Camera Calibration Based on Mirror Reflections

KOSUKE TAKAHASHI¹,a) SHOHEI NOBUHARA¹,b)

Abstract: This paper addresses the use of mirror reflections for camera calibration. Camera calibration is an essential technique for analyzing the geometric and radiometric relationship between a 3D space and a 2D image. Most conventional camera calibration methods are based on a fundamental assumption: a camera can directly observe a reference object of known geometry. However, there are cases in which this assumption does not hold in practical scenarios. One approach to camera calibration in such cases is the use of a “mirror” as a supporting device. A mirror generates a virtual reference object that can be expressed using a small number of parameters. In addition, the 2D projection of the reflection object is equal to that of the known reference object from the virtual viewpoint. This paper utilizes these features and tackles two challenges of the geometric camera calibration; the first challenge is the intrinsic camera calibration when a known reference object is not available and the second challenge is the extrinsic camera calibration when the camera cannot directly observe a known reference object due to a physical constraint on the imaging system. The proposed algorithms introduce novel constraints, kaledoscopic projection constraint and orthogonality constraint, which are hold with the mirror reflections for solving these problems. Evaluations with synthesized and real data demonstrates that the proposed algorithms can work properly and report the robustness of it in comparison with conventional methods.

1. Introduction

A ray emitted from a light source or reflected from the surface of an object reaches an image sensor through a lens and is collected as an “image”. Understanding this generating process of an image, that is, describing to where a point in 3D space is projected with how much intensity, is a fundamental and important problem for various tasks in computer vision, such as 3D reconstruction and motion analysis. In order to analyze the generating process, to use an appropriate camera model and to estimate the camera parameters are called “camera calibration.”

Camera calibration has been a fundamental research topic in computer vision for many years. While the strict meaning of camera calibration depends on the configuration of the imaging system, camera calibration can be divided into two types: geometric calibration and radiometric calibration.

Geometric camera calibration: With this type, the “to where” a 3D point is projected onto a 2D image plane is estimated. That is, there is a geometric relationship between a 3D space and a 2D image plane. This type of calibration is thus appropriate for analyzing geometric transformation involving a 3D space and a 2D image plane.

Radiometric camera calibration: With this type, the “how much” intensity is estimated. It is appropriate for analyzing radiometric transformation involving a 3D space and a 2D image plane. This calibration includes removing shading caused by the lens [36] or noise on the image sensor [12], estimation of the radiometric response function [21], and so on.

This paper focuses on the analysis of geometric properties involving a 3D space and a 2D image plane, so it mainly discusses geometric camera calibration of a perspective camera model.

1 Kyoto University, Yoshida-honmachi, Kyoto, Japan
a) takahasi@vision.kyoto-u.ac.jp
b) nob@i.kyoto-u.ac.jp

Fig. 1 The overview of the geometric camera calibration.

As illustrated in Figure 1, geometric camera calibration comprises intrinsic camera calibration and extrinsic camera calibration. The former estimates the intrinsic parameters, which describe the camera properties: focal length, optical center, lens distortion, and so on. The latter estimates the extrinsic parameters, which describe the position and pose of the camera.

Geometric camera calibration has three basic steps: (1) capture a reference object for which the geometric features are known, (2) determine the correspondences between the features and their projections, and (3) estimate each parameter from the correspondences using the camera model.

For example, Tsai [38] estimated the intrinsic and extrinsic parameters from the correspondences between the 3D points for which the geometries were known in a world coordinate system and their 2D projections. Zhang [40] used a known reference point on a single plane as a reference object, e.g., a chessboard, to estimate the camera parameters.

These conventional methods share a fundamental assumption: the camera can directly observe a reference object for which the geometry is known. However, this assumption does not always hold in practical cases, so these methods do not always work properly. For example, while intrinsic camera calibration should be done using known reference objects in the whole of camera’s field-of-view for high precision estimation, a practical known reference object is not always available when capturing images on a certain scale as illustrated in Figure 2(a). In addition, Figure 2(b) illustrates the case in which a reference object is not observable from the camera. For a vision-based robot [6] with a
display-camera system [14, 18], although the extrinsic parameters between the camera and each part of the system (e.g., a robot arm and a body) are important, they are often unobservable directly from the camera due to a physical constraint on the imaging system. Thus, the previous methods can be problematic depending on the configuration of the imaging system and the situation.

The methods introduced in this paper for solving these problems use algorithms that enable the use of a mirror as a supporting device for calibration. Mirrors can generate reflections of a real object, and the reflections can be defined using a small number of parameters. This means that they can be recognized as a parametric 3D model. Furthermore, 2D projections of reflections are equal to those of the known reference object from the virtual viewpoint. In other words, the mirror extends the camera’s field-of-view. As shown in Figure 3, these problems are overcome by

• setting planar mirrors and generating multiple reflections consisting of a parametric 3D model with an unknown 3D point and its reflections (Figure 3(a)) and
• setting a planar mirror that enables the camera to observe a known reference object (Figure 3(b)).

The potential of using mirrors to solve these problems of camera calibration in practical situations is explored in this paper.

1.1 Problem Statement and Contributions

This paper focuses on geometric camera calibration in cases where the fundamental assumption, i.e., the camera can directly observe a reference object for which the geometry is known, does not hold and tackles the two problems; intrinsic camera calibration without a known reference object and extrinsic camera calibration with an unobservable reference object.

Intrinsics Camera Calibration Without Known Reference Object: In cases where a known reference object is not available, we introduce mirrors to generate reflections of a 3D point of unknown geometry and recognize its reflections as a reference object of a parametric 3D model, as shown in Figure 3(a).

In this paper, a novel intrinsic camera calibration algorithm is introduced that utilizes the parametric 3D model by multiple planar mirrors consisting of a kaleidoscopic imaging system. Three problems must be solved in order to realize this method: chamber assignment of the kaleidoscopic projections, estimation of the mirror parameters, and estimation of the intrinsic parameters. The key contribution of this work is the introduction of a novel geometric constraint, the kaleidoscopic projection constraint, which is satisfied by projections of high-order reflections.

This constraint provides multiple linear equations for the mirror parameters for a single 3D point and solves the three problems.

Extrinsic Camera Calibration with Unobservable Reference Object: In cases where the camera cannot directly observe a reference object due to a physical constraint on the imaging system, we introduce mirrors to enable the camera to directly observe the reflections of a known reference object (Figure 3(b)).

In this paper, a novel two types of mirror-based algorithms are introduced that estimates the extrinsic parameters between the camera and a reference object located outside its field-of-view.

The first algorithm utilizes a planar mirror. Since the poses and positions of the reference object and the planar mirror are unknown, the extrinsic parameters cannot be determined uniquely from a single image of the reflection. To overcome this problem, an orthogonality constraint that is satisfied among reflections by multiple mirror poses is introduced. This constraint is used to estimate the extrinsic parameters, that is, three reference points and three mirror poses, with the minimal configuration.

The mirror-based methods can be troublesome for preparing a mirror and calibrating it every time in a casual scenario, such as gaze correction in a video conference [18]. For such cases, the second algorithm utilizes the human cornea based on the fact that the surface of the human eye reflects light like a mirror, meaning that the human eye can be modeled as a spherical mirror. The introduction of a geometric model of the human cornea enables estimation of extrinsic parameters with a simple configuration, i.e., one mirror pose and three or five reference points.

The rest of this paper is organized as follows. Section 2 provides fundamental knowledge and related work on camera calibration and mirror geometry. Section 3 presents a novel algorithm for intrinsic camera calibration using multiple planar mirrors. Section 4 and Section 5 present novel methods for planar mirror-based and human-cornea-based extrinsic camera calibration. Section 6 concludes this paper and outlines future works.

2. Camera Calibration and Mirror Geometry

2.1 Geometric Camera Calibration of Perspective Camera

2.1.1 Perspective Camera Model

The perspective camera model is illustrated in Figure 4. Let $p^{[W]} = (X^{[W]}, Y^{[W]}, Z^{[W]})^T$ and $p^{[C]} = (X^{[C]}, Y^{[C]}, Z^{[C]})^T$ denote a 3D point $p$ in a world coordinate system and a camera coordinate system, respectively. They satisfy

$$p^{[C]} = Rp^{[W]} + t,$$

where $R$ is a rotation matrix and $t$ is a translation vector. Pa-
parameters $R$ and $t$ are extrinsic camera parameters representing the pose and position of the camera in the world coordinate system. The goal of extrinsic camera calibration is to estimate these parameters. Note that we hereinafter omit the superscript representing the coordinate system for cases in which $p$ is represented in the camera coordinate system.

