IPSJ Transactions on Computer Vision and Applications Vol.8 11–19 (Feb. 2016)

[DOI: 10.2197/ipsjtcva.8.11]

**Regular Paper**

# Mirror-based Camera Pose Estimation Using an Orthogonality Constraint

Kosuke Takahashi[1,†1,a)] Shohei Nobuhara[1,b)] Takashi Matsuyama[1,c)]

Received: April 9, 2015, Accepted: November 19, 2015

***Abstract:*** This paper is aimed at employing mirrors to estimate relative posture and position of camera, i.e., extrinsic parameters, against a 3D reference object that is not directly visible from the camera. The key contribution of this paper is to propose a novel formulation of extrinsic camera calibration based on orthogonality constraint which should be satisfied by all families of mirror-reflections of a single reference object. This allows us to obtain a larger number of equations which contribute to make the calibration more robust. We demonstrate the advantages of the proposed method in comparison with a state-of-the-art by qualitative and quantitative evaluations using synthesized and real data.

***Keywords:*** camera calibration, mirror, householder transformation, orthogonality constraint

## 1. Introduction

Determining the camera position and rotation against 3D reference objects is called as extrinsic camera calibration, and it has been a fundamental research field in computer vision for long years [9], [24]. This technique is widely used for a variety of applications using cameras, such as 3D shape reconstruction from multi-view images [1], [13], vision-based robot navigation [4], [5], augmented reality [3], and so on. Conventional extrinsic calibration techniques have a fundamental assumption that the reference object, such as a chessboard, can be observed by a camera directly. In other words, the reference object is located in the field of view of the camera.

However, there are some cases where this assumption does not hold, for example in calibrating the camera and the screen mounted on a same laptop computer, in calibrating the omnidirectional camera consisting of several cameras each of which does not share its field of view with the others, and so on. Hirayama et al. [11] estimate the interests of users who are watching a digital signage. They use the geometric relationship between the users' gaze estimated from captured image and contents on the display which is located out of a camera's field of view. As another example, Funk et al. [6] use a display-camera system as a controllable surface light source for 3D reconstruction. It uses changes in the appearance of an object under varying surface light source, i.e., display. For such cases, some studies have proposed algorithms using mirrors [2], [10], [12], [14], [15], [18], [19], [20]. These algorithms observe reflections of the reference object in the mirrors and estimate the extrinsic parameters from them. The applications of mirror-based extrinsic camera calibrations are not limited to solve the field-of-view issue only. One advantage of using mirror is to introduce a virtual camera which strictly has the identical intrinsic parameter and is synchronized to the original camera. This advantage is available for 3D shape reconstruction from a single camera and multiple mirrors [12], [23].

In this paper, we propose a novel algorithm of the mirror-based extrinsic camera calibration. In our configuration, we use one stationary camera, some reference points as a reference object and a planar mirror with no markers on it (**Fig. 1**). Our method calibrates the extrinsic parameter from 2D projections of more than two known 3D reference points observed via a mirror under more than two different unknown poses.

At the heart of our work is to introduce an orthogonality constraint which should be satisfied by all families of reflections of a single reference object into our formulation. Based on this constraint, we established a *mirrored-points-based* formulation to compute the extrinsic parameters, while conventional studies can be categorized as *mirrored-camera-based* methods. This *mirrored-points-based* formulation allows us to obtain a larger number of constraints which contribute to make the calibration more robust. Evaluations in Section 5 demonstrate the advantage of our method qualitatively and quantitatively.

An earlier version of this study was presented in CVPR 2012 [21]. Compared with Ref. [21], this paper provides a general computation scheme for PnP ($n \geq 3$) configurations, while

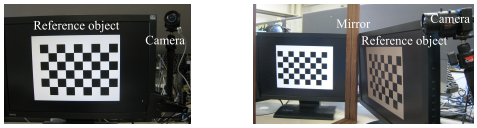


**Fig. 1** The set up of mirror-based extrinsic calibration.

[1] Graduate School of Informatics, Kyoto University, Kyoto 606–8501, Japan
[†1] Presently with NTT Media Intelligence Laboratories, Nippon Telegraph and Telephone Corporation
[a)] takahashi.kosuke@lab.ntt.co.jp
[b)] nob@vision.kuee.kyoto-u.ac.jp
[c)] tm@vision.kuee.kyoto-u.ac.jp

Ref. [21] is designed only for the $n = 3$ case. This generalization allows a practical calibration scenario where $n > 3$ points are detected from regular chess patterns. The improvement utilizing $n > 3$ points is evaluated using real datasets, in comparison with state-of-the-arts [10], [19], [20].

The rest of this paper is organized as follows. Section 2 provides a review on conventional techniques and clarifies the contribution of this paper. Section 3 describes the measurement model of extrinsic calibration and the orthogonality constraint used in our algorithm. Section 4 introduces the details of our algorithm. Section 5 provides qualitative and quantitative evaluations using synthesized and real data to demonstrate the advantages of our method against the state-of-the-art. Section 6 provides the discussions on the limitations of our method and Section 7 concludes this paper.

## 2. Related Work

Mirror-based calibration algorithms without using direct observations of 3D reference objects can be categorized into two major groups in terms of the assumptions on the mirror poses: (1) known poses, or (2) unknown poses.

The methods in the first group estimate the mirror poses explicitly, for example, from the clues located on the mirror plane, such as markers [12], [15]. Jang et al. [12] attached some markers on the mirror plane and calibrated the mirror poses from vanishing points estimated from the markers.

On the other hand, the methods in the second group estimate the mirror poses implicitly, from the reflections of the reference object for several mirror poses without any markers [10], [14], [19], [20]. Our proposed method uses a planar mirror without any markers, and is categorized into this second group.

As shown in **Table 1**, the conventional methods in the second group can be categorized into two subgroups: (2a) *mirrored-camera-based* approach [10], [14], [19], [20] or (2b) *mirrored-points-based* approach. This is based on whether the method models the mirrors duplicate the camera or the reference points.

The methods in (2a) utilize the reflection of the real camera. They model that virtual cameras reflected by mirrors observe the real reference points. These methods in (2a) can further be categorized into two subgroups in terms of whether they utilize the relationship of mirrored cameras or not.

Rodrigues et al. [19] and Sturm et al. [20] are categorized into the former subgroup. They focus on the fact that the planar motion between pairs of mirrored cameras can be described as a fixed-axis rotation with the intersection of the two mirror planes as the rotation axis. Rodrigues et al. recovered extrinsic parameters in a single estimation step by solving a system of linear equations. They propose two types of systems of linear equations.

**Table 1** Number of minimum mirror poses and the category.

| | Mirrors | Category |
|---|---|---|
| Kumar et al. [14] | 5 | mirrored camera |
| Sturm and Bonfort [20] | 3 | mirrored camera |
| Rodrigues et al. [19] | 3 | mirrored camera |
| Hesch et al. [10] | 3 | mirrored camera |
| Proposed | 3 | mirrored point |

While the first linear system has 4 unknown parameters, there are $3N_m$ constraints. In the second linear system, there are $3 + N_m$ unknown parameters and $4N_m$ constrains. However, Rodrigues et al. remark that more than 2 mirror poses are needed in order to estimate extrinsic parameters uniquely in their paper [19].

On the other hand, Kumar et al. [14] and Hesch et al. [10] do not use the relationship of mirrored cameras. Kumar et al. [14] obtained the camera poses by solving linear equations derived from orthogonality relations between the axes of the real camera and the ones of each mirrored camera. These equations have 15 unknown parameters describing the camera poses, while one mirror provides 3 equations. Hence they required at least 5 mirror poses.

