A New Mirror-based Extrinsic Camera Calibration
Using an Orthogonality Constraint

Kosuke Takahashi   Shohei Nobuhara   Takashi Matsuyama

Graduate School of Informatics, Kyoto University, Japan

{takahasi,nob,tm}@vision.kuee.kyoto-u.ac.jp

Abstract

This paper is aimed at calibrating the relative posture and position, i.e. *extrinsic parameters, of a stationary camera against a 3D reference object which is not directly visible from the camera. We capture the reference object via a mirror under three different unknown poses, and then calibrate the extrinsic parameters from 2D appearances of reflections of the reference object in the mirrors.*

The key contribution of this paper is to present a new algorithm which returns a unique solution of three P3P problems from three mirrored images. While each P3P problem has up to four solutions and therefore a set of three P3P problems has up to 64 solutions, our method can select a solution based on an orthogonality constraint which should be satisfied by all families of reflections of a single reference object. In addition we propose a new scheme to compute the extrinsic parameters by solving a large system of linear equations. These two points enable us to provide a unique and robust solution.

We demonstrate the advantages of the proposed method against a state-of-the-art by qualitative and quantitative evaluations using synthesized and real data.

1. Introduction

Extrinsic camera calibration against 3D reference objects has been a fundamental research field in computer vision for long years [7, 20]. It is widely used for applications using cameras such as 3D shape reconstruction from multi-view images [1, 11, 17], vision-based robot navigation [3,4], augmented reality [2], *etc*.

However standard calibration techniques have a fundamental limitation on the field-of-view of the camera; the camera should observe the 3D reference object directly. For the cases where this condition does not hold, some studies

---
[1]The project web page: http://vision.kuee.kyoto-u.ac.jp/~nob/proj/mirror/
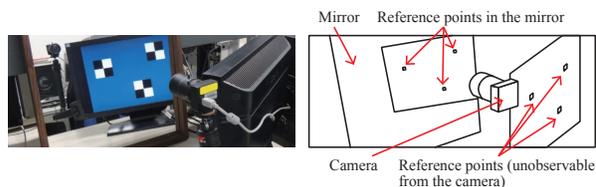
Figure 1. Mirror-based extrinsic calibration (Left: real setup, Right: illustration). The goal of this paper is to calibrate the camera extrinsic parameter against the reference object observed via a mirror under three different unknown poses.

proposed algorithm using mirrors [8, 10, 12, 13, 15, 16, 18]. They observe the reference object via mirrors, and then estimate the extrinsic parameters from the reflections of the reference objects in the mirrors.

The applications of mirror-based extrinsic camera calibrations are not limited to solve the field-of-view issue only. This technique is also useful for 3D shape reconstruction from a single camera and multiple mirrors [10, 19]. Multi-view videos captured by a single camera are definitely synchronized, and share a single intrinsic parameter.

The key contribution of this paper is to present an algorithm which returns a unique solution of three P3P problems [6] from three mirrored images. This is based on an orthogonality constraint which should be satisfied by all families of reflections of a single reference object. Based on this constraint, our method calibrates the extrinsic parameter from 2D projections of three known 3D reference points observed via a mirror under three different unknown poses (Figure 1). To the best of our knowledge, this is the minimum configuration to calibrate a stationary camera from mirrored images. Establishing a better algorithm for such a minimum configuration is important especially in RANSAC scenario since the fewer inputs the algorithm requires, the better RANSAC performes. Compared with a state-of-the-art [8] which also works for this configuration but can select a solution from 64 candidates only by a post-processing in terms of a reprojection error evaluation, our method can produce a better solution uniquely and is robust against ob-

servation noise as demonstrated quantitatively in Section 5.

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 a new orthogonality constraint used in our extrinsic calibration. Section 4 introduces our algorithm, and 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 concludes this paper with discussions on future work.

2. Related Work

Mirror-based calibration algorithms without using direct observation 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 first group measures mirror poses explicitly [10,13]. Jang *et al.* [10] attached markers on the mirror plane to calibrate the mirror poses from vanishing points estimated from the markers.

As shown in Table 1, the second group can be categorized into two subgroups: (2a) calibration from 4 (or more) known 3D reference object points [12, 16, 18] or (2b) calibration from 3 known 3D reference object points [8]. The biggest difference between (2a) and (2b) is whether the camera extrinsic parameters can be uniquely determined or not. This is known as a PnP problem [6]. Given $n \geq 4$ 2D projections of known 3D reference points, we can determine the camera extrinsic parameters uniquely. With $n = 3$ points, we have up to four possible solutions in general.

Our method is categorized as (2b). This configuration is also studied by Hesch *et al.* [8]. They first solved P3P problem and obtained 4 solutions per mirror pose, yielding $64 = 4 \times 4 \times 4$ possible combinations of 3D reflected reference points. Then they computed the extrinsic parameters for each combination. Finally they select a parameter which minimizes the reprojection error as a post-processing.

Compared with [8], we first introduce a new orthogonality constraint which should be satisfied by all families of reflections of a single reference object, and then show that this constraint enables us to select the best combination of 3D reflected reference points from 64 candidates without computing corresponding extrinsic parameters and their reprojection errors explicitly. In addition we propose a new scheme to compute the extrinsic parameters by solving a large system of linear equations. The evaluations in Section 5 demonstrate the advantage of our method qualitatively.

3. Measurement Model and Orthogonality Constraint

This section introduces our notations, measurement model, and an orthogonality constraint on reflections.

Table 1. Number of minimum mirror poses and reference points required.