Let $q = (u, v)$ denote the projection of $p^{(C)}$ in a pixel image coordinate system. This $q$ is given by a perspective projection:

$$\lambda q = Ap^{(C)} = \begin{bmatrix} f & 0 & cu \\ 0 & f & cv \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix},$$

(2)

where $\lambda$ denotes the homogeneous coordinate of $q$ and $\lambda$ is a scale parameter. The $f$ is focal length and $(c_u, c_v)$ are the optical centers expressed in pixels coordinates.

This projection model is extended by taking into account lens distortion. Suppose $\hat{p} = (x, y, 1)$ denotes the normalized image coordinates of $p^{(C)}$, and $\hat{q}$ denotes the distorted coordinates of $\hat{p}$. This $\hat{p}$ is expressed as:

$$\hat{x} = x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + 2 p_1 x y + p_2 (r^2 + 2x^2),$$

(3)

$$\hat{y} = y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + p_1 (r^2 + 2y^2) + 2 p_2 x y,$$

where $r^2 = x^2 + y^2$. The $k_i (i = 1, 2, 3)$ and $p_i (i = 1, 2)$ denote the coefficients of the radial and tangential factors of lens distortion, respectively.

A 2D observation $\hat{q}$ including lens distortion is given by the perspective projection of $\hat{p}$; that is, $\hat{q} = \hat{A} \hat{p}$. The ideal 2D projection $q$ without lens distortion can be computed from $\hat{q}$ by solving Eq. (3) numerically for $\hat{p}$ with intrinsic camera parameters $A$ and $d = (k_1, k_2, k_3, p_1, p_2)$. The goal of intrinsic camera calibration is to determine these parameters.

### 2.1.2 Related Works of Camera Calibration

Conventional studies on camera calibration fall into two basic groups: camera calibration with a known reference object and camera calibration without a known reference object.

**Camera Calibration with Known Reference Object:** As introduced in Section 1, geometric camera calibration is done by capturing a reference object for which the geometric features are known, determining the correspondences between the features and their projections, and estimating each parameter from the correspondences using the camera model.

The method proposed by Tsai [38] is based on the assumption of a pinhole perspective projection model and estimates the intrinsic and extrinsic parameters from 3D points for which the geometries are known and their 2D projections. The camera parameters are estimated using a least-squares fitting method and then optimized by minimizing the reprojection error. This method works for both planar and non-planar reference objects. The method proposed by Zhang [40] uses a planar-reference-object-based calibration algorithm and the same basic strategy: estimate the camera parameters and then optimize them. The reference points are assumed to be on a single plane, and each point is matched to its projection using homography. This method as well as that of Tsai [38] are the most commonly used calibration algorithms for estimating both intrinsic and extrinsic parameters using a known reference object.

**Camera Calibration with Unknown Reference Object:** While the methods above utilize a reference object of known geometry for calibration, there are cases in which such a reference object is not available. Previous studies addressed this problem by focusing on geometric features of the scene in the captured images and utilizing them for calibration. This approach is known as self-calibration [7, 20].

The most common approach to camera calibration without a known reference object is based on epipolar geometry satisfied by a 3D point and its projections on the multi-view images [1, 8, 11]. A fundamental matrix is estimated by using an 8-point algorithm along with several detected corresponding points and then decomposing the extrinsic camera parameters by using the essential matrix. The precision of correctly detecting the corresponding points greatly affects estimation precision. Methods using this approach therefore can fail to estimate the extrinsic parameters through a RANSAC procedure in case of wide baseline stereo, observing texture-less object, and so on.

A silhouette-based approach has been proposed for extrinsic calibration in such cases [4, 32]. The correspondences between special points on the silhouette boundaries, called frontier points [5], across the multiple views are established. These points are the projections of 3D points tangent to the epipolar plane. The epipolar geometry can be recovered from the correspondences of the frontier points.

While these approaches use typical geometric features included in the scene for calibration, our intrinsic camera calibration technique generates a parametric 3D model from unknown 3D points and their reflections by using multiple planar mirrors.

### 2.2 Mirror Geometry

#### 2.2.1 Planar Mirror Geometry

The planar mirror geometry is illustrated in Figure 5. Consider a 3D point $p$ and its reflection $p'$ from mirror $\pi$. Their ideal projections without lens distortion, $\hat{q}$ and $\hat{q}'$, are given by the perspective projection:

![Fig. 4 The perspective camera model. The effect of lens distortion is modeled as the non-linear transformation in the normalized image coordinates.

![Fig. 5 Planar mirror geometry: mirror $\pi$ with normal $n$ and distance $d$ reflects 3D point $p$ to $p'$; they are projected to $\hat{q}$ and $\hat{q}'$ respectively.](image-url)
\[ \lambda q = A p, \quad \lambda' q' = A' p', \]  
(4)

where \( \lambda \) and \( \lambda' \) are scale parameters.

Let \( n \) and \( d(>0) \) denote the normal and the distance of the mirror \( \pi \) satisfying \( n^\top x + d = 0 \), where \( x \) is a 3D position in the scene. Here the normal vector is oriented toward the camera center.

As illustrated in Figure 5, the distance \( t \) from \( p \) and \( p' \) to mirror \( \pi \) satisfies \( p = p' + 2dn \). The projection of \( p' \) to \( n \) gives \( t + d = -n^\top p' \). By eliminating \( t \) from these equations, we have

\[ p = -2(n^\top p + d)n + p', \]
\[ \Rightarrow p = S p' = \begin{bmatrix} H & -2dn \\ 0_{1 \times 3} & 1 \end{bmatrix} p', \]

where \( H \) is a \( 3 \times 3 \) Householder matrix given by \( H = I_{3 \times 3} - 2nn^\top \), \( 0_{m \times n} \) denotes the \( m \times n \) zero matrix, and \( I_{m \times n} \) denotes the \( m \times n \) identity matrix. This \( S \) also satisfies \( p' = S p_0 \).

### 2.2.2 Related Works on Mirrors in Computer Vision

Previous studies using mirrors can be categorized into two groups: those integrating mirrors as components into their imaging systems, and those using mirrors as supplemental devices for calibrating camera systems.

**Mirrors as Imaging Components**: Studies in the first group used mirrors as a component of the imaging system.

Non-planar mirrors are commonly used to widen the field-of-view. One of the most common usages of non-planar mirrors in an imaging system is for capturing omnidirectional images. Scaramuzza et al. [31] proposed a single-viewpoint omnidirectional camera with a hyperbolic mirror.

Planar mirrors, on the other hand, are commonly used for capturing multi-view images. Virtual cameras generated by planar mirrors have identical intrinsic parameters and are time-synchronized to the original camera. This can be a strong advantage for multi-view applications, such as reflectance analysis and 3D shape reconstruction. Mukaiaga et al. [22] introduced a hemispherical confocal imaging system using a turretback reflector, and Inoshita et al. [16] used it to measure a full-dimensional (8-D) BSSRDF (bidirectional subsurface scattering reflection distribution function). Nane and Nayer [24] proposed a computational stereo system using a single camera and various types of mirrors, such as planar, ellipsoidal, and so on.

In the context of kaleidoscopic imaging, Ihrke et al. [15] and Reshetouski and Ihrke [27, 28] developed a theory for modeling chamber detection, segmentation, bounce tracing, shape-from-silhouette, etc. The essential problems in this context related to calibrating a kaleidoscopic imaging system are chamber assignment and mirror parameter estimation.

Reshetouski and Ihrke [27] solved the chamber assignment problem by placing constraints on the apparent 2D distances between 2D projections while most other studies [28] assumed that the chamber assignment is done by hand. However, placing constraints on the apparent 2D distances limits the application of their method to specific mirror poses. The mirror parameters are estimated on a per-mirror basis in the conventional approaches [15, 28] without considering their kaleidoscopic, i.e. multiple reflection, relationships. That is, a chessboard is first detected [40] in each of the chambers, and then the mirror normals and the distances from the 3D chessboard positions in the camera frame are estimated. However, those algorithms do not make full use of the kaleidoscopic relationships of multiple reflections.

Our intrinsic camera calibration using planar mirrors technique uses novel algorithms that overcome these challenges: (i) a chamber assignment algorithm that can be applied to general mirror poses and (ii) a mirror parameter estimation algorithm that can estimate parameters satisfying kaleidoscopic reflection constraints.

**Mirrors as Supplemental Devices for Calibration**: Studies in the second group used mirrors for extrinsic calibration of the camera and a reference object outside the camera’s field-of-view. Previous mirror-based extrinsic calibration methods can be categorized in terms of the configuration, i.e. mirror shape, the required number of mirror poses, and the required number of reference points (See Table 1). Some of them are aimed at calibration using a simpler setup, which means reducing the number of required mirror poses and reference points, because a simpler setup is advantageous in terms of robustness and cost.

