

4 Camera-to-Base Transformation Analytical Solution

We address the problem of obtaining an analytical solution for all the unknown quantities in two steps: first, we obtain an analytical solution for the camera-to-base frame transformation, as well as for the mirrors' configurations. Once these quantities have been determined, we subsequently obtain an analytical solution for the 3D positions of the reconstruction points, as explained in Sect. 5.

4.1 Relationship to P*n*P

To obtain the analytical solution for the camera and mirror configurations, we exploit the similarity of our problem to the n-point perspective pose estimation problem (PnP). Specifically, in the standard formulation of PnP we seek the six-DOF transformation $\{{}^C_B\mathbf{R}, {}^C\mathbf{p}_B\}$ between a camera frame $\{C\}$, and a base frame $\{B\}$, given perspective measurements of N_f fiducial points, ${}^B\mathbf{p}_i$, $i = 1, \ldots, N_f$:

$$\mathbf{z}_i = \mathbf{h}({}^C\mathbf{p}_i) + \boldsymbol{\eta}_i, \quad \text{where} \quad \begin{bmatrix} {}^C\mathbf{p}_i \\ 1 \end{bmatrix} = \begin{bmatrix} {}^C_B\mathbf{R} & {}^C\mathbf{p}_B \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \begin{bmatrix} {}^B\mathbf{p}_i \\ 1 \end{bmatrix}. \tag{10}$$

By comparison of (10) to (5) and (9), the relationship between the mirror-based calibration and PnP problems becomes evident. Specifically, while in the PnP the only unknowns are $\{{}^C_B\mathbf{R}, {}^C\mathbf{p}_B\}$, in the mirror-based calibration problem we have additional unknowns, corresponding to the mirror configurations. All these unknowns, however, are "encoded" in the pair $\{\mathbf{A}_{N_v}, \mathbf{b}_{N_v}\}$, which appears in (5) and (9) in the same way as the pair $\{{}^C_B\mathbf{R}, {}^C\mathbf{p}_B\}$ does in (10). This similarity allows us to use the solution of the PnP, which is a well-studied problem, as a first step towards solving the multi-mirror calibration problem. Specifically, we first exploit the similarity to PnP to solve for the pair $\{\mathbf{A}_{N_v}, \mathbf{b}_{N_v}\}$, and next utilize (6) and (7) to solve for the camera and mirror configurations, as explained in Sect. 4.2.

In order to compute the pair $\{\mathbf{A}_{N_v}, \mathbf{b}_{N_v}\}$, we need to take the special properties of the matrix $\mathbf{A}_{N_v}$ into consideration. Specifically, $\mathbf{A}_{N_v}$ is the product of N_v Householder reflection matrices and one rotation matrix. As a result, $\mathbf{A}_{N_v}$ is unitary, and when N_v is even it is a rotation matrix (its determinant is equal to $+1$). Therefore, when N_v is even we can directly apply a PnP solution method to obtain $\mathbf{A}_{N_v}$ and $\mathbf{b}_{N_v}$. Any algorithm is suitable here; 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 [2].

When N_v is odd, the determinant of $\mathbf{A}_{N_v}$ is equal to -1, and therefore we cannot directly employ a PnP solution method. However, we can use a very simple transformation to bring the problem to a form in which the PnP algorithms can be directly applied. Specifically, we can transform $\mathbf{A}_{N_v}$ into a rotation matrix by applying an additional *known* reflection of our choice. For instance, if we change the sign of the y coordinates of all points in the image, this corresponds to applying a reflection across the xz-plane in the camera frame. Thus, the measurement equation in this case becomes $\mathbf{z} = \mathbf{h}({}^{\breve{C}}\mathbf{p}) + \boldsymbol{\eta}$, where

$$\begin{bmatrix} {}^{\breve{C}}\mathbf{p} \\ 1 \end{bmatrix} = \begin{bmatrix} (\mathbf{I}_3 - 2\mathbf{e}_2\mathbf{e}_2^T) & \mathbf{0}_{3\times 1} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \begin{bmatrix} {}^{C}\mathbf{p}' \\ 1 \end{bmatrix} = \begin{bmatrix} (\mathbf{I}_3 - 2\mathbf{e}_2\mathbf{e}_2^T) & \mathbf{0}_{3\times 1} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{A}_{N_v} & \mathbf{b}_{N_v} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \begin{bmatrix} {}^{B}\mathbf{p} \\ 1 \end{bmatrix}$$
$$= \begin{bmatrix} {}^{\breve{C}}_{B}\mathbf{R} & {}^{\breve{C}}\mathbf{p}_B \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \begin{bmatrix} {}^{B}\mathbf{p} \\ 1 \end{bmatrix}, \tag{11}$$

