Planar Catadioptric Stereo: Single and Multi-View Geometry for Calibration and Localization

Gian Luca Mariottini, Stefano Scheggi, Fabio Morbidi, Domenico Prattichizzo

***Abstract*— Planar Catadioptric Stereo vision sensors (PCS) combine a pinhole camera with two or more planar mirrors. PCS have recently received an increasing attention since a stereo view can be easily obtained without the need of exact multi-camera synchronization and calibration. In this paper we present a rigorous analytical treatment of the imaging geometry of PCS, propose new mirror calibration algorithms and introduce new multi-view properties that can be used for eye-in-hand camera localization. The effectiveness of the algorithms is shown via extensive simulation and real-data experiments on a robotic manipulator.**

I. Introduction

A. Motivation and related works

Stereoscopic vision has been playing a leading role in many fields, such as, e.g. 3-D imaging for entertainment and mapping, medicine, industrial robotics and augmented reality [1], [2]. Standard stereo sensors are composed by two coupled pinhole cameras and are generally expensive as well as difficult to synchronize and calibrate. In addition, the limited field of view considerably restricts their applicability domain. Several works have recently proposed the use of *catadioptric* vision sensors [3]–[6], which, combining both refracting (lens) and reflecting (mirrors) elements, naturally enlarge the field of view of classic pinhole cameras.

By using mirror reflections of a scene, stereo images can be captured with a single camera (catadioptric stereo). Single camera stereo has several advantages over traditional two-camera stereo: first, only a single set of internal camera calibration parameters needs to be determined. Second, no additional software or hardware is needed for exact camera synchronization [7].

This paper focuses on a special class of single camera stereo systems called *planar catadioptric stereo* vision sensors (PCS), consisting of a pinhole camera and two (or more) planar mirrors [8]. Industrial robotics could benefit from PCS: in fact, multiple planar mirrors disposed in the workspace, could be used to improve the pose accuracy of an eye-in-hand robotic arm performing high precision assembling and close manipulation tasks.

In [9], the authors carried out the first geometric study on 3-D reconstruction using reflective symmetry induced by planar mirrors. Several PCS have been designed and implemented in the last few years (see, e.g. [10], [11]) and some preliminary results on the geometry and calibration have been recently presented by Gluckman and Nayar in [12] (see [13] for the case of camera networks). The same authors have studied the set of constraints to be satisfied in order to obtain rectified stereo images from a catadioptric system [14]. An affine approximation for epipolar-based PCS rectification has been lately proposed in [15].

From the previous overview, we see that although some research has been done in this field, the specific properties of PCS have been addressed only sparsely in the literature and little attention has been devoted to the application of these sensors to image-based robot navigation.

G.L. Mariottini is with Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN 55455, USA. S. Scheggi, F. Morbidi and D. Prattichizzo are with Dipartimento di Ingegneria dell'Informazione, University of Siena, 53100 Siena, Italy.

Corresponding author: gianluca@cs.umn.edu

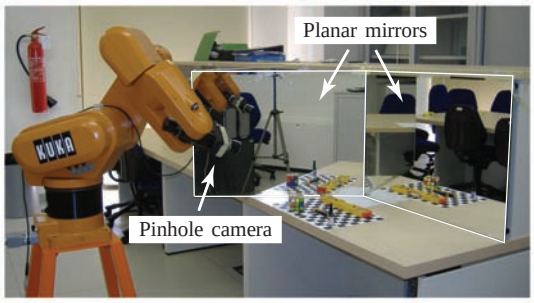

Fig. 1. PCS experimental setup composed of a pinhole camera mounted on the end-effector of a robotic manipulator and two planar mirrors.

B. Contributions

The original contribution of this paper is twofold:

- We propose a rigorous analytical treatment of the imaging properties of PCS composed by a pinhole camera and two planar mirrors (see Fig. 1) and introduce new algebraic results and geometric properties in the case of static and moving cameras (multi-view PCS).
- On the grounds of these results, we propose and compare a suite of original numeric algorithms for *mirror calibration* and *image-based camera localization* that can be used as the core of a position based visual servoing strategy.

Extensive simulation and real-data experiments performed on an eye-in-hand robot are presented in order to illustrate the theory and show the effectiveness of the proposed algorithms.

C. Organization

Sect. II reviews some basic results on perspective projection through planar mirrors [12]. Sect. III deals with the single and multi-view geometry for PCS. Sect. IV presents a suite of algorithms for mirror calibration and

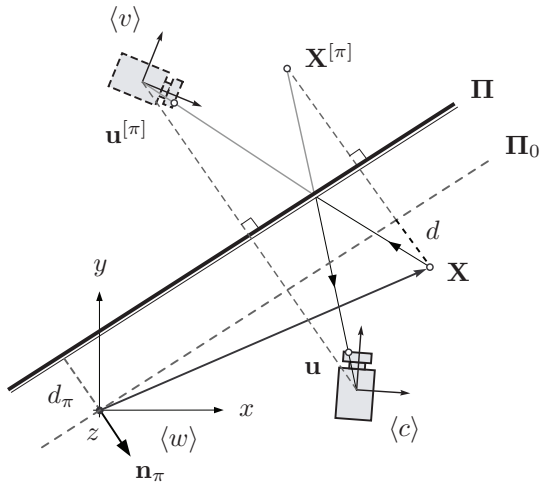

Fig. 2. Reflection through the planar mirror $\mathbf{\Pi}$: virtual point $\mathbf{X}^{[\pi]}$ and virtual camera $\langle v \rangle$.

image-based camera localization. These algorithms are tested via simulation and real-data experiments in Sect. V. In Sect. VI, the main contributions of the paper are summarized and future research directions are highlighted.

II. Planar mirrors and perspective projection

A. The virtual point and the reflection matrix

Let us consider the setup reported in Fig. 2 where a perspective camera $\langle c \rangle$ is in front of a planar mirror $\mathbf{\Pi}$ described by a normal vector $\mathbf{n}_\pi$ in the world frame $\langle w \rangle \equiv \{O; x, y, z\}$ and by its distance d_π from $\langle w \rangle$. A 3-D point $\mathbf{X} \triangleq [x\ y\ z]^T$ in $\langle w \rangle$ is in front of the mirror ($\widetilde{\mathbf{X}}$ indicates its extension in homogeneous coordinates). For the sake of clearness and without affecting the generality of our results, Fig. 2 shows a simplified planar camera-mirror arrangement. Note that the perspective image $\mathbf{u}$ (pixels) of $\mathbf{X}$ after its reflection by the planar mirror $\mathbf{\Pi}$ can be also calculated as the direct projection on $\langle c \rangle$ of the so-called *virtual point* $\mathbf{X}^{[\pi]}$ (see Fig. 2). In the next proposition $(\mathbf{R}_c^w, \mathbf{t}_c^w)$ denotes the rigid transformation from the camera frame $\langle c \rangle$ to the mirror frame $\langle w \rangle$.

