

Camera Calibration Based on Mirror Reflections

Kosuke Takahashi

Abstract

This dissertation 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 for which the geometry is known.* 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. In other words, the mirror extends the camera’s field-of-view. This dissertation utilizes these features and tackles two challenges of the geometric camera calibration.

The first challenge is intrinsic camera calibration when a known reference object is not available. This challenge is met by using multiple planar mirrors consisting of a kaleidoscopic imaging system used to generate reflections of an unknown 3D point. The reflections are used as a reference object which can be recognized as a parametric 3D model. The intrinsic calibration with the kaleidoscopic imaging system requires that three problems be solved: chamber assignment of the kaleidoscopic projections, estimation of the mirror parameters, and estimation of the intrinsic parameters. These problems are solved by the introduction of a novel geometric constraint, the *kaleidoscopic projection constraint*. This constraint provides multiple linear equations for the mirror parameters for a single 3D point and is satisfied by projections of high-order reflections. Evaluations with synthesized and real data demonstrates that the proposed technique can work properly and reports the robustness of it in comparison with conventional methods.

The second challenge is extrinsic camera calibration when the camera cannot directly observe a known reference object due to a physical constraint on the imaging system. This challenge is met by using two types of mirror reflections to enable the camera to directly observe the reflections of a known reference object. The first technique utilizes the planar mirror reflection. Since the pose and position 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* is introduced that is satisfied among reflections by using multiple mirror poses. Use of this constraint enables the extrinsic parameters to be estimated using the minimal configuration. The second technique utilizes the

human cornea reflection. The algorithm is based on the fact that the surface of the human eye reflects light like a mirror and thus can be modeled as a spherical mirror. The introduction of a geometric model of the human cornea thus provides a closed-form solution for the extrinsic parameters. Evaluations with synthesized and real data in practical scenario demonstrate the performance of each method in comparison with the conventional approaches.

Acknowledgement

Firstly I would like to express my deepest gratitude to my supervisors, Professor Takashi Matsuyama and Senior Lecturer Shohei Nobuhara. They have guided me to explore this interesting world of research by providing insightful advices and constant encouragement in these 8 years. I will never forget these exciting days.

I also wish to express my gratitude to my thesis committee, Professor Tetsuya Matsuda and Professor Michihiko Minoh, for taking time to review my thesis and give helpful comments.

Next, I would like to express my gratitude to Professor Akihiro Sugimoto at National Institute of Informatics. He supported my master thesis and the early stage of this work.

My thanks also go to Associate Professor Hiroaki Kawashima, secretaries of our lab, Ms. Chieko Funato, Ms. Aya Inoue, Ms. Asako Yoshimura, Ms. Kazuyo Hashimoto and Ms. Shiho Kimura, and all the members of our lab, to Mr. Kohei Haze, Mr. Hitoshi Itoh, Ms. Mai Nishimura, Dr. Kei Shimonishi, and Mr. Ryo Kawahara. I really enjoyed my daily life in the laboratory.

I deeply appreciate to Dr. Dan Mikami at NTT Media Intelligence Laboratories. He has encouraged and supported me since I joined his research team in NTT lab. This thesis cannot be completed without his support. Also I thank to all the members of NTT lab, especially to my colleagues Mr. Daisuke Ochi and Ms. Mariko Isogawa, for their daily assistance in NTT.

Special thanks go to my like-minded friends, Mr. Hiroshige Umino, Mr. Yuku Takahashi and Mr. Naoki Hatta. They are my best friends and their activities and products in their field have always encouraged me.

Furthermore, I wish to express my deepest thanks to my family, my father Etsuo Takahashi, my mother Tomoko Takahashi, and my sister Yuko Takahashi, for everything they have done for me.

Last but not least, I wish to express my greatest love and gratitude to my wife Itsumi.

Contents

1	Introduction	1
1.1	Background	1
1.2	Problem Statement	4
1.3	Contributions	5
1.4	Dissertation Structure	6
2	Camera Calibration and Mirror Geometry	7
2.1	Geometric Camera Calibration of Perspective Camera	7
2.1.1	Perspective Camera Model	7
2.1.2	Related Works of Camera Calibration	9
2.2	Mirrors in Computer Vision	10
2.2.1	Planar Mirror Geometry	10
2.2.2	Related Works on Mirrors in Computer Vision	11
3	Mirror-based Intrinsic Camera Calibration	15
3.1	Kaleidoscopic Imaging System	16
3.2	Problem Description	18
3.3	Kaleidoscopic Projection Constraint	19
3.4	Chamber Assignment	20
3.4.1	Base Structure	20
3.4.2	Geometric Constraints on Kaleidoscopic Projections	23
3.4.3	Discontinuity-aware Label Propagation	24
3.4.4	Chamber Assignment Algorithm	27
3.5	Mirror Parameters Estimation	28
3.5.1	Mirror Normals	28
3.5.2	Mirror Distances	29
3.5.3	Kaleidoscopic Bundleadjustment	31

3.6	Intrinsic Parameters Estimation	32
3.7	Performance Evaluations of Chamber Assignment	32
3.7.1	Experimental Environment	33
3.7.2	Quantitative evaluations with synthesized data	33
3.7.3	Qualitative evaluations with real data	35
3.8	Performance Evaluation of Mirror parameter estimation	36
3.8.1	Quantitative evaluations with synthesized data	39
3.8.2	Qualitative Evaluations with Real Data	41
3.9	Performance Evaluations of Intrinsic Camera Calibration	42
3.9.1	Quantitative Evaluations with Synthesized Data	42
3.9.2	Qualitative Evaluations with Real Data	45
3.10	Discussion	46
3.10.1	Ambiguity of chamber assignment	46
3.10.2	RANSAC or PROSAC Approach for Chamber Assignment	47
3.10.3	Degenerate Cases	47
3.10.4	3D Reconstruction with Kaleidoscopic Imaging System	48
3.11	Summary	49
4	Mirror-based	
	Extrinsic Camera Calibration	51
4.1	Measurement Model	52
4.2	Orthogonality Constraint on Mirror Reflections	53
4.3	Extrinsic Camera Calibration Using Orthogonality Constraint	54
4.3.1	Unique Solution of Three P3P Problems	54
4.3.2	Computing the axis vector from mirror planes	56
4.3.3	Computing the normal vector of a mirror plane	56
4.3.4	Computing Extrinsic Parameters	57
4.3.5	Linear Refinement of Rotation Matrix by Solving the Orthogonal Procrustes Problem	58
4.3.6	Non-Linear Refinement of Extrinsic Parameters	58
4.4	Performance Evaluations in $N_p = 3$ case	58
4.4.1	Quantitative Evaluations with Synthesized Data	59
4.4.2	Qualitative Evaluations with Real Data	62
4.4.3	Results	64

4.5	Performance Evaluations in $N_p > 3$ case	65
4.5.1	Quantitative Evaluations with Synthesized Data	65
4.5.2	Qualitative Evaluations with Real Data	68
4.6	Discussion	68
4.6.1	Degenerate Case	68
4.6.2	Sufficiency of the Orthogonality Constraint	70
4.7	Summary	71
5	Extrinsic Camera Calibration using Human Cornea Reflections	73
5.1	Measurement Model	74
5.2	Approach 1: Solving Absolute Orientation Problem	76
5.2.1	Estimating Parameters of Cornea Sphere Based on its Geometric Model	77
5.2.2	Equidistance Constraint	78
5.3	Approach 2: Direct Derivation Approach	79
5.3.1	Using Basis Vector Representation of 3D Reference Point Position	79
5.3.2	Derivation of Linear Equation for Estimating Extrinsic Parameters	80
5.3.3	Solving Non-linear Optimization Problem	81
5.4	Performance Evaluations of Approach 1	83
5.4.1	Quantitative Evaluations with Synthesized Data	83
5.4.2	Real Data	84
5.5	Performance Evaluations of Approach 2	87
5.5.1	Quantitative Evaluations with Synthesized Data	88
5.5.2	Real Data	89
5.6	Discussion	91
5.6.1	Effects of Differences among Individuals	92
5.6.2	Equidistance constraint in a real situation	92
5.6.3	Optimization of Approach 1	93
5.6.4	Validity of Using Reprojection Error as Criteria for Detecting Local Minimum	95
5.7	Summary	96
6	Conclusion	99
6.1	Summary	99
6.2	Future Works	100

Chapter 1

Introduction

1.1 Background

A ray omitted 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 a 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 [TLP02] or noise on the image sensor [HK94], estimation of the radiometric response function [MN99], and so on.

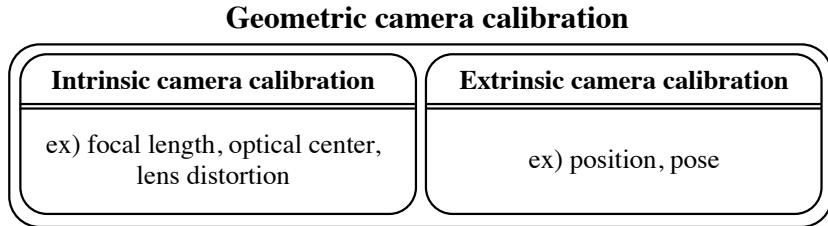


Figure 1.1: The overview of the geometric camera calibration.

This dissertation 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.

As illustrated in Figure 1.1, 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 [Tsa86] 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 [Zha00] used a known reference point on a single plane as a reference object, *e.g.*, a chessboard, to estimate the camera parameters. Lepetit *et al.* [LMNF09] obtained the extrinsic parameters by solving a PnP (perspective-n-point) problem using the correspondences between n points for which the 3D coordinates were known and their projections.

As described above, 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 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.

1.1. Background

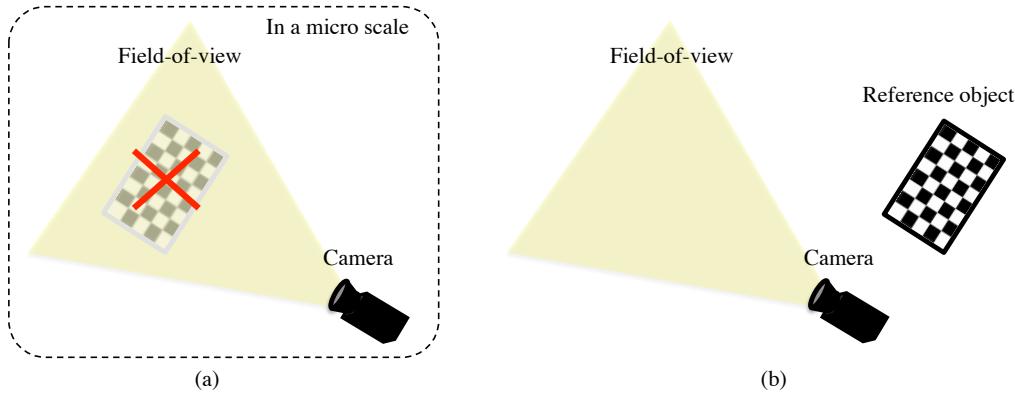


Figure 1.2: Example problems with previous calibration methods: (a) using a known reference object is not practical in terms of scale; (b) a reference object is not observable directly from the camera.

Figure 1.2(a) illustrates the case in which a known reference object is not practical in terms of scale. In addition, Figure 1.2(b) illustrates the case in which a reference object is not observable from the camera. For a vision-based robot [DK02] with a display-camera system [HDKM10, KPB⁺12, FY07], although the extrinsic parameters between the camera and each part of the system (a robot arm, the display, and so on) 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 method introduced in this dissertation for solving these two problems uses 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 illustrated in Figure 1.3, the 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 1.3(a)) and
- setting a planar mirror that enables the camera to observe a known reference object (Figure 1.3(b)).

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

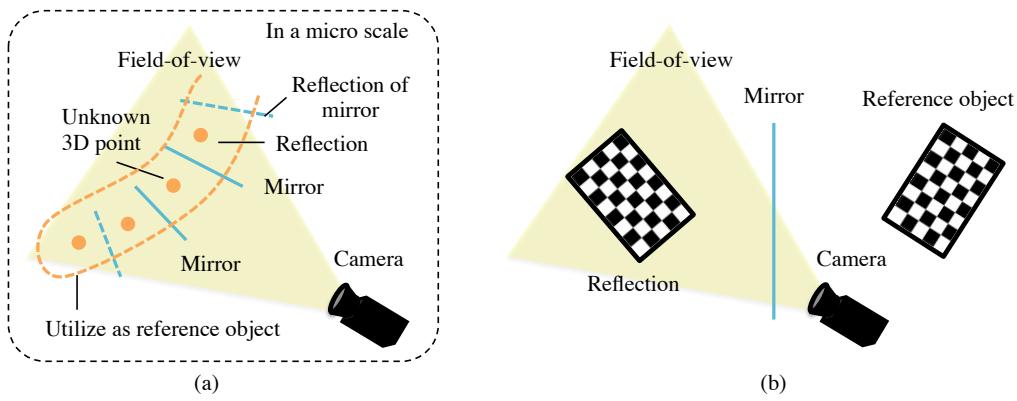


Figure 1.3: Mirror-based calibration uses (a) multiple reflections of a 3D point as a reference object of a parametric 3D model or (b) mirrors to enable the camera to directly observe the reflections of a known reference object.

1.2 Problem Statement

As explained in Section 1.1, this dissertation 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 following problems.

Intrinsic Camera Calibration Without Known Reference Object

The first problem occurs with intrinsic camera calibration. In cases where a known reference object is not available, we introduce mirrors to generate reflections of a 3D point for which the coordinates are unknown and recognize the reflections as a reference object of a parametric 3D model, as shown in Figure 1.3(a).

This intrinsic camera calibration is done using

- a perspective camera model with unknown intrinsic parameters,
- reference points for which the coordinates are unknown, and
- mirrors with unknown parameters, such as mirror normal and position.

The goal of this intrinsic camera calibration is to estimate the intrinsic parameters of the camera model from the projections of a reference point with unknown geometry and its reflections.

1.3. Contributions

Extrinsic Camera Calibration with Unobservable Reference Object

The second problem occurs with extrinsic camera calibration. 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, as illustrated in Figure 1.3(b).

This extrinsic camera calibration is done using

- a perspective camera model with known intrinsic parameters,
- a reference object for which the coordinates in a reference object coordinate system outside the camera’s field-of-view are known, and
- extrinsic camera parameters between the camera and the reference object that are unknown.

The goal of this extrinsic camera calibration is to estimate the extrinsic parameters between the camera and the known reference object from the projections of the mirrored reference object.

1.3 Contributions

Three novel techniques are introduced to address the two problems described in Section 1.2.

Intrinsic camera calibration using planar mirrors: A novel intrinsic camera calibration algorithm is introduced that utilizes a 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 using planar mirrors: A novel planar-mirror-based algorithm is introduced that estimates the extrinsic parameters between the camera and a reference object located outside its field-of-view. 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.

Extrinsic camera calibration using human cornea: The previously mirror-based methods for calibrating an imaging system in which the camera cannot observe a reference object, such as a display-camera system, can be troublesome for preparing a mirror and calibrating it every time in a casual scenario, such as gaze correction in a video conference [KPB⁺12]. A novel extrinsic camera calibration algorithm using the human cornea is introduced that makes use of 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.

1.4 Dissertation Structure

The rest of this dissertation is organized into six chapters. First, the background, problem statements, and contributions of this work are introduced in Chapter 1. In Chapter 2, fundamental knowledge and related work on camera calibration and mirror geometry are introduced. Chapter 3 presents a novel algorithm for intrinsic camera calibration using multiple planar mirrors. Chapter 4 presents a novel method for planar mirror-based extrinsic camera calibration between the camera and a reference object that is not directly observable by the camera. Chapter 5 discusses the recognition of the human cornea as a spherical mirror and the use of a novel human-cornea-based extrinsic camera calibration algorithm for extrinsic camera calibration in casual scenarios. The key points of the dissertation are summarized and future challenges are discussed in Chapter 6.

Chapter 2

Camera Calibration and Mirror Geometry

Camera calibration has been a fundamental research area in computer vision for many years. There have been numerous studies [NNM⁺15, GN05] on various types of calibration, such as geometric and radiometric calibration for various camera models (*e.g.* the perspective camera model and the generalized camera model). The assumption in this dissertation is analysis of the geometric properties involving a 3D space and a 2D image plane with the introduction of mirrors, hence it mainly focuses on geometric camera calibration of the perspective camera model.

This chapter first introduces the definition of geometric camera calibration of the perspective camera model used in this dissertation and then reviews related works. Next it describes the fundamental geometry of the key additional device, a mirror, and related works.

2.1 Geometric Camera Calibration of Perspective Camera

2.1.1 Perspective Camera Model

The perspective camera model is illustrated in Figure 2.1. Let $\mathbf{p}^{\{W\}} = (X^{\{W\}}, Y^{\{W\}}, Z^{\{W\}})^\top$ and $\mathbf{p}^{\{C\}} = (X^{\{C\}}, Y^{\{C\}}, Z^{\{C\}})^\top$ denote a 3D point \mathbf{p} in a world coordinate system and a camera coordinate system, respectively. They satisfy

$$\mathbf{p}^{\{C\}} = R\mathbf{p}^{\{W\}} + \mathbf{t}, \quad (2.1)$$

where R is a rotation matrix and \mathbf{t} is a translation vector. Parameters R and \mathbf{t} are *extrinsic camera parameters* representing the pose and position of

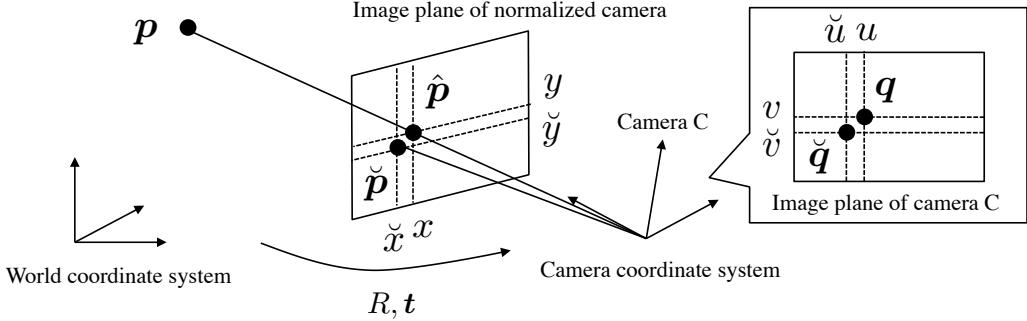


Figure 2.1: The perspective camera model. The effect of lens distortion is modeled as the non-linear transformation in the normalized image coordinates.

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 \mathbf{p} is represented in the camera coordinate system.