	Mirrors	Points
Sturm and Bonfort [18]	3	4
Rodorigues *et al.* [16]	3	4
Kumar *et al.* [12]	5	4
Hesch *et al.* [8]	3	3
Proposed	3	3
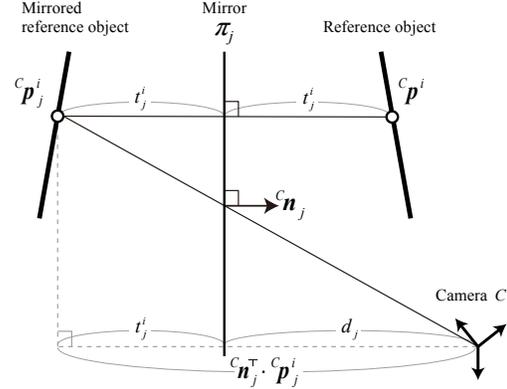
Figure 2. Measurement model

3.1. Notations and Measurement Model

As illustrated by Figure 2, we have a camera C, three planar mirrors π_j ($j = 1, 2, 3$), and three reference 3D points in the scene. Let $^X\boldsymbol{p}^i$ ($i = 1, 2, 3$) denote the positions of the reference points given a priori in its local coordinate system $\{X\}$. Here we use $^Y\boldsymbol{x}$ to describe the 3D position of $\boldsymbol{x}$ in $\{Y\}$. These positions are modeled as located at

$$^C\boldsymbol{p}^i = \boldsymbol{R} \cdot {}^X\boldsymbol{p}^i + \boldsymbol{T} \quad (i = 1, 2, 3), \tag{1}$$

in the camera coordinate system $\{C\}$ with a rotation matrix $\boldsymbol{R}$ and a translation vector $\boldsymbol{T}$. We model that the reflection of the reference point $^C\boldsymbol{p}^i$ mirrored by π_j appears as $^C\boldsymbol{p}^i_j$ in $\{C\}$. This mirrored point is projected to $\boldsymbol{q}^i_j$ in the image screen of C. Each mirror π_j is modeled by its normal vector $\boldsymbol{n}_j$ and its distance d_j from C. The goal of extrinsic calibration is to estimate $\boldsymbol{R}$ and $\boldsymbol{T}$ from $\boldsymbol{q}^i_j$ ($i, j = 1, 2, 3$).

In Figure 2, the distance t^i_j from the mirror π_j to $^C\boldsymbol{p}^i_j$ is equal to the distance from π_j to $^C\boldsymbol{p}^i$ by definition. This relationship is expressed by

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

This distance t^i_j also satisfies

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

By removing t^i_j from these two equations, we obtain

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

Algorithm 1 Extrinsic Calibration from Three Reference Points

---

Input: q_j^i $(i = 1, 2, 3, \quad j = 1, 2, 3)$
Output: R, T
   for each image $j = 1, 2, 3$ **do**
      Solve P3P and obtain four 3D positions of the reference points in mirror π_j: $P_{j_k} = \{{}^C p_{j_k}^i\}$ $(i = 1, 2, 3, k = 1, \ldots, 4)$.
   end for
   for each combination of 1_k, $2_{k'}$, and $3_{k''}$ $(k, k', k'' = 1, \ldots, 4)$ **do**
      Compute the eigenvalues of $M_{jj'}$ (Eq (10)) using P_{1_k}, $P_{2_{k'}}$ and $P_{3_{k''}}$.
   end for
   Select the combination of 1_k, $2_{k'}$ and $3_{k''}$ which minimizes Eq (11). Let P_1, P_2 and P_3 denote corresponding P_{1_k}, $P_{2_{k'}}$ and $P_{3_{k''}}$ hereafter.
   Compute m_{12}, m_{23} and m_{31} as the eigenvectors corresponding to smallest eigenvalues of M_{12}, M_{23} and M_{31} (Eq (10)) using P_1, P_2 and P_3.
   Compute n_1, n_2 and n_3 by Eq (12) using m_{12}, m_{23} and m_{31}.
   Compute R, T by solving Eq (13) using P_1, P_2, P_3, n_1, n_2 and n_3.

This relationship between ${}^C p^i$ and its reflection ${}^C p_j^i$ by mirror π_j is known as *Householder transformation*, and this is the fundamental equation which describes our measurement model.

3.2. Orthogonality Constraint on Mirrored Points

Consider a reference 3D point ${}^C p^i$ and its two mirrored 3D positions ${}^C p_j^i$ and ${}^C p_{j'}^i$ $(j \neq j')$ by two different mirrors π_j and $\pi_{j'}$ respectively. Let $m_{jj'} = n_j \times n_{j'}$ denote the axis vector which lies along the intersection of the two mirror planes. Then $m_{jj'}$ satisfies the following orthogonality constraint.

Proposition 3.1. $m_{jj'}$, the axis vector of mirrors π_j and $\pi_{j'}$ is orthogonal to the vector ${}^C p_j - {}^C p_{j'}$ between results of different Householder transformations (Eq (4)) of a point ${}^C p$. That is,

$$({}^C p_j - {}^C p_{j'})^\top \cdot m_{jj'} = 0. \tag{5}$$

Proof. Eq (4) indicates that (1) the 3D mirrored position of the ith reference point ${}^C p^i$ can be estimated from each of three mirrors π_j and the corresponding mirrored position ${}^C p_j^i$, and (2) the 3D positions estimated from different mirrors should be equal to each other. That is,

$$\begin{aligned}{}^C p^i &= -2(n_j^\top \cdot {}^C p_j^i + d_j)n_j + {}^C p_j^i \\ &= -2(n_{j'}^\top \cdot {}^C p_{j'}^i + d_{j'})n_{j'} + {}^C p_{j'}^i.\end{aligned} \tag{6}$$

The axis vector $m_{jj'}$ is a unit vector that lies along the intersection of the two mirror planes, and hence perpendicular to both mirror plane normals n_j and $n_{j'}$:

$$\begin{aligned} n_j^\top \cdot m_{jj'} &= 0, \\ n_{j'}^\top \cdot m_{jj'} &= 0. \end{aligned} \tag{7}$$

By multiplying $m_{jj'}$ on the right side of Eq (6), we obtain:

$$\begin{aligned} {}^C p_j^{i^\top} \cdot m_{jj'} &= {}^C p_{j'}^{i^\top} \cdot m_{jj'} \\ \Leftrightarrow ({}^C p_j^i - {}^C p_{j'}^i)^\top \cdot m_{jj'} &= 0. \end{aligned} \tag{8}$$

$\square$

The next section describes our extrinsic calibration method which utilizes this proposition.