Hesch et al. [10]'s scenario is equivalent to a P3P problem [8] in which a mirrored camera lies behind the mirror and observes the true points (not the reflections). Firstly, they compute the rotation matrix and translation vector between the reference object and the mirrored camera. Subsequently, they obtain the axis vectors, which are the intersection vectors of each pair of mirror planes, and compute the rotation matrix between the reference object and the real camera from them. Finally, they obtain the linear equation in order to estimate the translation vector between the reference object and the real camera and solve it. While these linear equations have 6 unknown parameters, they obtain 9 constraints from 3 mirror poses and the number of constraints do not increase depending on the number of mirror poses. Since Hesch et al. assumed that they only use the three reference points, they solved P3P problem and obtained 4 solutions per mirror pose, yielding $4^{N_m}$ possible combinations. Then they computed the extrinsic parameters for each combination and select the best parameters which minimize the reprojection error as a post-processing.

Our method is categorized into (2b). Firstly we introduce a new orthogonality constraint which should be satisfied by all families of reflections of a single reference object. Then we utilize it to estimate the axis vectors of each pair of mirror planes and obtain the mirror normals from them. Finally, we simultaneously obtain all rotation matrices and translation vectors by solving a large system of linear equations. In the case of $N_m$ mirror poses and $N_p$ reference points, this system has $12 + N_m$ unknown parameters and $3 \times N_m \times N_p$ constraints. While the number of constraints in most of mirrored camera-based approach depends on the number of mirrored camera, which is equal to the number of mirror poses, the number of constraints in *mirrored-points-based* approach depends on the number of reference points in addition to the number of mirror poses. Therefore, *mirrored-points-based* approach allows us to obtain a larger number of constraint equations which contribute to make the calibration more robust compared to the mirrored camera based approach.

## 3. Measurement Model

In this section, we introduce our notations, measurement model and an orthogonality constraint on reflections.

### 3.1 Notations and Measurement Model

As illustrated by **Fig. 2**, we denote a camera by $C$ and a mirror

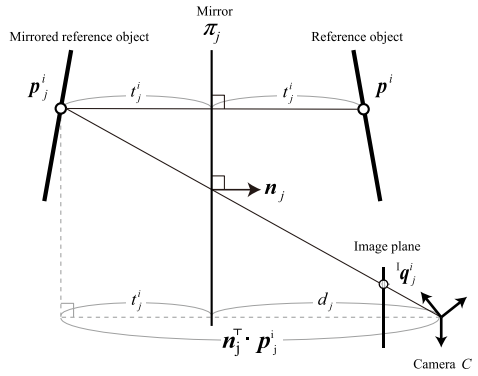


**Fig. 2** The measurement model.

pose by $\pi_j$ ($j = 1, \cdots, N_m$). We use $\{C\}$ to describe the camera $C$ coordinate system which is used as the world coordinate system in this paper. A vector $\boldsymbol{p}$ in the $Y$ coordinate system is expressed as ${}^Y\boldsymbol{p}$, while we may omit $Y$ if it is clear from the context.

Let ${}^X\boldsymbol{p}^i = (x_i, y_i, z_i)^\top$ ($i = 1, \cdots, N_p$) denote the positions of the reference points given a priori in its local coordinate system $X$. These positions are modeled as located at

$$\boldsymbol{p}^i = \boldsymbol{R} \cdot {}^X\boldsymbol{p}^i + \boldsymbol{T} \quad (i = 1, \cdots, N_p), \tag{1}$$

in $\{C\}$ with a rotation matrix $\boldsymbol{R}$ and a translation vector $\boldsymbol{T}$. The reflection of the $i$th reference point $\boldsymbol{p}^i$ mirrored by $\pi_j$ appears as $\boldsymbol{p}^i_j$ in $\{C\}$. These mirrored reference points are projected to the image screen of the real camera $C$ as ${}^I\boldsymbol{q}^i_j$. We model each mirror $\pi_j$ by its normal vector $\boldsymbol{n}_j$ and its distance $d_j$ from the camera $C$. The distance $t^i_j$ from the mirror $\pi_j$ to $\boldsymbol{p}^i_j$ is equal to the distance from $\pi_j$ to $\boldsymbol{p}^i$ by definition. The goal of the extrinsic calibration is to estimate $\boldsymbol{R}$ and $\boldsymbol{T}$ from projected reference points ${}^I\boldsymbol{q}^i_j$.

From Fig. 2, the relation between the reference point and its reflection is expressed as,

$$\boldsymbol{p}^i = 2t^i_j\boldsymbol{n}_j + \boldsymbol{p}^i_j. \tag{2}$$

The distance $t^i_j$ and $d_j$ also satisfy

$$t^i_j + d_j = -\boldsymbol{n}^\top_j \cdot \boldsymbol{p}^i_j. \tag{3}$$

By removing $t^i_j$ from these two equations, we have

$$\boldsymbol{p}^i = -2(\boldsymbol{n}^\top_j \cdot \boldsymbol{p}^i_j + d_j)\boldsymbol{n}_j + \boldsymbol{p}^i_j. \tag{4}$$

Equation (4) shows a relation between a reference point $\boldsymbol{p}^i$ and its reflection $\boldsymbol{p}^i_j$, known as *Householder transformation*.

In addition, by removing $\boldsymbol{p}^i$ from Eqs. (1) and (4), we obtain

$$\boldsymbol{R} \cdot {}^X\boldsymbol{p}^i + \boldsymbol{T} = -2(\boldsymbol{n}^\top_j \cdot \boldsymbol{p}^i_j + d_j)\boldsymbol{n}_j + \boldsymbol{p}^i_j. \tag{5}$$

This is the fundamental equation which describes our measurement model.

### 3.2 Orthogonality Constraint on Mirror Reflections

Consider a reference point $\boldsymbol{p}^i$ and its two mirrored points $\boldsymbol{p}^i_j$, $\boldsymbol{p}^i_{j'}$ by two different mirror planes $\pi_j$ and $\pi_{j'}$ respectively. The axis vector $\boldsymbol{m}_{jj'}$ lying along the intersection of the two mirror planes is expressed as the cross product of each mirror normals, $\boldsymbol{m}_{jj'} = \boldsymbol{n}_j \times \boldsymbol{n}_{j'}$. This axis vector $\boldsymbol{m}_{jj'}$ satisfies the following orthogonality constraint [20] (**Fig. 3**),

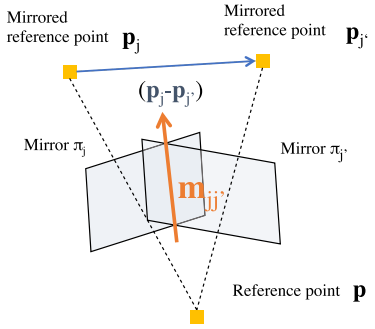


**Fig. 3** A figure illustrating orthogonality constraint.

$$(\boldsymbol{p}_j - \boldsymbol{p}_{j'})^\top \cdot \boldsymbol{m}_{jj'} = 0. \tag{6}$$

This is the key constraint of this paper. The next section provides our algorithm to estimate the extrinsic parameters which utilize this constraint.

## 4. Extrinsic Camera Calibration Using Orthogonality Constraint

In this section, we introduce our method which analytically determines the camera extrinsic parameter from the projections of $N_p$ reference points observed via $N_m$ different mirror poses.