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

$$\lambda \tilde{\mathbf{q}} = A \mathbf{p}^{\{C\}} = \begin{bmatrix} f & 0 & c_u \\ 0 & f & c_v \\ 0 & 0 & 1 \end{bmatrix} \mathbf{p}^{\{C\}}, \quad (2.2)$$

where $\tilde{\mathbf{q}}$ denotes the homogeneous coordinate of \mathbf{q} and λ 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 $\hat{\mathbf{p}} = (x, y, 1)$ denotes the normalized image coordinates of $\mathbf{p}^{\{C\}}$, and $\check{\mathbf{p}}$ denotes the distorted coordinates of $\hat{\mathbf{p}}$. This $\check{\mathbf{p}} = (\check{x}, \check{y}, 1)$ is expressed as

$$\begin{aligned} \check{x} &= x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + 2p_1 xy + p_2(r^2 + 2x^2) \\ \check{y} &= y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + p_1(r^2 + 2y^2) + 2p_2 xy, \end{aligned} \quad (2.3)$$

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 $\check{\mathbf{q}}$ including lens distortion is given by the perspective projection of $\check{\mathbf{p}}$; that is, $\lambda \check{\mathbf{q}} = A \check{\mathbf{p}}$. The ideal 2D projection \mathbf{q} without lens distortion can be computed from $\check{\mathbf{q}}$ by solving Eq. (2.3) numerically for $\hat{\mathbf{p}}$ with *intrinsic camera parameters* A and $\mathbf{d} = (k_1, k_2, k_3, p_1, p_2)$. 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 1.1, 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 [Tsa86] 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 [Zha00] 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 provides a closed-form solution that can be uniquely solved using three or more poses of the reference object. This method as well as that of Tsai [Tsa86] are the most commonly used calibration algorithms for estimating both intrinsic and extrinsic parameters using a known reference object.

Several PnP (perspective-n-point) problem-based methods have been proposed for estimating extrinsic parameters [LMNF09, GHTC03]. The goal of the PnP methods is to determine the extrinsic parameters of a camera given its intrinsic parameters and a set of n correspondences between 3D points and their 2D projections. If $n = 3$, the PnP problem is in its minimal configuration although the P3P problem has up to four solutions in general [GHTC03]. Lepetit et al. [LMNF09] proposed an efficient non-iterative solution with a computational cost of $O(n)$ for the $n \geq 4$ problem.

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 [MF92, FLM92].

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 [HZ00, FP09, AFS⁺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 [BAKPW16, SP10]. The correspondences between special points on the silhouette boundaries, called *frontier points* [CG00], 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. Puwein et al. [PBZP14] proposed using detected 2D human joints in multi-view images as common reference points and using these points to compute the extrinsic parameters.

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 Mirrors in Computer Vision

2.2.1 Planar Mirror Geometry

The planar mirror geometry is illustrated in Figure 2.2. Consider a 3D point \mathbf{p} and its reflection \mathbf{p}' from mirror π . Their ideal projections without lens distortion, \mathbf{q} and \mathbf{q}' , are given by the perspective projection:

$$\lambda\mathbf{q} = A\mathbf{p}, \quad \lambda'\mathbf{q}' = A\mathbf{p}', \quad (2.4)$$

where λ and λ' are scale parameters.

Let \mathbf{n} and $d(> 0)$ denote the normal and the distance of the mirror π satisfying

$$\mathbf{n}^\top \mathbf{x} + d = 0, \quad (2.5)$$

where \mathbf{x} is a 3D position in the scene. Here the normal vector is oriented toward the camera center.

2.2. Mirrors in Computer Vision

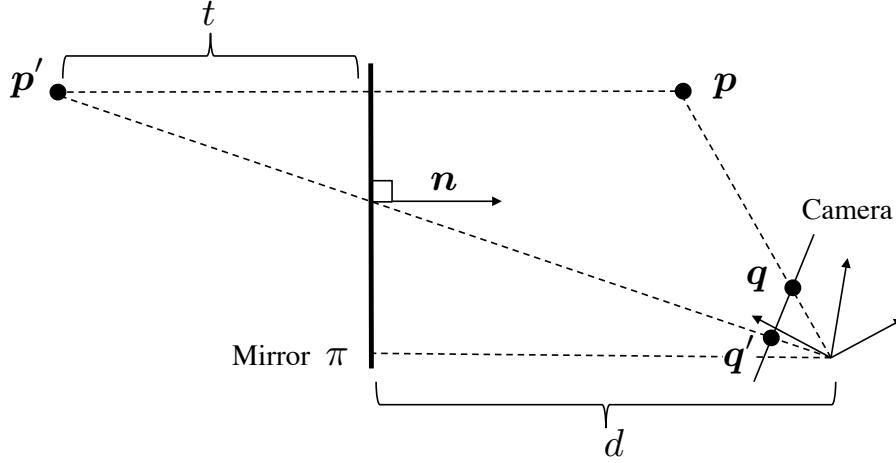


Figure 2.2: Planar mirror geometry: mirror π with normal \mathbf{n} and distance d reflects 3D point \mathbf{p} to \mathbf{p}' ; they are projected to \mathbf{q} and \mathbf{q}' respectively.

As illustrated in Figure 2.2, the distance t from \mathbf{p} and \mathbf{p}' to mirror π satisfies

$$\mathbf{p} = \mathbf{p}' + 2t\mathbf{n}. \quad (2.6)$$

The projection of \mathbf{p}' to \mathbf{n} gives

$$t + d = -\mathbf{n}^\top \mathbf{p}'. \quad (2.7)$$

By eliminating t from these two equations, we have

$$\mathbf{p} = -2(\mathbf{n}^\top \mathbf{p}' + d)\mathbf{n} + \mathbf{p}', \quad (2.8)$$

$$\Leftrightarrow \tilde{\mathbf{p}} = S\tilde{\mathbf{p}}' = \begin{bmatrix} H & -2d\mathbf{n} \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix} \tilde{\mathbf{p}}', \quad (2.9)$$

where H is a 3×3 Householder matrix given by $H = I_{3 \times 3} - 2\mathbf{n}\mathbf{n}^\top$, $\mathbf{0}_{m \times n}$ denotes the $m \times n$ zero matrix, and $I_{n \times n}$ denotes the $n \times n$ identity matrix. Note that this S also satisfies inverse transformation; that is $\tilde{\mathbf{p}}' = S\tilde{\mathbf{p}}_0$.

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 non-planar or planar 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.* [SMS06] proposed a single-viewpoint omnidirectional camera with a hyperbolic mirror. The extrinsic parameters are estimated by solving a two-step least-squares linear minimization problem from the observations of a calibration pattern.

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. Mukaigawa *et al.* [MTK⁺11] introduced a hemispherical confocal imaging system using a turtleback reflector, and Inoshita *et al.* [ITM⁺13] used it to measure a full-dimensional (8-D) BSSRDF (bidirectional subsurface scattering reflection distribution function)). Nane and Nayer [NN98] proposed a computational stereo system using a single camera and various types of mirrors, such as planar, ellipsoidal, hyperboloidal, and paraboloidal. They calibrated this stereo system for a planar mirror by computing the fundamental matrix from the point correspondences. Gluckman and Nayer [GN01] considered the case of two planar mirrors and devised an efficient calibration method based on the idea that the relative orientations of virtual cameras are restricted to the case of planar motion.

In the context of kaleidoscopic imaging exploiting high-order multiple reflections, Ihrke *et al.* [IRM⁺12] and Reshetouski and Ihrke [RI13, RiABR⁺13] 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 [RiABR⁺13] solved the chamber assignment problem by placing constraints on the apparent 2D distances between 2D projections while most other studies [RI13] 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 [IRM⁺12, RI13] without considering their kaleidoscopic, *i.e.* multiple reflection, relationships. That is, a chessboard is first detected [Zha00] in each of the chambers, and then the mirror normals and the distances from the 3D chessboard positions in the camera frame are estimated. This per-mirror calibration can also be done by applying the algorithms in Section 2.2.2. However, those algorithms do not make full use of the kaleidoscopic relationships of multiple reflections.

2.2. Mirrors in Computer Vision

Table 2.1: Features of previous mirror-based extrinsic calibration methods: mirror shape, number of reference points, number of mirror poses, and limitation.

	Shape	Points	Poses	Limitations
Kumar <i>et al.</i> [KIFP08]	Plane	5	3	-
Sturm <i>et al.</i> [SB06]	Plane	4	3	-
Rodrigues <i>et al.</i> [RBN10]	Plane	4	3	-
Hesch <i>et al.</i> [HMR09]	Plane	3	3	-
Proposed (Chapter 4)	Plane	3	3	-
Agrawal [Agr13]	Sphere	8	1	-
Nitschke <i>et al.</i> [NNT11]	(Cornea) Sphere	3	2	-
Proposed (Chapter 5.2)	(Cornea) Sphere	3	1	mirror position
Proposed (Chapter 5.3)	(Cornea) Sphere	5	1	-

Our intrinsic camera calibration using planar mirrors technique (introduced in Chapter 3) uses novel algorithms that overcome these challenges:

- a chamber assignment algorithm that can be applied to general mirror poses and
- 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 used (*i.e.* mirror shape), the required number of mirror poses, the required number of reference points, and limitation (See Table 2.1). 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. Kumar *et al.* [KIFP08] estimated the extrinsic planar mirror parameters by solving linear equations derived from orthogonality relations between the axes of the real camera and those of each mirrored camera. Sturm *et al.* [SB06] and Rodrigues *et al.* [RBN10] proposed a planar-mirror-based method for computing the pose of a reference object without a direct view. Hesch *et al.* [HMR09] assumed a camera-based robot with each body part not being visible from the camera

and proposed a linear method for estimating their relative poses and positions with a planar mirror in a P3P problem scenario. Our proposed method described in Chapter 3 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. [RBN10], the three mirror poses and three reference points are the minimal configuration for a planar-mirror-based method. In addition to using mirrors, Delaunoy *et al.* [DLJP14] used the reconstructed scene as a reference object and calibrated two cameras and a display on a mobile device with two mirror poses.

In contrast, Agrawal [Agr13] proposed a spherical-mirror-based calibration method that uses a matrix E similar to the essential matrix in the mirror geometry with a *coplanarity constraint* with eight point correspondences.

These methods need a planar or spherical mirror as an additional device, and using an additional device for every calibration instance can be troublesome in casual scenarios, such as gaze correction in video chatting [KPB⁺12].

Nitschke *et al.* [NNT11] 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 (corneas). The 3D positions of the reference points are estimated by finding the intersection of two rays connecting a reference point to the center of an eyeball. This method needs three reference points and both eyes, *i.e.* two spherical mirrors. This dissertation 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 5.2 and 5.3.

Chapter 3

Mirror-based Intrinsic Camera Calibration

This chapter provides a novel intrinsic camera calibration algorithm by introducing multiple planar mirrors and consisting a kaleidoscopic imaging system.

Intrinsic camera calibration is to estimate intrinsic parameters, such as focal length, optical center and lens distortion, and is one of the essential processes in case of analyzing geometrical relationship between 2D and 3D space. In general, it should be done with known reference objects covering the camera's field-of-view for high precision estimation. However, sometimes a practical reference object of known geometry is not available in case of capturing in a special scale, such as micro or large scale.

To conduct the intrinsic camera calibration in such situation, the proposed method introduces multiple planar mirrors which generate reflections of a 3D point whose coordinate is unknown and recognizes the reflections as a reference object of a parametric 3D model. The imaging system with multiple planar mirrors is called as *kaleidoscopic imaging system* in this dissertation (Figure 3.1).

In order to realize the intrinsic camera calibration with the kaleidoscopic imaging system, there are three problems to be solved, that is chamber assignment, mirror parameters estimation and intrinsic parameters estimation. The key contribution of this work is to introduce a novel geometric constraint *kaleidoscopic projection constraint* which is satisfied by projections of high-order reflections. This constraint provides multiple linear equations on mirror parameters per single 3D point and it enable the proposed method to solve above problems.

In this chapter, the kaleidoscopic imaging system is introduced in Section 3.1, and the problem description is provided in Section 3.2. The key con-

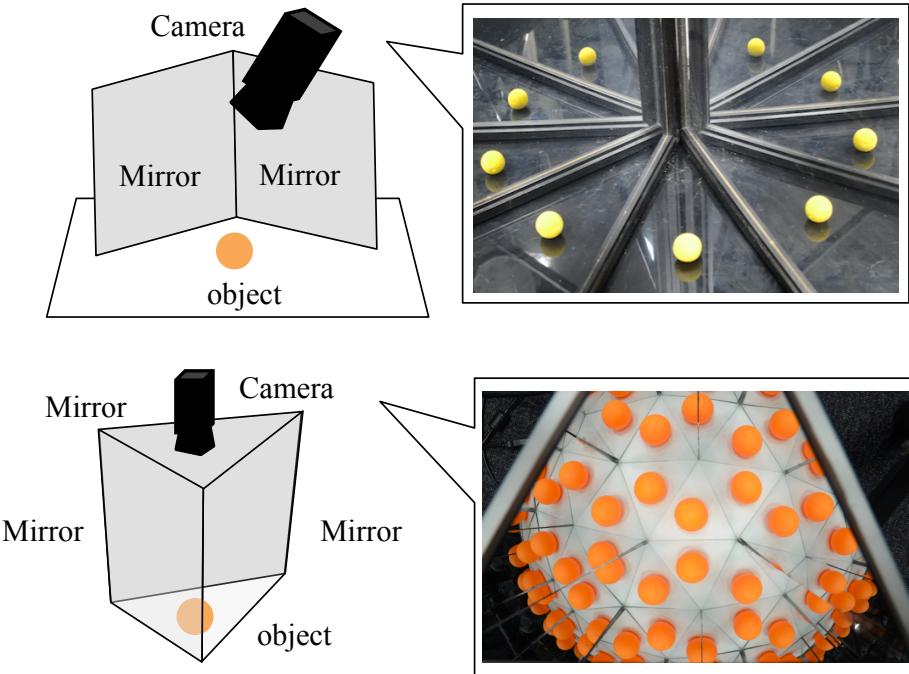


Figure 3.1: Kaleidoscopic imaging system consists of multiple planer mirrors. Up: two mirrors. Bottom: three mirrors. The projections of reflected target object in each chamber are recognized as the target observed by the virtual cameras, which are generated by the planer mirrors.

straint, *kaleidoscopic projection constraint*, is introduced in Section 3.3 and the proposed method for chamber assignment, mirror parameters estimation and intrinsic parameters estimation, are provided in Section 3.4, 3.5 and 3.6 respectively. The qualitative and quantitative evaluations are reported for the methods in Section 3.7, 3.8 and 3.9 respectively. Section 3.10 provides discussion on suggestion of improvement and limitations of the proposed method and Section 3.11 summarizes this chapter.

3.1 Kaleidoscopic Imaging System

Suppose the camera observes the target 3D point directly and indirectly via N_π mirrors as shown in Figure 3.1. Let \mathbf{p}_0 denote the original 3D point and \mathbf{p}_i denote the first reflection of \mathbf{p}_0 by the mirror π_i ($i = 1, \dots, N_\pi$) (Figure

3.1. Kaleidoscopic Imaging System

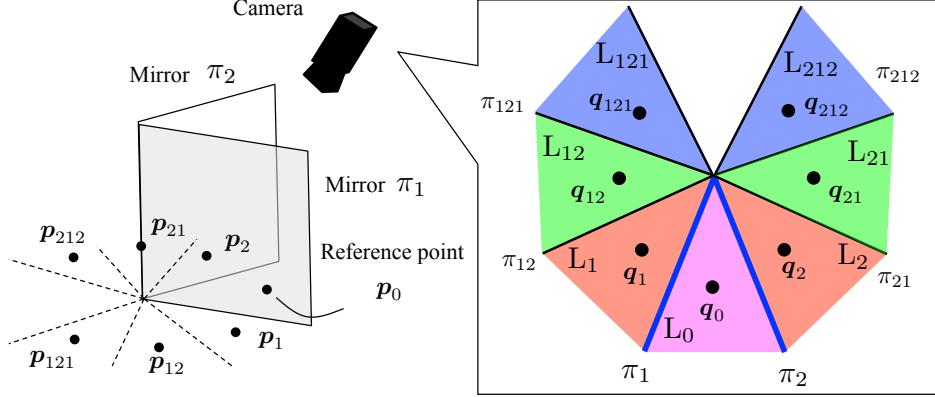


Figure 3.2: Chamber assignment. The magenta region indicates the base chamber. The red, green and blue regions indicate the chambers corresponding to the first, second, third reflections respectively.

3.2). The reflection \mathbf{p}_i is given by

$$\tilde{\mathbf{p}}_0 = S_i \tilde{\mathbf{p}}_i = \begin{bmatrix} H_i & -2d_i \mathbf{n}_i \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix} \tilde{\mathbf{p}}_i, \quad (3.1)$$

where \mathbf{n}_i and d_i denote the mirror normal and its distance respectively and H_i is given by $H_i = I_{3 \times 3} - 2\mathbf{n}_i \mathbf{n}_i^\top$ as introduced in Section 2.2.1.

Furthermore, such mirrors define virtual mirrors as a result of multiple reflections. Let π_{ij} ($i, j = 1, \dots, N_\pi$, $i \neq j$) denote the virtual mirror defined as a mirror of π_j by π_i , \mathbf{p}_{ij} denote the reflection of \mathbf{p}_i by π_{ij} , and L_{ij} denote the chamber where \mathbf{p}_{ij} is projected to. Also the matrices S_{ij} and H_{ij} for π_{ij} are given by

$$\begin{aligned} S_{ij} &= S_i S_j, \\ H_{ij} &= H_i H_j. \end{aligned} \quad (3.2)$$

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

$$\Pi_{k=1}^{N_k} S_{i_k} \quad (i_k = 1, 2, 3, i_k \neq i_{k+1}), \quad (3.3)$$

where N_k is the number of reflections.

Obviously the 3D subspaces where \mathbf{p}_0 and \mathbf{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 \mathbf{p}_x is denoted by $\mathbf{q}_x \in Q$ ($x = 0, 1, \dots, N_\pi, 12, 13, \dots$) in general. In this paper we denote the 2D region where \mathbf{q}_0 exists as the *base chamber* L_0 , and we use L_x to denote the chamber where \mathbf{q}_x exists.

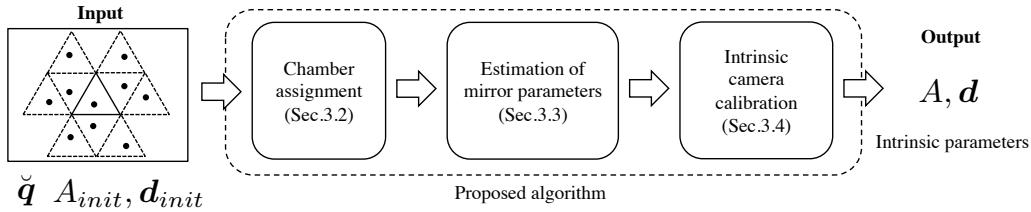


Figure 3.3: The outline of the proposed optimization strategy for intrinsic camera calibration.