In the case of calibrating a camera and a directly unobservable reference object, planar mirrors are commonly used.

Sturm et al. [33] and Rodrigues et al. [30] proposed a planar-mirror-based method for computing the pose of a reference object without a direct view. Hesch et al. [13] assumed a camera-based robot with each body part not being visible from the camera and estimated their relative poses and positions with a planar mirror in a P3P problem scenario. Our proposed method described in Section 4 is a planar-mirror-based method with a configuration comprising three mirror poses and three reference points based on an orthogonality constraint. As described by Rodrigues et al. [30], the three mirror poses and three reference points are the minimal configuration for a planar-mirror-based method.

These methods need a mirror as an additional device, and using such device for every calibration instance can be troublesome in casual scenarios, such as gaze correction in video chatting [18].

Nitschke et al. [26] presented a method that addresses this problem without the need for an additional device: the display-camera setup is calibrated using the reflections in the user’s eyes (corners).

This paper considers it important to avoid the use of additional hardware for calibration for casual display-camera systems and presents two methods using cornea-reflection-based calibration with simpler setups. They are described in Sections 5.

### 3. Mirror-based Intrinsic Camera Calibration

This section provides a novel intrinsic camera calibration algorithm by introducing multiple planar mirrors and consisting a kaleidoscopic imaging system (Figure 7).
3.1 Kaleidoscopic Imaging System

Suppose the camera observes the target 3D point directly and indirectly via $N_e$ mirrors as shown in Figure 7. Let $p_0$ denote the original 3D point and $p_i$ denote the first reflection of $p_0$ by the mirror $\pi_i$ ($i = 1, \cdots, N_e$) (Figure 8). The reflection $p_i$ is given by

$$p_0 = S_ip_i = \begin{bmatrix} H_i & -2d_in_i \\ 0_{1 \times 3} & 1 \end{bmatrix} p_i,$$

(7)

where $n_i$ and $d_i$ denote the mirror normal and its distance respectively, and $H_i$ is given by $H_i = I_{3 \times 3} - 2n_in_i^\top$.

Furthermore, such mirrors define virtual mirrors as a result of multiple reflections. Let $\pi_{ij}$ ($i, j = 1, \cdots, N_e$, $i \neq j$) denote the virtual mirror defined as a mirror of $\pi_i$ by $\pi_j$, $p_{ij}$ denote the reflection of $p_i$ by $\pi_{ij}$, and $L_{ij}$ denote the chamber where $p_{ij}$ is projected to. Also the matrices $S_{ij}$ and $H_{ij}$ for $\pi_{ij}$ are given by

$$S_{ij} = S_iS_j,$$

$$H_{ij} = H_iH_j.$$

(8)

The third and further reflections, virtual mirrors, and chambers are defined in the same manner:

$$\Pi_{k=1}^{N_e} S_i (i_k = 1, 2, 3, i_k \neq i_{k+1}),$$

(9)

where $N_i$ is the number of reflections.

Obviously the 3D subspaces where $p_0$ and $p_i$ can exist are mutually exclusive, and the captured image can be subdivided into regions called chambers corresponding to such subspaces. Suppose the perspective projections of $p_i$ is denoted by $q_i \in Q (x = 0, 1, \cdots, N_e, 12, 13, \cdots)$ in general. In this paper we denote the 2D region where $q_0$ exists as the base chamber $L_0$, and we use $L_x$ to denote the chamber where $q_i$ exists.

3.2 Problem Description

Conventional intrinsic camera calibration techniques have been conducted from a 3D model of “known” geometry and its 2D projections [40]. That is such approaches require a reference object whose surface has several feature points such that their 3D positions are provided a priori and they are uniquely identifiable in 2D images. Hence in cases of microscopic or large-scale environment, it is not trivial task to provide such 3D models in practice.

The key idea to solve this problem is to introduce a parametric 3D model whose 3D feature positions are defined by a small number of parameters. To realize the idea, this research utilizes multiple reflections of 3D points by planar mirrors. That is our calibration estimates the intrinsic parameters as well as the mirror parameters to identify the 3D model structure simultaneously.

The configuration we consider are as follows: (a) each of the detected observation is not assigned to the corresponding chamber, (b) it has two or more planar mirrors as a supporting device whose normals and distances are unknown, and (c) it has one perspective camera and its intrinsic parameters are unknown.

Figure 6 illustrates an outline of the proposed algorithm. In our algorithm, there are three problems to be solved. Consider a 2D point set $R = \{r_i\}$ detected from the captured image as candidates of $q_i$. The problems are:

- to assign the chamber label $L_i$ to $r_i \in R$ to identify to which chamber each of the projections $q_i$ belong (Section 3.4),
- to estimate the parameters of the real mirrors $\pi_i (i = 1, \cdots, N_e)$, i.e. normals $n_i$ and distances $d_i$ of them, from kaleidoscopic projections $q_i$ (Section 3.5), and
- to estimate the intrinsic parameters, i.e. $A$ and $d$, from $q_i$, and the mirror parameters as the model parameters (Section 3.6).

For solving these problems, we utilize a kaleidoscopic projection constraint introduced in the next section.

3.3 Kaleidoscopic Projection Constraint

Suppose the camera observes a 3D point of unknown geometry $p$. The mirror $\pi$ of matrix $S$ defined by the normal $n$ and the distance $d$ reflects $p$ to $p' = Sp$ (Eq (6)).

Based on the epipolar geometry [11, 39], $n$, $p$ and $p'$ are coplanar and satisfy $(n \times p')^\top p' = 0$. By substituting Eq (4), we obtain

$$q^\top A^{-1} [n]_e A^{-1} q' = 0,$$

(10)

where $[n]_e$ denotes the $3 \times 3$ skew-symmetric matrix representing the cross product by $n$ and this is the essential matrix of this mirror-based binocular geometry [39].

By representing the normalized image coordinates of $q$ and $q'$
by \((x, y, 1)^\top = A^{-1}q_{i}\) and \((x', y', 1)^\top = A^{-1}q'_i\) respectively, Eq (10) can be rewritten as
\[
(y - y' \quad x' - x \quad xy' - x'y')n = 0, \tag{11}
\]
We call this Eq (11) as kaleidoscopic projection constraint in this paper. This constraint is satisfied not only by single reflections but also high-order reflections as below.

**Single reflection:** Let \(p_0\) denote a 3D point and \(p_i\) denote the reflection by mirror \(\pi_i\). Since \(p_i\) is expressed as \(p_i = S_ip_0\), the normalized image coordinates of their projections, \(q_0\) and \(q_i\), obviously satisfy Eq (11) as:
\[
\begin{pmatrix} y_0 - y_i \\ x_i - x_0 \\ x_0y_i - x_iy_0 \end{pmatrix} n = 0, \tag{12}
\]
where \((x_0, y_0, 1)^\top = A^{-1}q_0\) and \((x_i, y_i, 1)^\top = A^{-1}q_i\).

**High-order reflections:** Let \(p_{ij}\) denote the reflection by mirror \(\pi_j\). This \(p_{ij}\) can be expressed as \(p_{ij} = S_iS_jp_0 = S_jS_ip_0\) based on Eq (8). Here \(p_j = S_jp_0\) holds as well, and we obtain \(p_{ij} = S_jS_ip_0 \iff p_{ij} = S_jp_j\). This equation means that \(p_{ij}\) can be recognized as the first reflection of \(p_j\) by \(\pi_j\) and hence the normalized image coordinates of their projections, \(q_{ij}\) and \(q_j\), also satisfy Eq (11) as:
\[
\begin{pmatrix} y_j - y_{ij} \\ x_j - x_i \\ x_iy_j - x_jy_i \end{pmatrix} n_i = 0, \tag{13}
\]
where \((x_{ij}, y_{ij}, 1)^\top = A^{-1}q_{ij}\).

In \(N_k\) reflections, we obtain the kaleidoscopic projection constraint between \(p_{ij} = S_{ij} \prod_{k=1}^{N_k-1} S_k p_0\) and \(p_{jk} = \prod_{k=1}^{N_k-1} S_k p_0\) in the same manner.

### 3.4 Chamber Assignment

Based on the kaleidoscopic projection constraint, we introduce a new algorithm that identifies the chamber label of each projection. Our algorithm utilizes an analysis-by-synthesis approach which iteratively draws a number of projections and evaluates their geometric consistency in terms of the kaleidoscopic projection in order to find the best chamber assignment.

In what follows the concept of base structure, i.e., minimal configuration for estimating the real mirror parameters using the kaleidoscopic projection constraint, is introduced. Our algorithm hypothesizes a number of base structure candidates from observed points and evaluates each of their consistencies as a kaleidoscopic projection. Notice that we introduce our algorithm in a two mirror case as an example here, and it can be extended to three or more mirror cases intuitively.