Algorithm 1 shows an overview of our calibration algorithm. Firstly, we solve the PnP problem from mirrored reference points projected to the image plane, ${}^I\boldsymbol{q}^i_j$, and obtain their 3D positions $\boldsymbol{p}^i_j$ by Ref. [21] ($N_p = 3$) or EPnP [17] ($N_p > 3$). Notice that the "handedness" of the extrinsic parameters obtained by solving PnP with mirrored reference points are flipped. However this does not affect the 3D position $\boldsymbol{p}^i_j$, and hence we ignore such flipped extrinsic parameters.

Secondly, based on the orthogonality constraint in the previous section, we estimate the axis vectors of each pair of mirror planes and obtain the mirror normals from them. Finally, we compute $\boldsymbol{R}$ and $\boldsymbol{T}$ by solving a large system of linear equations.

### 4.1 Computing the Axis Vector from Mirror Planes

As described in Section 3.2, the axis vector $\boldsymbol{m}_{jj'}$ ($j, j' = 1, \cdots, N_m, \quad j \neq j'$) and two mirrored points $\boldsymbol{p}^i_j$, $\boldsymbol{p}^i_{j'}$ ($i = 1, 2, \cdots, N_p$) satisfy the orthogonality constraint (Eq. (6)).

By applying this orthogonality constraint to $N_p$ mirrored reference points $\boldsymbol{p}^i_j$, we obtain:

$$\begin{pmatrix} (\boldsymbol{p}^1_j - \boldsymbol{p}^1_{j'})^\top \\ (\boldsymbol{p}^2_j - \boldsymbol{p}^2_{j'})^\top \\ \vdots \\ (\boldsymbol{p}^{N_p}_j - \boldsymbol{p}^{N_p}_{j'})^\top \end{pmatrix} \boldsymbol{m}_{jj'} = \boldsymbol{Q}_{jj'}\boldsymbol{m}_{jj'} = 0. \tag{7}$$

An axis vector $\boldsymbol{m}_{jj'}$ can be computed as the right-singular vector corresponding to the smallest singular value of $\boldsymbol{Q}_{jj'}$.

### 4.2 Computing the Normal Vector of a Mirror Plane

The axis vector $\boldsymbol{m}_{jj'}$ is perpendicular to the normal vectors $\boldsymbol{n}_j$ and $\boldsymbol{n}_{j'}$ of each mirror planes $\pi_j$ and $\pi_{j'}$ respectively. That is,

$$\begin{aligned} \boldsymbol{n}^\top_j \cdot \boldsymbol{m}_{jj'} &= 0, \\ \boldsymbol{n}^\top_{j'} \cdot \boldsymbol{m}_{jj'} &= 0. \end{aligned} \tag{8}$$

**Algorithm 1** An overview of extrinsic camera calibration using $N_m$ mirror poses and $N_p$ reference points

**Input:** ${}^l\boldsymbol{q}_j^i$ $(i = 1, \cdots, N_p, \quad j = 1, \cdots, N_m)$
**Output:** $\boldsymbol{R}, \boldsymbol{T}$
**for each** $N_m$ mirror poses $\pi_j$, $j = 1, \cdots, N_m$ **do**
Solve the PnP and obtain the mirrored points $\boldsymbol{p}_j^i$ by Ref. [21] ($N_p = 3$) or EPnP [17] ($N_p > 3$).
**end for**
**for each** $N_m$ mirror poses $\pi_j, \pi_{j'}, (j, j' = 1, \cdots, N_m, \quad j \neq j')$ **do**
Compute the axis vector $\boldsymbol{m}_{jj'}$ as the right-singular vector corresponding to the smallest singular value of $\boldsymbol{Q}_{jj'}$ in Eq. (7).
**end for**
**for each** $N_m$ mirror poses $\pi_j, (j = 1, \cdots, N_m)$ **do**
Compute the normal vector $\boldsymbol{n}_j$ of the mirror plane $\pi_j$ by Eq. (9).
**end for**
Compute the each column of rotation matrix $\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{r}_3$ and the translation vector by solving Eq. (10).
Refine $\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{r}_3$ by solving the orthogonal Procrustes problem [7].
Refine $\boldsymbol{R}, \boldsymbol{T}$ by applying non-linear optimization [22]

When using $N_m$ mirror poses, we obtain $(N_m - 1)$ equations of Eq. (8) for one normal vector $\boldsymbol{n}_j$. By collecting these equations, we have

$$\begin{aligned} &\boldsymbol{S}_j \boldsymbol{n}_j = 0, \\ &\boldsymbol{S}_j = (\boldsymbol{m}_{j1}\boldsymbol{m}_{j2}\cdots\boldsymbol{m}_{jj-1}\boldsymbol{m}_{jj+1}\cdots\boldsymbol{m}_{jN_m})^\top \end{aligned} \tag{9}$$

where $\boldsymbol{S}_j$ is a $(N_m - 1) \times 3$ matrix. A normal vector $\boldsymbol{n}_j$ can be computed as the right-singular vector corresponding to the smallest singular value of $\boldsymbol{S}_j$.

This equation also indicates that we have to provide $N_m \geq 3$ mirror poses in order to estimate $\boldsymbol{n}_j$, because the degree of freedom of $\boldsymbol{n}_j$ is 2.

### 4.3 Computing Extrinsic Parameters

Up to this point, we obtain the 3D positions of mirrored reference points $\boldsymbol{p}_j^i$ $(i = 1, \cdots, N_p, j = 1, \cdots, N_m)$ and mirror normals $\boldsymbol{n}_j$. The 3D positions of reference points ${}^X\boldsymbol{p}^i = (x_i, y_i, z_i)$ are supposed to be given a priori in its local coordinate system $X$. By substituting these known parameters into Eq. (5), we can derive a large system of linear equations:

$$\boldsymbol{A}\boldsymbol{Z} = \boldsymbol{B}. \tag{10}$$

where

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{I}_3 & 2\boldsymbol{n}_1 & \boldsymbol{0}_{3\times1} & \cdots & \boldsymbol{0}_{3\times1} & x_1\boldsymbol{I}_3 & y_1\boldsymbol{I}_3 & z_1\boldsymbol{I}_3 \\ \boldsymbol{I}_3 & 2\boldsymbol{n}_1 & \boldsymbol{0}_{3\times1} & \cdots & \boldsymbol{0}_{3\times1} & x_2\boldsymbol{I}_3 & y_2\boldsymbol{I}_3 & z_2\boldsymbol{I}_3 \\ & & & \vdots & & & & \\ \boldsymbol{I}_3 & 2\boldsymbol{n}_1 & \boldsymbol{0}_{3\times1} & \cdots & \boldsymbol{0}_{3\times1} & x_{N_p}\boldsymbol{I}_3 & y_{N_p}\boldsymbol{I}_3 & z_{N_p}\boldsymbol{I}_3 \\ \boldsymbol{I}_3 & \boldsymbol{0}_{3\times1} & 2\boldsymbol{n}_2 & \cdots & \boldsymbol{0}_{3\times1} & x_1\boldsymbol{I}_3 & y_1\boldsymbol{I}_3 & z_1\boldsymbol{I}_3 \\ \boldsymbol{I}_3 & \boldsymbol{0}_{3\times1} & 2\boldsymbol{n}_2 & \cdots & \boldsymbol{0}_{3\times1} & x_2\boldsymbol{I}_3 & y_2\boldsymbol{I}_3 & z_2\boldsymbol{I}_3 \\ & & & \vdots & & & & \\ \boldsymbol{I}_3 & \boldsymbol{0}_{3\times1} & 2\boldsymbol{n}_2 & \cdots & \boldsymbol{0}_{3\times1} & x_{N_p}\boldsymbol{I}_3 & y_{N_p}\boldsymbol{I}_3 & z_{N_p}\boldsymbol{I}_3 \\ & & & \vdots & & & & \\ \boldsymbol{I}_3 & \boldsymbol{0}_{3\times1} & \boldsymbol{0}_{3\times1} & \cdots & 2\boldsymbol{n}_{N_m} & x_1\boldsymbol{I}_3 & y_1\boldsymbol{I}_3 & z_1\boldsymbol{I}_3 \\ \boldsymbol{I}_3 & \boldsymbol{0}_{3\times1} & \boldsymbol{0}_{3\times1} & \cdots & 2\boldsymbol{n}_{N_m} & x_2\boldsymbol{I}_3 & y_2\boldsymbol{I}_3 & z_2\boldsymbol{I}_3 \\ & & & \vdots & & & & \\ \boldsymbol{I}_3 & \boldsymbol{0}_{3\times1} & \boldsymbol{0}_{3\times1} & \cdots & 2\boldsymbol{n}_{N_m} & x_{N_p}\boldsymbol{I}_3 & y_{N_p}\boldsymbol{I}_3 & z_{N_p}\boldsymbol{I}_3 \end{bmatrix}, \tag{11}$$