3.2 Problem Description

Conventionally, the intrinsic camera calibration techniques have been conducted from a 3D model of “known” geometry and its 2D projections[Zha00]. 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 outdoor 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. In order 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:

- each of the detected observation is not assigned to the corresponding chamber.
- it has two or more planar mirrors as a supporting device whose extrinsic parameters, *i.e.* mirror normals and distances, are unknown.
- it has one perspective camera and its intrinsic parameters are unknown.

Figure 3.3 illustrates an outline of the proposed algorithm. In our algorithm, there are three problems to be solved. Consider a 2D point set $R = \{\mathbf{r}_i\}$ detected from the captured image as candidates of \mathbf{q}_x . The problems are:

3.3. Kaleidoscopic Projection Constraint

- to assign the chamber label L_x to $\mathbf{r}_i \in R$ to identify to which chamber each of the projections \mathbf{q}_x belong (Section 3.4),
- to estimate the parameters of the real mirrors $\pi_i (i = 1, \dots, N_\pi)$, *i.e.* normals \mathbf{n}_i and distances d_i of them, from kaleidoscopic projections \mathbf{q}_x (Section 3.5), and
- to estimate the intrinsic camera parameters, *i.e.* A and \mathbf{d} , from $\check{\mathbf{q}}_x$ 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 \mathbf{p} . The mirror π of matrix S defined by the normal \mathbf{n} and the distance d reflects \mathbf{p} to $\mathbf{p}' = Sp$ (Eq (2.9)).

Based on the epipolar geometry[HZ00, YPH⁺13], \mathbf{n} , \mathbf{p} and \mathbf{p}' are coplanar and satisfy

$$(\mathbf{n} \times \mathbf{p})^\top \mathbf{p}' = 0. \quad (3.4)$$

By substituting \mathbf{p} and \mathbf{p}' by $\lambda A^{-1}\mathbf{q}$ and $\lambda' A^{-1}\mathbf{q}'$ respectively (Eq (2.4)), we obtain

$$\mathbf{q}^\top A^{-\top} [\mathbf{n}]_\times A^{-1} \mathbf{q}' = 0, \quad (3.5)$$

where $[\mathbf{n}]_\times$ denotes the 3×3 skew-symmetric matrix representing the cross product by \mathbf{n} and this is the essential matrix of this mirror-based binocular geometry[YPH⁺13].

By representing the normalized image coordinates of \mathbf{q} and \mathbf{q}' by $(x, y, 1)^\top = A^{-1}\mathbf{q}$ and $(x', y', 1)^\top = A^{-1}\mathbf{q}'$ respectively, Eq (3.5) can be rewritten as

$$(y - y' \quad x' - x \quad xy' - x'y) \mathbf{n} = 0. \quad (3.6)$$

We call this Eq (3.6) as *kaleidoscopic projection constraint* in this paper. This constraint is satisfied by not only single reflections but also high-order reflections as below.

Single reflection: Let \mathbf{p}_0 denote a 3D point and \mathbf{p}_i denote the reflection by mirror π_i . Since \mathbf{p}_i is expressed as $\mathbf{p}_i = S_i \mathbf{p}_0$, the normalized image coordinates of their projections, \mathbf{q}_0 and \mathbf{q}_i , obviously satisfy Eq (3.6) as:

$$(y_0 - y_i \quad x_i - x_0 \quad x_0 y_i - x_i y_0) \mathbf{n}_i = 0, \quad (3.7)$$

where $(x_0, y_0, 1)^\top = A^{-1}\mathbf{q}_0$ and $(x_i, y_i, 1)^\top = A^{-1}\mathbf{q}_i$.

High-order reflections: Let \mathbf{p}_{ij} ($i, j = 1, \dots, N_\pi, i \neq j$) denote the reflection of \mathbf{p}_i by π_{ij} . This \mathbf{p}_{ij} can be expressed as $\mathbf{p}_{ij} = S_{ij}\mathbf{p}_i = S_{ij}S_i\mathbf{p}_0 = S_iS_j\mathbf{p}_0$ based on Eq (3.2). Here $\mathbf{p}_j = S_j\mathbf{p}_0$ holds as well, and we obtain $\mathbf{p}_{ij} = S_iS_j\mathbf{p}_0 \Leftrightarrow \mathbf{p}_{ij} = S_i\mathbf{p}_j$. This equation means that \mathbf{p}_{ij} can be recognized as the first reflection of \mathbf{p}_j by π_i , and hence the normalized image coordinates of their projections, \mathbf{q}_{ij} and \mathbf{q}_j , also satisfy Eq (3.6) as:

$$(y_j - y'_{ij} \quad x_{ij} - x_j \quad x_j y_{ij} - x_{ij} y_j) \mathbf{n}_i = 0, \quad (3.8)$$

where $(x_{ij}, y_{ij}, 1)^\top = A^{-1}\mathbf{q}_{ij}$.

In N_k th reflections, we obtain the kaleidoscopic projection constraint between $\mathbf{p}_{i_k} = S_{i_{N_k}} \prod_{k=1}^{N_k-1} S_{i_k} \mathbf{p}_0$ and $\mathbf{p}_{i'_k} = \prod_{k=1}^{N_k-1} S_{i_k} \mathbf{p}_0$ in the same manner.

3.4 Chamber Assignment

Based on the kaleidoscopic projection constraint in the last section, we introduce a new algorithm that identifies the chamber label of each projections. 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 evaluate each of their consistencies as a kaleidoscopic projection. In this evaluation, additional geometric constraints that reject mirror parameters satisfying the kaleidoscopic projection constraint but physically infeasible are introduced.

Notice that we first introduce our algorithm using a two-mirror system (Figure 3.4(a)) as an example, and then extend it to the general case.

3.4.1 Base Structure

Suppose $2N_\pi$ points of the observed points R are selected and they could be hypothesized as $\mathbf{q}_0, \mathbf{q}_1, \dots$ correctly. The mirror normal \mathbf{n}_i has two degrees of freedom and can be linearly estimated by collecting more or equal to two linear constraints on it. In case of $N_\pi = 2$, the mirror normal \mathbf{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

3.4. Chamber Assignment

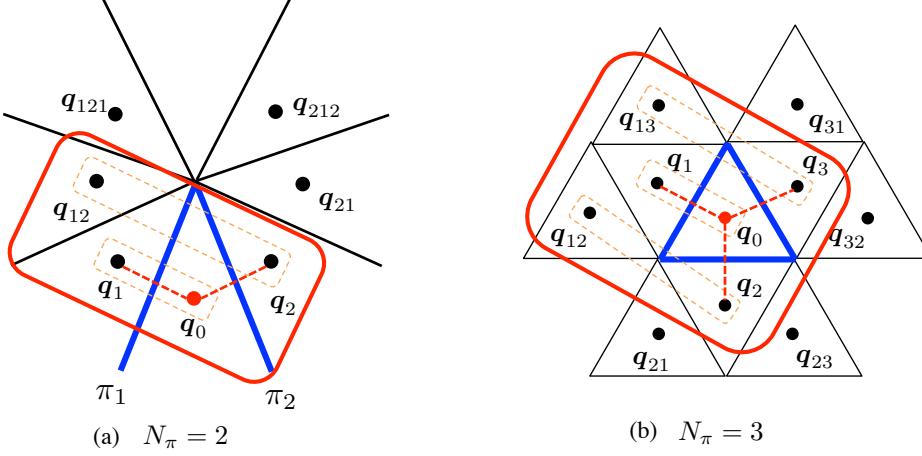


Figure 3.4: The red boxes show examples of base structures in case of (a) $N_\pi = 2$ and (b) $N_\pi = 3$. The red point indicates the point assumed as the base chamber and the dotted boxes indicate doublets. The red dotted lines indicate the reflection pairs and the blue lines indicate the discovered mirrors.

projection constraint (Eq (3.6)) using $\{\langle \mathbf{q}_0, \mathbf{q}_1 \rangle, \langle \mathbf{q}_2, \mathbf{q}_{12} \rangle\}$ in Figure 3.4(a):

$$\begin{bmatrix} y_0 - y_1 & x_0 - x_1 & x_0y_1 - x_1y_0 \\ y_2 - y_{12} & x_{12} - x_2 & x_0y_{12} - x_{12}y_0 \end{bmatrix} \mathbf{n}_1 = \mathbf{0}_{3 \times 1}, \quad (3.9)$$

where $\langle \mathbf{q}, \mathbf{q}' \rangle$ denotes a *doublet*, *i.e.* the pair of projections \mathbf{q} and \mathbf{q}' for Eq (3.6).

Using the estimated \mathbf{n}_1 and assuming $d_1 = 1$ without loss of generality, the 3D point \mathbf{p}_1 can be described as $\tilde{\mathbf{p}}_1 = S_1 \tilde{\mathbf{p}}_0$ by Eq (2.9). By substituting \mathbf{p}_0 and \mathbf{p}_1 in this equation by using \mathbf{q}_0 and \mathbf{q}_1 as expressed in Eq (2.4), the 3D point \mathbf{p}_0 and \mathbf{p}_1 can be triangulated by solving the following linear system for λ_0 and λ_1 :

$$\tilde{\mathbf{p}}_0 = S_1 \tilde{\mathbf{p}}_1, \quad (3.10)$$

$$\Leftrightarrow \lambda_0 A^{-1} \mathbf{q}_0 = H_1 \lambda_1 A^{-1} \mathbf{q}_1 - 2\mathbf{n}_1, \quad (3.11)$$

$$\Leftrightarrow [H_1 A^{-1} \mathbf{q}_0 \quad -A^{-1} \mathbf{q}_1] \begin{bmatrix} \lambda_0 \\ \lambda_1 \end{bmatrix} = 2\mathbf{n}_1. \quad (3.12)$$

Note that \mathbf{n}_1 can be determined up to sign from Eq (3.9). As illustrated in Figure 3.5, both of mirror parameters, (\mathbf{n}_1, d_1) and $(-\mathbf{n}_1, d_1)$, are possible configurations in terms of Eq (2.5), but one of them triangulates \mathbf{p}_i ($i = 0, 1$) behind the camera. As a result, by rejecting such configuration, we obtain a unique mirror parameter.

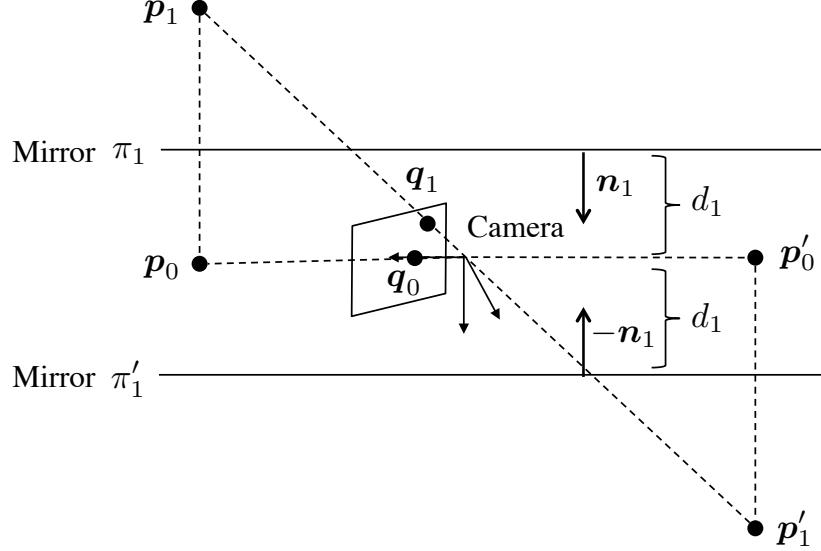


Figure 3.5: Sign ambiguity of \mathbf{n} and corresponding triangulations. $\mathbf{p}_i (i = 0, 1)$ and \mathbf{p}'_i are estimated from possible mirror parameters of π_1 and π'_1 , i.e. (\mathbf{n}_1, d_1) and $(-\mathbf{n}_1, d_1)$, respectively. Both of these mirror parameters satisfy Eq (2.5), and \mathbf{p}'_i appears as $-\mathbf{p}_i$.

Similarly the 3D points \mathbf{p}_2 and \mathbf{p}_{12} can be triangulated by solving the linear system for λ_2 and λ_{12} . Because \mathbf{p}_2 is the reflection of \mathbf{p}_0 by the mirror π_2 , the mirror normal \mathbf{n}_2 as well as the distance d_2 can be estimated as

$$\mathbf{n}_2 = \frac{\mathbf{p}_0 - \mathbf{p}_2}{|\mathbf{p}_0 - \mathbf{p}_2|}, \quad d_2 = -\mathbf{n}_2^\top \frac{\mathbf{p}_0 + \mathbf{p}_2}{2}. \quad (3.13)$$

This doublets pair $\{\langle \mathbf{q}_0, \mathbf{q}_1 \rangle, \langle \mathbf{q}_2, \mathbf{q}_{12} \rangle\}$ is a minimal configuration for linear estimation of the real mirror parameters in $N_\pi = 2$ 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, $\{\langle \mathbf{q}_0, \mathbf{q}_2 \rangle, \langle \mathbf{q}_1, \mathbf{q}_{21} \rangle\}$ is also a base structure for $N_\pi = 2$ case.

This procedure can be generalized intuitively as an $N_\pi (N_\pi > 2)$ pair method for N_π mirror system as follows. In case of N_π mirrors, we can observe a base structure which consists of N_π -tuple of doublets between a 3D point and its reflection by a mirror π_j up to the second reflections as $\langle \mathbf{q}_0, \mathbf{q}_j \rangle$ and $\langle \mathbf{q}_i, \mathbf{q}_{ji} \rangle$ ($1 \leq j \leq N_\pi$ $i = 1, \dots, j-1, j+1, \dots, N_\pi$). Using this base structure, \mathbf{n}_j can be estimated first, and then $\mathbf{p}_i (i = 0, \dots, N_\pi)$ as described for the two-mirror case. As a result, all the mirror normals and the distances can be estimated by assuming $d_1 = 1$.

For example, a base structure of the $N_\pi = 3$ system illustrated in Figure

3.4. Chamber Assignment

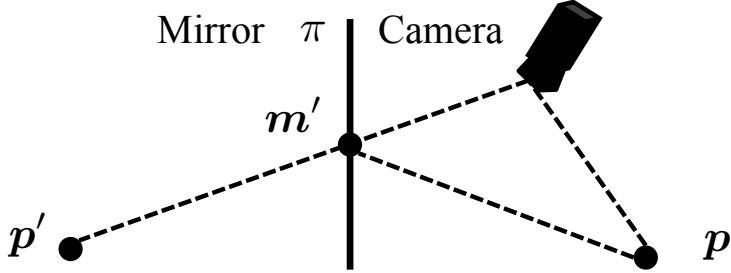


Figure 3.6: The 3D point \mathbf{p} is closer to the camera than its reflection \mathbf{p}' from triangle inequality.

3.4(b) is 3-tuple of doublets $\{\langle \mathbf{q}_0, \mathbf{q}_1 \rangle, \langle \mathbf{q}_2, \mathbf{q}_{12} \rangle, \langle \mathbf{q}_3, \mathbf{q}_{13} \rangle\}$. This can be seen as a 6-point algorithm by considering up to the second reflections.

3.4.2 Geometric Constraints on Kaleidoscopic Projections

The procedure in Section 3.4.1 assumed the selected points are correctly labeled as $\mathbf{q}_0, \mathbf{q}_1, \dots$. This section provides geometric constraints that evaluates a correctness of the hypothesized labeling.

In the mirror parameter estimation in Section 3.4.1, the smallest eigenvalue computed on solving Eq (3.9) for \mathbf{n}_1 indicates the feasibility of interpreting the set of N_π doublets as a base structure:

$$E = \frac{|e_3|}{\sum_{i=1,2,3} |e_i|}, \quad (3.14)$$

where e_i is the i th largest eigenvalue of the coefficient matrix in Eq (3.9). That is, once we select $2N_\pi$ projections in the image and hypothesize N_π doublets correctly as a base structure, this should be zero in ideal conditions without noise.

However, $E = 0$ is not a sufficient constraint to conclude the hypothesized base structure is physically feasible in terms of a kaleidoscopic projection. In addition to this condition, the hypothesized base structure should satisfy the following two propositions.

Proposition 1. *The 3D point \mathbf{p}_0 projected to \mathbf{q}_0 in the base chamber is the closest point to the camera among its reflections.*

Proof. As illustrated in Figure 3.6, the distance to the reflection \mathbf{p}' of \mathbf{p} is identical to the distance to the point of reflection \mathbf{m}' and the distance from

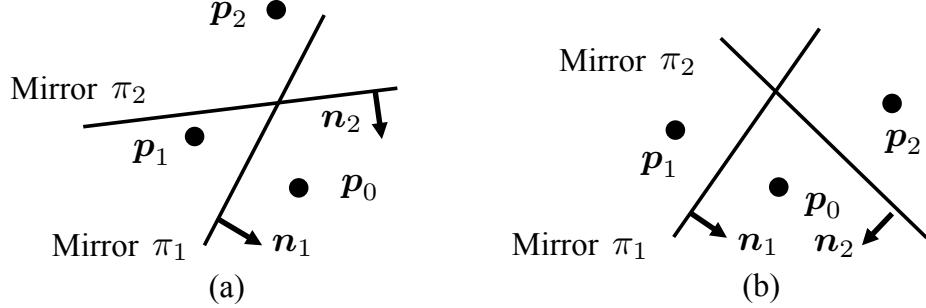


Figure 3.7: In order to obtain second reflections, the mirrors should be facing to each other, $\mathbf{n}_i^\top \mathbf{n}_j < 0$ as in (b). In $\mathbf{n}_i^\top \mathbf{n}_j \geq 0$ cases as in (a), the first reflection \mathbf{p}_2 on π_2 is reflected to the behind of the mirror π_1 and it can not be reflected by π_1 as the second reflection.

\mathbf{m}' to \mathbf{p} . From triangle inequality, \mathbf{p} is closer to the camera than \mathbf{p}' , i.e. $|\mathbf{p}| < |\mathbf{p}'|$.

By considering this single reflection for the case of \mathbf{p}_0 and \mathbf{p}_i with the original camera, the case of \mathbf{p}_i and \mathbf{p}_{ji} with the virtual camera reflected by the mirror π_j , and so forth, we have $|\mathbf{p}_0| < |\mathbf{p}_i| < |\mathbf{p}_{ji}| < \dots$. \square

Proposition 2. *The mirror normals should satisfy $\mathbf{n}_i^\top \mathbf{n}_j < 0$ ($i \neq j$).*

Proof. The mirror parameters estimated in Section 3.4.1 requires a projection of a second reflection. As illustrated in Figure 3.7, this is identical to guarantee that the mirrors are facing to each other:

$$\mathbf{n}_i^\top \mathbf{n}_j < 0. \quad (3.15)$$

\square

Proposition 1 rejects configurations as illustrated in Figure 3.8(a) where \mathbf{q}_1 is wrongly interpreted as the base chamber. Proposition 2 rejects configurations as illustrated in Figure 3.8(b) where \mathbf{q}_0 is correctly interpreted as the base chamber but π_2 reflects \mathbf{p}_{212} behind π_1 .