4. Extrinsic Calibration From Mirrored Images Using the Orthogonality Constraint

In this section we introduce our method which analytically determines the camera extrinsic parameter from projections of three reference 3D object points observed via three different mirror poses.

Algorithm 1 shows an overview of our calibration algorithm. We first solve P3P problem [6] for each camera, and obtain 64 possible combinations of the mirrored reference points. Then we select a single combination out of 64 possibilities by exploiting the orthogonality constraint described in Section 3.2. Once obtained the mirrored reference points, we compute R and T by solving a large system of linear equations.

4.1. Unique Solution of Three P3P Problems Using the Orthogonality Constraint

The original goal of P3P problem is to determine the 3D positions of the three points with respect to the camera frame $\{C\}$, ${}^C p_j^i$ $(i, j = 1, 2, 3)$, by using the corresponding perspective projection q_j^i on the image plane. As is well known, P3P problem has up to four solutions in general. We denote these four candidates of ${}^C p_j^i$ as $P_{j_k} = \{{}^C p_{j_k}^i\}$, $(k = 1, \ldots, 4)$. In our case, we use three different mirror poses and we obtain up to 64 possible combinations.

Here we propose an algorithm to select a unique combination of P_{j_k} $(j = 1, 2, 3)$ as P_1, P_2, and P_3 by using the orthogonality constraint in Section 3.2. By applying the orthogonality constraint for the mirrored positions of three reference points $^C\boldsymbol{p}^i$ $(i = 1, 2, 3)$, we have:

$$\begin{pmatrix} (^C\boldsymbol{p}_j^1 - {}^C\boldsymbol{p}_{j'}^1)^\top \\ (^C\boldsymbol{p}_j^2 - {}^C\boldsymbol{p}_{j'}^2)^\top \\ (^C\boldsymbol{p}_j^3 - {}^C\boldsymbol{p}_{j'}^3)^\top \end{pmatrix} \boldsymbol{m}_{jj'} = \boldsymbol{Q}_{jj'}\boldsymbol{m}_{jj'} = 0. \tag{9}$$

By multiplying $\boldsymbol{Q}_{jj'}^\top$ on the left side of Eq (9), we obtain

$$\boldsymbol{Q}_{jj'}^\top \boldsymbol{Q}_{jj'} \boldsymbol{m}_{jj'} = \boldsymbol{M}_{jj'} \boldsymbol{m}_{jj'} = 0. \tag{10}$$

Since $\boldsymbol{M}_{jj'}$ is a 3×3 positive semidefinite matrix and can be computed from the result of P3P problems, we can compute $\boldsymbol{m}_{jj'}$ as the eigenvector corresponding to the smallest eigenvalue of $\boldsymbol{M}_{jj'}$.

In ideal noiseless cases, the smallest eigenvalue of $\boldsymbol{M}_{jj'}$ should be exactly zero if the 3D points $^C\boldsymbol{p}_j^i$ and $^C\boldsymbol{p}_{j'}^i$ are results of Householder transformation since Eq (8) should hold strictly. This fact allows us to eliminate invalid combinations of P_{1_k}, $P_{2_{k'}}$ and $P_{3_{k''}}$ based on the size of the smallest eigenvalue of $\boldsymbol{M}_{jj'}$; if a combination of P_{1_k}, $P_{2_{k'}}$ and $P_{3_{k''}}$ is invalid in terms of Householder transformation, the value of the left hand side of (9) cannot be zero and hence the smallest eigenvalue is not zero even in ideal noiseless situations.

In the presence of observation noise in practice, we evaluate the size of the smallest eigenvalue by

$$\rho = \lambda_3/(\lambda_1 + \lambda_2 + \lambda_3), \tag{11}$$

where λ_1, λ_2, and λ_3 denote the largest, second, and the smallest eigenvalues of $\boldsymbol{M}_{jj'}$ respectively. Since one combination of P3P solutions has three $\boldsymbol{M}_{jj'}$: $\boldsymbol{M}_{12}$, $\boldsymbol{M}_{23}$ and $\boldsymbol{M}_{31}$, we compute the sum of ρ from these matrices. Using this value, we select P_{1_k}, $P_{2_{k'}}$ and $P_{3_{k''}}$ corresponding to the smallest sum of ρ as P_1, P_2 and P_3 which satisfy Householder transformation best.

Notice that the uniqueness of the combination which makes Eq (9) be zero is not proven theoretically here. However the evaluations by both synthesized and real data (Section 5) practically proves the validity of the solution selected by the above-mentioned process.