***Proposition** 1 (Planar mirror perspective projection):* Consider the setup in Fig. 2. The perspective projection $\widetilde{\mathbf{u}} = [u\ v\ 1]^T$ (pixels) of 3D point $\mathbf{X}$ in $\langle w \rangle$ after its mirroring on $\mathbf{\Pi}$ is given by,

$$\lambda\, \widetilde{\mathbf{u}} = \mathbf{K}\, [\mathbf{R}_c^w \ \ \mathbf{t}_c^w]\, \mathbf{D}^{[\pi]}\, \widetilde{\mathbf{X}} \quad \text{with} \ \ \lambda \in I\!R^+,$$

where $\mathbf{K}$ is the intrinsic camera calibration matrix and $\mathbf{D}^{[\pi]}$ is the (homogeneous) *reflection matrix* about the mirror $\mathbf{n}_\pi$ defined as,

$$\mathbf{D}^{[\pi]} = \begin{bmatrix} \mathbf{S}^{[\pi]} & 2\, d_\pi\, \mathbf{n}_\pi \\ \mathbf{0}^T & 1 \end{bmatrix}, \tag{1}$$

where $\mathbf{S}^{[\pi]} \triangleq \mathbf{I} - 2\, \mathbf{n}_\pi \mathbf{n}_\pi^T$. ■

Note that $\mathbf{S}^{[\pi]} \in \mathrm{O}(3)$, $\det(\mathbf{S}^{[\pi]}) = -1$ and that $(\mathbf{D}^{[\pi]})^{-1} = \mathbf{D}^{[\pi]}$.

B. The virtual camera and the projection equivalence

We introduce now the concept of *virtual camera* $\langle v \rangle$ whose reference frame is simply obtained reflecting $\langle c \rangle$ with respect to $\langle w \rangle$ about $\mathbf{\Pi}$ (see Fig. 2).

***Proposition** 2 (Camera frame mirroring):* Suppose given a planar mirror $(\mathbf{n}_\pi, d_\pi)$ and the cameras in $\langle c \rangle$ and $\langle v \rangle$ (dashed, in Fig. 2). Then the homogeneous matrix transforming vectors from $\langle v \rangle$ to $\langle w \rangle$ is given by

$$\mathbf{M}_w^v = \mathbf{D}^{[\pi]}\, \mathbf{H}_w^c \triangleq \begin{bmatrix} \mathbf{Q}_w^v & \mathbf{t}_w^v \\ \mathbf{0}^T & 1 \end{bmatrix}. \tag{2}$$

Proof: As shown in Prop. 1, $\mathbf{t}_w^v$ (from $\langle w \rangle$ to $\langle v \rangle$) coincides with the reflection of $\mathbf{t}_w^c$, i.e.,

$$\widetilde{\mathbf{t}}_w^v = \mathbf{D}^{[\pi]}\, \widetilde{\mathbf{t}}_w^c. \tag{3}$$

Let us now consider the versor $\mathbf{z}_{c(c)} = [0\ 0\ 1]^T$, where the subscript indicates that it is the z-axis of the camera frame and is expressed in $\langle c \rangle$. We can write $\widetilde{\mathbf{z}}_{v(w)} = \mathbf{D}^{[\pi]} \mathbf{H}_w^c\, \widetilde{\mathbf{z}}_{c(c)}$ and then $\mathbf{M}_w^v\, \widetilde{\mathbf{z}}_{v(v)} = \mathbf{D}^{[\pi]} \mathbf{H}_w^c\, \widetilde{\mathbf{z}}_{c(c)}$ where $\mathbf{z}_{v(v)} = [0\ 0\ 1]^T$ and $\mathbf{H}_w^c = \begin{bmatrix} \mathbf{R}_w^c & \mathbf{t}_w^c \\ \mathbf{0}^T & 1 \end{bmatrix}$. We easily obtain, for all i, $\begin{bmatrix} \mathbf{r}_{wi}^{v\,T} & 0 \end{bmatrix}^T + \widetilde{\mathbf{t}}_w^v = \mathbf{D}^{[\pi]} \big(\begin{bmatrix} \mathbf{r}_{wi}^{c\,T} & 0 \end{bmatrix}^T + \widetilde{\mathbf{t}}_w^c \big)$, where $\mathbf{r}_{wi}^v$ and $\mathbf{r}_{wi}^c$ are the ith columns of matrices $\mathbf{R}_w^v$ and $\mathbf{R}_w^c$, respectively. Exploiting the first three rows and using (3) we get,

$$\mathbf{Q}_w^v = \mathbf{S}^{[\pi]}\, \mathbf{R}_w^c. \tag{4}$$

From (4) and (3), we obtain (2). ■

To highlight the different nature of the homogeneous matrices in Prop. 2, two symbols, $\mathbf{M}$ and $\mathbf{H}$, have been used: in fact $\mathbf{H}_w^c \in \mathrm{SE}(3)$, while $\mathbf{M}_w^v \in \mathrm{E}(3)$ being $\det(\mathbf{Q}_w^v) = -1$.

***Proposition** 3 (Virtual camera projection equivalence):* Let $\mathbf{u}$ be the perspective projection in $\langle c \rangle$ of a 3-D point $\mathbf{X}$ after its reflection on $\mathbf{\Pi}$. Then, we have,

$$\mathbf{u} = \mathbf{u}^{[\pi]},$$

being $\mathbf{u}^{[\pi]}$ the perspective projection of $\mathbf{X}$ onto the image plane of the virtual camera $\langle v \rangle$. ■

Prop. 3 states that the perspective projection $\mathbf{u}$ of $\mathbf{X}_c^{[\pi]}$ (i.e., $\mathbf{X}^{[\pi]}$ centered in $\langle c \rangle$, see Fig. 2) coincides with the perspective projection $\mathbf{u}^{[\pi]}$ of $\mathbf{X}_v$ (i.e., $\mathbf{X}$ in $\langle v \rangle$).