3.4.3 Discontinuity-aware Label Propagation

Suppose a hypothesized base structure satisfies the above two propositions. This section introduces an algorithm that propagates the labeling to the projections not involved in the hypothesized base structure by synthesizing projections of all possible reflected points.

3.4. Chamber Assignment

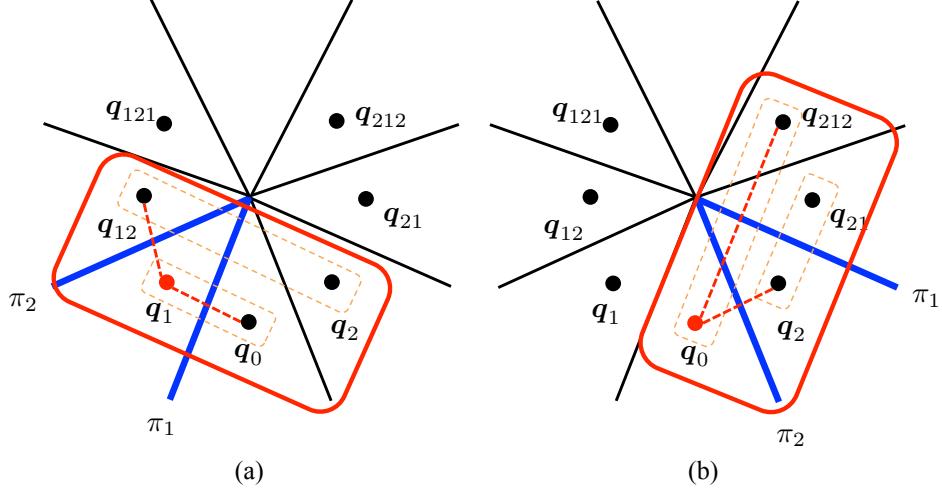


Figure 3.8: Configurations satisfying Eq (3.9) but not (a) Proposition 1 and (b) Proposition 2. The color codes are identical to Figure 3.2.

Given the mirror parameters \mathbf{n}_i, d_i ($i = 1, \dots, N_\pi$) and the triangulated 3D point \mathbf{p}_0 , the k th reflection and its projection can be computed by Eq (2.9) and Eq (2.4). However, if such reflection is projected outside of the corresponding chamber, it is not observable from the camera and cannot generate further reflections. This is called as *discontinuity*[RMSI11].

Intuitively detecting such discontinuity can simply be achieved by examining the chamber label at the projected pixel, but it requires knowing the pixel-wise chamber labeling that is not available up to this point. Instead, we introduce another detection approach based on a 3D geometry.

Discontinuity Detection

Consider a multiply-reflected 3D point $\tilde{\mathbf{p}}_{i_k \dots i_1} = \Pi_{j=i_1, \dots, i_k} S_j \tilde{\mathbf{p}}_0$. As pointed out in [RMSI11], if this is visible from the camera, the ray from the camera center to $\tilde{\mathbf{p}}_{i_k \dots i_1}$ should intersect with the mirror of the first reflection π_{i_1} as illustrated in Figure 3.9.

Let ℓ denote the ray to the target point. Since the system has N_π mirror planes, the above condition can be evaluated by computing the N_π intersections between each of the planes and ℓ , and by testing if the intersection with the mirror π_{i_1} in question is the closest intersection among the intersections in front of the camera.

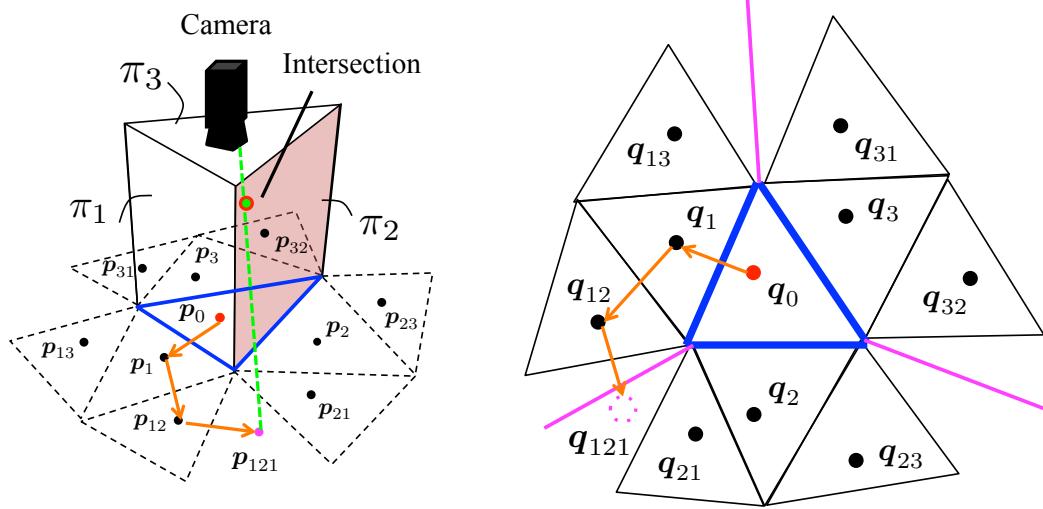


Figure 3.9: Discontinuity. The third reflection \mathbf{p}_{121} is not visible from the camera since the viewing ray (the green dotted line in the left) intersects with not the mirror π_1 but with π_2 . The boundary of such visibility appears as the discontinuity boundary in the image (the magenta lines in the right)[RMSI11].

Label Propagation for a Hypothesized Base Structure

Suppose projections of the visible reflections by considering the discontinuity are synthesized as $\hat{Q} = \{\hat{\mathbf{q}}_{i_k \dots i_1}\}$ ($k \geq 1, 1 \leq i_x \leq N_\pi$).

The goal of the label propagation is to find correspondences between the synthesized point set \hat{Q} and the observed point set $R = \{\mathbf{r}_i\}$ as a sort of bipartite matching. Suppose the matching cost to associate $\hat{\mathbf{q}}_{i_k \dots i_1}$ and \mathbf{r}_i is modeled by the 2D distance between them. The matching should minimize the total matching cost

$$\mathcal{E} = \sum_{\hat{\mathbf{q}}_{i_k \dots i_1} \in \hat{Q}} |\hat{\mathbf{q}}_{i_k \dots i_1} - \hat{\mathbf{r}}_{i_k \dots i_1}|, \quad (3.16)$$

where $\hat{\mathbf{r}}_{i_k \dots i_1} \in R$ is the point selected as the corresponding point of $\hat{\mathbf{q}}_{i_k \dots i_1}$ by assigning label $L_{i_k \dots i_1}$.

Since solving this combination optimization for each of the trials in our analysis-by-synthesis is computationally expensive, we approximated this process by the nearest neighbor search which simply assigns the nearest candidate within a threshold distance for each of $\hat{\mathbf{q}}_{i_k \dots i_1}$. As a result, doublets can share a candidate point in R by multiple synthesized points in \hat{Q} , but we found this approximation is acceptable to some extent because of the sparse

3.4. Chamber Assignment

Algorithm 1 Chamber assignment algorithm

Require: $\mathbf{r}_i (i = 0, 1, \dots, |R|)$

Ensure: $\Lambda_{\text{out}} = \{L_{\mathbf{r}_0}, L_{\mathbf{r}_1}, \dots, L_{\mathbf{r}_{|R|}}\}$

for all base structure **do**

 compute mirror parameters $\mathbf{n}_j, d_j (j = 1, \dots, N_\pi)$ by solving Eq (3.9) and Eq (3.13).

if $\exists i, |\mathbf{p}_0| > |\mathbf{p}_i|$ **then**

 continue; # Proposition 1

end if

if $\mathbf{n}_i^\top \mathbf{n}_j > 0$ **then**

 continue; # Proposition 2

end if

 assign chamber labels for each \mathbf{r}_i as described in Section 3.4.3 and obtain label set Λ_r .

 compute the recall score \mathcal{R} of Λ_r by Eq (3.17).

end for

select the label set with a highest recall score as Λ_{out} .

distribution of the points in R and \hat{Q} . This point is discussed later in Section 3.10.

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

$$\mathcal{R} = \frac{|R_c|}{|\hat{Q}|}, \quad (3.17)$$

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

3.4.4 Chamber Assignment Algorithm

Algorithm 1 shows the flow of the proposed chamber assignment algorithm. For each possible base structures, it examines if it satisfies the kaleidoscopic projection constraint expressed by Eq (3.9), the base chamber constraint by Proposition 1, and the mirror angle constraint by Proposition 2. Once the base structure passes these verifications, its recall ratio \mathcal{R} is computed by Eq (3.17). Finally the best estimate of the chamber assignment is returned by finding the base structure of the highest recall ratio.

It should be noted that this algorithm evaluates all possible combinations to evaluate the behavior of the proposed geometric constraints in Section

3.4.1 thoroughly. Possible efficient implementations are discussed in Section 3.10.2.

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 3.4.1 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.

Notice that the algorithm is first introduced by utilizing up to the second reflections, but they can be extended to third or further reflections intuitively as described later.

3.5.1 Mirror Normals

As illustrated in Figure 3.10 (a), suppose a 3D point \mathbf{p}_0 is projected to \mathbf{q}_0 in the base chamber, and its mirror \mathbf{p}_i by π_i is projected to \mathbf{q}_i in the chamber L_i . Likewise, the second mirror \mathbf{p}_{ij} by π_{ij} is projected to \mathbf{q}_{ij} in the chamber L_{ij} , and so forth. Here, kaleidoscopic projection constraints are satisfied by two pairs of projections on each mirror π_1 and π_2 . From these constraints, \mathbf{n}_1 and \mathbf{n}_2 can be estimated by solving

$$\begin{bmatrix} y_0 - y_1 & x_1 - x_0 & x_0y_1 - x_1y_0 \\ y_2 - y_{12} & x_{12} - x_2 & x_2y_{12} - x_{12}y_2 \end{bmatrix} \mathbf{n}_1 = \mathbf{0}_{3 \times 1}. \quad (3.18)$$

and

$$\begin{bmatrix} y_0 - y_2 & x_2 - x_0 & x_0y_2 - x_2y_0 \\ y_2 - y_{21} & x_{21} - x_1 & x_1y_{21} - x_{21}y_1 \end{bmatrix} \mathbf{n}_2 = \mathbf{0}_{3 \times 1}. \quad (3.19)$$

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 \mathbf{n}_1 and \mathbf{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 (2.9) coincides with the observed 2D position as introduced in Section 3.5.3 later. This expression,

3.5. Mirror Parameters Estimation

however, is nonlinear in the normals \mathbf{n}_i ($i = 1, 2$) (e.g. $\mathbf{p}_{12} = S_1 S_2 \mathbf{p}_0$). On the other hand, projections of such multiple reflections can be associated as a result of single reflection by Eq (3.8) directly (e.g. \mathbf{n}_1 with \mathbf{q}_{12} and \mathbf{q}_2 as the projections of \mathbf{p}_{12} and $S_2 \mathbf{p}_0$ 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 \mathbf{q}_{12} , \mathbf{q}_2 and \mathbf{n}_1 in Eq (3.18) assumes $\mathbf{p}_2 = g S_2 \mathbf{p}_0$ being satisfied, and it is enforced by $(A^{-1} \mathbf{q}_2 \times A^{-1} \mathbf{q}_0)^\top \mathbf{n}_2 = 0$ in the first row of Eq (3.19). Inversely, on estimating \mathbf{n}_1 by Eq (3.18), it enforces $\mathbf{p}_1 = S_1 \mathbf{p}_0$ for Eq (3.19).

It should be noted that this algorithm can be extended to third or further reflections intuitively. For example, if \mathbf{p}_{21} and its reflection by π_1 is observable as $\lambda_{121} \mathbf{q}_{121} = A \mathbf{p}_{121} = A S_1 \mathbf{p}_{21}$, then it provides

$$(y_{21} - y_{121}, x_{121} - x_{21}, x_{21}y_{121} - x_{121}y_{21}) \mathbf{n}_1 = 0, \quad (3.20)$$

and can be integrated with Eq (3.18).

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

$$\begin{bmatrix} y_0 - y_1 & x_1 - x_0 & x_0 y_1 - x_1 y_0 \\ y_2 - y_{12} & x_{12} - x_2 & x_2 y_{12} - x_{12} y_2 \\ y_3 - y_{13} & x_{13} - x_3 & x_3 y_{13} - x_{13} y_3 \end{bmatrix} \mathbf{n}_1 = \mathbf{0}_{3 \times 1}, \quad (3.21)$$

$$\begin{bmatrix} y_0 - y_2 & x_2 - x_0 & x_0 y_2 - x_2 y_0 \\ y_3 - y_{23} & x_{23} - x_3 & x_3 y_{23} - x_{23} y_3 \\ y_2 - y_{21} & x_{21} - x_1 & x_1 y_{21} - x_{21} y_1 \end{bmatrix} \mathbf{n}_2 = \mathbf{0}_{3 \times 1}, \quad (3.22)$$

and

$$\begin{bmatrix} y_0 - y_3 & x_3 - x_0 & x_0 y_3 - x_3 y_0 \\ y_1 - y_{31} & x_{31} - x_1 & x_1 y_{31} - x_{31} y_1 \\ y_2 - y_{32} & x_{32} - x_2 & x_2 y_{32} - x_{32} y_2 \end{bmatrix} \mathbf{n}_3 = \mathbf{0}_{3 \times 1}, \quad (3.23)$$

instead of Eqs (3.18) and (3.19) from point correspondences in Figure 3.10 (b).

3.5.2 Mirror Distances

Once the mirror normals \mathbf{n}_1 and \mathbf{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 (2.4) indicates that a 3D point \mathbf{p}_i and its projection \mathbf{q}_i should satisfy the colinearity constraint:

$$(A^{-1} \mathbf{q}_i) \times \mathbf{p}_i = \mathbf{x}_i \times \mathbf{p}_i = \mathbf{0}_{3 \times 1}, \quad (3.24)$$

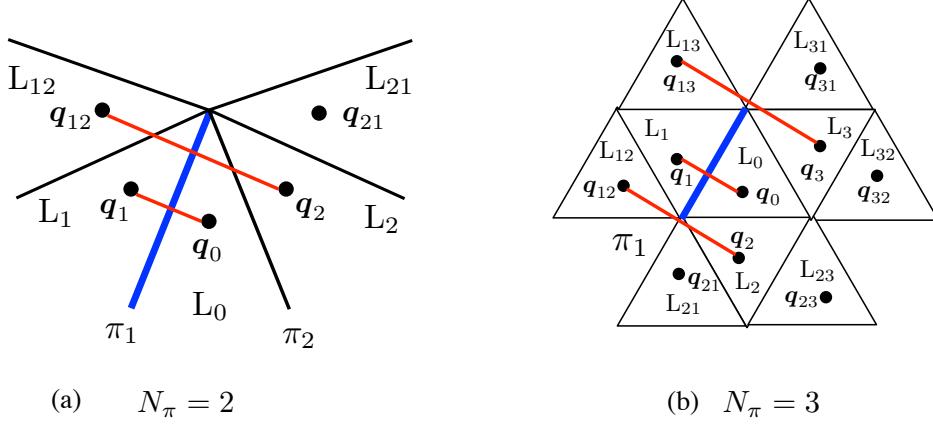


Figure 3.10: Corresponding points in (a) $N_\pi = 2$ and (b) $N_\pi = 3$ case. (a) Two pairs $\langle \mathbf{q}_0, \mathbf{q}_1 \rangle$ and $\langle \mathbf{q}_2, \mathbf{q}_{12} \rangle$ (red) are available on mirror π_1 (blue). (b) Three pairs $\langle \mathbf{q}_0, \mathbf{q}_1 \rangle$, $\langle \mathbf{q}_2, \mathbf{q}_{12} \rangle$ and $\langle \mathbf{q}_3, \mathbf{q}_{13} \rangle$ (red) are available on mirror π_1 (blue)

where $\mathbf{x}_i = (x_i \ y_i \ 1)^\top$ is the normalized camera coordinate of \mathbf{q}_i . Since the mirrored points \mathbf{p}_i ($i = 1, 2$) are then given by Eq (2.9) as

$$\mathbf{p}_i = H_i \mathbf{p}_0 - 2d_i \mathbf{n}_i, \quad (3.25)$$

we obtain

$$\begin{aligned} \mathbf{x}_i \times \mathbf{p}_i &= \mathbf{x}_i \times (H_i \mathbf{p}_0 - 2d_i \mathbf{n}_i), \\ &= [\mathbf{x}_i] \times [H_i \ -2\mathbf{n}_i] \begin{bmatrix} \mathbf{p}_0 \\ d_i \end{bmatrix}, \\ &= \mathbf{0}_{3 \times 1}. \end{aligned} \quad (3.26)$$

Similarly, the second reflection \mathbf{p}_{ij} is also collinear with its projection \mathbf{q}_{ij} :

$$\begin{aligned} &(A^{-1} \mathbf{q}_{ij}) \times \mathbf{p}_{ij}, \\ &= [\mathbf{x}_{ij}] \times (H_i \mathbf{p}_j - 2d_i \mathbf{n}_i), \\ &= [\mathbf{x}_{ij}] \times (H_i (H_j \mathbf{p}_0 - 2d_j \mathbf{n}_j) - 2d_i \mathbf{n}_i), \\ &= [\mathbf{x}_{ij}] \times [H_i H_j \ -2\mathbf{n}_i \ -2H_i \mathbf{n}_j] \begin{bmatrix} \mathbf{p}_0 \\ d_i \\ d_j \end{bmatrix}, \\ &= \mathbf{0}_{3 \times 1}. \end{aligned} \quad (3.27)$$

3.5. Mirror Parameters Estimation

By using these constraints, we obtain a linear system of \mathbf{p}_0 , d_1 and d_2 :

$$\begin{aligned} & \begin{bmatrix} [\mathbf{x}_0] \times & \mathbf{0}_{3 \times 1} & \mathbf{0}_{3 \times 1} \\ h_1 & -2[\mathbf{x}_1] \times \mathbf{n}_1 & \mathbf{0}_{3 \times 1} \\ h_2 & \mathbf{0}_{3 \times 1} & -2[\mathbf{x}_2] \times \mathbf{n}_2 \\ h'_{1,2} & -2[\mathbf{x}_{12}] \times \mathbf{n}_1 & -2h''_{1,2} \\ h'_{2,1} & -2h''_{2,1} & -2[\mathbf{x}_{21}] \times \mathbf{n}_2 \end{bmatrix} \begin{bmatrix} \mathbf{p}_0 \\ d_1 \\ d_2 \end{bmatrix}, \\ & = K \begin{bmatrix} \mathbf{p}_0 \\ d_1 \\ d_2 \end{bmatrix} = \mathbf{0}_{15 \times 1}, \end{aligned} \quad (3.28)$$

where $h_i = [\mathbf{x}_i] \times H_i$, $h'_{i,j} = [\mathbf{x}_{ij}] \times H_i H_j$, $h''_{i,j} = [\mathbf{x}_{ij}] \times H_i \mathbf{n}_j$. By computing the eigenvector corresponding to the smallest eigenvalue of $K^\top K$, $(\mathbf{p}_0, d_1, d_2)^\top$ can be determined up to a scale factor. In this paper, we choose the scale that normalizes $d_1 = 1$.