### 3.4.1 Base Structure

Suppose \(2N_r\) points of the observed points are selected and they could be hypothesized as \(q_0, q_1, \ldots\) correctly. The mirror normal \(n_i\) has two degrees of freedom and can be linearly estimated by collecting two or more linear constraints on it. In case of \(N_r = 2\), the mirror normal \(n_1\) can be estimated as the eigenvector corresponding to the smallest eigenvalue of the coefficient matrix of the following system defined by the kaleidoscopic projection constraint using \(\{q_0, q_1\}\) and \(\{q_2, q_3\}\) in Figure 9(a):
\[
\begin{bmatrix}
|y_0 - y_1 | & |x_0 - x_1 | & |x_0y_1 - x_1y_0 | \\
|y_2 - y_1 | & |x_2 - x_1 | & |x_0y_2 - x_1y_2 |
\end{bmatrix} n_1 = 0_{2 \times 1}, \tag{14}
\]
where \(\langle q, q' \rangle\) denotes a doublet, the pair of \(q\) and \(q'\) for Eq (11).

Using the estimated \(n_1\) and assuming \(d_1 = 1\) without loss of generality, the 3D point \(p_1\) can be described as \(p_1 = S_1\hat{p}_0\) by Eq (6). By substituting \(p_0\) and \(p_1\) in this equation by using \(q_0\) and \(q_1\) as expressed in Eq (4), the 3D point \(p_0\) and \(p_1\) can be triangulated by solving the following linear system for \(\lambda_0\) and \(\lambda_1\):
\[
p_0 = S_1\hat{p}_1, \tag{15}
\]
\[
\Rightarrow \begin{bmatrix}
H_1A^{-1}q_0 & -A^{-1}q_1
\end{bmatrix} \begin{bmatrix} \lambda_0 \\ \lambda_1 \end{bmatrix} = 2n_1. \tag{16}
\]

Similarly the 3D points \(p_2\) and \(p_{12}\) can be triangulated by solving the linear system for \(\lambda_2\) and \(\lambda_{12}\). Because \(p_1\) is the reflection of \(p_0\) by the mirror \(\pi_2\), the mirror normal \(n_2\) as well as the distance \(d_2\) can be estimated as
\[
n_2 = \frac{p_0 - p_2}{|p_0 - p_2|}, \quad d_2 = -n_2^\top \frac{p_0 + p_2}{2}. \tag{17}
\]

This doublet pair \(\{q_0, q_1\}, \{q_2, q_{12}\}\) is a minimal configuration for linear estimation of the real mirror parameters in \(N_r = 2\) case and we call such minimal configuration as a base structure of our chamber assignment. Notice that the above doublet pair is not the unique base structure. That is, \(\{q_0, q_2\}, \{q_1, q_{21}\}\) is also a base structure for \(N_r = 2\) case.

### 3.4.2 Chamber Assignment Algorithm

Given the mirror parameters and the triangulated 3D point from one base structure, the 4th reflection and its projection can be synthesized by Eq (6) and Eq (4) with known chamber labels. These labels are assigned by finding correspondences between the synthesized point \(\hat{Q}\) and the observed point set \(R = \{|r_i|\}\) as a sort of bipartite matching.

Based on this assignment, the proposed method introduces a
recall ratio $\mathcal{R}$ which measures how many of the synthesized projections that are supposed to be visible have been assigned detected points:

$$\mathcal{R} = \frac{|\mathcal{R}|}{|\hat{Q}|}, \quad (18)$$

where $\mathcal{R} \subseteq R$ is the set of detected points assigned labels, $|\hat{Q}|$ and $|\mathcal{R}|$ denote the size of the set $\hat{Q}$ and $\mathcal{R}$, respectively.

The proposed chamber assignment algorithm examines if each possible base structure satisfies the kaleidoscopic projection constraint expressed by Eq (14) and other geometric constraints, which are detailed in [34]. Once the base structure passes these verifications, its recall ratio $\mathcal{R}$ is computed by Eq (18). Finally the best estimate of the chamber assignment is returned by finding the base structure of the highest recall ratio.

### 3.5 Mirror Parameters Estimation

This section introduces a novel algorithm of mirror parameters estimation given the chamber assignment for kaleidoscopic projections of single 3D point.

While the algorithm in Section 3.4 can also estimate the mirror parameters, it is a per-mirror estimation and it is not guaranteed to estimate mirror parameters consistent with projections of higher order reflections. Instead of such mirror-wise estimations, this section proposes a new linear algorithm which calibrates the kaleidoscopic mirror parameters simultaneously by observing a single 3D point in the scene.

#### 3.5.1 Mirror Normals

As illustrated in Figure 10 (a), suppose a 3D point $p_0$ is projected to $q_0$ in the base chamber, and its mirror $p_i$ by $\pi_i$ is projected to $q_i$ in the chamber $L_i$. Likewise, the second mirror $p_j$ by $\pi_j$ is projected to $q_j$ in the chamber $L_j$, and so forth. Here, kaleidoscopic projection constraints are satisfied by two pairs of projections on each mirror $\pi_1$ and $\pi_2$. From these constraints, $n_1$ and $n_2$ can be estimated by solving

$$\begin{bmatrix} y_0 - y_1 & x_0 - x_1 & x_0y_1 - x_1y_0 \\ y_2 - y_1 & x_2 - x_1 & x_2y_1 - x_1y_2 \end{bmatrix} n_1 = 0_{3 \times 1}, \quad (19)$$

and

$$\begin{bmatrix} y_0 - y_2 & x_0 - x_2 & x_0y_2 - x_2y_0 \\ y_2 - y_1 & x_2 - x_1 & x_2y_1 - x_1y_2 \end{bmatrix} n_2 = 0_{3 \times 1}, \quad (20)$$

An important observation in this simple algorithm is the fact that (1) this is a linear algorithm while it utilizes multiple reflections, and (2) the estimated normals $n_1$ and $n_2$ are enforced to be consistent with each other while they are computed on a per-mirror basis apparently.

The first point is realized by using not the multiple reflections of a 3D position but their 2D projections. Intuitively a reasonable formalization of kaleidoscopic projection is to define a real 3D point in the scene, and then to express each of the projections of its reflections by Eq (6) coincides with the observed 2D position as introduced in Section 3.5.3 later. This expression, however, is nonlinear in the normals $n_i$ ($i = 1, 2$) (e.g. $p_{12} = S_1S_2p_0$). On the other hand, projections of such multiple reflections can be associated as a result of single reflection by Eq (13) directly (e.g. $n_1$ with $q_{12}$ and $q_2$ as the projections of $p_{12}$ and $S_2p_0$ respectively).

As a result, we can utilize 2D projections of multiple reflections in the linear systems above.

This explains the second point as well. The above constraint on $q_{12}, q_2$ and $n_1$ in Eq (19) assumes $p_2 = S_2p_0$ being satisfied, and it is enforced by $(A^{-1}q_{12} \times A^{-1}q_0)^\top n_2 = 0$ in the first row of Eq (20). Inversely, on estimating $n_1$ by Eq (19), it enforces $p_1 = S_1p_0$ for Eq (20).

Note that this algorithm can be extended to third or further reflections intuitively. For example, if $p_{21}$ and its reflection by $\pi_1$ are observable as $A_{12}q_{21} = Ap_{21} = AS_1p_{21}$, then they provide

$$(y_{21} - y_{121}, x_{21} - x_{121}, x_{21}y_{21} - x_{121}y_{21}) n_1 = 0, \quad (21)$$

and can be integrated with Eq (19).

Also, this algorithm can be extended to $N_e \geq 3$ cases. In case of $N_e = 3$, for example, we solve

$$\begin{bmatrix} y_0 - y_1 & x_0 - x_1 & x_0y_1 - x_1y_0 \\ y_2 - y_1 & x_2 - x_1 & x_2y_1 - x_1y_2 \\ y_3 - y_1 & x_3 - x_1 & x_3y_1 - x_1y_3 \end{bmatrix} n_1 = 0_{3 \times 1}, \quad (22)$$

instead of Eqs (19) from point correspondences in Figure 10 (b).

#### 3.5.2 Mirror Distances

Once the mirror normals $n_1$ and $n_2$ are given linearly, the mirror distances $d_1$ and $d_2$ can also be estimated linearly as follows.