$$\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{T}^\top & d_1 & d_2 & \cdots & d_{N_m} & \boldsymbol{r}_1^\top & \boldsymbol{r}_2^\top & \boldsymbol{r}_3^\top \end{bmatrix}^\top, \tag{12}$$

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}_1 & \boldsymbol{B}_2 & \cdots & \boldsymbol{B}_{N_m} \end{bmatrix}^\top, \tag{13}$$

$$\boldsymbol{B}_j = \begin{bmatrix} \boldsymbol{b}_j^1 & \boldsymbol{b}_j^2 & \cdots & \boldsymbol{b}_j^{N_p} \end{bmatrix}, \tag{14}$$

$$\boldsymbol{b}_j^i = (-2\boldsymbol{n}_j^\top \boldsymbol{p}_j^i \boldsymbol{n}_j + {}^C\boldsymbol{p}_j^i)^\top \tag{15}$$

The vectors $\boldsymbol{r}_1$, $\boldsymbol{r}_2$ and $\boldsymbol{r}_3$ denote the first, second and third column of the rotation matrix $\boldsymbol{R}$. From $N_m$ mirror poses and $N_p$ reference points, we have $12+N_m$ unknown parameters and $3\times N_m\times N_p$ equations. Hence, when $N_m \geq 3$ and $3 \times N_m \times N_p > 12 + N_m$, we can solve the Eq. (10) by $\boldsymbol{Z} = \boldsymbol{A}^*\boldsymbol{B}$, where $\boldsymbol{A}^*$ is the pseudo-inverse matrix of $\boldsymbol{A}$.

In case of that reference points are on a single plane, the 3D position of reference points in its local coordinate system can be expressed as ${}^X\boldsymbol{p}^i = (x_i, y_i, 0)^\top$ and we cannot compute the third column vector $\boldsymbol{r}_3$ of the rotation matrix. In this case, we compute $\boldsymbol{r}_3$ as the cross product of first and second column vector $\boldsymbol{r}_1$, $\boldsymbol{r}_2$, that is $\boldsymbol{r}_3 = \boldsymbol{r}_1 \times \boldsymbol{r}_2$.

### 4.4 Linear Refinement of Rotation Matrix by Solving the Orthogonal Procrustes Problem

Now we obtain columns of rotation matrix $\boldsymbol{r}_1$, $\boldsymbol{r}_2$, $\boldsymbol{r}_3$ and translation vector $\boldsymbol{T}$ linearly, but $\boldsymbol{r}_1$, $\boldsymbol{r}_2$ and $\boldsymbol{r}_3$ do not necessarily satisfy the following constraints as a rotation matrix due to noise:

$$\begin{aligned} &|\boldsymbol{r}_1| = |\boldsymbol{r}_2| = |\boldsymbol{r}_3| = 1, \\ &\boldsymbol{r}_1^\top\boldsymbol{r}_2 = \boldsymbol{r}_2^\top\boldsymbol{r}_3 = \boldsymbol{r}_3^\top\boldsymbol{r}_1 = 0. \end{aligned} \tag{16}$$

Here, we solve the orthogonal Procrustes problem [7] and obtain a rotation matrix which satisfies Eq. (16) and is closest to the original linear solution as proposed in Zhang's method [24]. That is $\boldsymbol{R} = \boldsymbol{U}\boldsymbol{V}^\top$, where $\boldsymbol{U}$ and $\boldsymbol{V}$ are given by as the SVD of the original matrix $(\boldsymbol{r}_1 \quad \boldsymbol{r}_2 \quad \boldsymbol{r}_3) = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$.

### 4.5 Non-linear Refinement of Extrinsic Parameters

In general, obtained extrinsic parameters can be refined by non-linear optimization [22]. Here, we minimize following reprojection error function,

$$\boldsymbol{E}_{opt} = \sum_{j=1}^{N_m}\sum_{i=1}^{N_p} |\boldsymbol{q}_j^i - \breve{\boldsymbol{q}}_j^i(\boldsymbol{R}, \boldsymbol{T}, \boldsymbol{n}_j, d_j)|, \tag{17}$$

where $\breve{\boldsymbol{q}}_j^i(\boldsymbol{R}, \boldsymbol{T}, \boldsymbol{n}_j, d_j)$ denote the reprojected point calculated

from estimated parameters. We solved this non-linear optimization problem of Eq. (17) with Levenberg-Marquardt algorithm.

## 5. Evaluations

In this section, we show experimental results using synthesized and real data in order to evaluate the performance of our method quantitatively and qualitatively. In both cases, we compare our method with state-of-the-arts proposed by Sturm et al. [20] and by Rodrigues et al. [19] with non-linear refinement.

### 5.1 Evaluations Using Synthesized Data

#### 5.1.1 Experiment Environment

To synthesize data, we used the following experiment setup by default. The matrix of intrinsic parameters, $\boldsymbol{K}$, consists of $(fx, fy, cx, cy)$; $fx$ and $fy$ represents the focal length in pixels, and $cx$ and $cy$ represent the 2D coordinates of the principle point. We set them to $(500, 500, 300, 250)$ in this evaluation respectively.

The normal vectors $\boldsymbol{n}_j$ $(j = 1, \cdots, N_m)$ of mirror poses $\pi_j$ are set to $(\sin\theta_z \sin\theta_x + \cos\theta_x \cos\theta_z \sin\theta_y, \sin\theta_x \cos\theta_z + \cos\theta_x \sin\theta_z \sin\theta_y, \cos\theta_x \cos\theta_y)$ where $\theta_k$ $(k = x, y, z)$ is the angle respect to each axis, and drawn randomly within the ranges of $(-20 \le \theta_x \le 20,\ 160 \le \theta_y \le 200,\ -20 \le \theta_z \le 20)$. The distance between each mirror plane and camera center was set to 300 mm.

The reference object consists of $N_p$ reference points forming a grid pattern and the distance between each reference point is 50 mm. The center of $X$ is located at the centroid of these points.