Notice that Eq (3.28) apparently has 15 equations, but only 10 of them are linearly independent. This is simply because each of the cross products by Eqs (3.24) and (3.27) 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 $N_\pi \geq 3$ cases as well. In $N_\pi = 3$, considering the reflection of \mathbf{p}_{23} by π_1 as $\lambda_{123}\mathbf{q}_{123} = A\mathbf{p}_{123} = AS_1\mathbf{p}_{23}$, we have

$$[\mathbf{x}_{123}] \times \begin{bmatrix} (H_1 H_2 H_3)^\top \\ -2\mathbf{n}_1^\top \\ -2(H_1 \mathbf{n}_2)^\top \\ -2(H_1 H_2 \mathbf{n}_3)^\top \end{bmatrix}^\top \begin{bmatrix} \mathbf{p}_0 \\ d_1 \\ d_2 \\ d_3 \end{bmatrix} = \mathbf{0}_{3 \times 1}. \quad (3.29)$$

3.5.3 Kaleidoscopic Bundleadjustment

Once estimated the mirror normals \mathbf{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:

$$K' \mathbf{p}_0 = -K'' \boldsymbol{\delta}, \quad (3.30)$$

as $\mathbf{p}_0^* = -(K'^\top K')^{-1} K'^\top K'' \boldsymbol{\delta}$, where $\boldsymbol{\delta} = (d_1, d_2)^\top$, K' is the 15×3 matrix corresponding to the first three columns of K :

$$K' = [[\mathbf{x}_0]^\top, h_1^\top, h_2^\top, h'_{1,2}^\top, h'_{2,1}^\top]^\top, \quad (3.31)$$

and K'' is the 15×3 matrix corresponding to the 4th and 5th columns of K :

$$K'' = \begin{bmatrix} \mathbf{0}_{3 \times 1} & \mathbf{0}_{3 \times 1} \\ -2[\mathbf{x}_1]_{\times} \mathbf{n}_1 & \mathbf{0}_{3 \times 1} \\ \mathbf{0}_{3 \times 1} & -2[\mathbf{x}_2]_{\times} \mathbf{n}_2 \\ -2[\mathbf{x}_{12}]_{\times} \mathbf{n}_1 & -2h''_{1,2} \\ -2h''_{2,1} & -2[\mathbf{x}_{21}]_{\times} \mathbf{n}_2 \end{bmatrix}. \quad (3.32)$$

By reprojecting this \mathbf{p}_0^* to each of the chambers as

$$\begin{aligned} \lambda \hat{\mathbf{q}}_0 &= A \mathbf{p}_0^*, \\ \lambda \hat{\mathbf{q}}_i &= AS_i \mathbf{p}_0^* \quad (i = 1, 2), \\ \lambda \hat{\mathbf{q}}_{i,j} &= AS_i S_j \mathbf{p}_0^* \quad (i, j = 1, 2, i \neq j), \end{aligned} \quad (3.33)$$

we obtain a reprojection error as

$$\mathbf{E}(\mathbf{n}_1, \mathbf{n}_2, d_1, d_2) = [\mathbf{q}_0 - \hat{\mathbf{q}}_0, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}'_{1,2}, \mathbf{e}'_{2,1}]^\top, \quad (3.34)$$

where $\mathbf{e}_i = \mathbf{q}_i - \hat{\mathbf{q}}_i$ and $\mathbf{e}'_{i,j} = \mathbf{q}'_{i,j} - \hat{\mathbf{q}}'_{i,j}$. By minimizing $\|\mathbf{E}(\cdot)\|^2$ nonlinearly over $\mathbf{n}_1, \mathbf{n}_2, d_1, d_2$, 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 $\check{\mathbf{q}}^*(A, \mathbf{d}, \mathbf{n}, \delta)$ denotes the reprojection point computed from A , \mathbf{d} , \mathbf{n} and δ , the proposed method defines the reprojection error as,

$$\mathbf{E}(A, \mathbf{d}, \mathbf{n}, \delta) = \Sigma_x \|\check{\mathbf{q}}_x - \check{\mathbf{q}}^*_x(A, \mathbf{d}, \mathbf{n}, \delta)\|^2. \quad (3.35)$$

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

3.7 Performance Evaluations of Chamber Assignment

This section provides evaluations of the proposed chamber assignment algorithm with synthesized and real data.

3.7. Performance Evaluations of Chamber Assignment

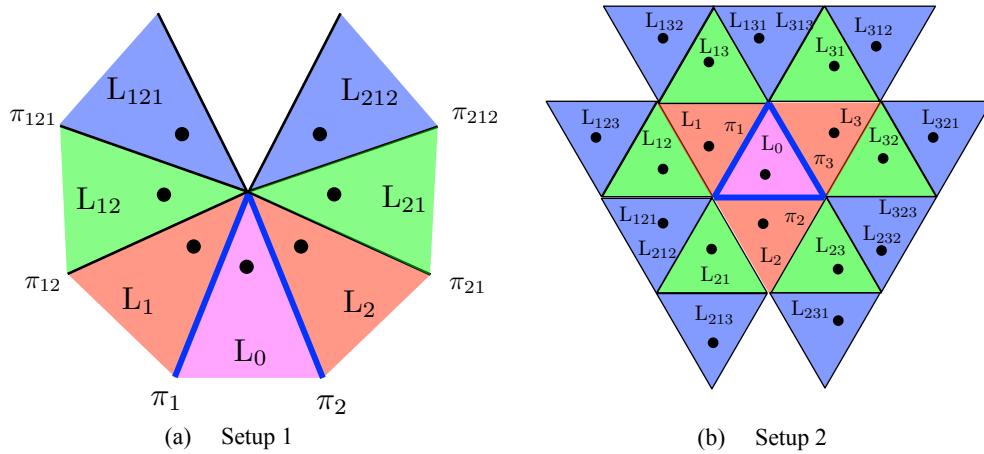


Figure 3.11: Mirror configurations. The red, green and blue regions indicate the chambers corresponding to the first, second, third reflections respectively. The bold blue lines indicate the real mirror positions.

3.7.1 Experimental Environment

The performance of the chamber labeling is evaluated in the following $N_\pi = 2$ and $N_\pi = 3$ scenarios (Figure 3.1).

1. two-mirror system using up to third reflections $r_i(i = 0, \dots, 7)$ (Figure 3.11(a))
 2. three-mirror system using up to second reflections $r_i(i = 0, \dots, 10)$ (Figure 3.11(b))

In the latter case with three mirrors, the mirrors are tilted at 5 degrees approximately in order to evaluate the performance with mirrors non-orthogonal to each other.

The performance is evaluated by the accuracy of labeling defined by

$$E_m = N_{m:\text{correct}}/N_m, \quad (3.36)$$

where N_m is the number of m th reflections and $N_{m:\text{correct}}$ is the number of projections labeled correctly.

3.7.2 Quantitative evaluations with synthesized data

Figures 3.12 and 3.13 report the average accuracy of our labeling in cases of $N_\pi = 2$ and $N_\pi = 3$ respectively, under different conditions: (a) with Proposition 1 only, (b) with Proposition 2 only, and (c) with Propositions

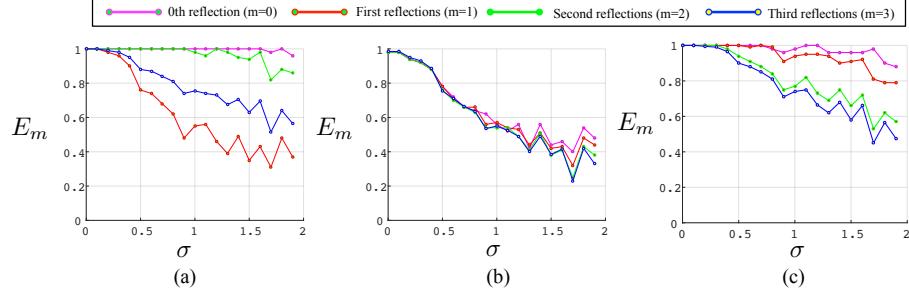


Figure 3.12: The average accuracy of labeling in $N_\pi = 2$ scenario. (a) with Proposition 1, (b) with Proposition 2, (c) with Propositions 1 and 2

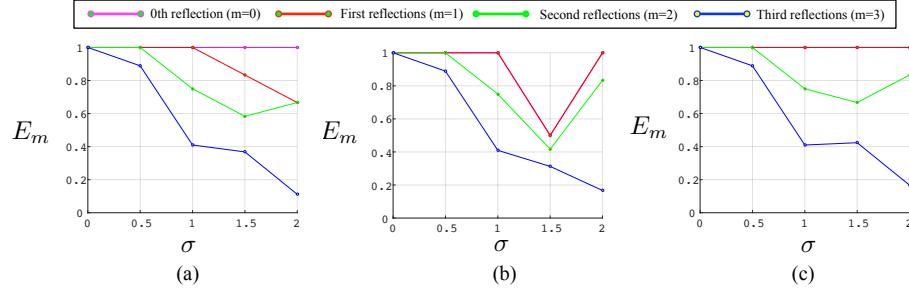


Figure 3.13: The average accuracy of labeling in $N_\pi = 3$ scenario. (a) with Proposition 1, (b) with Proposition 2, (c) with Propositions 1 and 2

1 and 2. In these figures, the magenta, red, green, and blue plots indicate the accuracy of labeling 0th, 1st, 2nd, and 3rd reflections respectively. σ_r denotes the standard deviation of zero-mean Gaussian pixel noise injected to the positions of the input points $\mathbf{r} \in R$, and the average accuracy is computed from the results of 50 trials at each noise level.

Figure 3.14 shows failure cases in Figures 3.12(a) and 3.12(b). In Figure 3.14(a), the mirror π_1 is reconstructed between \mathbf{r}_2 and \mathbf{r}_4 , and π_2 is reconstructed between \mathbf{r}_0 and \mathbf{r}_2 . These mirrors correspond to π_{21} and π_2 in the original configuration (Figure 3.11(a)), and such chamber assignment can result in a good reprojection error and a good recall ratio (Eq (3.17)) but violating Proposition 2. In case of Figure 3.14(b), the labeling is valid in terms of the reprojection error and the mirror angle, but not valid in terms of Proposition 1.

From these results, we can conclude that (1) our method can estimate the correct labeling in the ideal case without noise, and (2) the two propositions can improve the accuracy. In particular, since our algorithm evaluates all possible candidates, the perfect assignments for $\sigma_r = 0$ indicate that the suf-

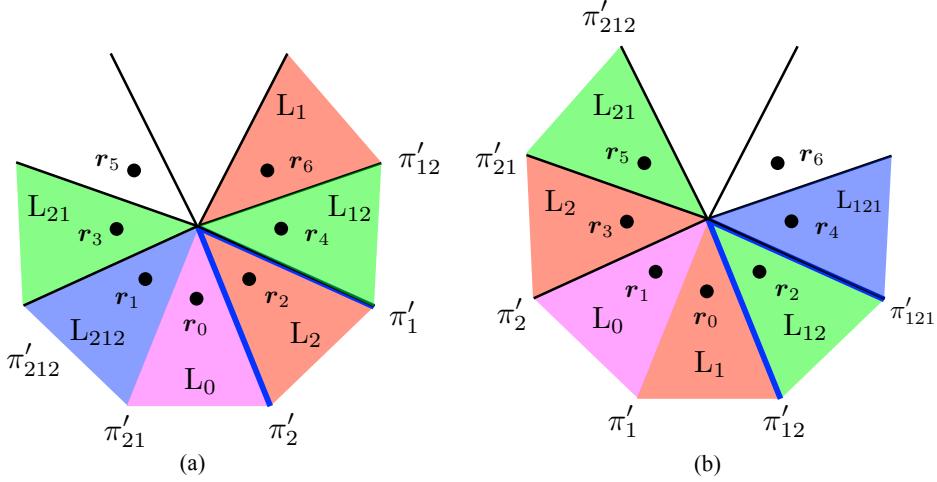


Figure 3.14: Failure cases. (a): The mirror π'_1 is reconstructed between r_2 and r_4 , and π'_2 is reconstructed between r_0 and r_2 . (b): The labeling is not valid in terms of the base chamber selection.

ficiency of the proposed conditions experimentally. Also these results proves experimentally that our method can work with non-orthogonal mirrors while the state-of-the-art algorithm[RiABR⁺13] assumes mirrors to be orthogonal to a common ground plane.

It should be noted that one of the main reasons for the degraded accuracy in the noisy condition is the performance of the mirror parameter estimation defined in Section 3.4.1. Since it minimizes the number of input points, the accuracy of the estimated mirror parameters can be sensitive to noise, and hence the projections of the reflections computed using such mirror parameters can fall far from the expected candidate points.

3.7.3 Qualitative evaluations with real data

Figure 3.15 shows our mirror-based imaging system. We use planar first surface mirrors and captured images with Nikon D600 (resolution 6016×4016). In this evaluation, the intrinsic parameter A of the camera is calibrated by Zhang's method [Zha00] in advance.

Figure 3.16 shows the chamber assignment results by our method for different mirror numbers and different number of reflections. We used projections of a single corner point in the images. The labels L_0, L_1, \dots indicate the assigned chambers, and the target objects are superimposed by colors according to the number of reflections. In the case of $N_\pi = 3$, each of the mirrors is tilted at about 5 degrees.

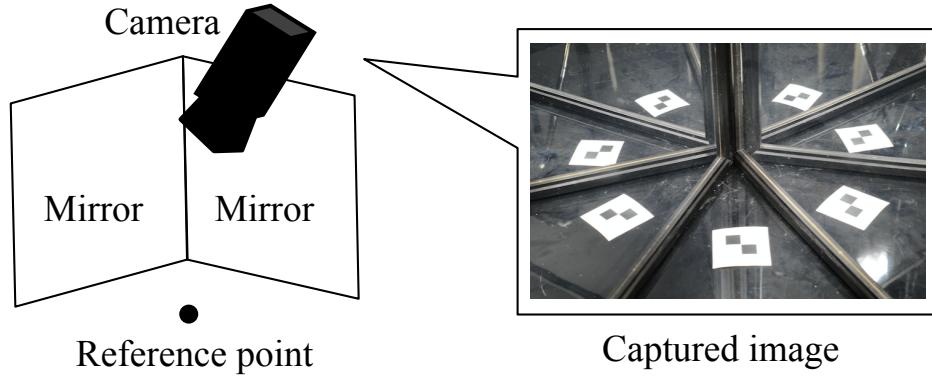


Figure 3.15: A illustration of a capture system (left) and a captured image (right). It consists of planar first surface mirrors and a camera.

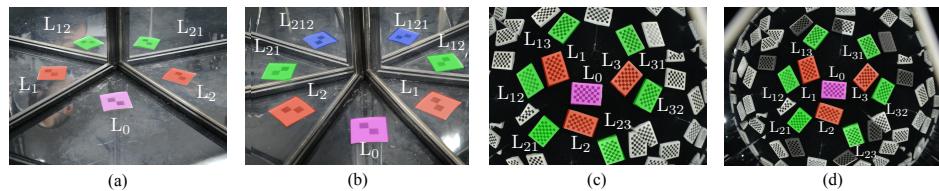


Figure 3.16: Chamber assignment results. The labels denote the assigned chambers and the magenta, red, green and blue colors are superimposed on the direct, the first, the second, and the third reflections.

From these results, we can observe that our method can successfully estimate the chamber assignment automatically in practice.

3.8 Performance Evaluation of Mirror parameter estimation

This section provides evaluations of the performance of the proposed method in terms of mirror parameter estimation. The proposed method was compared with following two conventional algorithms that both utilize a reference object of known geometry as shown in Figure 3.17. Notice that since Takahashi *et al.* [TNM12], which is introduced in Chapter 4, require more than two mirrors, we evaluate in $N_\pi = 3$ configuration.

[Baseline] Since the 3D geometry of the reference object is known, the 3D positions of the real image $\mathbf{p}_0^{(l)}$ and their reflections $\mathbf{p}_i^{(l)}$ and $\mathbf{p}_{i,j}^{(l)}$ can be estimated by solving PnP [LMNF09]. Here the superscript (l) indicates the

3.8. Performance Evaluation of Mirror parameter estimation

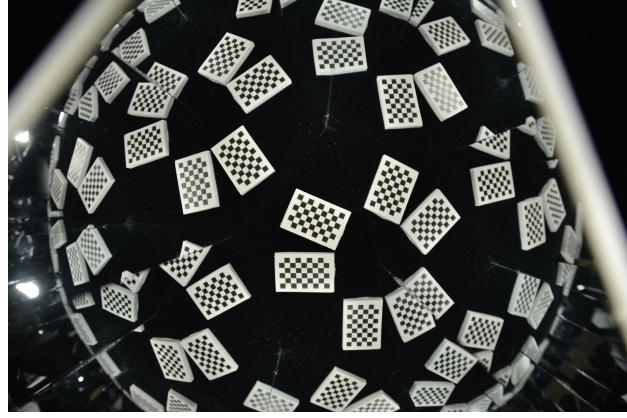


Figure 3.17: A capture of a chessboard used as the reference object for conventional methods.

l th landmark in the reference object. Once N_l such landmark 3D positions are given, then the mirror normals can be computed simply by

$$\begin{aligned}\mathbf{n}_1 &= \sum_l^{N_l} \mathbf{l}_{1,2,3}^{(l)} / \left\| \sum_l^{N_l} \mathbf{l}_{1,2,3}^{(l)} \right\|, \\ \mathbf{n}_2 &= \sum_l^{N_l} \mathbf{l}_{2,3,1}^{(l)} / \left\| \sum_l^{N_l} \mathbf{l}_{2,3,1}^{(l)} \right\|, \\ \mathbf{n}_3 &= \sum_l^{N_l} \mathbf{l}_{3,1,2}^{(l)} / \left\| \sum_l^{N_l} \mathbf{l}_{3,1,2}^{(l)} \right\|,\end{aligned}\quad (3.37)$$

where $\mathbf{l}_{i,j,k}^{(l)} = \mathbf{p}_i^{(l)} - \mathbf{p}_0^{(l)} + \mathbf{p}_{ij}^{(l)} - \mathbf{p}_j^{(l)} + \mathbf{p}_{ik}^{(l)} - \mathbf{p}_k^{(l)}$, and then the mirror distances can be computed by

$$\begin{aligned}d_1 &= \frac{1}{6N_l} \mathbf{n}_1^\top \sum_l^{N_l} \left(\sum_{i=0}^3 (\mathbf{p}_i^{(l)}) + \mathbf{p}_{12}^{(l)} + \mathbf{p}_{13}^{(l)} \right), \\ d_2 &= \frac{1}{6N_l} \mathbf{n}_2^\top \sum_l^{N_l} \left(\sum_{i=0}^3 (\mathbf{p}_i^{(l)}) + \mathbf{p}_{23}^{(l)} + \mathbf{p}_{21}^{(l)} \right), \\ d_3 &= \frac{1}{6N_l} \mathbf{n}_3^\top \sum_l^{N_l} \left(\sum_{i=0}^3 (\mathbf{p}_i^{(l)}) + \mathbf{p}_{31}^{(l)} + \mathbf{p}_{32}^{(l)} \right).\end{aligned}\quad (3.38)$$

Notice that the above PnP procedure requires a non-linear reprojection error minimization process in practice.