**Kaleidoscopic Re-projection Constraint:** The perspective projection Eq (4) indicates that a 3D point $p_i$ and its projection $q_i$ should satisfy the colinearity constraint:

$$(A^{-1}q_i) \times p_i = x_i \times p_i = 0_{3 \times 1}, \quad (23)$$

where $x_i = (x_i, y_i, 1)^\top$ is the normalized camera coordinate of $q_i$. Since the mirrored points $p_i$ ($i = 1, 2$) are given by Eq (6) as $p_i = H_ip_0 - 2d_in_i$, we obtain

$$x_i \times p_i = [x_i]_x \begin{bmatrix} H_i & -2n_i \end{bmatrix} \begin{bmatrix} p_0 \\ d_i \end{bmatrix} = 0_{3 \times 1}. \quad (24)$$

Similarly, $p_{ij}$ is also collinear with its projection $q_{ij}$:

$$(A^{-1}q_{ij}) \times p_{ij} = [x_{ij}]_x \begin{bmatrix} H_jH_j & -2n_i \end{bmatrix} \begin{bmatrix} p_0 \\ d_j \end{bmatrix} = 0_{3 \times 1}. \quad (25)$$

Based on them, we obtain a linear system of $p_0, d_1$ and $d_2$:
where \( h_i = [x_i], h_i' = [x_i], h_i'' = [x_i], H_i n_j \). By computing the eigenvector corresponding to the smallest eigenvalue of \( K'K \), \( (p_0, d_1, d_2)\) can be determined up to a scale factor. In this paper, we choose the scale that normalizes \( d_1 = 1 \).

Notice that Eq (26) apparently has 15 equations, but only 10 of them are linearly independent. This is simply because each of the cross products by Eqs (23) and (25) has only two independent constraints by definition.

Also, as discussed in Section 3.5.1, the above algorithm can be extended to third or further reflections and \( N_r \geq 3 \) cases as well. In \( N_r = 3 \), considering the reflection of \( p_{23} \) by \( \pi_4 \) as \( \lambda_{123}q_{123} = A p_{123} = AS A p_{23} \), we have

\[
[x_{123}]_x = \begin{bmatrix} (H_i H_i H_i) \end{bmatrix}^\top \begin{bmatrix} p_0 \end{bmatrix} = \begin{bmatrix} d_1 \end{bmatrix} = 0_{15 \times 1}, \tag{27}
\]

**3.5.3 Kaleidoscopic Bundleadjustment**

Once estimated the mirror normals \( n_i \) and the distances \( d_i (i = 1, 2) \) linearly, the triangulation from kaleidoscopic projections of a single 3D point can be given in a DLT manner by solving:

\[
K' p_0 = K' \delta, \tag{28}
\]

as \( p_0 = - (K'^\top K')^{-1} K'^\top K' \delta \), where \( \delta = (d_1, d_2) \). \( K' \) is the 15x3 matrix corresponding to the first three columns of \( K \):

\[
K' = \begin{bmatrix} x_0 & x_1 \end{bmatrix}_x h_i^\top, h_i' h_i'' h_i'^\top h_i''^\top, \tag{29}
\]

and \( K'' \) is the 15x3 matrix corresponding to the 4th and 5th columns of \( K \):

\[
K'' = \begin{bmatrix} 0_{3 \times 1} & 0_{3 \times 1} \\
-2[x_0, n_1] & 0_{3 \times 1} \\
-2[x_1, n_1] & -2[x_2, n_2] \\
-2[x_2, n_1] & -2[x_1, n_2] \\
\end{bmatrix}, \tag{30}
\]

By reprojecting this \( p_0 \) to each of the chambers as

\[
\lambda q_i = A p_0, \quad \lambda q_i = A S q_i (i = 1, 2), \quad \lambda q_{i,j} = A S q_j p_0 (i, j = 1, 2, i \neq j), \tag{31}
\]

we obtain a reprojection error as

\[
E(n_1, n_2, d_1, d_2) = q_0 - q_1, e_1, e_2, e_1' e_2', \tag{32}
\]

where \( e_i = q_i - q_0 \) and \( e_{i,j}' = q_{i,j}' - q_{i,j}' \). By minimizing \( \|E(\cdot)\|^2 \) nonlinearly over \( n_1, n_2, d_1, d_2 \), we obtain a best estimate of the mirror normals and the distances.

**3.6 Intrinsic Parameters Estimation**

As described in Section 3.5, the mirror parameters can be computed from the kaleidoscopic observations undistorted with the intrinsic parameters. Finally, the proposed method optimizes these parameters by minimizing the reprojection error as follow.

Suppose \( \hat{q} = (A, d, n, \delta) \) denotes the reprojected point from \( A, d, n \) and \( \delta \), the proposed method defines the reprojection error as,

\[
E(A, d, n, \delta) = \Sigma \|\hat{q} - q\|^2. \tag{33}
\]

By minimizing \( E(\cdot) \) nonlinearly over \( A, d, n \) and \( \delta \), we obtain a best estimate of the intrinsic parameters.

**3.7 Evaluations of Intrinsic Camera Calibration**

**3.7.1 Quantitative Evaluations with Synthesized Data**

**Experimental Environment** The performance of intrinsic camera calibration with synthesized data is evaluated in the following configuration.

The kaleidoscopic imaging system in this evaluation consists of three mirrors \( \pi_i (i = 0, 1, 2) \) whose normal vectors \( n_i \) of each mirror are set to \( (\cos \theta \cos \phi, \sin \theta \cos \phi, \sin \phi) \) with \( (\theta, \phi) = (-8, 0) \) for \( n_0 \), \( (\theta, \phi) = (186, 60) \) for \( n_1 \), and \( (\theta, \phi) = (190, -60) \) for \( n_2 \). The distance of them \( d_i \) are set to \( d_0 = 50 \text{mm}, d_1 = 55 \text{mm} \) and \( d_2 = 54 \text{mm} \). The image size is set to \( 1920 \times 1080 \). The ground truth of intrinsic parameters \((f, u_0, v_0)\) and \( d = (k_1, k_2, k_3, p_1, p_2) \) are set to \( (2700, 960, 540) \) and \( (0.001, -0.001, 0.001, 0.002, -0.002) \). Notice that the chamber labels are assigned correctly in this evaluation.

The evaluations in this section utilize the absolute error of each parameters as error metrics, that is \( E_p = |p - p_0| \) where \( p \) denotes each intrinsic camera parameters and \( p_0 \) denotes its ground truth.

**Results** In this evaluation, the performance of our proposed method is compared with the most common intrinsic camera calibration algorithm of Zhang [40] (Baseline 1). For comparison, we set a \( 3 \times 4 \) chessboard whose distance of each chess corner is 10mm in the base chamber whose distance from the camera is 160mm in the kaleidoscopic imaging system. As illustrated in Figure 11(1), we utilize the projections of its reflections in 10 chambers, that is until second reflection, as input for each method.

In addition, we compare the results by [40] with the ideal configuration as a reference data, that is the input data consists of 30 observations of above chessboard scattered covering the camera’s field-of-view in the 3D space randomly as illustrated in Figure 11(b) (Baseline 2).
Figure 12 shows average estimation errors of each intrinsic camera parameters over 30 trials at different noise $\sigma_q$. The $\sigma_q$ denotes the standard deviation of zero-mean Gaussian pixel noise injected to the observations $q$. In each trial, the initial values of $A$ and $d$ are generated by adding random noise whose level is less than 5% to the ground truth of each parameters. The red, and blue are results by the proposed method and Baseline 1 and green line is the result by Baseline 2.

In Figure 12, we can see that the proposed method outperforms Baseline 1 and is comparable with the Baseline 2. Especially, distortion parameters estimated by the proposed method have higher precision that those by the other methods. We consider that this improvement is caused by the properties of the kaledoscopic imaging system, i.e., the reference objects are scattered in an isotropic manner by mirror reflections and the reflections are strictly constrained each other with mirror parameters.

From these results, we can conclude that our method can estimate intrinsic parameters in an ideal and a noised environment robustly compared with baseline methods.

### 3.7.2 Qualitative Evaluations with Real Data

**Experimental Environment** As illustrated in Figure 13, the capture setup consists of one camera (Nikon D7000, 4948 × 3280 resolution) with a single-focus lens of 28mm and three planar mirrors. In this evaluation, we compare our method with Zhang [40] (Baseline 1) as with 3.7.1. The reference object for Baseline 1 is a 5 × 8 chessboard whose distance of each chess corner is 4.45mm. As an ideal configuration for Zhang [40], we utilize a larger reference object which has 7 × 10 chessboard with 2.05mm corners (Baseline 2).