We represent the ground truth of rotation matrix as a product of three elemental rotation matrices, that is $\boldsymbol{R} = \boldsymbol{R}_1(\theta_1)\boldsymbol{R}_2(\theta_2)\boldsymbol{R}_3(\theta_3)$, and we set random values to each angles $\theta_1$, $\theta_2$ and $\theta_3$ within $[-10 : 10]$ respectively. The position $\boldsymbol{T}$ is generated of each trial by assigning a random value within $[-5 : 5]$ to each $x$, $y$ and $z$ element of $\boldsymbol{T}$.

In this experiment, we evaluate the performance of each method under various conditions of the following parameters.

**(a)** $\sigma$: the standard deviation of Gaussian pixel noise of zero-mean.
**(b)** $N_p$: the number of reference points.
**(c)** $N_m$: the number of mirror poses.

**Table 2** describes the min, max, increment step and default value of the parameters. We computed the average of the estimation errors of 100 trials for each of combinations. While changing these parameters respectively, the other parameters are set to values in Default column, that is the minimum setup for Sturm et al. [20] and Rodrigues et al. [19]. As to the comparison with $N_p = 3$ and $N_m = 3$ configuration, please see Ref. [21].

#### 5.1.2 Error Metrics

Throughout this evaluation, we used the following metrics to measure the performance of the calibration methods.

The estimation error of $\boldsymbol{R}$ is defined as the Riemannian distance [16]:

**Table 2** The range of changing parameters.

| Parameter | Min | Max | Step | Default |
|---|---|---|---|---|
| $\sigma$ | 0 | 2 | 0.1 | 1 |
| $N_p$ | 4 | 20 | 1 | 4 |
| $N_m$ | 3 | 20 | 1 | 3 |

$$E_R = \frac{1}{\sqrt{2}} \| \mathrm{Log}(\boldsymbol{R}^\top \boldsymbol{R}_g) \|_F \tag{18}$$

$$\mathrm{Log}\,\boldsymbol{R}' = \begin{cases} 0 & (\theta = 0), \\ \frac{\theta}{2\sin\theta}(\boldsymbol{R}' - \boldsymbol{R}'^\top) & (\theta \neq 0), \end{cases} \tag{19}$$

where $\theta = \cos^{-1}\left(\frac{\mathrm{Tr}(\boldsymbol{R}')-1}{2}\right)$.

The estimation error of $\boldsymbol{T}$ is defined as the RMS error:

$$E_T = \sqrt{|\boldsymbol{T} - \boldsymbol{T}_g|^2/3}. \tag{20}$$

The reprojection error is defined as follows:

$$E_P = \frac{1}{N_m \times N_p} \sum_{j=1}^{N_m} \left( \sum_{i=1}^{N_p} (\breve{\boldsymbol{q}}_j^i - \boldsymbol{q}_j^i) \right), \tag{21}$$

where $\boldsymbol{q}_j^i$ is the observation and $\breve{\boldsymbol{q}}_j^i$ is the reprojected point calculated from estimated parameters.

#### 5.1.3 Results

**Figure 4** shows results for different standard deviations $\sigma$ of pixel noise added to the observations. The averages of each estimation error of proposed method are smaller than those of Sturm et al. [20] and Rodrigues et al. [19]. This fact indicates that our method can estimate better initial values from same inputs, because of the larger number of constraints involved in the estimation. Notice that there exist some trials in each of which all the methods result in a same optimal value regardless of the differences between initial values returned by their linear methods. Besides, there exists some trials where all the methods fall in local minima. This is the reason for the spikes.

**Figure 5** shows results for difference number of reference points. We added the results of Hesch et al. [10], which is method for $N_p = 3$ scenario, as a reference. From these results, we can observe that the number of reference points affects the performance of each method drastically. These improvements are thought to be due to the improvement of estimation of mirrored reference point by PnP. In fact, Noguer et al. [17] shows that the result of PnP in increasing number of points follows similar pattern of this experiments in their paper.

**Figure 6** shows results in changing the number of mirror poses $N_m$. While Sturm's method does not improve with increasing number of mirror pose, we can see that proposed method and Rodrigues's method improves. This is considered to be due to the scalability of formulation for mirror pose, that is the number of equations for estimating extrinsic parameters in proposed method and Rodrigues's method changes depends on the number of mirror poses.

These results prove that proposed method works robustly with observation noise and has the scalability for the number of reference points and mirror poses.

### 5.2 Evaluations with Real Data

#### 5.2.1 Experiment Environment

We evaluated the performance of our proposed method with real data assuming calibration of digital-signage as introduced in Section 1. **Figure 7** shows an overview of the setup for calibration. We used two cameras (Pointgrey Flea3) $C_1$ and $C_2$, a 20-inch flat panel display and a sputtering mirror. The goal of this

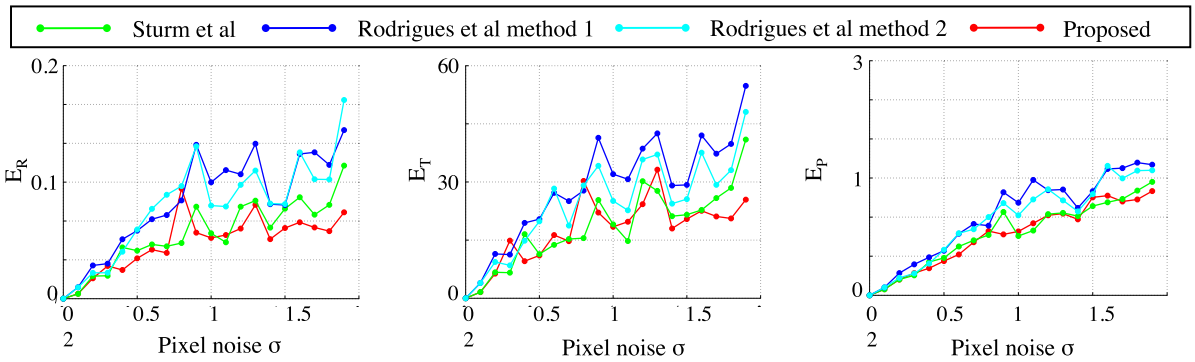


**Fig. 4** Estimation error of each parameter in changing the standard deviation $\sigma$ of pixel noise added to the input.