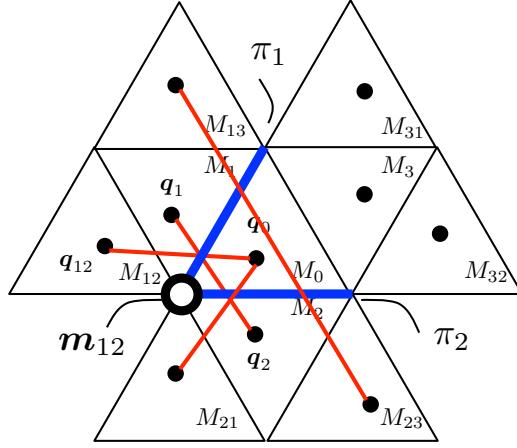


Figure 3.18: Corresponding points for the orthogonality constraint [TNM12]. Four pairs $\langle \mathbf{p}_1, \mathbf{p}_2 \rangle$, $\langle \mathbf{p}_0, \mathbf{p}_{21} \rangle$, $\langle \mathbf{p}_{12}, \mathbf{p}_0 \rangle$, and $\langle \mathbf{p}_{13}, \mathbf{p}_{23} \rangle$ are available for the intersection $\mathbf{m}_{12} = \mathbf{n}_1 \times \mathbf{n}_2$.

[Takahashi *et al.* [TNM12]] As pointed out by Takahashi *et al.* [TNM12], two 3D points \mathbf{p}_i and \mathbf{p}_j defined as reflections of a 3D point by different mirrors of normal \mathbf{n}_i and \mathbf{n}_j respectively satisfy an orthogonality constraint:

$$(\mathbf{p}_i - \mathbf{p}_j)^\top (\mathbf{n}_i \times \mathbf{n}_j) = (\mathbf{p}_i - \mathbf{p}_j)^\top \mathbf{m}_{ij} = 0. \quad (3.39)$$

As illustrated by Figure 3.18, this constraint on \mathbf{m}_{12} holds for four pairs $\langle \mathbf{p}_1, \mathbf{p}_2 \rangle$, $\langle \mathbf{p}_0, \mathbf{p}_{21} \rangle$, $\langle \mathbf{p}_{12}, \mathbf{p}_0 \rangle$, and $\langle \mathbf{p}_{13}, \mathbf{p}_{23} \rangle$ as the reflections of \mathbf{p}_0 , \mathbf{p}_1 , \mathbf{p}_2 , and \mathbf{p}_3 respectively. Similarly, $\langle \mathbf{p}_2, \mathbf{p}_3 \rangle$, $\langle \mathbf{p}_{21}, \mathbf{p}_{31} \rangle$, $\langle \mathbf{p}_0, \mathbf{p}_{32} \rangle$, and $\langle \mathbf{p}_{23}, \mathbf{p}_0 \rangle$ can be used for computing $\mathbf{m}_{23} = \mathbf{n}_2 \times \mathbf{n}_3$, and $\langle \mathbf{p}_3, \mathbf{p}_1 \rangle$, $\langle \mathbf{p}_{31}, \mathbf{p}_0 \rangle$, $\langle \mathbf{p}_{32}, \mathbf{p}_{12} \rangle$, and $\langle \mathbf{p}_0, \mathbf{p}_{13} \rangle$ can be used for $\mathbf{m}_{31} = \mathbf{n}_3 \times \mathbf{n}_1$. Once obtained the intersection vectors \mathbf{m}_{12} , \mathbf{m}_{23} and \mathbf{m}_{31} , the mirror normals and the distances can be estimated linearly as described in [TNM12].

The following three error metrics are used in this section in order to evaluate the performance of the proposed method in comparison with the above-mentioned conventional approaches quantitatively. The average estimation error of normal E_n measures the average angular difference from the ground truth by

$$E_n = \frac{1}{3} \sum_{i=1}^3 |\cos^{-1}(\mathbf{n}_i^\top \check{\mathbf{n}}_i)|, \quad (3.40)$$

where $\check{\mathbf{n}}_i$ ($i = 1, 2, 3$) denotes the ground truth of the normal \mathbf{n}_i . The average estimation error of distance E_d is defined as the average L_1 -norm to the

3.8. Performance Evaluation of Mirror parameter estimation

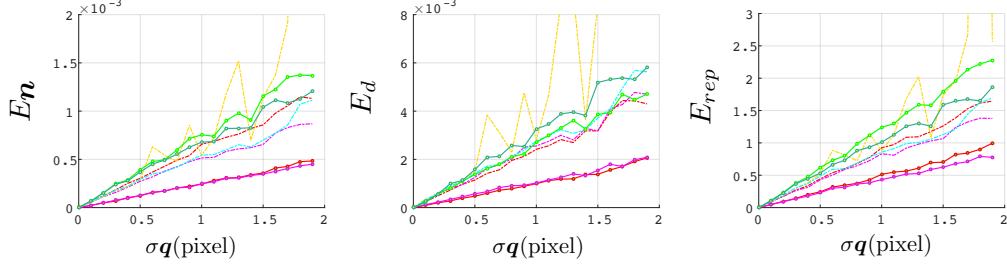


Figure 3.19: Estimation errors at different noise levels σq . Legends are provided in Figure 3.22.

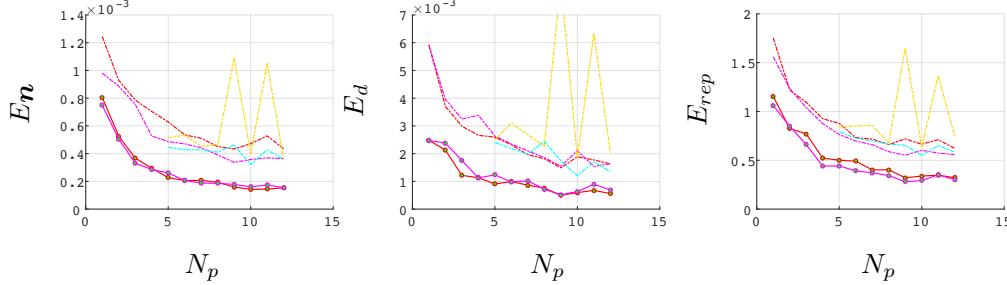


Figure 3.20: Estimation errors at different numbers of reference points N_p . Legends are provided in Figure 3.22.

ground truth:

$$E_d = \frac{1}{3} \sum_{i=1}^3 |d_i - \check{d}_i|, \quad (3.41)$$

where \check{d}_i ($i = 1, 2, 3$) denotes the ground truth of the distance d_i . Also, the average reprojection error E_{rep} is defined as:

$$E_{rep} = \frac{1}{10N_l} \sum_{l=1}^{N_l} \left| \mathbf{E}^{(l)}(\mathbf{n}_1, \mathbf{n}_2, \mathbf{n}_3, d_1, d_2, d_3) \right|, \quad (3.42)$$

where $\mathbf{E}^{(l)}(\cdot)$ denotes the reprojection error $\mathbf{E}(\cdot)$ defined by Eq (3.34) at l th point.

3.8.1 Quantitative evaluations with synthesized data

This section provides a quantitative performance evaluation using synthesized dataset. A virtual camera and three mirrors are arranged according to the real setup (Figure 3.28). By virtually capturing 3D points simulating a

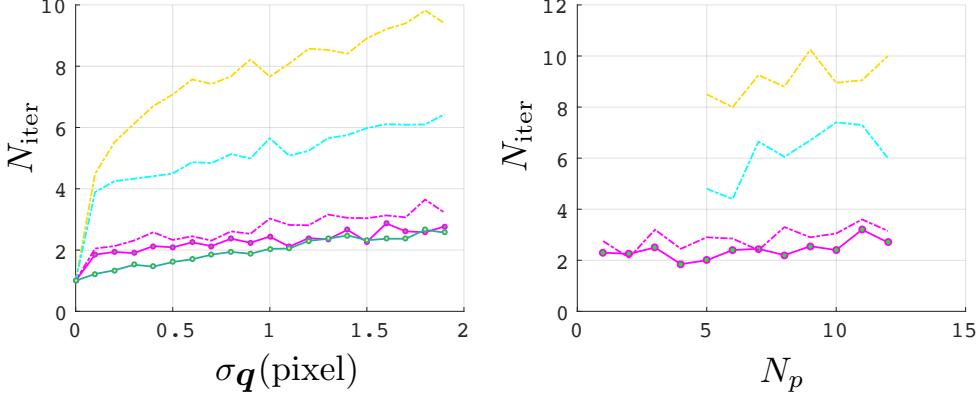


Figure 3.21: Number of iterations at different σ_q with $N_p = 5$ (left) and at different N_p with $\sigma_q = 1$ (right). Legends are provided in Figure 3.22.

Line	Method	Model	Bundle Adjustment	Point
—	Proposed	Non-planar	-	5
—	Proposed	Non-planar	✓	5
---	Proposed	Planar	-	5
- - -	Proposed	Planar	✓	5
—	Proposed	-	-	1
—	Proposed	-	✓	1
---	Takahashi et al.	Planar	✓	5
- - -	Baseline	Planar	✓	5

Figure 3.22: Legends for each configurations.

reference object, the corresponding 2D kaleidoscopic projections used as the ground truth are generated first, and then random pixel noise is injected to them at each trial of calibration.

Figures 3.19, 3.20 and 3.21 report average estimation errors $E_{\mathbf{n}}$, E_d , E_{rep} over 100 trials at different noise levels and different numbers of reference points. In these figures σ_q denotes the standard deviation of zero-mean Gaussian pixel noise, N_p denotes the number of 3D points used in the calibration, and N_{iter} denotes the number of iterations required by the kaleidoscopic bundle adjustment.

As shown in Figure 3.22, the magenta and red lines denote the results by the proposed method with and without the non-linear optimization (Section 3.5.3). They use kaleidoscopic projections of non-planar random five 3D

points, while the dashed red and magenta lines are the results with planar five points simulating the chessboard (Figure 3.17). The light and dark green lines are the results with a single 3D point generated randomly followed by the non-linear optimization or not.

The yellow and cyan dashed lines are the results by Takahashi *et al.* [TNM12] and the baseline with the same five points for the red and magenta dashed lines. Notice that the baseline and Takahashi *et al.* [TNM12] without the final non-linear optimization could not achieve comparable results (typically $E_{\text{rep}} \gg 10$ pixel). Also these methods using 3D reference positions without applying non-linear refinement after a linear PnP[LMNF09] could not estimate valid initial parameters for the final non-linear optimization. Therefore, they are omitted in these figures. On the other hand, the final non-linear optimization for our method does not improve the result drastically. This is because our algorithm originally utilizes the reprojection error constraint.

From these results, we can conclude that (1) the proposed method can achieve comparable estimation linearly even with a single 3D point (dark green), and (2) the proposed method (red and magenta) with the same number 3D points used in the conventional methods (yellow and cyan) performs better, even without the final non-linear optimization.

Also in particular in the cases of $\sigma_q \geq 1$, we can observe Takahashi *et al.* (yellow) do not show robust behavior. This is because the method degenerates obviously if the intersection vectors \mathbf{m}_{12} , \mathbf{m}_{23} and \mathbf{m}_{31} are parallel since the normal is recovered by $\mathbf{n}_i = \mathbf{m}_{ij} \times \mathbf{m}_{ki}$. Therefore if the estimated 3D reference points by PnP return the intersection vectors close to such a singular configuration due to noise, then it will not perform robustly[TNM12, Agr13].

3.8.2 Qualitative Evaluations with Real Data

In this evaluation, we utilized a camera and three planar mirrors. The intrinsic parameter A of the camera (Nikon D600, 6016×4016 resolution) is calibrated beforehand[Zha00].

Figures 3.17 shows a captured image of a chessboard, and Figure 3.23 shows the mirror normals and distances calibrated by the proposed method and the conventional methods. While the estimated mirror parameters look close to each other, the reprojection errors E_{rep} of the proposed, the baseline, and Takahashi *et al.* were 3.37, 4.75, and 13.6 pixels respectively. These reprojection errors are higher than simulation results and this is because of the localization accuracy of corresponding points and nonplanarity of mirrors. From these results, we can conclude that the proposed method performs reasonably in the real situation.

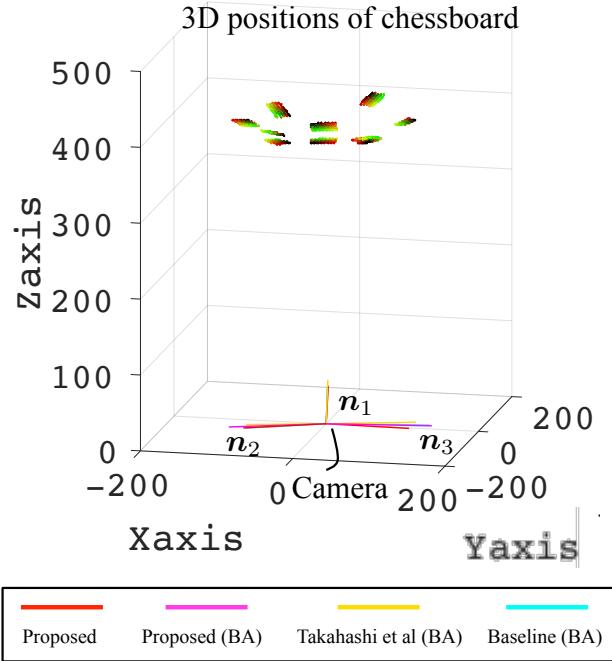


Figure 3.23: Calibration results. The colored lines in the bottom illustrate $d\mathbf{n}$ (*i.e.* the foot of perpendicular from the camera center) of each mirrors. The 10 patterns in the top illustrate the 3D points estimated by PnP.

3.9 Performance Evaluations of Intrinsic Camera Calibration

3.9.1 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 \mathbf{n}_i of each mirror are set to $(\cos\theta\cos\phi, \sin\theta, \cos\theta\sin\phi)$ with $(\theta, \phi) = (-8, 0)$ for \mathbf{n}_0 , $(\theta, \phi) = (186, 60)$ for \mathbf{n}_1 , and $(\theta, \phi) = (190, -60)$ for \mathbf{n}_2 . The distance of them d_i are set to $d_0 = 50\text{mm}$, $d_1 = 53\text{mm}$ and $d_2 = 54\text{mm}$. The image size is set to 1920×1080 . The ground truth of intrinsic parameters (f, u_0, v_0) and $\mathbf{d} = (k_1, k_2, k_3, p_1, p_2)$ 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 param-

3.9. Performance Evaluations of Intrinsic Camera Calibration

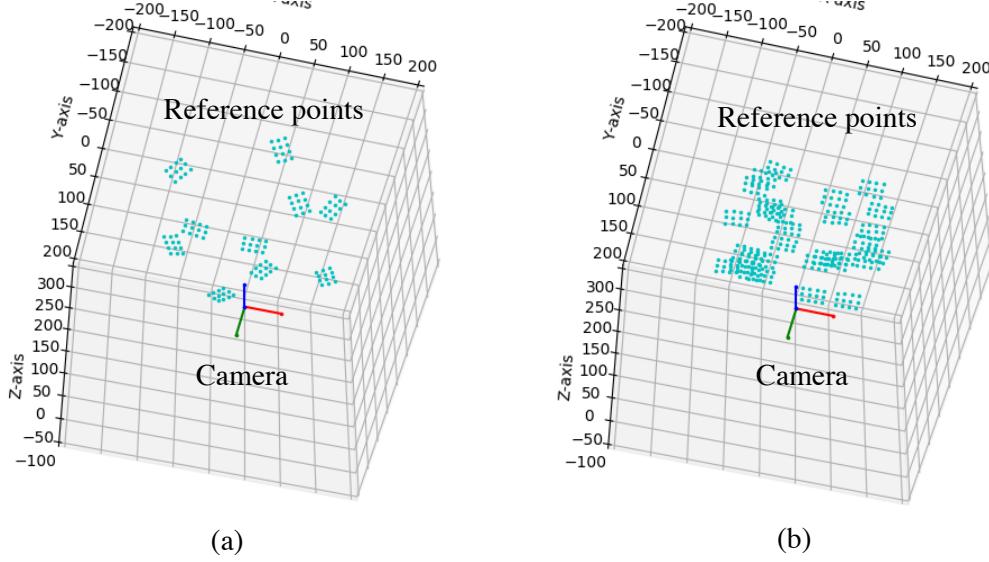


Figure 3.24: Configurations of intrinsic camera calibration. (a) utilizes the 3×4 chessboards and its reflections in 10 chambers. (b) utilizes 30 random poses of the chessboards.

eters as error metrics, that is $E_p = |p - p_g|$ where p denotes each intrinsic camera parameters and p_g 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[Zha00] (Baseline 1). For comparison, we set a 3×4 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 3.24(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 [Zha00] 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 3.24(b) (Baseline 2).

Figure 3.25 shows average estimation errors of each intrinsic camera parameters over 30 trials at different noise σ_q . The σ_q denotes the standard deviation of zero-mean Gaussian pixel noise injected to the observations \mathbf{q} . In each trial, the initial values of A and \mathbf{d} are generated by adding random noise whose level is less than 5% to the ground truth of each parameters.