**Results** Table 2 reports the each parameters estimated by each method. Notice that proposed method and [40] with the kaleidoscopic imaging system utilize 10 projections of the reference object, that is projections in each chambers (Figure 13(a)), and citezhang2000flexible with the ideal configuration utilizes 15 images of the larger reference object (Figure 13(b)). From these results, while the intrinsic parameters estimated by Baseline 1 are

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Proposed</th>
<th>Baseline 1</th>
<th>Baseline 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f$</td>
<td>5882.2</td>
<td>12318.3</td>
<td>6283.4</td>
</tr>
<tr>
<td>$u_0$</td>
<td>2474.4</td>
<td>2462.7</td>
<td>2427.6</td>
</tr>
<tr>
<td>$v_0$</td>
<td>1639.4</td>
<td>1622.9</td>
<td>1686.2</td>
</tr>
<tr>
<td>$k_1$</td>
<td>-0.0103</td>
<td>-0.3381</td>
<td>-0.1059</td>
</tr>
<tr>
<td>$k_2$</td>
<td>0.0143</td>
<td>-86.1886</td>
<td>0.3035</td>
</tr>
<tr>
<td>$k_3$</td>
<td>-0.0020</td>
<td>-0.3820</td>
<td>-0.9600</td>
</tr>
<tr>
<td>$p_1$</td>
<td>0.0056</td>
<td>0.0740</td>
<td>0.00109</td>
</tr>
<tr>
<td>$p_2$</td>
<td>0.0036</td>
<td>0.0625</td>
<td>0.0006</td>
</tr>
</tbody>
</table>

### Table 3 Reprojection error by each method for evaluating robustness.

<table>
<thead>
<tr>
<th>Data</th>
<th>Proposed</th>
<th>Baseline 1</th>
<th>Baseline 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.526</td>
<td>0.487</td>
<td>0.524</td>
</tr>
<tr>
<td>2</td>
<td>1.735</td>
<td>13.014</td>
<td>0.933</td>
</tr>
</tbody>
</table>
different greatly compared with those by Baseline 2 which is considered as the ground truth, the parameters estimated by proposed method are close to those by Baseline 2.

Table 3 shows the average of reprojection errors by each method for evaluating their robustness. Data 1 consists of 10 observations of 5 × 8 chessboard used for estimating intrinsic parameters by proposed method and Baseline 1. Data 2 consists of 5 images of large 7 × 10 chessboard used for estimating intrinsic parameters by Baseline 2. As to the Baseline 1, based on the fact that the reprojection errors with Data 2 get worse apparently, the intrinsic parameters estimated by Baseline 1 does not have robustness. On the other hand, the reprojection errors by our proposed method with both Data 1 and Data 2 are small enough and this shows that it estimates robust intrinsic parameters.

From above results, we can state that our proposed method works properly in the real configuration.

3.7.3 Degenerate Cases

The both proposed algorithms of chamber assignment and mirror parameters estimation are based on the kaleidoscopic projection constraint (Eq (11)) satisfied by more than or equal to second reflections. Therefore these algorithms do not work in the following two cases. (1) If the two mirror are parallel, the mirror normals are not computable by solving Eq (14), Eq (19) and Eq (20) because the constraints are linearly dependent. (2) If the second reflections are not observable due to the angle of view or discontinuities, the mirror normals are not computable. Especially, in case of using more than three mirrors, discontinuities are more likely to happen in general, and finding the second reflections itself become difficult (Figure 14).

3.7.4 Application: 3D Reconstruction

The kaleidoscopic system can be recognized as a virtual multi-view capture system. Here we evaluated the feasibility of 3D reconstruction with estimated mirror parameters.

Figure 15 shows our kaleidoscopic capture setup and reconstructed result. In this evaluation, the intrinsic parameter A of the camera (Nikon D600, 6016×4016 resolution) is calibrated beforehand [40]. As a target object, we utilized a cat (about 4×5×1 cm) with three planar first surface mirrors. The projector (MicroVision SHOWWX+ Laser Pico Projector, 848×480 resolution) is used to cast line patterns to the object for simplifying the correspondence search problem in a light-sectioning fashion, and the projector itself is not involved in the calibration w.r.t. the camera and the mirrors. In these evaluations, the mirror parameters are estimated by the algorithm introduced in Section 3.5.

Figure 15 shows a 3D rendering of the estimated 3D shape using the mirror parameters calibrated by the proposed method, while the residual reprojection error indicates the parameters can be further improved for example through the 3D shape reconstruction process itself [8]. From these results, we can conclude that the proposed provides a sufficiently accurate calibration for 3D shape reconstruction.

4. Mirror-based Extrinsic Camera Calibration

This section provides a novel planar-mirror based extrinsic camera calibration algorithm in case where a camera cannot observe a reference object directly.

4.1 Measurement Model

As illustrated by Figure 16, we denote a camera by C and a mirror by \( \pi_j \) \( (j = 1, \cdots, N_e) \). We use \( \{C\} \) to describe the camera C coordinate system which is used as the world coordinate system in this section.

Let \( p^{(X)}_i \) \( (i = 1, \cdots, N_p) \) denote the positions of the reference points given a priori in its local coordinate system \( X \). These positions are modeled as

\[
p^{(C)}_i = R \cdot p^{(X)}_i + t \quad (i = 1, \cdots, N_p),
\]

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

Based on the planar mirror geometry introduced in Section 2.2, the reference point \( p^{(C)}_i \) and its reflection \( p^{(C)}_j \) satisfy

\[
p^{(C)}_i = -2(n_j^T p^{(C)}_j + d_j n_j) + p^{(C)}_j.
\]

By removing \( p^{(C)}_j \) from Eq (34) and Eq (35), we obtain

\[
R \cdot p^{(X)}_i + t = -2(n_j^T p^{(C)}_j + d_j n_j) + p^{(C)}_j.
\]

This is the fundamental equation which describes our measurement model.

4.2 Orthogonality Constraint on Mirror Reflections

Consider a reference point \( p \) and its two mirrored points \( p'_{j_1}, p'_{j_2} \) by two different mirror planes \( \pi_{j_1}, \pi_{j_2} \) respectively. The axis vector \( m_{j_1} \times m_{j_2} \) lying along the intersection of the two mirror planes is expressed as the cross product of each mirror normals,
The extrinsic parameters obtained by solving PnP with mirrored poses based on the orthogonality constraint projected to the image plane, \( m_{ij} = n_i \times n_j \). This axis vector \( m_{ij} \) satisfies the following orthogonality constraint \([33]\) (Figure 17),

\[
(p_j - p_j')^T \cdot m_{ij} = 0. \tag{37}
\]

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

### 4.3 Extrinsic Camera Calibration Using Orthogonality Constraint

This section introduces the proposed algorithm which analytically determines the camera extrinsic parameter from the projections of \( N_p \) reference points observed via \( N_r \) different mirror poses based on the orthogonality constraint.

Algorithm 1 shows an overview of our calibration algorithm. Firstly, we solve the PnP problem from mirrored reference points and obtain their 3D positions \( p_j \) by EPNP \([19]\) (\( N_r > 3 \) case). 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 \( p_j' \), and hence we ignore such flipped extrinsic parameters. Secondly, we estimate the axis vectors of each pair of mirror planes and obtain the mirror normals from them based on the orthogonality constraint. Finally, we compute \( R \) and \( t \) by solving a large system of linear equations.

#### 4.3.1 Computing the axis vector from mirror planes

As described in Sec 4.2, the axis vector \( m_{ij}(j, j' = 1, \cdots, N_r, j \neq j') \) and two mirrored points \( p_j', p_j''(i = 1, 2, \cdots, N_p) \) satisfy the orthogonality constraint (Eq \(37)\).

By applying this orthogonality constraint to \( N_p \) mirrored reference points \( p_j' \), we obtain:

\[
\begin{bmatrix}
(p_j' - p_j)^T \\
(p_j'' - p_j')^T \\
\vdots \\
(p_j' - p_j')^T
\end{bmatrix}
\begin{bmatrix}
m_{ij}
\end{bmatrix} = \begin{bmatrix} Q_{ij} \end{bmatrix} \cdot \begin{bmatrix} m_{ij} \end{bmatrix} = 0. \tag{38}
\]

An axis vector \( m_{ij} \) can be computed as the right-singular vector corresponding to the smallest singular value of \( Q_{ij} \).

#### 4.3.2 Computing the normal vector of a mirror plane

The axis vector \( m_{ij} \) is perpendicular to the normal vectors \( n_j \) and \( n_j' \) of each mirror planes \( \pi_j \) and \( \pi_j' \) respectively. That is,

\[
n_j' \cdot m_{ij} = 0, \tag{39}
\]

\[
n_j' \cdot m_{ij} = 0.
\]

In using \( N_r \) mirror poses, we obtain \((N_r - 1)\) equations of Eq \(39)\) for one normal vector \( n_j \). By collecting these equations, we have

\[
S_j \cdot n_j = 0,
\]

\[
S_j = (m_{ij} m_{i2} \cdots m_{ij-1} m_{ij+1} \cdots m_{jN_p})^T. \tag{40}
\]

where \( S_j \) is a \((N_r - 1) \times 3 \) matrix. A normal vector \( n_j \) can be computed as the right-singular vector corresponding to the smallest singular value of \( S_j \).