where $\mathbf{e}_2 = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T$. Note that the matrix ${}^{\breve{C}}_{B}\mathbf{R}$ is a rotation matrix, not a reflection. Thus, after negating the y-coordinate of all the points in the image, we can solve the PnP to obtain solution(s) for the unknown transformation $\{{}^{\breve{C}}_{B}\mathbf{R}, {}^{\breve{C}}\mathbf{p}_B\}$. Subsequently, given the PnP solution, we recover $\mathbf{A}_{N_v}$ and $\mathbf{b}_{N_v}$ through the following relationship which follows directly from (11)

$$\begin{bmatrix} (\mathbf{I}_3 - 2\mathbf{e}_2\mathbf{e}_2^T) & \mathbf{0}_{3\times 1} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \begin{bmatrix} {}^{\breve{C}}_{B}\mathbf{R} & {}^{\breve{C}}\mathbf{p}_B \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{N_v} & \mathbf{b}_{N_v} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix}. \tag{12}$$

4.2 Analytical Solution for the Camera and Mirror Configurations

We next describe how we use the computed $\{\mathbf{A}_{N_v}, \mathbf{b}_{N_v}\}$, to analytically solve for the camera-to-base transformation and the configuration of the mirrors. Before presenting our analytical solution, we discuss the conditions necessary for such a solution to exist.

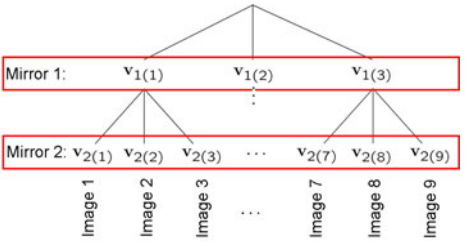

Fig. 3. Mirror configurations for the 2-mirror case depicted as a ternary tree

We start by noting that, in order to compute a discrete set of solutions to the PnP problem in the preceding section, at least three non-collinear points are required. For a unique solution, at least four points in a general configuration are necessary [3]. By analogy, in the mirror-based calibration problem at least $N_f = 3$ known points are needed in order to obtain solutions for $\{\mathbf{A}_{N_v}, \mathbf{b}_{N_v}\}$, and when $N_f \geq 4$ the solution is unique in each image (barring degenerate configurations).

We now examine the number of unknowns in the system, and compare it with the number of available constraint equations. Note that, $\mathbf{A}_{N_v}$ is a unitary matrix, and only has 3 degrees of freedom. Thus determining the pair $\{\mathbf{A}_{N_v}, \mathbf{b}_{N_v}\}$ from $N_f \geq 3$ points in each image only provides us with 6 independent constraint equations (3 for the unitary matrix $\mathbf{A}_{N_v}$, and 3 for the vector $\mathbf{b}_{N_v}$), which we can utilize to solve for the camera and mirror configurations. Thus, from N_c images, we can obtain $6N_c$ independent constraints.

On the other hand, if N_v mirrors are used, then each of the mirrors introduces 3 unknowns in the system (the elements of vector $\mathbf{v}_i$). Moreover, the 6-DOF camera-to-base transformation introduces 6 additional unknowns. Therefore, if in each of the N_c images each mirror moves to a new configuration, the total number of unknowns is equal to $3N_vN_c+6$. Thus, if $N_v > 1$, moving each mirror

to a new configuration in each image results in a problem where the number of unknowns, $3N_vN_c + 6$, is larger than the number of constraints available, $6N_c$. Therefore, some restrictions on the mirrors' motion must be imposed, for a unique solution to exist.

Our strategy is to move the mirrors in a specific order so as to isolate different system unknowns in each observation. Fig. 3 depicts our approach for the two-mirror case. For each configuration of mirror 1, we move mirror 2 in three different locations and record an image. 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 ($\mathbf{v}_{\ell(m)}$, denotes the m-th configuration of mirror ℓ). The mirror-configuration tree helps visualize the order in which the mirrors move and the unknown quantities are computed.