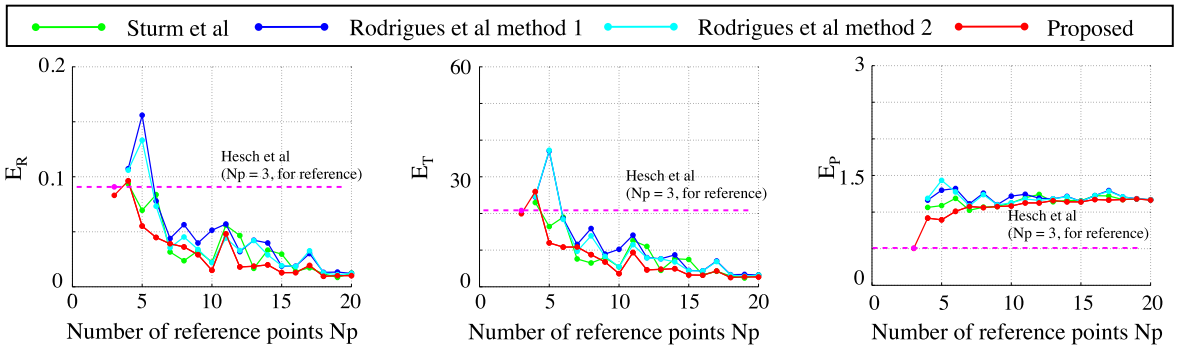


**Fig. 5** Estimation error of each parameter in changing the number of reference points $N_p$.

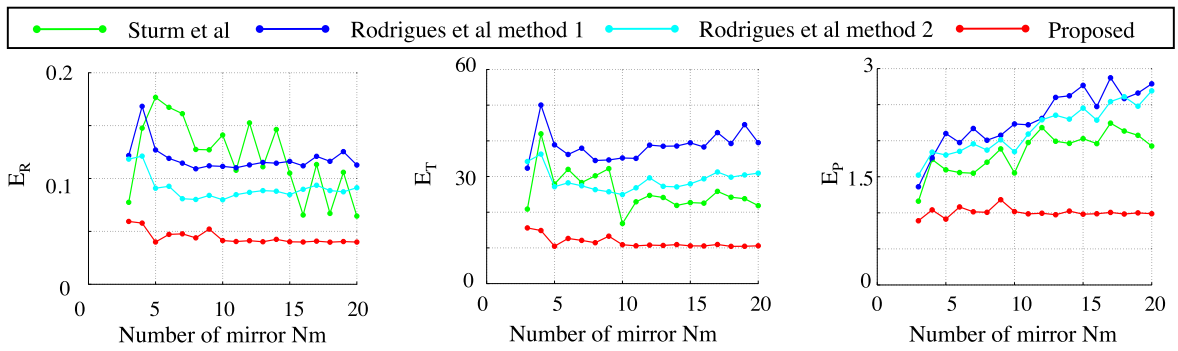


**Fig. 6** Estimation error of each parameter in changing the number of mirror poses $N_m$.

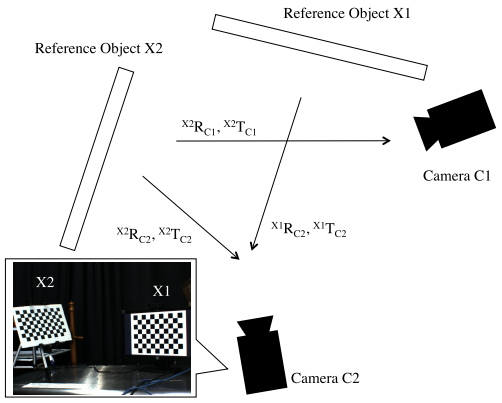


**Fig. 7** Environment for calibration.

experiment is to calibrate the extrinsic parameters of $C_1$ against a $7 \times 10$ chess pattern $X_1$ rendered in the display. Notice that we used only $N_p$ reference points for calibration. The length between each reference point is 82.5 mm. $C_1$ is located where it cannot observe $X_1$ directly. It captures $N_m$ UXGA images of different mirror poses $\pi_j$ $(j = 1, \cdots, N_m)$ for calibration.

#### 5.2.2 Baseline Calibration

In order to evaluate the performance with real data, we prepared a baseline extrinsic parameter of camera $C_1$ against $X_1$. We locate $C_2$ where it can observe both $X_1$ and the mirror directly. Using the observation of $X_1$, we can obtain the extrinsic parameter of $C_2$ against $X_1$ by Zhang's method [24]. In addition, by attaching another chess pattern $X_2$ on the mirror, we can obtain the extrinsic parameters of $C_1$ and $C_2$ against $X_2$ for evaluation.

As illustrated by Fig. 7, suppose ${}^{X_l}\boldsymbol{R}_{C_k}$, ${}^{X_l}\boldsymbol{T}_{C_k}$ denote the rotations and translations from $X_l$ to $C_k$ respectively. That is, the 3D position of a 3D point ${}^{X_l}\boldsymbol{p}^i$ of $X_l$ in $C_k$ is given by

$$ {}^{C_k}\boldsymbol{p}^i = {}^{X_l}\boldsymbol{R}_{C_k} \cdot {}^{X_l}\boldsymbol{p}^i + {}^{X}\boldsymbol{T}_{C_k} \ (k = 1, 2, \ l = 1, 2). \tag{22} $$

We can calibrate ${}^{X_1}\boldsymbol{R}_{C_1}$, ${}^{X_1}\boldsymbol{T}_{C_1}$ by the proposed method as well as by Refs. [19] and [20]. In addition, by calibrating ${}^{X_1}\boldsymbol{R}_{C_2}$, ${}^{X_1}\boldsymbol{T}_{C_2}$, ${}^{X_2}\boldsymbol{R}_{C_1}$, ${}^{X_2}\boldsymbol{T}_{C_1}$, ${}^{X_2}\boldsymbol{R}_{C_2}$ and ${}^{X_2}\boldsymbol{T}_{C_2}$ by Zhang's method, these param-

eters also provide ${}^{X_1}\boldsymbol{R}_{C_1}$, ${}^{X_1}\boldsymbol{T}_{C_1}$ as a baseline calibration:

$$\begin{aligned} {}^{X_1}\boldsymbol{R}_{C_1} &= {}^{X_2}\boldsymbol{R}_{C_1}\,{}^{X_2}\boldsymbol{R}_{C_2}^{\top}\,{}^{X_1}\boldsymbol{R}_{C_2} \\ {}^{X_1}\boldsymbol{T}_{C_1} &= {}^{X_2}\boldsymbol{R}_{C_1}\,{}^{X_2}\boldsymbol{R}_{C_2}^{\top}({}^{X_1}\boldsymbol{T}_{C_2} - {}^{X_2}\boldsymbol{T}_{C_2}) + {}^{X_2}\boldsymbol{T}_{C_1}. \end{aligned} \tag{23}$$

### 5.2.3 Results

**Figure 8** renders the estimated positions of the reference object by each method with $N_p = 4$ and $N_m = 3$ configuration. We can see that the reference objects estimated by each method are located near the baseline result. This precision is acceptable for applications using display-camera system such as gaze detection for digital-signage.

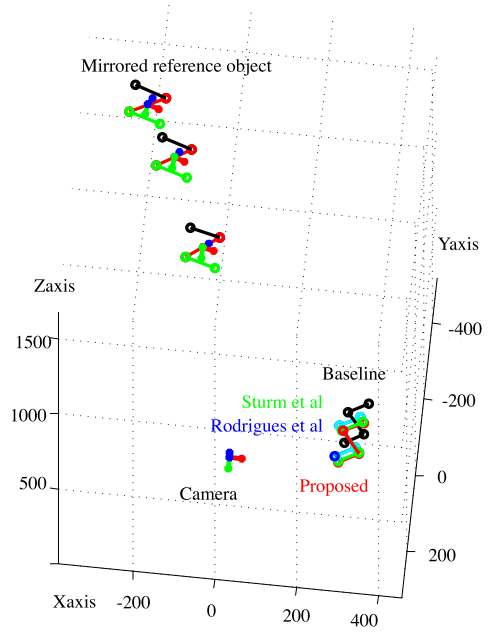


**Fig. 8** Estimated positions of the reference object by the proposed method (red), by Ref. [19] (blue and cyan), by Ref. [20] (green) and by Eq. (23) (black). Notice that this figure renders estimate positions of $\boldsymbol{p}$ and $X_1$ in $\{C\}$, and therefore $C_1$ is located at $(0,0,0)^{\top}$.

**Figures 9** and **10** shows results in changing the number of reference points $N_p$ and the number of mirror poses $N_m$ respectively. Notice that the estimated parameters by each method are almost identical, and therefore we can observe only one line in Fig. 10.