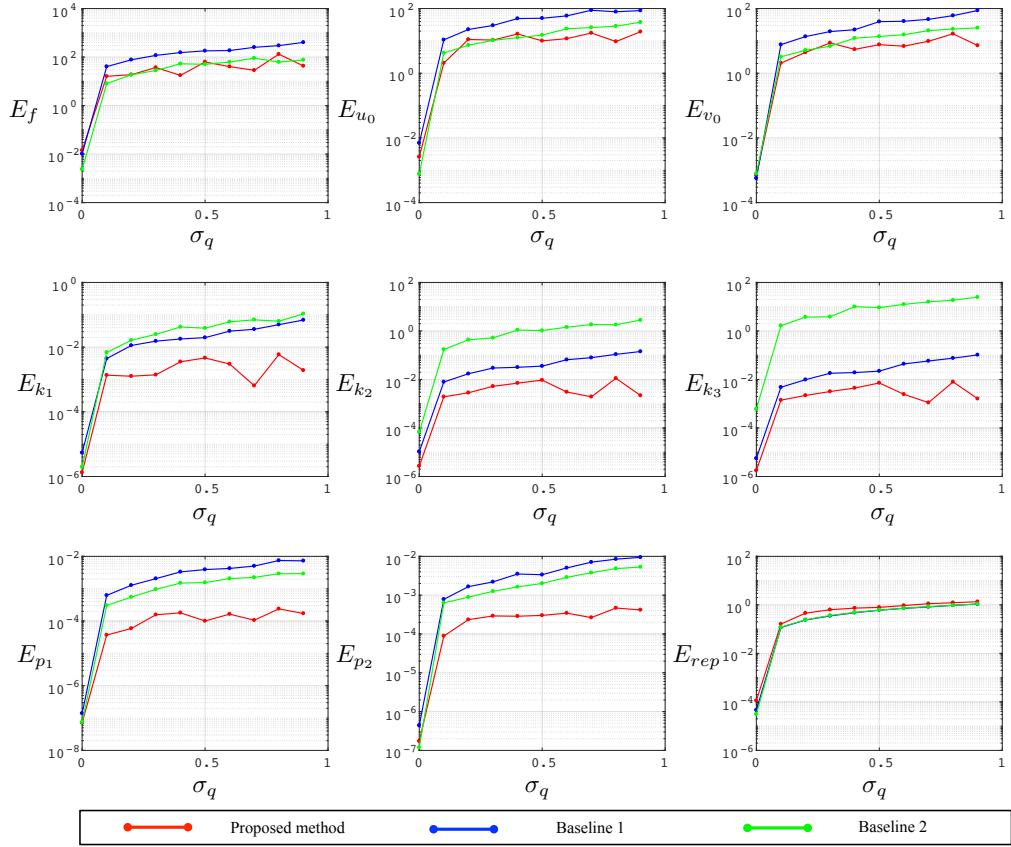


Figure 3.25: Estimation errors of each intrinsic parameter and reprojection error at different noise levels σ_q .

The red, and blue are results by the proposed method and Baseline 1 and green line is the result by Baseline 2.

In Figure 3.25, 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 than those by the other methods. We consider that this improvement is caused by the properties of the kaleidoscopic 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.

3.9. Performance Evaluations of Intrinsic Camera Calibration

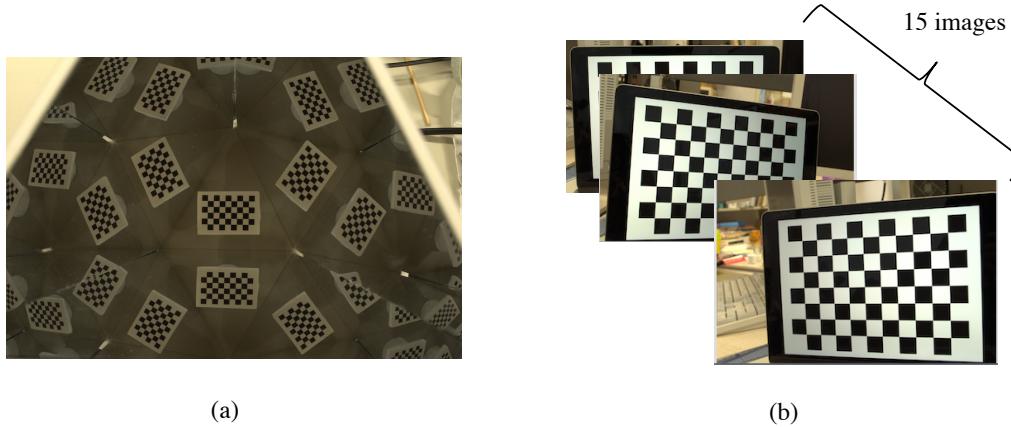


Figure 3.26: Configurations of intrinsic camera calibration with real data. (a) utilizes the 5×8 chessboards and its reflections in 10 chambers. (b) utilizes 15 random poses of the larger 7×10 chessboards.

Table 3.1: Intrinsic parameters estimated by each method.

Parameters	Proposed	Baseline 1	Baseline 2
f	5882.2	12318.3	6283.4
u_0	2474.4	2462.7	2427.6
v_0	1639.4	1622.9	1686.2
k_1	-0.0103	-0.3381	-0.1059
k_2	0.0143	-86.1886	0.3035
k_3	-0.0020	2007.4	-0.9600
p_1	0.0056	0.0740	0.00109
p_2	0.0036	0.0625	0.0006

3.9.2 Qualitative Evaluations with Real Data

Experimental Environment

As illustrated in Figure 3.26, 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[Zha00] (Baseline 1) as with 3.9.1. 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[Zha00], we utilize a larger reference object which has 7×10 chessboard with 20.5mm corners (Baseline 2).

Table 3.2: Reprojection error by each method for evaluating robustness.

Data	Proposed	Baseline 1	Baseline 2
1	0.526	0.487	0.524
2	1.735	13.014	0.933

Results

Table 3.1 reports the each parameters estimated by each method. Notice that proposed method and [Zha00] with the kaleidoscopic imaging system utilize 10 projections of the reference object, that is projections in each chambers (Figure 3.26(a)), and [Zha00] with the ideal configuration utilizes 15 images of the larger reference object (Figure 3.26(b)). From these results, while the intrinsic parameters estimated by Baseline 1 are 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 3.2 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 our proposed method estimates robust intrinsic parameters.

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

3.10 Discussion

3.10.1 Ambiguity of chamber assignment

In the case of the 8-point algorithm for the regular two-view extrinsic calibration[HZ00], the linear algorithm return four possible combinations of the rotation and the translation, and we can choose the right combination by examining if triangulated 3D points appear in front of the cameras. The mirror normal estimation in Section 3.4.1 is a special case of the 8-point algorithm, and this has such sign ambiguity on the mirror normal as described in Section 3.4.1. This ambiguity is also solved by considering the result of triangulation. In

3.10. Discussion

other words, estimating the essential matrix is identical to estimating mirror normal.

In addition to the sign ambiguity, the normal estimation for kaleidoscopic system has another family of ambiguity due to multiple reflections. As introduced in Section 3.4.1, particular combinations of kaleidoscopic projections can return physically infeasible solutions, and they can be rejected by additional geometric constraints as done for the 8-point algorithm. However, there exists another class of solutions due to a *sparse sampling* of the observations.

Consider a base structure by the pairs $\langle \mathbf{q}_0, \mathbf{q}_{12} \rangle$ and $\langle \mathbf{q}_2, \mathbf{q}_{121} \rangle$ in Figure 3.2. This configuration can estimate the mirror parameters successfully, one between \mathbf{p}_0 and \mathbf{p}_2 , and the other between \mathbf{p}_0 and \mathbf{p}_{12} . While the latter is a virtual mirror, this interpretation satisfies all the constraints in Section 3.4.1. In other words, we can assemble a mirror system of this configuration in practice.

To solve this problem, Section 3.4.3 utilizes the recall ratio (Eq (3.17)) so that our algorithm returns the solution which reproduces as many as possible candidate points \mathbf{r} observed in the image.

3.10.2 RANSAC or PROSAC Approach for Chamber Assignment

While the proposed algorithm examines all possible base structures as introduced in Algorithm 1 to evaluate the performance thoroughly, we can also consider a RANSAC or PROSAC approach[CM05]. For example, we can first hypothesize the base chamber from N_r candidates, and then can consider only $N_\pi + 1$ nearest points around it for estimating the mirror parameters. Designing and evaluating such approach is one of our future work.

3.10.3 Degenerate Cases

The both proposed algorithms of chamber assignment and mirror parameters estimation are based on the kaleidoscopic projection constraint (Eq (3.6)) 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 (3.9), Eq (3.18) and Eq (3.19) 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 3.27).

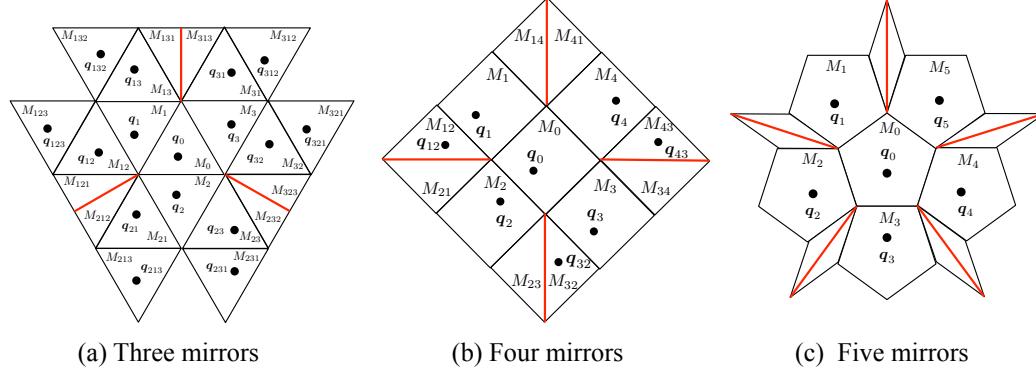


Figure 3.27: Kaleidoscopic imaging system using (a) three, (b) four, and (c) five mirrors. Discontinuities (red lines) appear on the boundaries of overlapping chambers.

3.10.4 3D Reconstruction with Kaleidoscopic Imaging System

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 3.28 shows our kaleidoscopic capture setup. In this evaluation, the intrinsic parameter A of the camera (Nikon D600, 6016×4016 resolution) is calibrated beforehand[Zha00]. As a target object, we utilized a *cat* (about $4 \times 5 \times 1$ 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 (Figure 3.28 left), and the projector itself is not involved in the calibration *w.r.t.* the camera and the mirrors. In this evaluations, the mirror parameters estimated in Section 3.8.2 were utilized.

Figure 3.29 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[FP09]. From these results, we can conclude that the proposed provides a sufficiently accurate calibration for 3D shape reconstruction.

3.11. Summary

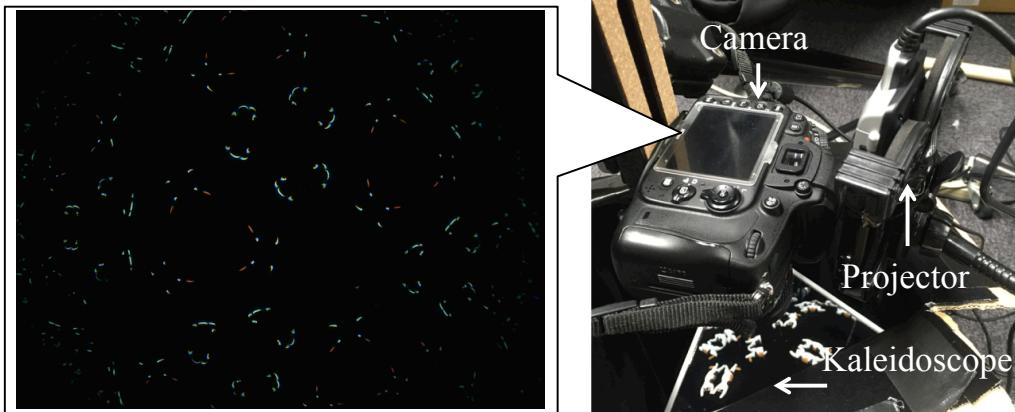


Figure 3.28: A kaleidoscopic capture setup for 3D reconstruction. It consists of three first surface mirrors, a camera and a laser projector.

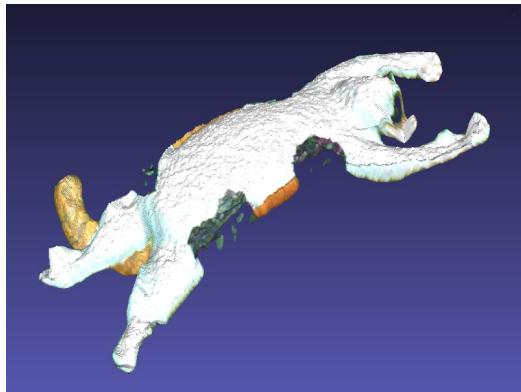


Figure 3.29: Reconstructed 3D shape of *cat* object.

3.11 Summary

This chapter proposes a novel algorithm of full-automatic intrinsic camera calibration algorithm introducing multiple mirrors and consisting a kaleidoscopic imaging system. In order to realize this algorithm, there are three challenges to be solved, *i.e.* (1) chamber assignment, (2) mirror parameters estimation and (3) intrinsic parameters estimation. For solving these problems, the proposed method introduces a *kaleidoscopic projection constraint* which is satisfied by projections of high-order reflections. This constraint enables the proposed method to conduct a 3D validation on candidates of chamber assignments and to derive mirror parameters linearly as introduced in Section 3.4 and Section 3.5. Evaluations with synthesized data and real data prove that the proposed method works properly in terms of chamber

assignment, mirror parameters estimation and intrinsic parameters estimation.

In this chapter, the proposed method mainly focuses on the perspective camera model with simple lens distortion model as described in Eq (2.3). Along with the development of imaging systems, a wide variety of camera models[NNM⁺15, GN05, KB06] and lens models[BL95, SN00] have been proposed. The adaptation to such configuration should be investigated as a future work.

Chapter 4

Mirror-based Extrinsic Camera Calibration

This chapter provides a novel planar-mirror based extrinsic camera calibration algorithm.

Extrinsic camera calibration is to estimate extrinsic parameters, *i.e.* pose and position of the camera against the reference object, and is widely used for a variety of 3D applications, such as 3D shape reconstruction from multi-view images[KRN97, AFS⁺10], vision-based robot navigation[DK02, FSP10], augmented reality[ABB⁺01], and so on. While conventional extrinsic camera calibration techniques[Zha00, LMNF09] assume that the reference object, *e.g.* a chessboard, can be observed by a camera directly, there are some cases where this assumption does not hold. For example, in case of utilizing a vision-based robot[HMR09] and display-camera system[HDKM10, FY07], although the extrinsic parameters between a camera and a reference object are important for their application, they are often unobservable from the camera directly.

In order to conduct the extrinsic camera calibration in such situation, the proposed method introduces 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 the single image of the reflection.

For this problem, the proposed method introduces an *orthogonality constraint* which is satisfied among reflections by multiple mirror poses. Based on this constraint, the proposed method established mirrored-point-based formulation which enables us to obtain a larger number of constraints which contribute to make the calibration more robust even with a simple configuration.

In this chapter, the measurement model of the proposed method is introduced in Section 4.1. Section 4.2 describes the key constraint of the proposed

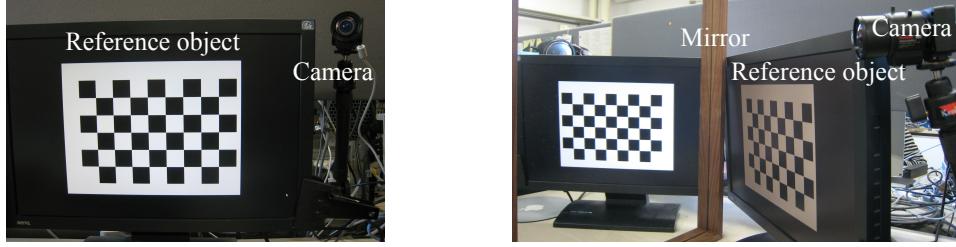


Figure 4.1: The set up of mirror-based extrinsic calibration.

method, that is orthogonality constraint on mirror reflections, and 4.3 provides the proposed extrinsic camera calibration algorithm based on it. The proposed method in case of utilizing three reference points and more than three reference points are evaluated in Section 4.4 and Section 4.5 respectively. Section 4.6 provides discussions on the limitation of our method and Section 4.7 summarizes this chapter.

4.1 Measurement Model

As illustrated by Figure 4.2, we denote a camera by C and a mirror by π_j ($j = 1, \dots, N_\pi$). We use $\{C\}$ to describe the camera C coordinate system which is used as the world coordinate system in this chapter. A vector \mathbf{p} in the Y coordinate system is expressed as $\mathbf{p}^{\{Y\}}$, while we may omit Y if it is clear from the context.

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

$$\mathbf{p}^{i\{C\}} = R \cdot \mathbf{p}^{i\{X\}} + \mathbf{t} \quad (i = 1, \dots, N_p), \quad (4.1)$$

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

Based on the planar mirror geometry introduced in Section 2.2.1, the

4.2. Orthogonality Constraint on Mirror Reflections

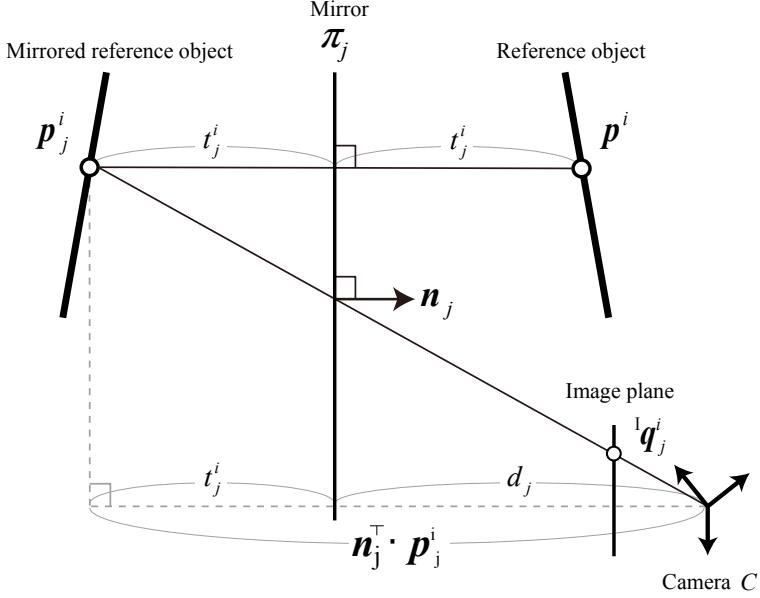


Figure 4.2: The measurement model

reference point $\mathbf{p}^{i\{C\}}$ and its reflection $\mathbf{p}_j^{i\{C\}}$ satisfy,

$$\mathbf{p}^{i\{C\}} = -2(\mathbf{n}_j^\top \mathbf{p}_j^{i\{C\}} + d_j)\mathbf{n}_j + \mathbf{p}_j^{i\{C\}}. \quad (4.2)$$

In addition, by removing $\mathbf{p}^{i\{C\}}$ from Eq (4.1) and Eq (4.2), we obtain

$$R \cdot \mathbf{p}^{i\{W\}} + \mathbf{t} = -2(\mathbf{n}_j^\top \cdot \mathbf{p}_j^{i\{C\}} + d_j)\mathbf{n}_j + \mathbf{p}_j^{i\{C\}}. \quad (4.3)$$

This is the fundamental equation which describes our measurement model.