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 $6 + 3 \times (3 + 3^2 + \ldots + 3^{N_v}) = 6 + \frac{9}{2}(3^{N_v} - 1)$ unknowns, and 6×3^{N_v} constraints, which is an over-determined problem. We stress that, even though the problem is overdetermined, using a smaller number of images is not possible. At least three different configurations of each mirror are necessary, in order to compute a unique solution for the camera-to-base transformation. If only two configurations per mirror are used, then a continuum of solutions exists [10]. This dictates the proposed mirror motion strategy.

We next describe our algorithm for determining all the mirror-configuration vectors, $\mathbf{v}_{\ell(m)}$, as well as the transformation $\{{}^C_B\mathbf{R}, {}^C\mathbf{p}_B\}$. This algorithm is a recursive one: first, all the mirror configurations for mirror N_v are determined, then we proceed to mirror $N_v - 1$, and so on.

Determining the configuration of the N_v-th mirror: Specifically, once $\{\mathbf{A}_{N_v(m)}, \mathbf{b}_{N_v(m)}\}$, $m = 1, \ldots, 3^{N_v}$, are obtained using the PnP solution, we exploit the structure of (6) and (7) to compute the vectors $\mathbf{v}_{N_v(m)}$, $m = 1, \ldots, 3^{N_v}$. We proceed to compute these vectors in sets of three, corresponding to those images for which the configurations of mirrors 1 through $N_v - 1$ remain fixed.

To demonstrate the procedure, we focus on $\mathbf{v}_{N_v(m)}$, $m = 1, 2, 3$, which are the first three configurations for mirror N_v. In this case

$$\mathbf{A}_{N_v(m)} = \mathbf{M}_{N_v(m)}\mathbf{A}_{N_v-1(1)}, \quad m = 1, 2, 3, \tag{13}$$

where $\mathbf{M}_{\ell(m)}$ denotes the Householder reflection matrix for the m-th configuration of mirror ℓ. For each pair (m, m') of mirror-N_v configurations, if we let $\mathbf{r}_{mm'}$ be a unit vector perpendicular to both $\mathbf{v}_{N_v(m)}$ and $\mathbf{v}_{N_v(m')}$, we obtain

$$\begin{aligned}\mathbf{A}_{N_v(m)}\mathbf{A}^T_{N_v(m')}\mathbf{r}_{mm'} &= \mathbf{M}_{N_v(m)}\mathbf{A}_{N_v-1(1)}\mathbf{A}^T_{N_v-1(1)}\mathbf{M}^T_{N_v(m')}\mathbf{r}_{mm'} \\ &= \mathbf{M}_{N_v(m)}\mathbf{M}^T_{N_v(m')}\mathbf{r}_{mm'} = \mathbf{r}_{mm'},\end{aligned} \tag{14}$$

where we exploited $\mathbf{A}_{N_v-1(1)}\mathbf{A}^T_{N_v-1(1)} = \mathbf{I}_3$, and $\mathbf{v}^T_{N_v(m)}\mathbf{r}_{mm'} = \mathbf{v}^T_{N_v(m')}\mathbf{r}_{mm'} = 0$. The above result states that $\mathbf{r}_{mm'}$ is the eigenvector of $\mathbf{A}_{N_v(m)}\mathbf{A}^T_{N_v(m')}$ corresponding to the unit eigenvalue. Since $\mathbf{A}_{N_v(m)}\mathbf{A}^T_{N_v(m')}$ is a known matrix, we

can compute $\mathbf{r}_{mm'}$ up to sign. Moreover, since the vectors $\mathbf{r}_{mm'}$ and $\mathbf{r}_{mm''}$, for $m' \neq m''$, are both perpendicular to $\mathbf{v}_{N_v(m)}$, we can use the following expressions to obtain the vectors $\mathbf{v}_{N_v(m)}$, $m = 1, 2, 3$, up to scale:

$$\mathbf{v}_{N_v(1)} = c_{N_v(1)}\,\mathbf{r}_{13} \times \mathbf{r}_{12}, \quad \mathbf{v}_{N_v(2)} = c_{N_v(2)}\,\mathbf{r}_{21} \times \mathbf{r}_{23}, \quad \mathbf{v}_{N_v(3)} = c_{N_v(3)}\,\mathbf{r}_{13} \times \mathbf{r}_{23}, \tag{15}$$

where $c_{N_v(m)}$, $m = 1, 2, 3$ are unknown scalars. In order to determine these scalars, we note that they appear linearly in (7), along with the vector $\mathbf{b}_{N_v-1(1)}$. Using these equations we can formulate the over-determined linear system:

$$\begin{bmatrix} \mathbf{M}_{N_v(1)} & 2\mathbf{r}_{13} \times \mathbf{r}_{12} & \mathbf{0}_{3\times 1} & \mathbf{0}_{3\times 1} \\ \mathbf{M}_{N_v(2)} & \mathbf{0}_{3\times 1} & 2\mathbf{r}_{21} \times \mathbf{r}_{23} & \mathbf{0}_{3\times 1} \\ \mathbf{M}_{N_v(3)} & \mathbf{0}_{3\times 1} & \mathbf{0}_{3\times 1} & 2\mathbf{r}_{13} \times \mathbf{r}_{23} \end{bmatrix} \begin{bmatrix} \mathbf{b}_{N_v-1(1)} \\ c_{N_v(1)} \\ c_{N_v(2)} \\ c_{N_v(3)} \end{bmatrix} = \begin{bmatrix} \mathbf{b}_{N_v(1)} \\ \mathbf{b}_{N_v(2)} \\ \mathbf{b}_{N_v(3)} \end{bmatrix}. \tag{16}$$

The solution of this system provides us with the scale factors $c_{N_v(m)}$, $m = 1, 2, 3$, which, in turn, allows us to fully determine the vectors $\mathbf{v}_{N_v(m)}$, $m = 1, 2, 3$, using (15). Following the same procedure, we determine the configuration, $\mathbf{v}_{N_v(m)}$, of mirror N_v for the remaining images, $m = 4, \ldots, 3^{N_v}$.

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 analytical 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 analytical solution by following the above procedure for each of the PnP solutions, and select the solution which yields the minimum residual in the solution of (16). 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 the realistic case where noise is present, we have found that choosing the solution with the minimum error is a suitable way of rejecting invalid solutions.

Solving for the remaining mirror vectors, and the camera-to-base transformation: Once a solution for the vectors $\mathbf{v}_{N_v(m)}$ has been computed, we proceed to eliminate the unknowns corresponding to mirror N_v from the problem, using [see (6) and (7)]:

$$\mathbf{A}_{N_v-1(j)} = \mathbf{M}^{-1}_{N_v(m)} \mathbf{A}_{N_v(m)} \tag{17}$$

$$\mathbf{b}_{N_v-1(j)} = \mathbf{M}^{-1}_{N_v(m)} \left(\mathbf{b}_{N_v(m)} - 2\,\mathbf{v}_{N_v(m)} \right). \tag{18}$$

for $m = 1, \ldots, 3^{N_v}$, $j = \lceil m/3 \rceil$ (where $\lceil \cdot \rceil$ denotes the round-up operation). Note that since three different images (three different values of m) correspond to the same index j, we can obtain three estimates of $\mathbf{A}_{N_v-1(j)}$ and $\mathbf{b}_{N_v-1(j)}$. Due to the presence of noise, these estimates will generally be slightly different. To obtain a single, "average" estimate for $\mathbf{A}_{N_v-1(j)}$ and $\mathbf{b}_{N_v-1(j)}$, we employ a least-squares procedure similar to the one presented in [10]. Proceeding recursively as above, we can compute all mirrors' configurations. Moreover, as explained in Sect. 3.1, the camera-to-base transformation $\mathbf{A}_0$ and $\mathbf{b}_0$, can be obtained using the same process.

5 Analytical Solution for Scene Reconstruction

In the previous sections, we discussed 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 coordinates of all reconstruction points, whose coordinates are *not known a priori*. We describe an analytical method to compute their coordinates, which will be subsequently refined in an MLE for determining a more precise estimate while accounting for measurement noise. We assume that each image contains the reflections of at least three fiducial points and N_r reconstruction points. The fiducial points are utilized to determine the mirror vectors as well as the camera-to-base transformation (see Sect. 4.2), and the observations of the N_r reconstruction points are utilized to compute a 3D point-cloud representation of important objects in the scene (e.g., the robot chassis) with respect to the base frame.