4.2. Linear Solution of Rotation and Translation

Up to this point, we obtained the axis vectors $\boldsymbol{m}_{12}$, $\boldsymbol{m}_{23}$ and $\boldsymbol{m}_{31}$ up to signs as the eigenvectors corresponding to the smallest eigenvalues of $\boldsymbol{M}_{12}$, $\boldsymbol{M}_{23}$ and $\boldsymbol{M}_{31}$ respectively. Using these axis vectors, we can compute the normal direction of the mirrors by:

$$\begin{aligned} \boldsymbol{n}_1 &= (\boldsymbol{m}_{12} \times \boldsymbol{m}_{31})/(\| \boldsymbol{m}_{12} \times \boldsymbol{m}_{31} \|), \\ \boldsymbol{n}_2 &= (\boldsymbol{m}_{23} \times \boldsymbol{m}_{12})/(\| \boldsymbol{m}_{23} \times \boldsymbol{m}_{12} \|), \\ \boldsymbol{n}_3 &= (\boldsymbol{m}_{31} \times \boldsymbol{m}_{23})/(\| \boldsymbol{m}_{31} \times \boldsymbol{m}_{23} \|). \end{aligned} \tag{12}$$

Here we can correct the signs of $\boldsymbol{n}_1$, $\boldsymbol{n}_2$ and $\boldsymbol{n}_3$ by exploiting the fact that the mirrors are facing towards the camera, *i.e.* the Z values of these vectors are negative in $\{C\}$, as illustrated by Figure 2.

Using $\boldsymbol{n}_1$, $\boldsymbol{n}_2$ and $\boldsymbol{n}_3$, we can derive a large system of linear equations by removing $^C\boldsymbol{p}^i$ from Eq (1) and (4):

$$\boldsymbol{AZ} = \boldsymbol{B}, \tag{13}$$

where

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{I}_3 & 2\boldsymbol{n}_1 & \boldsymbol{0}_{3\times 1} & \boldsymbol{0}_{3\times 1} & x_1\boldsymbol{I}_3 & y_1\boldsymbol{I}_3 \\ \boldsymbol{I}_3 & 2\boldsymbol{n}_1 & \boldsymbol{0}_{3\times 1} & \boldsymbol{0}_{3\times 1} & x_2\boldsymbol{I}_3 & y_2\boldsymbol{I}_3 \\ \boldsymbol{I}_3 & 2\boldsymbol{n}_1 & \boldsymbol{0}_{3\times 1} & \boldsymbol{0}_{3\times 1} & x_3\boldsymbol{I}_3 & y_3\boldsymbol{I}_3 \\ \boldsymbol{I}_3 & \boldsymbol{0}_{3\times 1} & 2\boldsymbol{n}_2 & \boldsymbol{0}_{3\times 1} & x_1\boldsymbol{I}_3 & y_1\boldsymbol{I}_3 \\ \boldsymbol{I}_3 & \boldsymbol{0}_{3\times 1} & 2\boldsymbol{n}_2 & \boldsymbol{0}_{3\times 1} & x_2\boldsymbol{I}_3 & y_2\boldsymbol{I}_3 \\ \boldsymbol{I}_3 & \boldsymbol{0}_{3\times 1} & 2\boldsymbol{n}_2 & \boldsymbol{0}_{3\times 1} & x_3\boldsymbol{I}_3 & y_3\boldsymbol{I}_3 \\ \boldsymbol{I}_3 & \boldsymbol{0}_{3\times 1} & \boldsymbol{0}_{3\times 1} & 2\boldsymbol{n}_3 & x_1\boldsymbol{I}_3 & y_1\boldsymbol{I}_3 \\ \boldsymbol{I}_3 & \boldsymbol{0}_{3\times 1} & \boldsymbol{0}_{3\times 1} & 2\boldsymbol{n}_3 & x_2\boldsymbol{I}_3 & y_2\boldsymbol{I}_3 \\ \boldsymbol{I}_3 & \boldsymbol{0}_{3\times 1} & \boldsymbol{0}_{3\times 1} & 2\boldsymbol{n}_3 & x_3\boldsymbol{I}_3 & y_3\boldsymbol{I}_3 \end{bmatrix}, \tag{14}$$

$$\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{T}^\top & d_1 & d_2 & d_3 & \boldsymbol{r}_1^\top & \boldsymbol{r}_2^\top \end{bmatrix}^\top, \tag{15}$$

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{b}_1^1 & \boldsymbol{b}_1^2 & \boldsymbol{b}_1^3 & \boldsymbol{b}_2^1 & \boldsymbol{b}_2^2 & \boldsymbol{b}_2^3 & \boldsymbol{b}_3^1 & \boldsymbol{b}_3^2 & \boldsymbol{b}_3^3 \end{bmatrix}^\top, \tag{16}$$

$$\boldsymbol{b}_j^i = (-2\boldsymbol{n}_j^\top{}^C\boldsymbol{p}_j^i\boldsymbol{n}_j + {}^C\boldsymbol{p}_j^i)^\top. \tag{17}$$

In these equations, we used $^X\boldsymbol{p}^i = (x_i, y_i, 0)^\top$ without loss of generality. $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are the first and second column vectors of $\boldsymbol{R}$, that is, $\boldsymbol{R} = (\boldsymbol{r}_1 \, \boldsymbol{r}_2 \, \boldsymbol{r}_3)$.

Since this system has 12 unknowns ($\boldsymbol{Z}$) and 27 constraints (rows of $\boldsymbol{A}$ and $\boldsymbol{B}$), the least-squares solution for $\boldsymbol{Z}$ can be computed by $\boldsymbol{Z} = \boldsymbol{A}^*\boldsymbol{B}$, where $\boldsymbol{A}^*$ is the Moore-Penrose pseudo-inverse of $\boldsymbol{A}$. The third colmun vector of rotation matrix $\boldsymbol{r}_3$ can be computed as follows.

$$\boldsymbol{r}_3 = (\boldsymbol{r}_1 \times \boldsymbol{r}_2)/(\|\boldsymbol{r}_1\| \cdot \|\boldsymbol{r}_2\|), \tag{18}$$

However, the obtained $\boldsymbol{r}_1$, $\boldsymbol{r}_2$ and $\boldsymbol{r}_3$ are not guaranteed to satisfy the following constraints to form a valid rotation matrix:

$$\begin{aligned} | \boldsymbol{r}_1 | = | \boldsymbol{r}_2 | = | \boldsymbol{r}_3 | = 1, \\ \boldsymbol{r}_1^\top \boldsymbol{r}_2 = \boldsymbol{r}_2^\top \boldsymbol{r}_3 = \boldsymbol{r}_3^\top \boldsymbol{r}_1 = 0. \end{aligned} \tag{19}$$