***Remark** 1 (Reflective epipolar geometry):* The epipolar geometry between $\langle c \rangle$ and $\langle v \rangle$ is well defined, i.e., $(\widetilde{\mathbf{u}}^{[\pi]})^T \mathbf{K}^{-T}\, \mathbf{E}^{[\pi]}\, \mathbf{K}^{-1}\, \widetilde{\mathbf{u}} = 0$ where the essential matrix $\mathbf{E}^{[\pi]} = 2\, d_{\pi(c)} [\mathbf{n}_{\pi(c)}]_\times$, with $d_{\pi(c)}$ and $\mathbf{n}_{\pi(c)}$ representing distance and the normal of the mirror measured from $\langle c \rangle$[1]. The epipoles on $\langle c \rangle$ and $\langle v \rangle$ are equal and $\widetilde{\mathbf{e}}_c \sim \mathbf{n}_{\pi(c)}$.

III. Single and multi-view geometry for PCS

A. Single-view geometry

In this section we will assume to have two planar mirrors $\mathbf{n}_1$ and $\mathbf{n}_2$ as in Fig. 3. In this case two virtual cameras $\langle v_1 \rangle$ and $\langle v_2 \rangle$ exist, as well as suitable epipolar geometry relating $\langle c \rangle$ with both $\langle v_1 \rangle$ and $\langle v_2 \rangle$. The generalization to multiple

[1] $[\,\cdot\,]_\times$ is the matrix form of the cross product operator such that for any vectors $\mathbf{a}$ and $\mathbf{b}$, $\mathbf{a} \times \mathbf{b} = [\mathbf{a}]_\times\, \mathbf{b}$.

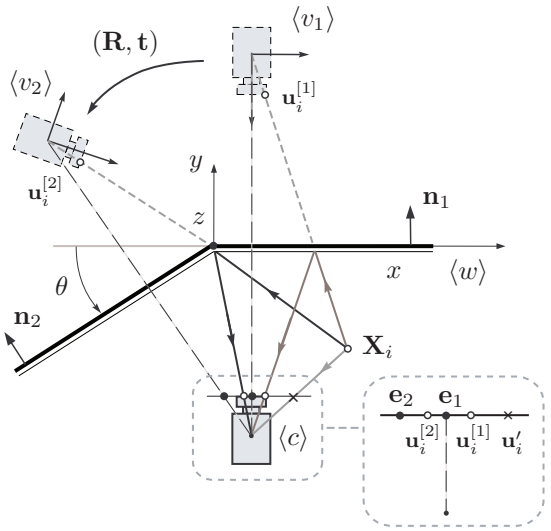

Fig. 3. The camera $\langle c \rangle$ and the virtual cameras $\langle v_1 \rangle$ and $\langle v_2 \rangle$.

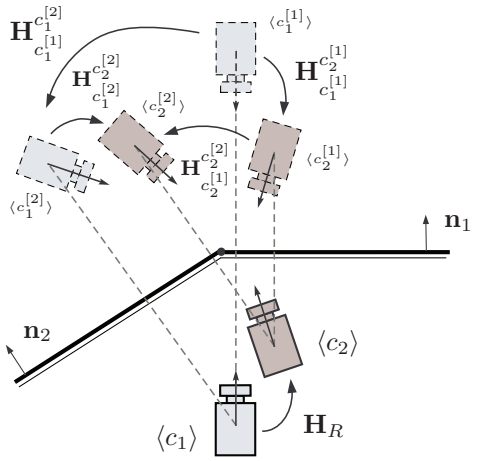

Fig. 4. Multi-view geometry for the cameras $\langle c_1 \rangle$ and $\langle c_2 \rangle$.

mirrors is straightforward and it will be not discussed in this paper. Let be given the image points $\mathbf{u}_i^{[1]}$, $\mathbf{u}_i^{[2]}$, $i = 1, \ldots, n$ in $\langle c \rangle$, projections of a set of $n \geq 8$, 3-D points $\mathbf{X}_i$ reflected onto $\mathbf{n}_1$ and $\mathbf{n}_2$, respectively (in the present context, for convenience, the subscript π in $\mathbf{n}_1$ and $\mathbf{n}_2$ will be henceforth neglected). Note that, while the subscript i is the point index, the superscript inside the brackets will always refer to the mirror number through which that 3-D point is reflected. From Prop. 2 we can write:

$$\mathbf{M}_w^{v_1} = \mathbf{D}^{[1]}\mathbf{H}_w^c, \quad (5)$$

$$\mathbf{M}_w^{v_2} = \mathbf{D}^{[2]}\mathbf{H}_w^c, \quad (6)$$

where $\mathbf{D}^{[\cdot]}$ is as in (1). By solving (5) and (6) with respect to $\mathbf{H}_w^c$, it yields:

$$\mathbf{M}_w^{v_2}\,\mathbf{M}_{v_1}^w = \mathbf{D}^{[2]}\mathbf{D}^{[1]} \triangleq \begin{bmatrix} \mathbf{R}_D & \mathbf{t}_D \\ \mathbf{0}^T & 1 \end{bmatrix}. \quad (7)$$

From (7), $\mathbf{R}_D$ and $\mathbf{t}_D$ are given by,

$$\mathbf{R}_D = \mathbf{I} + 4(\mathbf{n}_1 \cdot \mathbf{n}_2)\,\mathbf{n}_1\mathbf{n}_2^T - 2\,\mathbf{n}_1\mathbf{n}_1^T - 2\,\mathbf{n}_2\mathbf{n}_2^T,$$
$$\mathbf{t}_D = 2\,d_1\mathbf{n}_1 - 2(d_1(\mathbf{n}_1 \cdot \mathbf{n}_2) + d_2)\mathbf{n}_2.$$

Note that $(\mathbf{R}_D, \mathbf{t}_D)$ only depend on the mirror setup (i.e., $\mathbf{n}_1$, $\mathbf{n}_2$ and d_1, d_2). Since $\mathbf{u}_i^{[1]}$ and $\mathbf{u}_i^{[2]}$ in $\langle c \rangle$ are corresponding, the epipolar geometry between $\langle v_1 \rangle$ and $\langle v_2 \rangle$ does exist, i.e., $\widetilde{\mathbf{u}}_i^{[2]\,T}\mathbf{F}\,\widetilde{\mathbf{u}}_i^{[1]} = 0$, where $\mathbf{F}$ is the fundamental matrix, from which we can compute both the epipoles $\mathbf{e}_1$ and $\mathbf{e}_2$ and the epipolar lines [16]. In addition, given the internal camera calibration matrix $\mathbf{K}$, from $\mathbf{F}$ we can compute the essential matrix $\mathbf{E} = [\mathbf{t}]_\times\mathbf{R}$ where $(\mathbf{R}, \mathbf{t})$ is the rigid body motion between $\langle v_1 \rangle$ and $\langle v_2 \rangle$ (see Fig. 3). Once $\mathbf{E}$ is known, it can be decomposed to retrieve the rotation matrix $\mathbf{R}$ and the vector $\mathbf{t}$ (up to a scale factor), [16].