This equation also indicates that we have to provide \( N_r \geq 3 \) mirror poses in order to estimate \( n_j \), because the degree of freedom of \( n_i \) is 2.

### 4.3.3 Computing Extrinsic Parameters

Up to this point, we obtain the 3D positions of mirrored reference points \( p_j'(i = 1, \cdots, N_p, j = 1, \cdots, N_r) \) and mirror normals \( n_j \). The 3D positions of reference points \( p_j = (x_i, y_i, z_i) \) are supposed to be given a priori in its local coordinate system \( X \). By substituting these known parameters into Eq \(36)\), we can derive a large system of linear equations:

\[
AZ = B. \tag{41}
\]

where

\[
A = \begin{bmatrix}
I_3 & 2n_1 & 0_{3 \times 1} & \cdots & 0_{3 \times 1} & x_1 I_3 & y_1 I_3 & z_1 I_3 \\
I_3 & 2n_1 & 0_{3 \times 1} & \cdots & 0_{3 \times 1} & x_2 I_3 & y_2 I_3 & z_2 I_3 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
I_3 & 3n_1 & 0_{3 \times 1} & \cdots & 0_{3 \times 1} & x_N r_3 & y_N r_3 & z_N r_3 \\
I_3 & 0_{3 \times 1} & 2n_2 & \cdots & 0_{3 \times 1} & x_1 I_3 & y_1 I_3 & z_1 I_3 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
I_3 & 0_{3 \times 1} & 2n_2 & \cdots & 0_{3 \times 1} & x_N r_3 & y_N r_3 & z_N r_3 \\
I_3 & 0_{3 \times 1} & 3n_1 & \cdots & 2n_N r_3 & x_N r_3 & y_N r_3 & z_N r_3
\end{bmatrix},
\]

\[
Z = \begin{bmatrix}
1^T & d_1 & d_2 & \cdots & d_{N_r} & r_1^T & r_2^T & r_3^T
\end{bmatrix}^T, \tag{43}
\]

\[
B = \begin{bmatrix}
B_1 & B_2 & \cdots & B_{N_p}
\end{bmatrix}^T, \tag{44}
\]

\[
B_j = \begin{bmatrix}
b_j^1 & b_j^2 & \cdots & b_j^N_p
\end{bmatrix}^T, \tag{45}
\]

\[
b_j = (-2n_j^T p_j + p_j')^T. \tag{46}
\]
The vectors $r_1$, $r_2$ and $r_3$ denote the first, second and third column of the rotation matrix $R$. From $N_p$ mirror poses and $N_p$ reference poses, we have $12 + N_p$ unknown parameters and $3 \times N_p \times N_p$ equations. Hence, when $N_p \geq 3$ and $3 \times N_p \times N_p > 12 + N_p$, we can solve the Eq (41) by $Z = A^* B$, where $A^*$ is the pseudo-inverse matrix of $A$.

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 $g^{(3)} = (x, y, 0)^T$ and we cannot compute the third column vector $r_3$ of the rotation matrix. In this case, we compute $r_3$ as the cross product of first and second column vector $r_1, r_2$, that is $r_3 = r_1 \times r_2$.

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

Now we obtain columns of rotation matrix $r_1, r_2, r_3$ and translation vector $t$ linearly, but $r_1, r_2$, and $r_3$ do not necessarily satisfy the following constraints as a rotation matrix due to noise:

\[
\begin{align*}
|r_1| &= |r_2| = |r_3| = 1, \\
r_1^T r_1 &= r_2^T r_2 = r_3^T r_3 = 0.
\end{align*}
\]

Here, we solve the orthogonal Procrustes problem [9] and obtain a rotation matrix which satisfies Eq (47) and is closest to the original linear solution as proposed in Zhang’s method [40]. That is $R = UV^T$, where $U$ and $V$ are given by as the SVD of the original matrix $(r_1 \ r_2 \ r_3) = U \Sigma V^T$.

### 4.4 Non-Linear Refinement of Extrinsic Parameters

In general, obtained extrinsic parameters can be refined by non-linear optimization [37]. Here, we minimize following reprojection error function,

\[
E_{rep} = \sum_{j=1}^{N_p} \sum_{i=1}^{N_f} |q_{ij} - \tilde{q}_{ij}(R, t, n_j, d_j)|,
\]

where $q_{ij}(R, t, n_j, d_j)$ denotes the reprojected point calculated from estimated parameters. We solved this non-linear optimization problem of Eq (48) with Levenberg-Marquardt algorithm.

### 4.5 Performance Evaluations

This section provides experimental evaluations using synthesized and real data in $N_p > 3$ case. In these evaluations, we compare our method with state-of-the-arts proposed by Sturm et al. [33] and by Rodrigues et al. [30] with non-linear refinement.

#### 4.5.1 Quantitative Evaluations with Synthesized Data

**Experiment Environment:** To synthesize data, we used the following experiment setup by default. The matrix of intrinsic parameters, $K$, consists of $(f_x, f_y, c_x, c_y)$; $f_x$ and $f_y$ represents the focal length in pixels, and $c_x$ and $c_y$ represent the 2D coordinates of the principle point. We set them to $(500, 500, 300, 250)$ in this evaluation respectively.

The normal vectors $n_j (j = 1, \ldots, N_p)$ of mirror poses $\pi_j$ are set to $(\sin \theta_1, \sin \theta_2 \cos \theta_3, \sin \theta_2 \sin \theta_3, \cos \theta_1 \cos \theta_2 \sin \theta_3, \cos \theta_1 \cos \theta_2 \cos \theta_3)$ where $\theta_0 (k = x, y, z)$ is the angle respect to each axis, and drawn randomly within the ranges of $[-20 \leq \theta_x \leq 20, 160 \leq \theta_y \leq 200, -20 \leq \theta_z \leq 20]$. The distance between each mirror plane and camera center was set to 300 mm.

The reference object consists of $N_p$ reference poses forming a grid pattern and the distance between each reference point is 50mm. 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 $R = R_1(\theta_1)R_2(\theta_2)R_3(\theta_3)$, and we set random values to each angles $\theta_1, \theta_2$, and $\theta_3$ within $[-10 : 10]$ respectively. The position $t$ is generated of each trial by assigning a random value within $[-5 : 5]$ to each element of $t$.

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 poses.

(c) $N_x$: the number of mirror poses.

Table 4 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 [33] and [30].

Throughout this evaluation, we used three metrics to measure the performance of each method; the Riemannian distance for $R$.
(\(E_\beta\)), the root mean square error for \(t\) (\(E_t\)), and reprojection error (\(E_P\)). The definitions of these metrics are detailed in [34].

**Results:** Figure 18 (Top) shows results for different standard deviations \(\sigma\) of pixel noise, (Middle) the number of reference points \(N_p\), and (Bottom) the number of mirrors \(N_m\).

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

### 4.5.2 Qualitative Evaluations with Real Data

**Experiment Environment:** We evaluated the performance of our proposed method with real data assuming calibration of a display-camera system, such as digital-signage, laptop computer and so on.

As illustrated in Figure 20 (Left), 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 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 \(N_p\) UXGA images of different mirror poses \(\pi_j (j = 1, \cdots, N_p)\) for calibration.

**Results:** Figure 19 (Top) and (Bottom) 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 observe only one line in Figure 19 (Bottom).

Figure 20 (Right) renders the estimated positions of the refer-
ence object by each method with \(N_p = 4\) and \(N_{\pi} = 3\) configuration. We can see that the reference objects estimated by each method are located near the baseline result, which is computed by Zhang’s method [40] via multiple mirrors and multiple reference objects. This precision is acceptable for applications using display-camera system such as gaze detection for digital-signage or gaze correction [18] in a video conference scenario.

From these results, we can observe that our method performs better than conventional methods [13, 30, 33] in real situation qualitatively and quantitatively.

### 4.6 Degenerate Case

Our algorithm does not work if it cannot compute enough axis vectors \(m_{jj}'\) 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 (40). These (1) and (2) cases has been originally observed by Sturm et al. [33]. (3) If reference points and the intersection of two mirrors \(\pi_i\) and \(\pi_j\) are on a same plane, the axis vector \(m_{ij}'\) is not be computable by solving Eq (38) though \(m_{ij}'\) does exist physically because of the following reason which makes the two rows of \(M_{ij}'\) corresponding to \(\pi_i\) and \(\pi_j\) be linearly dependent.