In order to enforce this constraint, here we solve the orthogonal Procrustes problem [5, 20]. Given two matrices $\boldsymbol{R}$ and $\boldsymbol{S}$ in $\mathbb{R}^{3\times 3}$, an orthognal $\hat{\boldsymbol{R}} \in \mathbb{R}^{3\times 3}$ which minimizes the frobenius norm $|\boldsymbol{R} - \boldsymbol{S}\hat{\boldsymbol{R}}|_F$ is given by

$$\hat{\boldsymbol{R}} = \boldsymbol{U}\boldsymbol{V}^\top, \tag{20}$$

where $\boldsymbol{U}$ and $\boldsymbol{V}$ are the left and right matrices of SVD of $\boldsymbol{S}^\top \boldsymbol{R} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$. In our case, by substituiting $\boldsymbol{S} = \boldsymbol{I}_3$, we have $\hat{\boldsymbol{R}}$ which satisfies the above constraint and closest to $\boldsymbol{R}$.
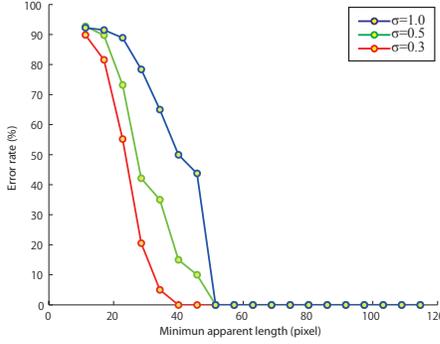
Figure 3. Error rate against apparent observation size. If the reference object is observed smaller than a certain size, the selection process (Section 4.1) can chose a wrong solution.

5. Evaluations

This section shows experiments with synthesized and real data in order to evaluate the performance of our method quantitatively and qualitatively. In the both cases, we compare our method against a state-of-the-art proposed by Hesch *et al.* [8] which works in the same situation: three reference 3D points observed via three different mirror poses. Notice that the original method by Hesch *et al.* [8] first computes up to 64 extrinsic parameters analytically, and select the best one by comparing their reprojection errors. Then it refines the extrinsic parameter by a non-linear optimization process which minimizes the reprojection error. In the comparisons below, we use the extrinsic parameter before the non-liner refinement.

5.1. Synthesized Data

To synthesize the data, we used the following values by default. The intrinsic parameter is set to

$$K = \begin{bmatrix} 500 & 0 & 300 \\ 0 & 500 & 250 \\ 0 & 0 & 1 \end{bmatrix}. \tag{21}$$

The original distance from the camera to mirrors d is set to 300. The mirror poses n_j is set to $(\cos\theta_x \cos\theta_y, \cos\theta_x \sin\theta_y, \sin\theta_z)^\top$ with $(\theta_x, \theta_y, \theta_z) = (0, 150, 0)$ for n_1, $(-30, 180, 0)$ for n_2, and $(0, 210, 0)$ for n_3. The reference points are defined as $^X p^1 = (0,0,0)^\top$, $^X p^2 = (225, 0, 0)^\top$ and $^X p^3 = (0, 225, 0)^\top$. The rotation parameter R of the camera is set to the identity matrix. The position T is generated for each trial by assigning a random value within $[0:20]$ to each x, y and z element of T.

Error Rate Our algorithm selects the best combination of P3P solutions based on the values of Eq (11). If the apparent size of the projected reference object is small against the observation noise level, the quality of P3P solutions become unstable, and hence the value of Eq (11) becomes un-

reliable. As a result, our algorithm can fail to select the best solution from 64 candidates in terms of the reprojection error.

Figure 3 shows how likely the proposed algorithm fails to select the correct solution under different apparent sizes of 2D projections of the 3D reference object via mirror and under different observation noise. Here the apparent size is defined as the minimum distance between 2D projections of the reference points. We added Gaussian noise with zero-mean and standard deviation $\sigma = 0.3, 0.5, 1.0$ to the synthesized 2D observation q_j^i. From this result, we can conclude that our method can work properly if the reference object is observed larger than 50 pixels in presence of Gaussian noise with zero-mean and standard deviation $\sigma = 1$.

Notice that we did not have such failure cases in the evaluations hereafter.

Quantitative Evaluation In order to evaluate the performance quantitatively, we used synthesized data and evaluated the results by comparing estimation errors of R, T, n and the reprojection errors. Here the estimation error of R is defined as the Riemannian distance [14] between R and synthesized ground truth value R_g:

$$E_R = \frac{1}{\sqrt{2}} \parallel \mathrm{Log}(R^\top R_g) \parallel_F \tag{22}$$

$$\mathrm{Log}R' = \begin{cases} 0 & (\theta = 0), \\ \frac{\theta}{2\sin\theta}(R' - R'^\top) & (\theta \neq 0). \end{cases} \tag{23}$$

where $\theta = \cos^{-1}(\frac{trR'-1}{2})$. The estimation error of T against the ground truth T_g is defined as the RMS error:

$$E_T = \sqrt{\|T - T_g\|^2/3}. \tag{24}$$

The estimation error of n_j is defined as the angle against the ground truth $n_{g,j}$:

$$E_n = \frac{1}{3} \sum_{j=1}^{3} \arccos(n_j^\top n_{g,j}). \tag{25}$$

The reprojection error per point is defined as

$$E_P = \frac{1}{3 \times 3} \sum_{j=1,2,3} \left(\sum_{i=1,2,3} |\breve{q}_j^i - q_j^i| \right), \tag{26}$$

where q_j^i is the synthesized 2D observation of ith reference point reflected by jth mirror. $\breve{q}_j^i$ is the corresponding 2D reprojection of the mirrored reference points computed by Eq (1), Eq (4), and the synthesized intrinsic camera parameter K.