B. Multi-view geometry

As from Fig. 4, in this section we suppose given *two* views $\langle c_1 \rangle$ and $\langle c_2 \rangle$ of a set of 3-D scene points, mirrored through two planar mirrors.

Also, let $\mathbf{H}_R$ be the homogeneous transformation matrix between $\langle c_1 \rangle$ and $\langle c_2 \rangle$ and let $\mathbf{H}_{c_1^{[2]}}^{c_2^{[2]}}$, $\mathbf{H}_{c_2^{[2]}}^{c_2^{[1]}}$, $\mathbf{H}_{c_2^{[1]}}^{c_1^{[2]}}$, $\mathbf{H}_{c_1^{[1]}}^{c_1^{[2]}}$ be the four transformation matrices for all the pairs of virtual views.

***Proposition** 4 (Multiple cameras mirroring):* The following equation holds true,

$$\mathbf{H}_{c_2^{[1]}}^{c_2^{[2]}} = \mathbf{H}_R^{-1}\,\mathbf{H}_{c_1^{[1]}}^{c_1^{[2]}}\,\mathbf{H}_R. \quad (8)$$

Proof: It is easy to verify that the following equalities, concerning the relative motion between $\langle c_1 \rangle$ and $\langle c_2 \rangle$, hold true: $\mathbf{H}_R \triangleq \mathbf{H}_{c_1}^{c_2} = \mathbf{H}_{c_1^{[1]}}^{c_2^{[1]}} = \mathbf{H}_{c_1^{[2]}}^{c_2^{[2]}}$, being,

$$\mathbf{H}_{c_1}^{c_2} = \left(\mathbf{M}_w^{c_1^{[1]}}\right)^{-1}\mathbf{M}_w^{c_2^{[1]}}. \quad (9)$$

From (14), it yields,

$$\mathbf{D}^{[2]}\mathbf{D}^{[1]} = \mathbf{M}_w^{c_1^{[1]}}\,\mathbf{H}_{c_1^{[1]}}^{c_1^{[2]}}\,\mathbf{M}_{c_1^{[1]}}^w, \quad (10)$$

$$\mathbf{D}^{[2]}\mathbf{D}^{[1]} = \mathbf{M}_w^{c_2^{[1]}}\,\mathbf{H}_{c_2^{[1]}}^{c_2^{[2]}}\,\mathbf{M}_{c_2^{[1]}}^w. \quad (11)$$

Retrieving $\mathbf{M}_w^{c_1^{[1]}}$ from (9), we get,

$$\mathbf{M}_w^{c_1^{[1]}} = \mathbf{M}_w^{c_2^{[1]}}\,\mathbf{H}_{c_2}^{c_1} \quad (12)$$

and, analogously, retrieving $\mathbf{M}_{c_1^{[1]}}^w$ from (9) yields,

$$\mathbf{M}_{c_1^{[1]}}^w = \mathbf{H}_{c_1}^{c_2}\mathbf{M}_{c_2^{[1]}}^w. \quad (13)$$

Substituting (12) and (13) in (10) we obtain $\mathbf{D}^{[2]}\mathbf{D}^{[1]} = \mathbf{M}_w^{c_2^{[1]}}\,\mathbf{H}_{c_2}^{c_1}\,\mathbf{H}_{c_1^{[1]}}^{c_1^{[2]}}\mathbf{H}_{c_1}^{c_2}\,\mathbf{M}_{c_2^{[1]}}^w$. Comparing this last equation with (11), we finally obtain (8). ■

Note that equation (8) allows the computation of the rigid motion $\mathbf{H}_R$ also when the epipolar geometry between $\langle c_1 \rangle$ and $\langle c_2 \rangle$ is not well-defined (e.g. small baseline case). In fact, the epipolar geometry between the virtual cameras is always well-defined by construction.

***Proposition** 5 (Virtual cameras mirroring):* The following equality holds true,

$$\mathbf{M}_w^{v_1}\mathbf{H}_E = \mathbf{D}^{[2]}\mathbf{D}^{[1]}\mathbf{M}_w^{v_1}, \quad (14)$$

where $\mathbf{H}_E \triangleq \mathbf{H}_{v_1}^{v_2} = \left[\begin{smallmatrix}\mathbf{R} & \mathbf{t}\\ \mathbf{0}^T & 1\end{smallmatrix}\right]$ is the homogeneous transformation matrix representing the rigid-body motion between the virtual cameras $\langle v_1\rangle$ and $\langle v_2\rangle$.

Proof: From (7) it directly follows that,

$$\mathbf{M}_w^{v_2} = \mathbf{D}^{[2]}\mathbf{D}^{[1]}\mathbf{M}_w^{v_1}, \tag{15}$$

and since $\mathbf{H}_E = \mathbf{M}_{v_1}^{w}\mathbf{M}_w^{v_2}$, we have $\mathbf{M}_w^{v_2} = \mathbf{M}_w^{v_1}\mathbf{H}_E$. Substituting into (15), we obtain (14). ■

***Remark** 2:* Without loss of generality, we can assume that the $z-$axis of $\langle w\rangle$ coincides with the mirrors screw axis and the $x-$axis lies on mirror 1, so that $\mathbf{D}^{[1]}$ becomes,

$$\mathbf{D}^{[1]} = \begin{bmatrix}1 & 0 & 0 & 0\\ 0 & -1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}. \tag{16}$$

Moreover, we assume a calibrated mirror setup, such that $\mathbf{S}^{[2]}$ (the 3×3 reflection matrix in $\mathbf{D}^{[2]}$) is known. Then, (14) can be decomposed as,

$$\mathbf{Q}_w^{v_1}\,\mathbf{R} = \mathbf{S}^{[2]}\mathbf{S}^{[1]}\mathbf{Q}_w^{v_1} \tag{17}$$

$$\mathbf{Q}_w^{v_1}\,\mathbf{t} + (\mathbf{I} - \mathbf{S}^{[2]}\mathbf{S}^{[1]})\mathbf{t}_w^{v_1} = \mathbf{0}. \tag{18}$$

Note that the translation vectors in (18) are assumed to be unknown and that only the rotation matrix $\mathbf{R}$ can be obtained from the $\mathbf{E}$-matrix decomposition.