**Proposition 1.** If two reference points \(p'\) and \(p''\) and the intersection of two mirrors \(\pi_i\) and \(\pi_j\) are on a same plane, the two lines connecting results of different Householder transformations of \(p'\) and \(p''\), i.e., the lines connecting \(p'_i\) to \(p''_i\) and \(p'_j\) to \(p''_j\), are parallel (Figure 21).

**Proof.** Suppose the line connecting \(p'\) and \(p''\) intersects with the intersection of the two mirrors at \(O\) as shown in Figure 21. By definition of the reflection, the distance from \(p'\) to \(O\) is equal to the one from \(p'_i\) to \(O\). Similarly, the distance from \(p''\) to \(O\) is equal to the one from \(p''_j\) to \(O\). Also, these distances are equal to the ones from \(p'\) to \(O\), and from \(p''\) to \(O\) respectively. Here \(\Delta p'p_i'p_j\) and \(\Delta Op''p_i'p_j\) are isosceles triangles sharing the apex \(O\). Therefore, the two lines \(p'_i\) to \(p'_j\) and \(p''_i\) to \(p''_j\) are parallel. \(\square\)

These three degenerate cases can be detected by observing the rank of \(M_{ij}'\) in Eq (38). If the rank is less than 2, we can discard the mirror pair and try with more mirrored images in practice.

---

**Fig. 20** (Left:) A capture setup. (Right:) estimated positions of the reference object by the proposed method (red), by [30](blue and cyan), by [33](green) and by [40](black). Notice that this figure renders estimate positions of \(p\) and \(X_i\) in \(C\), and therefore \(C_i\) is located at \((0, 0, 0)^T\).

**Fig. 19** Estimation error of each parameter with real data in changing (Top) the number of reference points \(N_p\) and (Bottom) the number of mirror poses \(N_{\pi}\).

**Fig. 21** Degenerate case
5. Extrinsic Camera Calibration using Human Cornea Reflections

This section provides a human cornea reflection based extrinsic calibration algorithm of a camera and a reference object located out of the camera’s filed-of-view.

5.1 Measurement Model

As illustrated in Figure 22, we denote a center of the spherical mirror, its radius, and a reflection point of a reference point \( p_i \) on the cornea sphere by \( S, r \) and \( m \), respectively. Let us assume a unit vector from \( m \) to \( p_i \) expressed as \( u_i \), \( p_i \) is expressed as \( p_i = k_i u_i + m_i \), where \( k_i \) is the distance between \( m_i \) and \( p_i \). This \( p_i \) also satisfies \( p_i = R p_i^{(X)} + t \), where \( R \) and \( t \) are the extrinsic parameters between the camera and the reference object. From these equations, we obtain the following equation:

\[
R p_i^{(X)} + t = k_i u_i + m_i. \tag{49}
\]

This Eq.(49) has 10 unknown parameters, that is \( R, t, S \) and \( r \), and it is defined as the measurement model in this configuration.

5.2 Approach: Solving Absolute Orientation Problem

Determining extrinsic parameters between two coordinate systems, such as \([C]\) and \([X]\), through the use of a set of corresponding points in each coordinate system is known as the Absolute Orientation Problem. By solving this problem, we can obtain the extrinsic parameters from at least three point correspondences. Since the 3D positions of reference points \( p_i^{(X)} \) in \([X]\) are supposed to be given a priori, we estimate the 3D positions of reference point \( p_i \) in \([C]\) by estimating \( k_{in,p_i}, S \) and \( r \). In order to estimate them, we introduce the following two constraints, a geometric model of the cornea sphere and the equidistance constraint.

5.2.1 Cornea Sphere Parameters Estimation Based on its Geometric Model

Based on the geometric model [23], the radius of the cornea is recognized as the average of it based on [29], and the limbus projection is modeled as an ellipse represented by five parameters: the center, \( i_k \), the major and minor radii, \( r_{max} \) and \( r_{min} \), respectively, and rotation angle \( \phi \). Since the depth of a tilted limbus is much smaller than the distance between camera and the cornea sphere, we assume weakly perspective projection. Under this assumption, the 3D position of the center of limbus \( L \) is expressed as \( L = dK^{-1}i_L \), where \( d \) denotes the distance between the center of camera, \( O \), and the center of limbus \( L \), and is expressed as \( d = f \cdot r_{d}/r_{max}. \) \( f \) and \( K \) represent the focal length in pixels and intrinsic parameters, respectively. Gaze direction \( g \) is approximated by the optical axis of the eye, and is theoretically determined by \( g = [\sin \tau \sin \phi, -\sin \tau \cos \phi, -\cos \tau]^T \), where \( \tau = \pm \arccos(r_{min}/r_{max}) \); \( \tau \) corresponds to the tilt of the limbus plane with respect to the image plane. Since the center of cornea sphere, \( S \), is located at distance \( d_{LS} = \sqrt{r_s^2-r_L^2} = \sqrt{7.72^2-5.62^2} \approx 5.3 \text{ mm} \), the radius of the cornea sphere from \( L \), we compute \( S \) as \( S = L - d_{LS}g \). In this way, we estimate \( S \) from the ellipse parameters of the limbus projected onto the image plane.

5.2.2 Equidistance Constraint

To obtain \( k_i \), we introduce the Equidistance Constraint. The Equidistance Constraint states that the distance from reference point \( p_i \) to the center of cornea sphere \( S \) is equal to the distance from the center of the camera, \( O \), to \( S \).

If this equidistance constraint is satisfied, triangle \( \angle OSP_i \) is an isosceles triangle that satisfies \( |OS| = |p_i - S| \). Let the unit vector representing the bisector of \( \angle OSP_i \) be denoted as \( l_i \), and a point on this bisector be denoted as \( a_i \). Triangle \( \angle Oa_ip_i \) is also an isosceles triangle that satisfies \( |O - a_i| = |p_i - a_i| \). Additionally, among the point set \( O, a_i \) and \( p_i \), the laws of reflection can be established at \( a_i \) where \( I_i \) is used as the normal vector. Therefore, when \( a_i \) is the intersection of the bisector and the surface of cornea sphere, \( a_i \) is equal to the \( m_i \) of reference point \( p_i \).

Equidistance Constraint therefore, when user sets his/her center of cornea sphere such that the equidistance constraint does hold, triangle \( \angle Oa_ip_i \) should be an isosceles triangle that satisfies \( |p_i - m_i| = |O - m_i| \), that is \( k_i = k_i' \).

By introducing above constraints, we compute 3D positions of reference point \( p_i \) and obtain extrinsic parameters \( R \) and \( t \) by solving the Absolute Orientation Problem. This algorithm works with the minimal configuration where three reference points and one mirror pose.

5.3 Performance Evaluations

Figure 23 (Left) shows the configuration of the evaluation. The camera is set at the top of display which has three reference points. Figure 23 (b) renders the estimated positions of the reference points. It is difficult to obtain the ground truth of extrinsic parameters in any real configuration, so we used the baseline method introduced in Section 4 as the reference parameters. From this result, we can see that the positions estimated by the proposed method are almost identical to those of baseline. Notice that the difference in these rotation matrices for x-axis, y-axis and z-axis are 2.44, 6.88, and 0.49 degrees, respectively (\( E_R = 0.1278 \)), and \( E_t \) is 27.4989 mm in Figure 23.
6. Conclusion

The geometric camera calibration algorithms presented in this paper are based on mirror reflections. Particular focus was placed on cases in which a fundamental assumption of conventional camera calibration methods, that is, the camera can directly observe a reference object for which the geometry is known, does not hold. Intrinsic and extrinsic calibration algorithms using a mirror as a supporting device were introduced that enable calibration in such cases.

For cases in which a known reference object is not available for intrinsic camera calibration, the use of multiple planar mirrors was introduced. These mirrors consist of a kaleidoscopic imaging system that uses reflections as a reference object in the parametric 3D model described in Section 3. The proposed algorithm is focused mainly on the perspective camera model with simple lens distortion, as described in Eq. (3). Along with the development of imaging systems, a wide variety of camera models [10, 25] and lens models [3] have been proposed. Adaptation to these models should be investigated as the future works.

For cases in which the camera cannot directly observe a reference object for intrinsic camera calibration due to a physical constraint on the imaging system, planar-mirror-based and human-cornea-based extrinsic camera calibration algorithms were introduced in Section 4 and Section 5. The proposed algorithms use single reflection of a reference object. However, in the case of calibration of multiple widely scattered cameras or an omnidirectional camera, the reference object cannot be observable via a single reflection. Our future works include extrinsic camera calibration via multiple reflections for such cases.

References