Figure 4 shows results of quantitative evaluations. The top row (a) shows the performance against pixel noise. We
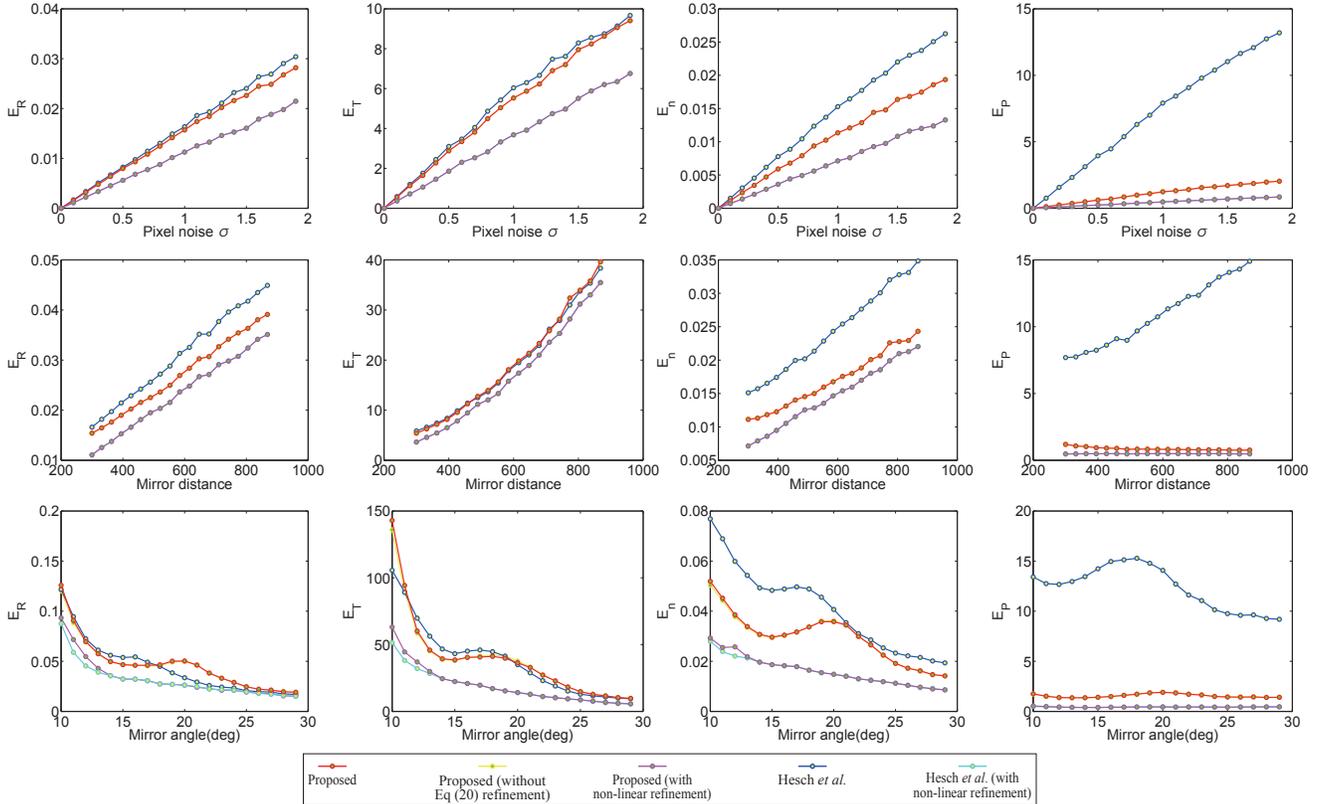
Figure 4. Calibration errors against different (a) pixel noise, (b) distance between the mirrors and camera, and (c) mirror poses. Note that (1) all trials of (b) and (c) have Gaussian pixel noise with zero-mean and standard deviation $\sigma = 1.0$, and (2) the plots of red and yellow are almost overlapped. This applies to the magenta and cyan plots as well.

added Gaussian noise with zero-mean and standard deviation $\sigma = 0, \ldots, 2$ to the synthesized 2D observation $\boldsymbol{q}_j^i$. The middle row (b) shows the performance against different camera-to-mirror distances. We set the distance from 300 to 900. The bottom row (c) shows the performance against different mirror angles. We changed the mirror poses up to ± 30 degree. For (b) and (c) we added Gaussian noise with zero-mean and standard deviation $\sigma = 1.0$. The plots of each row show, from left to right, the average values of E_R, E_T, E_n and E_P over 1000 trials for each noise level, distance, and angle. The plots in magenta and cyan show results by the proposed and [8] followed by the non-linear optimization of reprojection errors. These results quantitatively prove that (1) our method outperforms [8] in terms of reprojection errors, and (2) that improvement is mainly provided by the improvement on $\boldsymbol{n}$. Plus, our result is comparable with the results after non-linear optimization. Even though both non-linear optimizations started from ours and [8] converge to almost same results (the magenta and cyan curves), this fact suggests that our method can provide a better initial value to the optimizer and help it to converge faster. In addition, the refinement by Eq (20) does not contribute much in terms of these estimation accuracies. In this sense, this
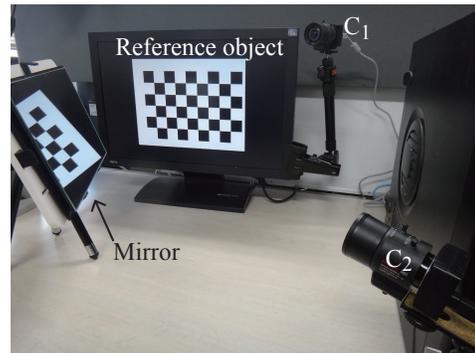
Figure 5. Calibration setup

refinement can be regarded just as a numerical process to ensure $\boldsymbol{R}$ be a rotation matrix, not to improve the result.