IV. Algorithms for mirror calibration and image-based camera localization

In this section we present a suite of algorithms for mirror calibration (i.e., computation of the angle θ between the mirrors) and for image-based localization of a moving camera, based on the properties presented in Sects. II and III.

A. Mirror calibration

Consider the setup in Fig. 3 in which at least $n \geq 8$ points are directly observed by the camera $\langle c\rangle$ at $\mathbf{u}'_i$, $i = 1,\ldots,n$. The same points are also reflected by the mirrors $\mathbf{n}_1$ and $\mathbf{n}_2$ at $\mathbf{u}_i^{[1]}$ and $\mathbf{u}_i^{[2]}$ (white dots). In this scenario two fundamental matrices $\mathbf{F}_1$ and $\mathbf{F}_2$ exist and they are computed from the corresponding pairs $(\mathbf{u}_i^{[1]}, \mathbf{u}'_i)$ and $(\mathbf{u}_i^{[2]}, \mathbf{u}'_i)$, respectively (cf. Prop. 3). From $\mathbf{F}_1$ and $\mathbf{F}_2$ we can then find the epipoles $\mathbf{e}_1$ and $\mathbf{e}_2$ as their right null-spaces (black dots). Since the direction of each epipole $\mathbf{e}_j$, $j = 1, 2$, is always parallel to $\mathbf{n}_j$, we obtain the following:

***Proposition** 6 (Calibration with epipoles):* The angle θ between the mirrors $\mathbf{n}_1$ and $\mathbf{n}_2$ is given by,

$$\theta = \arccos\left(\frac{(\mathbf{K}^{-1}\,\widetilde{\mathbf{e}}_1)^T(\mathbf{K}^{-1}\,\widetilde{\mathbf{e}}_2)}{\|\mathbf{K}^{-1}\,\widetilde{\mathbf{e}}_1\|\,\|\mathbf{K}^{-1}\,\widetilde{\mathbf{e}}_2\|}\right).$$

In the next proposition we uses the epipoles between the virtual cameras to solve the mirror calibration problem (see Fig. 5). Let $\mathbf{F}^{[21]}$ be the fundamental matrix computed from the corresponding points $(\mathbf{u}_i^{[1]}, \mathbf{u}_i^{[2]})$ and γ the angle between the virtual epipoles $\mathbf{e}^{[12]}$ and $\mathbf{e}^{[21]}$. It is easy to verify that $\gamma = \arccos\left(\frac{(\mathbf{K}^{-1}\,\widetilde{\mathbf{e}}^{[12]})^T(\mathbf{K}^{-1}\,\widetilde{\mathbf{e}}^{[21]})}{\|\mathbf{K}^{-1}\,\widetilde{\mathbf{e}}^{[12]}\|\,\|\mathbf{K}^{-1}\,\widetilde{\mathbf{e}}^{[21]}\|}\right)$.

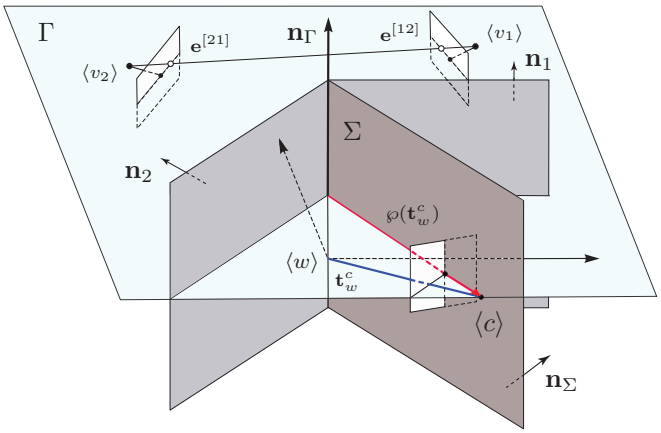

Fig. 5. Vector $\wp(\mathbf{t}_w^c)$ and virtual cameras epipoles $\mathbf{e}^{[12]}$ and $\mathbf{e}^{[21]}$.

***Proposition** 7 (Calibration with virtual epipoles):* The angle θ between the mirrors $\mathbf{n}_1$ and $\mathbf{n}_2$ is given by,

$$\theta = \frac{\pi - \gamma}{2}.$$

B. Image-based camera localization

This section presents a set of algorithms for the estimation of the rigid-motion $(\mathbf{R}_w^c, \mathbf{t}_w^c)$ from image measurements. In particular, Prop. 8 provides a method to estimate $\mathbf{R}_w^c$ and Prop. 9 two methods to compute the projection of $\mathbf{t}_w^c$ on the plane defined by the camera centers.

Let Γ be the plane defined by the center of the cameras $\langle c\rangle$, $\langle v_1\rangle$ and $\langle v_2\rangle$ and $\mathbf{n}_{\Gamma(c)}$ the normal vector to Γ written in $\langle c\rangle$ (see Fig. 5) . Let $\mathbf{e}_1$ be the epipole with unitary norm defined in Prop. 6 and let $\mathbf{a}_{(c)} \triangleq \mathbf{e}_1 \times \mathbf{n}_{\Gamma(c)}$.

***Proposition** 8 ($\mathbf{R}_w^c$ estimation):* For every rigid-motion $(\mathbf{R}_w^c, \mathbf{t}_w^c)$, the following equation holds true,

$$\mathbf{R}_w^c = \begin{bmatrix}0 & 0 & -1\\ 0 & 1 & 0\\ -1 & 0 & 0\end{bmatrix}\left[\mathbf{n}_{\Gamma(c)}\;\; \mathbf{e}_1\;\; \mathbf{a}_{(c)}\right]^{-1}. \tag{19}$$