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

$$s_j {}^{C}\hat{\mathbf{p}}'_j = \mathbf{A}_{N_v(j)} {}^{B}\mathbf{p} + \mathbf{b}_{N_v(j)}, \quad j = 1, \ldots, N_c, \tag{19}$$

where the scalar s_j is the *unknown* distance to the reflected point in image j. This measurement model is equivalent to the perspective projection model defined in Sect. 3.1. Equation (19) is linear in the unknowns s_j and ${}^{B}\mathbf{p}$. When ${}^{B}\mathbf{p}$ is observed in at least two images (i.e., $N_c \geq 2$), we can form an overdetermined set of linear equations ($N_c + 3$ unknowns and $3N_c$ constraints), which is solved to obtain the distance to the reconstruction point in each image, as well as the point's coordinates in the base frame:

$$\begin{bmatrix} \mathbf{A}_{N_v(1)} & -{}^{C}\hat{\mathbf{p}}'_1 & \ldots & \mathbf{0}_{3\times 1} \\ \mathbf{A}_{N_v(2)} & \mathbf{0}_{3\times 1} & \ldots & \mathbf{0}_{3\times 1} \\ \vdots & & \ddots & \vdots \\ \mathbf{A}_{N_v(N_c)} & \mathbf{0}_{3\times 1} & \ldots & -{}^{C}\hat{\mathbf{p}}'_{N_c} \end{bmatrix} \begin{bmatrix} {}^{B}\mathbf{p} \\ s_1 \\ \vdots \\ s_{N_c} \end{bmatrix} = \begin{bmatrix} -\mathbf{b}_{N_v(1)} \\ -\mathbf{b}_{N_v(2)} \\ \vdots \\ -\mathbf{b}_{N_v(N_c)} \end{bmatrix}. \tag{20}$$

6 MLE Refinement of Analytical Solutions

After the analytical solutions for the mirror configurations, camera-to-base transformation, and position of the reconstruction points have been computed, we refine them by applying maximum likelihood estimation. The vector of all unknown parameters is given by:

$$\mathbf{x} = \begin{bmatrix} {}^{C}\mathbf{p}_B^T & {}^{C}\bar{q}_B^T & {}^{C}\mathbf{v}_{1(1)}^T & \ldots & {}^{C}\mathbf{v}_{N_v(3^{N_v})}^T & {}^{B}\mathbf{p}_1^T & \ldots & {}^{B}\mathbf{p}_{N_r}^T \end{bmatrix}^T, \tag{21}$$

where N_r is the number of reconstruction points and ${}^{C}\bar{q}_B$ is the unit quaternion of 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

$$L(\mathcal{Z};\mathbf{x}) = \prod_{i=1}^{N_p}\prod_{j=1}^{N_c} p(\mathbf{z}_{ij};\mathbf{x}) = \prod_{i=1}^{N_p}\prod_{j=1}^{N_c} \frac{1}{2\pi\sigma_\eta^2}\exp\Big[-\frac{(\mathbf{z}_{ij}-\mathbf{h}_{ij}(\mathbf{x}))^T(\mathbf{z}_{ij}-\mathbf{h}_{ij}(\mathbf{x}))}{2\sigma_\eta^2}\Big], \quad (22)$$

where $\mathbf{h}_{ij}(\mathbf{x})$ is the measurement function defined in (9), and $N_p = N_r + N_f$ is the total number of reconstruction and fiducial points. Maximizing the likelihood, in the presence of i.i.d. Gaussian noise, is equivalent to minimizing the following non-linear least-squares cost function:

$$J(\mathbf{x}) = \sum_{i,j}(\mathbf{z}_{ij} - \mathbf{h}_{ij}(\mathbf{x}))^T(\mathbf{z}_{ij} - \mathbf{h}_{ij}(\mathbf{x})), \quad (23)$$

which is done iteratively using the Levenberg-Marquardt (LM) algorithm for estimating the parameter vector in (21).