From these results, we can observe that our method performs better than conventional methods [10], [19], [20] in real situation qualitatively and quantitatively.

## 6. Discussion

### 6.1 Degenerate Case

Our algorithm does not work if it cannot compute enough axis vectors $\boldsymbol{m}_{jj'}$ for estimating mirror normals. This happens in the following three cases. (1) If two mirrors are parallel, then the intersection of them does not exist and therefore not be computable. (2) If all the mirror planes intersect at single axis in 3D, the mirror normals cannot be computable by solving Eq. (9). These (1) and (2) cases has been originally observed by Sturm et al. [20]. (3) If reference points and the intersection of two mirrors $\pi_j$ and $\pi_{j'}$ are on a same plane, the axis vector $\boldsymbol{m}_{jj'}$ is not be computable by solving Eq. (7) though $\boldsymbol{m}_{jj'}$ does exist physically because of the following reason which makes the two rows of $M'_{jj'}$ corresponding to $\pi_j$ and $\pi_{j'}$ be linearly dependent.

**Proposition 6.1.** *If two reference points $\boldsymbol{p}^i$ and $\boldsymbol{p}^{i'}$ and the intersection of two mirrors $\pi_j$ and $\pi_{j'}$ are on a same plane, the two lines connecting results of different Householder transformations of $\boldsymbol{p}^i$ and $\boldsymbol{p}^{i'}$, i.e., the lines connecting $\boldsymbol{p}^i_j$ to $\boldsymbol{p}^i_{j'}$ and $\boldsymbol{p}^{i'}_j$ to $\boldsymbol{p}^{i'}_{j'}$, are parallel* (**Fig. 11**).

*Proof.* Suppose the line connecting $\boldsymbol{p}^i$ and $\boldsymbol{p}^{i'}$ intersects with the intersection of the two mirrors at $\boldsymbol{O}$ as shown in Fig. 11. By def-

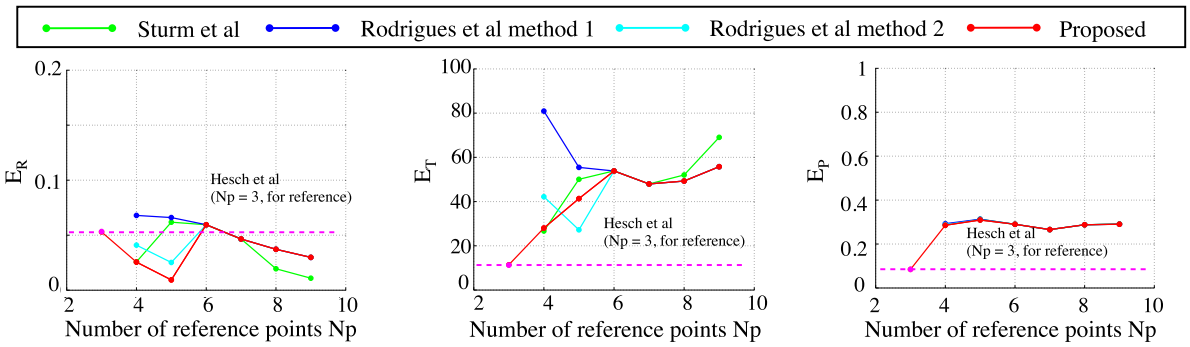


**Fig. 9** Estimation error of each parameter with real data in changing the number of reference points $N_p$.

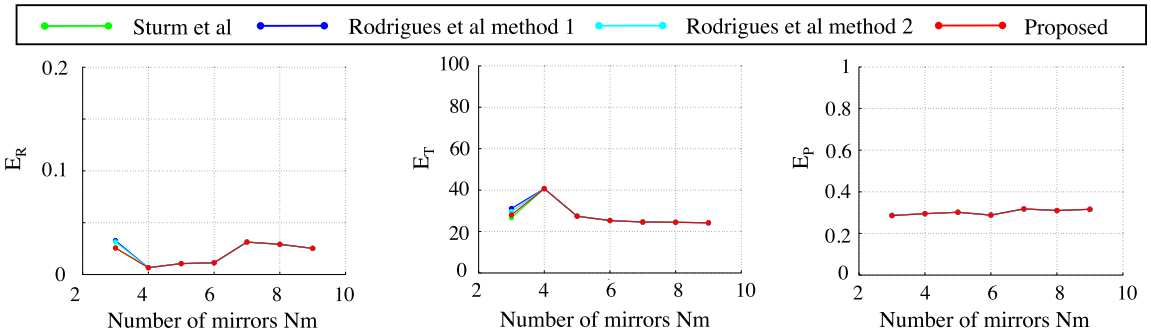


**Fig. 10** Estimation error of each parameter with real data in changing the number of mirror poses $N_m$.

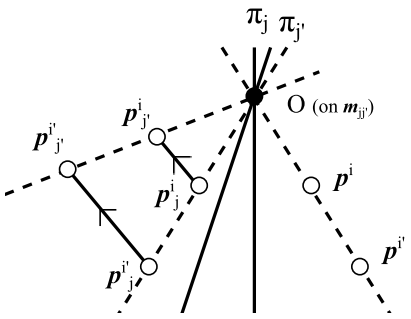


**Fig. 11** Degenerate case.

inition of the reflection, the distance from $\boldsymbol{p}^i$ to $\boldsymbol{O}$ is equal to the one from $\boldsymbol{p}^i_j$ to $\boldsymbol{O}$. Similarly, the distance from $\boldsymbol{p}^{i'}$ to $\boldsymbol{O}$ is equal to the one from $\boldsymbol{p}^{i'}_j$ to $\boldsymbol{O}$. Also, these distances are equal to the ones from $\boldsymbol{p}^i_{j'}$ to $\boldsymbol{O}$, and from $\boldsymbol{p}^{i'}_{j'}$ to $\boldsymbol{O}$ respectively. Here $\triangle \boldsymbol{O}\boldsymbol{p}^i_j\boldsymbol{p}^i_{j'}$ and $\triangle \boldsymbol{O}\boldsymbol{p}^{i'}_j\boldsymbol{p}^{i'}_{j'}$ are isosceles triangles sharing the apex $\angle \boldsymbol{p}^i_j\boldsymbol{O}\boldsymbol{p}^i_{j'}$. Therefore, the two lines $\boldsymbol{p}^i_j$ to $\boldsymbol{p}^i_{j'}$ and $\boldsymbol{p}^{i'}_j$ to $\boldsymbol{p}^{i'}_{j'}$ are parallel. □

These three degenerate cases can be detected by observing the rank of $M_{jj'}$ in Eq. (7). If the rank is less than 2, we can discard the mirror pair and try with more mirrored images in practice.

### 6.2 Sufficiency of the Orthogonality Constraint

The orthogonality constraint holds for two reflections of a single reference point and the axis vector, as a necessary condition. Obviously, this does not constrain the position of the mirror (the parameter $d$ of Eq. (5)). That is a mirror of another distance satisfies Eq. (6) as long as it has the same intersection direction.

This fact indicates that the orthogonality constraint itself does not serve as the sufficient condition to determine all of the mirror parameters. Instead, given three mirrors, it becomes the sufficient condition to obtain the mirror normals uniquely as described in Sections 4.1 and 4.2. Using the estimated normals, we can define the linear equations (Eq. (10)) based on the measurement model (Eq. (5)).