Proof: For every pose of $\langle c\rangle$, the camera center lies on the same plane Γ having normal vector $\mathbf{n}_{\Gamma(w)} = [0\;\;0\;\;1]^T$ in $\langle w\rangle$. Hence, also $\widetilde{\mathbf{e}}^{[12]}$ and $\widetilde{\mathbf{e}}^{[21]}$ lie on Γ. Let $\boldsymbol{\ell}$ be the line joining the epipoles $\widetilde{\mathbf{e}}^{[12]}$ and $\widetilde{\mathbf{e}}^{[21]}$ expressed in $\langle c\rangle$, then $\boldsymbol{\ell} = \widetilde{\mathbf{e}}^{[12]} \times \widetilde{\mathbf{e}}^{[21]}$. Since $\boldsymbol{\ell}$ is the projection of the plane Γ in $\langle c\rangle$, $\mathbf{n}_{\Gamma(c)} = \mathbf{K}^T\boldsymbol{\ell}$ and,

$$-\mathbf{n}_{\Gamma(w)} = \mathbf{R}_w^c\,\mathbf{n}_{\Gamma(c)}. \tag{20}$$

Similarly, being $\mathbf{e}_1$ a unit norm vector obtained as the right null-space of the fundamental matrix $\mathbf{F}_1$ between $\langle c\rangle$ and $\langle v_1\rangle$, $\mathbf{e}_{1(w)} = [0\;\;1\;\;0]^T$ in $\langle w\rangle$. Then,

$$\mathbf{e}_{1(w)} = \mathbf{R}_w^c\mathbf{e}_1. \tag{21}$$

Finally, since $\mathbf{e}_{1(w)} \perp -\mathbf{n}_{\Gamma(w)}$ and $\mathbf{e}_{1(c)} \perp \mathbf{n}_{\Gamma(c)}$, being $\mathbf{a}_{(w)} = \mathbf{e}_{1(w)} \times -\mathbf{n}_{\Gamma(w)}$ and $\mathbf{a}_{(c)} = \mathbf{e}_{1(c)} \times \mathbf{n}_{\Gamma(c)}$, we obtain $\mathbf{a}_{(w)} = \mathbf{R}_w^c\,\mathbf{a}_{(c)}$. Using this last equation together with (20) and (21), we have that $[-\mathbf{n}_{\Gamma(w)}\;\;\mathbf{e}_{1(w)}\;\;\mathbf{a}_{(w)}] = \mathbf{R}_w^c\,[\mathbf{n}_{\Gamma(c)}\;\;\mathbf{e}_1\;\;\mathbf{a}_{(c)}]$, from which (19) follows. ■

Consider the setup in Fig. 5 and let $\wp(\mathbf{t}_w^c) \in I\!R^2$ contain the first two components of the projection of $\mathbf{t}_w^c$ onto the plane Γ. Prop. 9 provides two methods to estimate the direction of $\wp(\mathbf{t}_w^c)$, i.e., $\frac{\wp(\mathbf{t}_w^c)}{\|\wp(\mathbf{t}_w^c)\|} \triangleq d(\wp(\mathbf{t}_w^c))$. The first one uses the fundamental matrix $\mathbf{F}^{[12]}$ between $\langle v_1 \rangle$ and $\langle v_2 \rangle$ and the second one the Sylvester equation derived from (14). Let Σ be the plane defined by the mirrors *screw axis* (i.e., the 3-D line of intersection between the mirrors) and the center of $\langle c \rangle$, and let $\mathbf{n}_\Sigma$ be the normal to Σ, (see Fig. 5).

In the next proposition, $\mathbf{R}_z(\gamma)$ denotes a rotation matrix of an angle γ about the z-axis.

***Proposition** 9 (Estimation of $d(\wp(\mathbf{t}_w^c))$):* The direction of $\wp(\mathbf{t}_w^c)$ is given by,

$$d(\wp(\mathbf{t}_w^c)) = \frac{\mathbf{t}_w^{c\,*}}{\|\mathbf{t}_w^{c\,*}\|}, \tag{22}$$

where $\mathbf{t}_w^{c\,*}$ can be computed as in (i) or in (ii), as follows:

$$(i) \qquad \mathbf{t}_w^{c\,*} = [\mathbf{a}(1) \quad \mathbf{a}(2)]^T, \tag{23}$$

with $\mathbf{a} = \mathbf{R}_z(-90^\circ)\,\mathbf{n}_\Sigma$, $\mathbf{n}_\Sigma(1) > 0$;

$$(ii)\ \mathbf{t}_w^{c\,*} = \Big[\frac{\mathbf{a}(1)(1-\mathbf{n}_2^2(2))+\mathbf{n}_2(1)\mathbf{n}_2(2)\mathbf{a}(2)}{2\,\mathbf{n}_2^2(1)} \quad \frac{\mathbf{a}(2)\mathbf{n}_2(1)-\mathbf{a}(1)\mathbf{n}_2(2)}{2\,\mathbf{n}_2(1)}\Big]^T, \tag{24}$$

with $\mathbf{a} = -\mathbf{R}_w^c \mathbf{t}_{v_2}^{v_1}$.

Proof: Let us start by proving part (i) of the statement. $\mathbf{t}_w^c$ and $\wp(\mathbf{t}_w^c)$ lie on the plane Σ with normal vector $\mathbf{n}_\Sigma = [\mathbf{n}_\Sigma(1) \quad \mathbf{n}_\Sigma(2) \quad 0]^T$ in $\langle w \rangle$. Since all the corresponding epipolar lines computed from the fundamental matrix $\mathbf{F}^{[12]}$ intersect at the image projection of the screw axis $\mathbf{m}$, [12], we have $\mathbf{n}_\Sigma = \mathbf{R}_w^c(\mathbf{K}^T\mathbf{m})$, where $\mathbf{K}^T\mathbf{m}$ is the normal vector to Σ in $\langle c \rangle$. Since $\mathbf{n}_\Sigma \perp \wp(\mathbf{t}_w^c)$, it is then sufficient to rotate $\mathbf{n}_\Sigma$ of -90° around the z-axis in order to obtain $\mathbf{t}_w^{c\,*}$ in (23). To prove part (ii) of the statement, consider the following Sylvester equation derived from (14),

$$\mathbf{H}_w^c \mathbf{H}_{v_2}^{v_1} = \mathbf{D}^{[2]}\mathbf{D}^{[1]}\mathbf{H}_w^c, \tag{25}$$

where $\mathbf{D}^{[1]}$, $\mathbf{D}^{[2]}$ are as in (16). Let $\mathbf{R}_w^c$ be given (computed, e.g. using Prop. 8) and $\mathbf{D}^{[2]}$, $\mathbf{n}_2$ be estimated using one of the algorithms in Sect. IV-A. Let $\mathbf{H}_E$ also be given (cf. Prop. 5). From (25) we have $\mathbf{R}_w^c \mathbf{t}_{v_2}^{v_1} + \mathbf{t}_w^c = \mathbf{S}^{[2]}\mathbf{S}^{[1]}\mathbf{t}_w^c$. Collecting $\mathbf{t}_w^c$ on the left-hand side of this equation, we get $(\mathbf{I} - \mathbf{S}^{[2]}\mathbf{S}^{[1]})\,\mathbf{t}_w^c = -\mathbf{R}_w^c \mathbf{t}_{v_2}^{v_1}$, from which, after few maniputations, we obtain (24). Note that the normalization of $\mathbf{t}_w^{c\,*}$ in (22), removes the ambiguity due to the up to scale estimation of $\mathbf{t}_{v_2}^{v_1}$. ■