7 Simulations

In this section, we present simulation results that demonstrate the feasibility of computing the camera-to-base transformation using the proposed approach.

We evaluate the accuracy of the analytically computed camera-to-base transformation (see Sect. 4.2) as well as the uncertainty in the MLE estimates. In particular, we investigate how the performance is affected by pixel noise, and by the distance between the camera and the mirrors. We consider a *base* case, in which three fiducial points placed at the corners of a right triangle, with sides measuring $20 \times 20 \times 20\sqrt{2}$ cm, are observed in 200 images. The points are seen via their reflections in two planar mirrors, which are placed at distances of 0.5 m in front of and behind the camera, and rotated by 30 deg in two directions.

In Fig. 4, we plot the errors in the position and attitude estimates of the analytical solution along with those of the MLE. To evaluate the analytical 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.

- As expected, by increasing the pixel noise, the accuracy of the analytical solution, as well as of the MLE decreases.
- As the distance between the camera and the mirrors increases, the accuracy also decreases. We note that the effect is more pronounced than in the single-mirror scenario [10], since in the two-mirror simulation *both* mirrors are moving farther away from the camera, and the effective depth to the scene is increasing at twice the rate.

Note that using the analytical 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 (typically 4) when compared to using a naïve initial guess (typically 18). This shows that the availability of a precise analytical solution improves the speed and robustness of the overall estimation process.

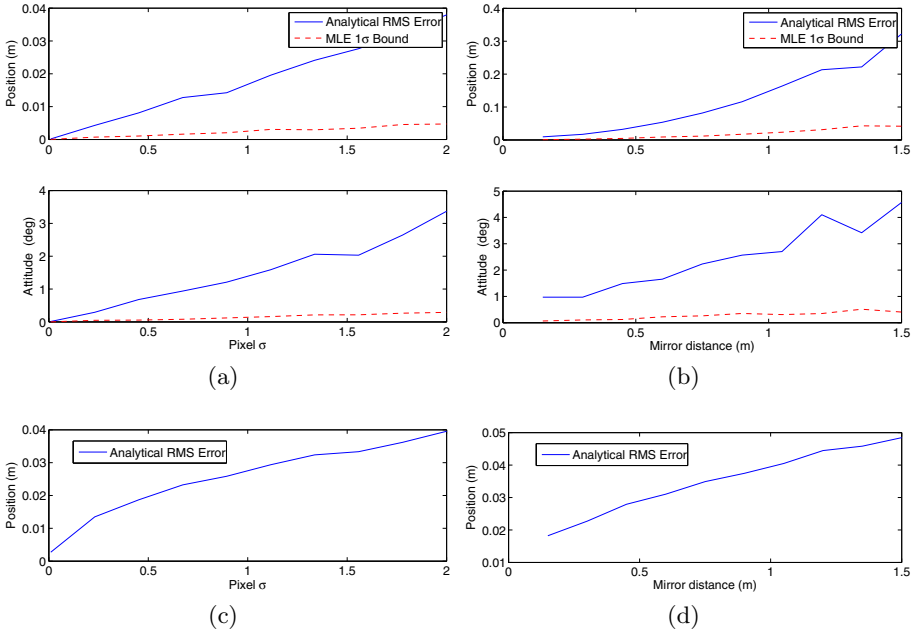

Fig. 4. Two-mirror case: Analytical solution and MLE accuracy for attitude and position plotted versus: (a) pixel noise and (b) mirror distance. Plots (c) and (d) depict the average reconstruction error for the least-accurately reconstructed point versus pixel noise and distance.

In order to evaluate the accuracy of the 3D reconstruction, we randomly populated the simulation environment with 60 points placed near the fiducial markers. The average RMS error (over 100 simulations), for the least-accurately reconstructed point, is plotted versus pixel noise and mirror distance [see Fig. 4(c) and (d)]. Note that even for large distances, or pixel disturbances, the analytical reconstruction error is 5 cm or less.

8 Experiments

The proposed method was also evaluated in real-world experiments to assess its performance and effectiveness in practice. In particular, we consider the case of two mirrors 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

(a) (b)

Fig. 5. (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. The fiducial points are only visible through the two-mirror reflection.

front mirror [see Fig.s 5(a) and 5(b)]. Instead, the markers were only visible via their double reflections, first in the rear mirror, then in the front mirror.