## 7. Conclusion

In this paper, we proposed a new algorithm to extrinsically calibrate a camera to a 3D reference object that is not directly visible from the camera. We introduced an orthogonality constraint which should be satisfied by all families of reflections of a same reference object and established a *mirrored-points-based* formulation. This formulation allow us to obtain a larger number of constraints which contribute to make the calibration more robust even with a simple configuration, that is using fewer reference points and fewer mirror poses. The evaluations of the extrinsic calibration by synthesized and real data showed our improvement on the accuracy and robustness against state-of-the-arts quantitatively and qualitatively.

**Acknowledgments** This study is partially supported by JSPS Kakenhi 25540068.

### References

[1] Agarwal, S., Furukawa, Y., Snavely, N., Curless, B., Seitz, S.M. and Szeliski, R.: Reconstructing Rome, *IEEE Computer*, Vol.43, pp.40–47 (2010).
[2] Agrawal, A.: Extrinsic Camera Calibration without a Direct View Using Spherical Mirror, *Proc. ICCV*, pp.2368–2375 (2013).
[3] Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S. and MacIntyre, B.: Recent advances in augmented reality, *IEEE Computer Graphics and Applications*, Vol.21, No.6, pp.34–47 (2001).
[4] Desouza, G. and Kak, A.: Vision for mobile robot navigation: A survey, *TPAMI*, Vol.24, No.2, pp.237–267 (2002).
[5] Fraundorfer, F., Scaramuzza, D. and Pollefeys, M.: A constricted bundle adjustment parameterization for relative scale estimation in visual odometry, *Proc. ICRA*, pp.1899–1904 (2010).
[6] Funk, N. and Yang, Y.-H.: Using a Raster Display for Photometric Stereo, *4th Canadian Conference on Computer and Robot Vision* (*CRV '07*), pp.201–207 (2007).
[7] Golub, G. and van Loan., C.: *Matrix Computations*, The Johns Hopkins University Press, Baltimore, Maryland, 3rd edition (1996).
[8] Haralick, B.M., Lee, C.-N., Ottenberg, K. and Nölle, M.: Review and analysis of solutions of the three point perspective pose estimation problem, *IJCV*, Vol.13, pp.331–356 (1994).
[9] Hartley, R.I. and Zisserman, A.: *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2nd edition (2004).
[10] Hesch, J.A., Mourikis, A.I. and Roumeliotis, S.I.: *Algorithmic Foundation of Robotics VIII*, Springer Tracts in Advanced Robotics, Vol.57, chapter Mirror-Based Extrinsic Camera Calibration, pp.285–299, Springer-Verlag (2009).
[11] Hirayama, T., Dodane, J.-B., Kawashima, H. and Matsuyama, T.: Estimates of user interest using timing structures between proactive content-display updates and eye movements, *IEICE Trans. Inf. Syst.*, Vol.93, No.6, pp.1470–1478 (2010).
[12] Jang, K.H., Lee, D.H. and Jung, S.K.: A moving planar mirror based approach for cultural reconstruction: Research Articles, *Comput. Animat. Virtual Worlds*, Vol.15, pp.415–423 (2004).
[13] Kanade, T., Rander, P. and Narayanan, P.: Virtualized reality: Constructing virtual worlds from real scenes, *IEEE Multimedia*, Vol.4, No.1, pp.34–47 (1997).
[14] Kumar, R., Ilie, A., Frahm, J.-M. and Pollefeys, M.: Simple calibration of non-overlapping cameras with a mirror, *Proc. CVPR*, pp.1–7 (2008).
[15] Lébraly, P., Deymier, C., Ait-Aider, O., Royer, E. and Dhome, M.: Flexible extrinsic calibration of non-overlapping cameras using a planar mirror: Application to vision-based robotics, *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems* (*IROS*), pp.5640–5647, IEEE (2010).
[16] Moakher, M.: Means and Averaging in the Group of Rotations, *SIAM J. Matrix Anal. Appl.*, Vol.24, pp.1–16 (2002).
[17] Moreno-Noguer, F., Lepetit, V. and Fua, P.: Accurate Non-Iterative O(n) Solution to the PnP Problem, *Proc. ICCV*, pp.1–8 (2007).
[18] Nayar, S.: Catadioptric omnidirectional camera, *Proc. CVPR*, pp.482–488 (1997).
[19] Rodrigues, R., Barreto, P. and Nunes, U.: Camera Pose Estimation Using Images of Planar Mirror Reflections, *Proc. ECCV*, pp.382–395 (2010).
[20] Sturm, P. and Bonfort, T.: How to Compute the Pose of an Object without a Direct View, *Proc. ACCV*, pp.21–31 (2006).
[21] Takahashi, K., Nobuhara, S. and Matsuyama, T.: A new mirror-based extrinsic camera calibration using an orthogonality constraint, *Proc. CVPR*, pp.1051–1058 (2012).
[22] Triggs, B., McLauchlan, P., Hartley, R. and Fitzgibbon, A.: Bundle Adjustment — A Modern Synthesis, *Vision Algorithms:* Triggs, B., Zisserman, A. and Szeliski, R. (eds.), *Theory and Practice*, Lecture Notes in Computer Science, Vol.1883, pp.298–372, Springer Berlin Heidelberg (2000).
[23] Ying, X., Peng, K., Ren, R. and Zha, H.: Geometric properties of multiple reflections in catadioptric camera with two planar mirrors, *Proc. CVPR*, pp.1126–1132 (2010).
[24] Zhang, Z.: A flexible new technique for camera calibration, *TPAMI*, pp.1330–1334 (2000).

**Kosuke Takahashi** received his B.Sc. degree in engineering and M.Sc. in informatics from Kyoto University, Japan, in 2010 and 2012, respectively. He is currently a researcher at NTT Media Intelligence Laboratories. His research interest includes computer vision. He received "Best Open Source Code" award Second Prize in CVPR 2012.

**Shohei Nobuhara** received his B.Sc. in Engeneering, M.Sc. and Ph.D. in Informatics from Kyoto University, Japan, in 2000, 2002, and 2005 respectively. Since 2010, he has been a senior lecturer at Kyoto University. His research interest includes computer vision and 3D video. He is a member of IPSJ, IEICE, and IEEE.

**Takashi Matsuyama** received his B.Eng., M.Eng., and D.Eng. degrees in electrical engineering from Kyoto University, Japan, in 1974, 1976, and 1980, respectively. He is currently a professor in the Department of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University. He served as the director of the Academic Center for Computing and Media Studies (2002–2006), the director general of the Institute for Information Management and Communications (2005–2010), and a vice president of Kyoto University (2008–2010). He has been studying cooperative distribute sensing-control-reasoning systems over 30 years. Their application fields include knowledge-based image understanding, visual surveillance, 3D video, human-computer interaction, and smart energy management. He wrote more than 100 journal papers and more than 20 books including three research monographs: A Structural Analysis of Complex Aerial Photographs, PLENUM, 1980, SIGMA: A Knowledge-Based Aerial Image Understanding System, PLENUM, 1990, and 3D Video and its Applications, Springer, 2012. He organized and served as the leaders of 4 5-years national research projects and won more than 20 best paper awards from Japanese and international academic societies including the Marr Prize at ICCV '95 and the commendation for Science and Technology by the Minister of Education, Culture, Sports, Science and Technology in 2009. He is on the editorial board of the Pattern Recognition Journal. He was awarded Fellowships from the International Association for Pattern Recognition, the Information Processing Society of Japan, and the Institute of Electronics, Information, and Communication Engineers Japan.

(Communicated by *Tat-Jun Chin*)