5.2. Real Data

Figure 5 shows an overview of the setup for calibration. We have two cameras (PointGrey Chameleon CMLN01352C) C_1 and C_2, a flat-panel display and a mirror in the scene. The goal is to calibrate the extrinsic parameter of C_1 against a 5×8 chess pattern X_1 rendered in the dis-
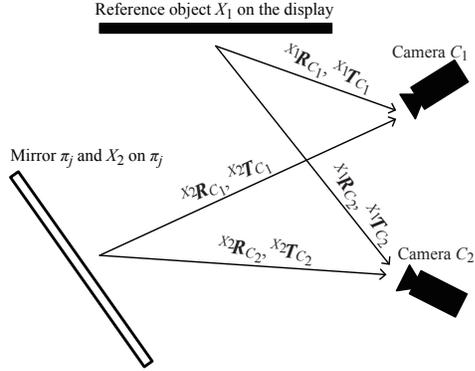
Figure 6. Illustration of relationships between the cameras, display and mirror

Table 2. Comparison of the proposed method and [8] against Eq (28)

	Hesch *et al*. [8]	Proposed
E_R (Eq (22))	0.239	0.057
E_T (Eq (24))	60.57	48.04
E_P (Eq (26))	3.653	0.170

play. C_1 is located where it cannot observe X_1 directly. It captures three VGA images of three different mirror poses π_j ($j = 1, 2, 3$) for calibration. C_2 is located 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 [20]. 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 Figure 6, suppose $^{X_l}R_{C_k}$, $^{X_l}T_{C_k}$ denote the rotations and translations from X_l to C_k respectively. That is, the 3D position of a 3D point $^{X_l}p^i$ of X_l in C_k is given by

$$^{C_k}p^i = {}^{X_l}R_{C_k} \cdot {}^{X_l}p^i + {}^{X_l}T_{C_k} \ (k = 1, 2, l = 1, 2). \quad (27)$$

We can calibrate $^{X_1}R_{C_1}$, $^{X_1}T_{C_1}$ by the proposed method as well as by [8]. In addition, by calibrating $^{X_1}R_{C_2}$, $^{X_1}T_{C_2}$, $^{X_2}R_{C_1}$, $^{X_2}T_{C_1}$, $^{X_2}R_{C_2}$ and $^{X_2}T_{C_2}$ by Zhang's method, these parameters also provide $^{X_1}R_{C_1}$, $^{X_1}T_{C_1}$ as a baseline calibration:

$$\begin{aligned} ^{X_1}R_{C_1} &= {}^{X_2}R_{C_1}{}^{X_2}R_{C_2}^{\top}{}^{X_1}R_{C_2} \\ ^{X_1}T_{C_1} &= {}^{X_2}R_{C_1}{}^{X_2}R_{C_2}^{\top}({}^{X_1}T_{C_2} - {}^{X_2}T_{C_2}) + {}^{X_2}T_{C_1}. \end{aligned} \quad (28)$$

Figure 7 shows results by the proposed method, by [8], and by Eq (28). This figure renders estimate positions of ^Cp and X_1 in $\{C\}$, and therefore C_1 is located at $(0, 0, 0)^{\top}$. Table 2 reports quantitative comparisons of them. E_R and E_T are the differences against Eq (28). E_P is the reprojection errors against the original observations. Notice that
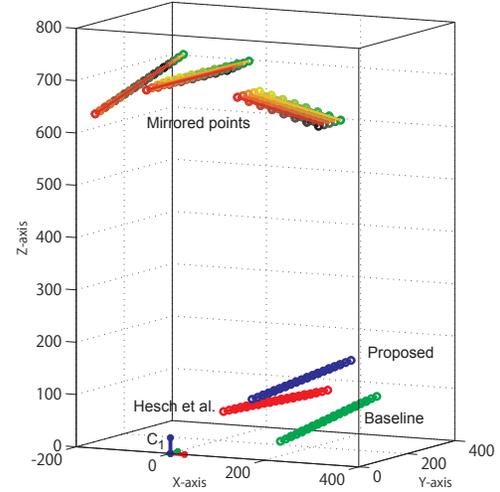
Figure 7. Estimated positions of the reference object by the proposed method (blue), by [8] (red), and by Eq (28) (green). Notice that this figure renders estimate positions of ^Cp and X_1 in $\{C\}$, and therefore C_1 is located at $(0, 0, 0)^{\top}$.
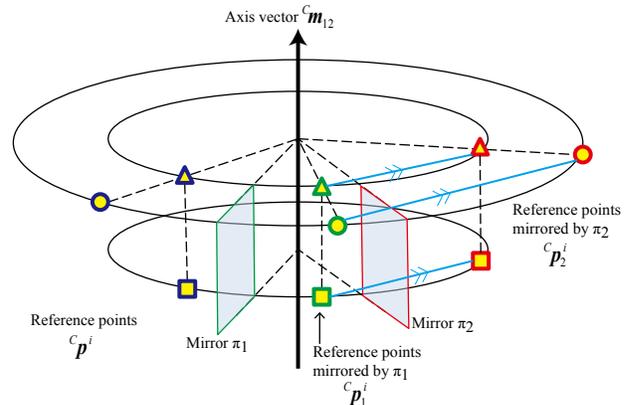
Figure 8. Degenerate case

even though the chess pattern provides more than 3 reference points, we used only three points of them for both our method and [8]. From these results, we can conclude that our method performs better than [8].