V. Simulation and Experimental Results

A. Simulations

Simulation experiments have been performed with the Epipolar Geometry Toolbox [17]. The setup consists of a pinhole camera with calibration matrix,

$$\mathbf{K} = \begin{bmatrix} 951.8 & 0 & 640.66 \\ 0 & 951.8 & 605.11 \\ 0 & 0 & 1 \end{bmatrix}, \tag{26}$$

and two planar mirrors with normal vectors $\mathbf{n}_1 = [0 \ -1 \ 0]^T$ and $\mathbf{n}_2 = [-0.866 \quad 0.5 \quad 0]^T$, corresponding to an angular displacement $\theta = 60^\circ$. The camera observes directly 20 random points: the matching is supposed to be solved by manually selecting the correspondences. The rigid-motion between the camera and world frame is $\mathbf{R}_w^c = \mathbf{R}_z(40^\circ)\,\mathbf{R}_y(-20^\circ)\,\mathbf{R}_x(30^\circ)$ and $\mathbf{t}_w^c = [0.6 \ -2 \ -0.2]^T$ m. In Figs. 6(a)-(c), we ran 100 iterations of the M-estimator [18], (used to estimate the fundamental matrix), for increasing values of the image pixel noise standard deviation. Fig. 6(a) reports the mirror angle estimation error $|\widehat{\theta} - \theta|$ for the two algorithms discussed in Sect. IV-A. The estimation with the virtual cameras epipoles is more accurate: the mean error corresponding to the maximum noise power is about 1.5°. Fig. 6(b) shows the $\mathbf{R}_w^c$ estimation error (cf. Prop. 8) computed as the roll-pitch-yaw angle error. As we can see in the figure, the proposed method is robust to noisy data: in fact the maximum mean error is around 2°. Fig. 6(c) finally reports the error $\varepsilon(\widehat{\mathbf{t}}_w^c) \triangleq \|d(\wp(\widehat{\mathbf{t}}_w^c)) - \ d(\wp(\mathbf{t}_w^c))\|$, for the two algorithms in Prop. 9: (i) *Screw axis* and (ii) *Sylvester equation*. From Fig. 6(c), we note that the second method is less sensitive to noise since it does not rely on the direct computation of the epipoles.

	$\mathbf{R}_w^c$ [deg.]			$\mathbf{t}_w^c$ [m]
	Roll	*Pitch*	*Yaw*	
a	−4.00	−3.03	−37.07	$[0.147 \ -0.824 \ \ 0.377]^T$
b	8.16	−3.49	−43.13	$[0.157 \ -0.824 \ \ 0.610]^T$
c	13.99	−20.87	−40.57	$[0.432 \ -0.960 \ \ 0.532]^T$
d	−6.38	−9.01	−51.60	$[0.106 \ -0.613 \ \ 0.408]^T$
e	−3.63	−3.46	−43.73	$[0.215 \ -0.824 \ \ 0.552]^T$
f	7.74	−13.55	−35.60	$[0.313 \ -0.960 \ \ 0.535]^T$

TABLE I

Displacements considered in Figs. 6(e) and 6(f).

B. Experiments

In order to test the proposed methods in real scenarios we performed a series of experiments using the setup shown in Fig. 1, where a Lumenera® LU071C camera mounted on the end-effector of a 6 axes KUKA® KR 3 manipulator, observes a structured scene composed by several objects mirrored by two planar mirrors. The calibration matrix $\mathbf{K}$ is the same as in (26). The observed objects are in a box of $[-0.085, 0.34] \times [-0.15, -0.44]$ m in $\langle w \rangle$. Fig. 6(d) shows the estimation error $|\widehat{\theta} - \theta|$ of the algorithms in Sect. IV-A, when $\theta \in \{30^\circ, 45^\circ, 60^\circ\}$. The values reported in Fig. 6(d) are the average of 5 experiments we conducted for each θ. In accordance with the simulation results, the method based on the virtual cameras epipoles is more accurate, with a mean error of about 6°. Fig. 6(e) shows the $\mathbf{R}_w^c$ estimation error, computed as the roll-pitch-yaw angle error, for the rotational displacements reported in Table I. The mirror angle $\theta = 60^\circ$. As pointed out in Sect. V-A, the proposed algorithm has a reduced sensitivity to noise: in fact, the maximum error is about 5.5°. Finally, Fig. 6(f) depicts the estimation error $\varepsilon(\widehat{\mathbf{t}}_w^c)$ for the translational displacements reported in Table I. As in Sect. V-A, the algorithm based on the Sylvester equation exhibits the best performance.