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 the Kanade-Lucas-Tomasi Feature Tracker (KLT) [13]. The front mirror was moved continuously through the image sequence, while the rear mirror was moved in three configurations (total 300 images per configuration). The analytically computed camera-to-base transformation is ${}^{C}\mathbf{p}_{B} = \begin{bmatrix}11.26 & -4.48 & -52.48\end{bmatrix}^{T}$ cm, and ${}^{C}\mathbf{q}_{B} = \begin{bmatrix}-0.5005 & 0.5063 & -0.4931 & 0.4998\end{bmatrix}^{T}$, which is very close to the manually determined estimate.

We initialized the MLE with the analytically computed quantities (both the camera-to-base transformation and the mirror vectors in all images), and the Levenberg-Marquardt minimization converged after three iterations. The final estimate for translation and orientation was ${}^{C}\mathbf{p}_{B} = \begin{bmatrix}10.54 & -4.42 & -53.01\end{bmatrix}^{T}$ 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.

9 Conclusions and Future Work

In this paper, we presented a method for point-based extrinsic camera calibration and 3D scene reconstruction in the challenging case when the points of interest lie outside the camera's direct field of view. To address this issue, we utilize one or more moving planar mirrors to extend the area which the camera can view. We do not assume prior knowledge about the mirror size or placement with respect to the camera. Instead, the only information we exploit are the reflections of

fiducial points, whose coordinates are known *a priori*, and reconstruction points, whose coordinates are unknown and must be calculated from the measurements. We introduced an analytical approach to determine the mirror configurations, the camera-to-base transformation, and the base-frame coordinates of the reconstruction points. Subsequently, we refined the analytically computed quantities using an MLE to produce high-accuracy estimates, along with a measure of the uncertainty in each parameter's estimate. We carried out simulation trials to verify the correctness of the proposed algorithm, as well as to evaluate its sensitivity to various system parameters. Furthermore, we validated the real-world performance of our approach, demonstrating its effectiveness and reliability in practical implementations.

In our ongoing work, we are investigating the feasibility of multi-mirror strategies for complete robot-body 3D reconstruction. Furthermore, 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 are estimated along with the camera-to-base transformation and the mirror configurations.

References

1. Merritt, E.L.: Explicitly three-point resection in space. Photogrammetric Engineering XV, 649–655 (1949)
2. Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applicatlons to image analysis and automated cartography. Communications of the ACM 24, 381–395 (1981)
3. Haralick, R.M., Lee, C.N., Ottenberg, K., Nölle, M.: Review and analysis of solutions of the three point perspective pose estimation problem. Int. Journal of Computer Vision 13, 331–356 (1994)
4. Gluckman, J., Nayar, S.K.: Planar catadioptric stereo: Geometry and calibration. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, Ft. Collins, CO, pp. 22–28 (1999)
5. Jang, G., Kim, S., Kweon, I.: Single camera catadioptic stereo system. In: Proc. of the Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras, Beijing, China (2005)
6. Ramsgaard, B.K., Balslev, I., Arnspang, J.: Mirror-based trinocular systems in robot-vision. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, Barcelona, Spain, pp. 499–502 (2000)
7. Nayar, S.K.: 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, pp. 245–254 (1988)
8. Jang, K.H., Lee, D.H., Jung, S.K.: A moving planar mirror based approach for cultural reconstruction. Computer Animation and Virtual Worlds 15, 415–423 (2004)
9. Kumar, R.K., Ilie, A., Frahm, J.M., Pollefeys, M.: Simple calibration of non-overlapping cameras with a mirror. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, Anchorage, AK (2008)

10. Hesch, J.A., Mourikis, A.I., Roumeliotis, S.I.: Mirror-based extrinsic camera calibration. In: Proc. of the Int. Workshop on the Algorithmic Foundations of Robotics, Guanajuato, Mexico, pp. 285–299 (2008)
11. Hesch, J.A., Mourikis, A.I., Roumeliotis, S.I.: 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, pp. 3865–3871 (2008)
12. Bouguet, J.Y.: Camera calibration toolbox for matlab (2006)
13. Shi, J., Tomasi, C.: Good features to track. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, Washington, DC, pp. 593–600 (1994)