5.3. Discussions

Degenerate cases   Our algorithm does not work if it cannot compute the axis vectors $m_{jj'}$ (Section 3.2). This happens in the following two cases. (1) As illustrated by Figure 8, we cannot compute $m_{jj'}$ if $m_{jj'}$, the intersection of two mirrors, is on the plane formed by the three reference points. In this case, the intersection of two mirrors does exist physically, but not be computable by solving Eq (9). In addition, (2) if two mirrors are parallel, then the intersection of two mirrors does not exist and therefore not be computable.

These two degenerate cases can be detected by observing the rank of $M_{jj'}$ in Eq (10). If the rank is less than 2, we can

discard the mirror pair and try with 4-th or more mirrored images in practice.

Comparison with [8] The differences between ours and Hesch *et al.* [8] are twofold. The first difference is how they select a solution from 64 candidates given by three P3P problems. While our method can select a solution directly from the results of P3P problems (Section 4.1), [8] selects a solution only by comparing reprojection errors of 64 extrinsic calibration parameters. The other difference is the number of constraints involved to compute R and T. The method by [8] computes R candidates from each mirror by a 3×3 matrix computation, and then average them to output a single R. T is computed by solving a system of linear equations using 9 constraints. On the other hand, our method computes R and T by solving a larger system of linear equations with 27 constraints (Eq (13)). In general solving a larger system can be more robust against observation noise than solving smaller systems. This difference appears as the robustness against observation noise as reported by Figure 4. Note that this does not mean the selection criterion of [8] based on the reprojection error computation is not correct. The criterion is correct, but because of the above-mentioned difference on the number of involved equations for computation, ours returns better results than [8].

6. Conclusion

In this paper, we proposed a new algorithm which finds a unique solution of three P3P problems from three mirrored images of three reference 3D points. Our algorithm is based on an orthogonality constraint which should be satisfied by results of Householder transformations of a same reference object. Based on this algorithm, we proposed a new extrinsic calibration algorithm for a stationary camera using three reference 3D points observed via three different mirror poses. The evaluations by synthesized and real data demonstrated our improvement on the calibration accuracy and robustness against the state-of-the-art quantitatively and qualitatively.

Future work includes further studies on the proof of the uniqueness (Section 4.1), workarounds of the degenerated cases (Section 5.3), and calibration from multiple reflections [9] using the orthogonality constraint.

Acknowledgements

References

[1] S. Agarwal, Y. Furukawa, N. Snavely, B. Curless, S. M. Seitz, and R. Szeliski. Reconstructing rome. *IEEE Computer*, 43:40–47, 2010.

[2] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre. Recent advances in augmented reality. *Computer Graphics and Applications, IEEE*, 21(6):34 –47, 2001.

[3] G. Desouza and A. Kak. Vision for mobile robot navigation: a survey. *IEEE TPAMI*, 24(2):237 –267, 2002.

[4] F. Fraundorfer, D. Scaramuzza, and M. Pollefeys. A constricted bundle adjustment parameterization for relative scale estimation in visual odometry. In *Proc. of ICRA*, pages 1899 –1904, 2010.

[5] G. Golub and C. van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, Maryland, third edition, 1996.

[6] B. M. Haralick, C.-N. Lee, K. Ottenberg, and M. Nölle. Review and analysis of solutions of the three point perspective pose estimation problem. *IJCV*, 13:331–356, 1994.

[7] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004.

[8] J. A. Hesch, A. I. Mourikis, and S. I. Roumeliotis. *Algorithmic Foundation of Robotics VIII*, volume 57 of *Springer Tracts in Advanced Robotics*, chapter Mirror-Based Extrinsic Camera Calibration, pages 285–299. 2009.

[9] J. A. Hesch, A. I. Mourikis, and S. I. Roumeliotis. Extrinsic camera calibration using multiple reflections. In *Proc. of ECCV*, 2010.

[10] K. H. Jang, D. H. Lee, and S. K. Jung. A moving planar mirror based approach for cultural reconstruction: Research articles. *Comput. Animat. Virtual Worlds*, 15:415–423, 2004.

[11] T. Kanade, P. Rander, and P. Narayanan. Virtualized reality: constructing virtual worlds from real scenes. *IEEE Multimedia*, 4(1):34 –47, 1997.

[12] R. Kumar, A. Ilie, J.-M. Frahm, and M. Pollefeys. Simple calibration of non-overlapping cameras with a mirror. In *Proc. of CVPR*, pages 1–7, 2008.

[13] P. Lébraly, C. Deymier, O. Ait-Aider, and M. D. E. Royer. Flexible extrinsic calibration of non-overlapping cameras using a planar mirror: Application to vision-based robotics. In *Proc. of IROS*, pages 5640–5647, 2010.

[14] M. Moakher. Means and averaging in the group of rotations. *SIAM J. Matrix Anal. Appl.*, 24(1):1–16, 2002.

[15] S. Nayar. Catadioptric omnidirectional camera. In *Proc. of CVPR*, pages 482–488, 1997.

[16] R. Rodrigues, P. Barreto, and U. Nunes. Camera pose estimation using images of planar mirror reflections. In *Proc. of ECCV*, pages 382–395, 2010.

[17] J. Starck, A. Maki, S. Nobuhara, A. Hilton, and T. Matsuyama. The multiple-camera 3-d production studio. *IEEE TCSVT*, 19(6):856 –869, 2009.

[18] P. Sturm and T. Bonfort. How to compute the pose of an object without a direct view. In *Proc. of ACCV*, pages 21–31, 2006.

[19] X. Ying, K. Peng, R. Ren, and H. Zha. Geometric properties of multiple reflections in catadioptric camera with two planar mirrors. In *Proc. of CVPR*, pages 1126 –1132, 2010.

[20] Z. Zhang. A flexible new technique for camera calibration. *IEEE TPAMI*, pages 1330–1334, 2000.