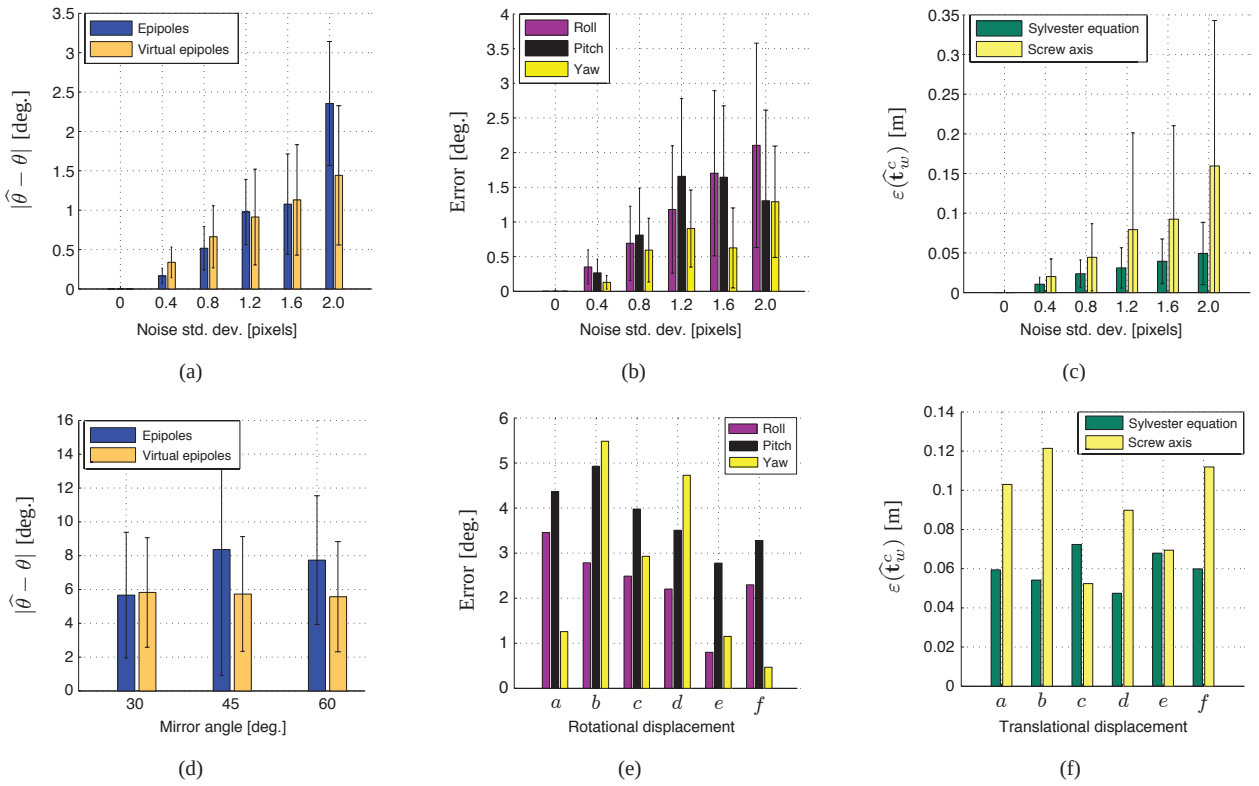

Fig. 6. **Simulations**: (a) Mirror angle estimation error for the algorithms in Sect IV-A; (b) $\mathbf{R}_w^c$ estimation error for the algorithm in Prop. 8; (c) Error $\varepsilon(\widehat{\mathbf{t}}_w^c)$ for the algorithms in Prop. 9. **Experiments**: (d) Mirror angle estimation error for $\theta \in \{30°, 45°, 60°\}$; (e) $\mathbf{R}_w^c$ estimation error for the rotational displacements reported in Table I; (f) Error $\varepsilon(\widehat{\mathbf{t}}_w^c)$ for the translational displacements reported in Table I.

VI. Conclusions and Future Work

The paper presents a systematic study of the imaging geometry of *Planar Catadioptric Stereo* vision sensors (PCS). New algebraic and geometric properties are presented in the cases of static and moving pinhole cameras (multi-view PCS). On the basis of these new results, we propose a suite of algorithms for mirror calibration and camera localization using image data solely.

The extension of our results to sensors composed by multiple cameras and multiple planar mirrors is a subject of on-going research. Future investigations will also study the connection between the imaging geometry of two mirrors PCS and trifocal geometry [16].

Acknowledgement

The authors would like to thank Francesco Chinello for his help in the preparation of the experimental setup.

References

[1] Y. Ma, S. Soatto, J. Košecká, and S.S. Sastry. *An Invitation to 3-D Vision: From Images to Geometric Models*. Interdisciplinary Applied Mathematics. Springer, 2003.

[2] J. Konrad. View Reconstruction for 3-D Video Entertainment: Issues, Algorithms and Applications. In *Proc. IEEE Int. Conf. Image Proc. and its Appl.*, volume 1, pages 8–12, 1999.

[3] S.K. Nayar. Catadioptric Omnidirectional Camera. In *Proc. IEEE Conf. Comp. Vis. and Pattern Rec.*, pages 482–488, 1997.

[4] C. Geyer and K. Daniilidis. Catadioptric Camera Calibration. In *Proc. 7th IEEE Int. Conf. Comp. Vis.*, volume 1, pages 398–404, 1999.

[5] T. Svoboda and T. Pajdla. Epipolar Geometry for Central Catadioptric Cameras. *Int. J. Comput. Vision*, 49(1):23–37, 2002.

[6] G.L. Mariottini and D. Prattichizzo. Image-based Visual Servoing with Central Catadioptric Camera. *Int. J. Robot. Res.*, 27(1):41–57, 2008.

[7] W. Hong, A. Yang Yang, K. Huang, and Y. Ma. On Symmetry and Multiple-View Geometry: Structure, Pose, and Calibration from a Single Image. *Int. J. Comput. Vision*, 60(3):241–265, 2004.

[8] S.A. Nene and S.K. Nayar. Stereo with Mirrors. In *Proc. 6th IEEE Int. Conf. Comp. Vis.*, pages 1087–1094, 1998.

[9] H. Mitsumoto, S. Tamura, K. Okazaki, N. Kajimi, and Y. Fukui. 3-D Reconstruction using Mirror Images based on a Plane Symmetry Recovering Method. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(9):941–946, 1992.

[10] T.P. Pachidis and J. Lygouras. A Pseudo Stereo Vision System as a Sensor for Real-Time Path Control of a Robot. In *Proc. 19th IEEE Conf. Instr. and Measurem. Tech.*, volume 2, pages 1589–1594, 2002.

[11] T. E. Boult and S. Nayar. *A Handy Stereo Camera*. University of Columbia and Colorado, `http://vast.uccs.edu/~tboult/VSAM`, 2000.

[12] J. Gluckman and S. K. Nayar. Planar Catadioptric Stereo: Geometry and Calibration. In *Proc. IEEE Conf. Comp. Vis. and Pattern Rec.*, volume 1, pages 22–28, 1999.

[13] R.K. Kumar, A. Ilie, J.-M. Frahm, and M. Pollefeys. Simple Calibration of Non-overlapping Cameras with a Mirror. In *Proc. IEEE Conf. Comp. Vis. and Pattern Rec.*, pages 1–7, 2008.

[14] J. Gluckman and S.K. Nayar. Rectified Catadioptric Stereo Sensors. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(2):224–236, 2002.

[15] H.-H.P. Wu. Rectification of Stereoscopic Video for Planar Catadioptric Stereo Systems. *IEEE Trans. Circ. Syst. Video Tech.*, 17(6):686–698, 2007.

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

[17] G.L. Mariottini and D. Prattichizzo. EGT for Multiple View Geometry and Visual Servoing: Robotics Vision with Pinhole and Panoramic Cameras. *IEEE Robot. Autom. Mag.*, 12(4):26–39, 2005.

[18] X. Armangué and J. Salvi. Overall View Regarding Fundamental Matrix Estimation. *Image Vision Comput.*, 21(2):205–220, 2003.