4.2 Orthogonality Constraint on Mirror Reflections

Consider a reference point \mathbf{p}^i and its two mirrored points \mathbf{p}_j^i , $\mathbf{p}_{j'}^i$ by two different mirror planes π_j and $\pi_{j'}$ respectively. The axis vector $\mathbf{m}_{jj'}$ lying along the intersection of the two mirror planes is expressed as the cross product of each mirror normals, $\mathbf{m}_{jj'} = \mathbf{n}_j \times \mathbf{n}_{j'}$. This axis vector $\mathbf{m}_{jj'}$ satisfies the following *orthogonality constraint* [SB06] (Figure 4.3),

$$(\mathbf{p}_j - \mathbf{p}_{j'})^\top \cdot \mathbf{m}_{jj'} = 0. \quad (4.4)$$

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

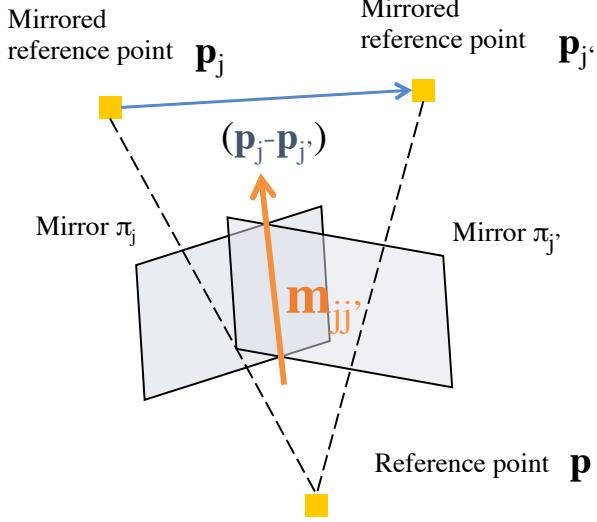


Figure 4.3: An illustration of orthogonality constraint.

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_π different mirror poses based on the orthogonality constraint.

Algorithm 2 shows an overview of our calibration algorithm. Firstly, we solve the PnP problem from mirrored reference points projected to the image plane, ${}^I\mathbf{q}_j^i$, and obtain their 3D positions \mathbf{p}_j^i as introduced in Section 4.3.1 ($N_p = 3$ case) or by EPnP [LMNF09] ($N_p > 3$ 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 \mathbf{p}_j^i , 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 \mathbf{t} by solving a large system of linear equations.

4.3.1 Unique Solution of Three P3P Problems

The original goal of P3P problem is to determine the 3D positions of the three points with respect to the camera frame $\{C\}$, \mathbf{p}_j^i ($i, j = 1, 2, 3$), by using the corresponding perspective projection \mathbf{q}_j^i on the image plane. As is well known, P3P problem has up to four solutions in general. We denote

4.3. Extrinsic Camera Calibration Using Orthogonality Constraint

Algorithm 2 An overview of extrinsic camera calibration using N_π mirror poses and N_p reference points

Require: ${}^I\mathbf{q}_j^i$ ($i = 1, \dots, N_p$, $j = 1, \dots, N_\pi$)
Ensure: R, \mathbf{t}

for all N_π mirror poses $\pi_j, j = 1, \dots, N_\pi$ **do**
 Solve the PnP and obtain the mirrored points \mathbf{p}_j^i ($N_p = 3$) or EPnP [LMNF09] ($N_p > 3$).
end for

for all N_π mirror poses $\pi_j, \pi_{j'}, (j, j' = 1, \dots, N_\pi, j \neq j')$ **do**
 Compute the axis vector $\mathbf{m}_{jj'}$ as the right-singular vector corresponding to the smallest singular value of $\mathbf{Q}_{jj'}$ in Eq (4.8).
end for

for all N_π mirror poses $\pi_j, (j = 1, \dots, N_\pi)$ **do**
 Compute the normal vector \mathbf{n}_j of the mirror plane π_j by Eq (4.10).
end for

Compute the each column of rotation matrix $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$ and the translation vector by solving Eq (4.11).
 Refine $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$ by solving the orthogonal Procrustes problem[GVL96].
 Refine R, \mathbf{t} by applying non-linear optimization [TMHF00]

these four candidates of \mathbf{p}_j^i as $P_{jk} = \{\mathbf{p}_{jk}^i\}$, ($k = 1, \dots, 4$). 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_{jk} ($j = 1, 2, 3$) as P_1, P_2 , and P_3 by using the orthogonality constraint. By applying the orthogonality constraint for the mirrored positions of three reference points \mathbf{p}_j^i ($i = 1, 2, 3$), we have:

$$\begin{pmatrix} (\mathbf{p}_j^1 - \mathbf{p}_{j'}^1)^\top \\ (\mathbf{p}_j^2 - \mathbf{p}_{j'}^2)^\top \\ (C\mathbf{p}_j^3 - \mathbf{p}_{j'}^3)^\top \end{pmatrix} \mathbf{m}_{jj'} = \mathbf{Q}_{jj'} \mathbf{m}_{jj'} = 0. \quad (4.5)$$

By multiplying $\mathbf{Q}_{jj'}^\top$ on the left side of Eq (4.5), we obtain

$$\mathbf{Q}_{jj'}^\top \mathbf{Q}_{jj'} \mathbf{m}_{jj'} = \mathbf{M}_{jj'} \mathbf{m}_{jj'} = 0. \quad (4.6)$$

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

In ideal noiseless cases, the smallest eigenvalue of $\mathbf{M}_{jj'}$ should be exactly zero if the 3D points \mathbf{p}_j^i and $\mathbf{p}_{j'}^i$ are results of Householder transformation since Eq (4.4) 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 $\mathbf{M}_{jj'}$; 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 (4.5) 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

$$\rho = \lambda_3 / (\lambda_1 + \lambda_2 + \lambda_3), \quad (4.7)$$

where λ_1 , λ_2 , and λ_3 denote the largest, second, and the smallest eigenvalues of $\mathbf{M}_{jj'}$ respectively. Since one combination of P3P solutions has three $\mathbf{M}_{jj'}$: \mathbf{M}_{12} , \mathbf{M}_{23} and \mathbf{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 (4.5) be zero is not proven theoretically here. However the evaluations by both synthesized and real data (Section 4.5) practically proves the validity of the solution selected by the above-mentioned process.

4.3.2 Computing the axis vector from mirror planes

As described in Sec 4.2, the axis vector $\mathbf{m}_{jj'}(j, j' = 1, \dots, N_\pi, j \neq j')$ and two mirrored points \mathbf{p}_j^i , $\mathbf{p}_{j'}^i(i = 1, 2, \dots, N_p)$ satisfy the orthogonality constraint (Eq (4.4)).

By applying this orthogonality constraint to N_p mirrored reference points \mathbf{p}_j^i , we obtain:

$$\begin{pmatrix} (\mathbf{p}_j^1 - \mathbf{p}_{j'}^1)^\top \\ (\mathbf{p}_j^2 - \mathbf{p}_{j'}^2)^\top \\ \vdots \\ (\mathbf{p}_j^{N_p} - \mathbf{p}_{j'}^{N_p})^\top \end{pmatrix} \mathbf{m}_{jj'} = \mathbf{Q}_{jj'} \mathbf{m}_{jj'} = 0. \quad (4.8)$$

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

4.3.3 Computing the normal vector of a mirror plane

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

$$\begin{aligned} \mathbf{n}_j^\top \cdot \mathbf{m}_{jj'} &= 0, \\ \mathbf{n}_{j'}^\top \cdot \mathbf{m}_{jj'} &= 0. \end{aligned} \quad (4.9)$$

4.3. Extrinsic Camera Calibration Using Orthogonality Constraint

When using N_π mirror poses, we obtain $(N_\pi - 1)$ equations of Eq (4.9) for one normal vector \mathbf{n}_j . By collecting these equations, we have

$$\begin{aligned} \mathbf{S}_j \mathbf{n}_j &= 0, \\ \mathbf{S}_j &= (\mathbf{m}_{j1} \mathbf{m}_{j2} \cdots \mathbf{m}_{jj-1} \mathbf{m}_{jj+1} \cdots \mathbf{m}_{jN_\pi})^\top \end{aligned} \quad (4.10)$$

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

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

4.3.4 Computing Extrinsic Parameters

Up to this point, we obtain the 3D positions of mirrored reference points $\mathbf{p}_j^i (i = 1, \dots, N_p, j = 1, \dots, N_\pi)$ and mirror normals \mathbf{n}_j . The 3D positions of reference points $\mathbf{p}^{i\{W\}} = (x_i, y_i, z_i)$ are supposed to be given a priori in its local coordinate system X . By substituting these known parameters into Eq (4.3), we can derive a large system of linear equations:

$$A\mathbf{Z} = \mathbf{B}. \quad (4.11)$$

where

$$A = \left[\begin{array}{ccccccc} \mathbf{I}_3 & 2\mathbf{n}_1 & \mathbf{0}_{3 \times 1} & \cdots & \mathbf{0}_{3 \times 1} & x_1 \mathbf{I}_3 & y_1 \mathbf{I}_3 & z_1 \mathbf{I}_3 \\ \mathbf{I}_3 & 2\mathbf{n}_1 & \mathbf{0}_{3 \times 1} & \cdots & \mathbf{0}_{3 \times 1} & x_2 \mathbf{I}_3 & y_2 \mathbf{I}_3 & z_2 \mathbf{I}_3 \\ & & & & & \vdots & & \\ \mathbf{I}_3 & 2\mathbf{n}_1 & \mathbf{0}_{3 \times 1} & \cdots & \mathbf{0}_{3 \times 1} & x_{N_p} \mathbf{I}_3 & y_{N_p} \mathbf{I}_3 & z_{N_p} \mathbf{I}_3 \\ \mathbf{I}_3 & \mathbf{0}_{3 \times 1} & 2\mathbf{n}_2 & \cdots & \mathbf{0}_{3 \times 1} & x_1 \mathbf{I}_3 & y_1 \mathbf{I}_3 & z_1 \mathbf{I}_3 \\ \mathbf{I}_3 & \mathbf{0}_{3 \times 1} & 2\mathbf{n}_2 & \cdots & \mathbf{0}_{3 \times 1} & x_2 \mathbf{I}_3 & y_2 \mathbf{I}_3 & z_2 \mathbf{I}_3 \\ & & & & & \vdots & & \\ \mathbf{I}_3 & \mathbf{0}_{3 \times 1} & 2\mathbf{n}_2 & \cdots & \mathbf{0}_{3 \times 1} & x_{N_p} \mathbf{I}_3 & y_{N_p} \mathbf{I}_3 & z_{N_p} \mathbf{I}_3 \\ & & & & & \vdots & & \\ \mathbf{I}_3 & \mathbf{0}_{3 \times 1} & \mathbf{0}_{3 \times 1} & \cdots & 2\mathbf{n}_{N_\pi} & x_1 \mathbf{I}_3 & y_1 \mathbf{I}_3 & z_1 \mathbf{I}_3 \\ \mathbf{I}_3 & \mathbf{0}_{3 \times 1} & \mathbf{0}_{3 \times 1} & \cdots & 2\mathbf{n}_{N_\pi} & x_2 \mathbf{I}_3 & y_2 \mathbf{I}_3 & z_2 \mathbf{I}_3 \\ & & & & & \vdots & & \\ \mathbf{I}_3 & \mathbf{0}_{3 \times 1} & \mathbf{0}_{3 \times 1} & \cdots & 2\mathbf{n}_{N_\pi} & x_{N_p} \mathbf{I}_3 & y_{N_p} \mathbf{I}_3 & z_{N_p} \mathbf{I}_3 \end{array} \right], \quad (4.12)$$

$$\mathbf{Z} = [\mathbf{t}^\top \ d_1 \ d_2 \ \cdots \ d_{N_\pi} \ \mathbf{r}_1^\top \ \mathbf{r}_2^\top \ \mathbf{r}_3^\top]^\top, \quad (4.13)$$

$$\mathbf{B} = [\mathbf{B}_1 \ \mathbf{B}_2 \ \cdots \ \mathbf{B}_{N_\pi}]^\top, \quad (4.14)$$

$$\mathbf{B}_j = [\mathbf{b}_j^1 \ \mathbf{b}_j^2 \ \cdots \ \mathbf{b}_j^{N_p}], \quad (4.15)$$

$$\mathbf{b}_j^i = (-2\mathbf{n}_j^\top \mathbf{p}_j^i \mathbf{n}_j + \mathbf{p}_j^i)^\top. \quad (4.16)$$

The vectors \mathbf{r}_1 , \mathbf{r}_2 and \mathbf{r}_3 denote the first, second and third column of the rotation matrix R . From N_π mirror poses and N_p reference points, we have $12 + N_\pi$ unknown parameters and $3 \times N_\pi \times N_p$ equations. Hence, when $N_\pi \geq 3$ and $3 \times N_\pi \times N_p > 12 + N_\pi$, we can solve the Eq (4.11) by $\mathbf{Z} = \mathbf{A}^* \mathbf{B}$, where \mathbf{A}^* is the pseudo-inverse matrix of \mathbf{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 $\mathbf{p}^{i\{W\}} = (x_i, y_i, 0)^\top$ and we cannot compute the third column vector \mathbf{r}_3 of the rotation matrix. In this case, we compute \mathbf{r}_3 as the cross product of first and second column vector $\mathbf{r}_1, \mathbf{r}_2$, that is $\mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2$.

4.3.5 Linear Refinement of Rotation Matrix by Solving the Orthogonal Procrustes Problem

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

$$\begin{aligned} |\mathbf{r}_1| &= |\mathbf{r}_2| = |\mathbf{r}_3| = 1, \\ \mathbf{r}_1^\top \mathbf{r}_2 &= \mathbf{r}_2^\top \mathbf{r}_3 = \mathbf{r}_3^\top \mathbf{r}_1 = 0. \end{aligned} \quad (4.17)$$

Here, we solve the orthogonal Procrustes problem [GVL96] and obtain a rotation matrix which satisfies Eq (4.17) and is closest to the original linear solution as proposed in Zhang's method[Zha00]. That is $R = \mathbf{U}\mathbf{V}^\top$, where \mathbf{U} and \mathbf{V} are given by as the SVD of the original matrix $(\mathbf{r}_1 \ \mathbf{r}_2 \ \mathbf{r}_3) = \mathbf{U}\Sigma\mathbf{V}^\top$.

4.3.6 Non-Linear Refinement of Extrinsic Parameters

In general, obtained extrinsic parameters can be refined by non-linear optimization[TMHF00]. Here, we minimize following reprojection error function,

$$\mathbf{E}_{opt} = \sum_{j=1}^{N_\pi} \sum_{i=1}^{N_p} |\mathbf{q}_j^i - \check{\mathbf{q}}_j^i(R, \mathbf{t}, \mathbf{n}_j, d_j)|, \quad (4.18)$$

where $\check{\mathbf{q}}_j^i(R, \mathbf{t}, \mathbf{n}_j, d_j)$ denote the reprojected point calculated from estimated parameters. We solved this non-linear optimization problem of Eq (4.18) with Levenberg-Marquardt algorithm.

4.4 Performance Evaluations in $N_p = 3$ case

This section provides experimental evaluations using synthesized and real data in $N_p = 3$ case. In these evaluations, we compare our method with

4.4. Performance Evaluations in $N_p = 3$ case

state-of-the-arts for $N_p = 3$ configuration proposed by Hesch *et al.* [HMR09]. Notice that the original method by Hesch *et al.* [HMR09] 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.

4.4.1 Quantitative Evaluations with Synthesized Data Experiment Environment

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

$$K = \begin{bmatrix} 500 & 0 & 300 \\ 0 & 500 & 250 \\ 0 & 0 & 1 \end{bmatrix}. \quad (4.19)$$

The normal vectors $\mathbf{n}_j (j = 1, \dots, N_\pi)$ of mirror poses π_j are set to $(\sin \theta_z \sin \theta_x + \cos \theta_x \cos \theta_z \sin \theta_y, \sin \theta_x \cos \theta_z + \cos \theta_x \sin \theta_z \sin \theta_y, \cos \theta_x \cos \theta_y)$ with $(\theta_x, \theta_y, \theta_z) = (0, 150, 0)$ for \mathbf{n}_1 , $(-30, 180, 0)$ for \mathbf{n}_2 , and $(0, 210, 0)$ for \mathbf{n}_3 . The reference points are defined as $\mathbf{p}^{1\{X\}} = (0, 0, 0)^\top$, $\mathbf{p}^{2\{X\}} = (225, 0, 0)^\top$ and $\mathbf{p}^{3\{X\}} = (0, 225, 0)^\top$. The rotation parameter R of the camera is set to the identity matrix. The position \mathbf{t} is generated for each trial by assigning a random value within $[0 : 20]$ to each x, y and z element of \mathbf{t} .

Error Rate

Our algorithm selects the best combination of P3P solutions based on the values of Eq (4.7). 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 (4.7) becomes unreliable. As a result, our algorithm can fail to select the best solution from 64 candidates in terms of the reprojection error.

Figure 4.4 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 \mathbf{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.

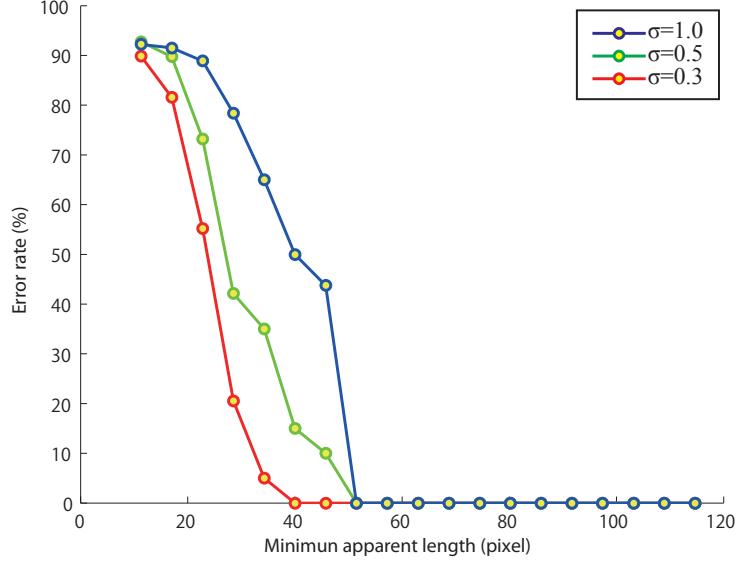


Figure 4.4: Error rate against apparent observation size. If the reference object is observed smaller than a certain size, the selection process can choose a wrong solution.

Error metrics

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

The estimation error of R is defined as the Riemannian distance [Moa02]:

$$E_R = \frac{1}{N_\pi} \sum_{j=1}^{N_\pi} \frac{1}{\sqrt{2}} \| \text{Log}(R_j^\top R_{g,j}) \|_F \quad (4.20)$$

$$\text{Log}R' = \begin{cases} 0 & (\theta = 0), \\ \frac{\theta}{2 \sin \theta} (R' - R'^\top) & (\theta \neq 0), \end{cases} \quad (4.21)$$

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

The estimation error of t is defined as the root mean square error:

$$E_t = \frac{1}{N_\pi} \sum_{j=1}^{N_\pi} \sqrt{\| t_j - t_{g,j} \|^2 / 3}. \quad (4.22)$$

The estimation error of n is defined as the angle against the ground truth n_g :

$$E_n = \frac{1}{N_\pi} \sum_{j=1}^{N_\pi} \arccos(n_j^\top n_{g,j}). \quad (4.23)$$

4.4. Performance Evaluations in $N_p = 3$ case

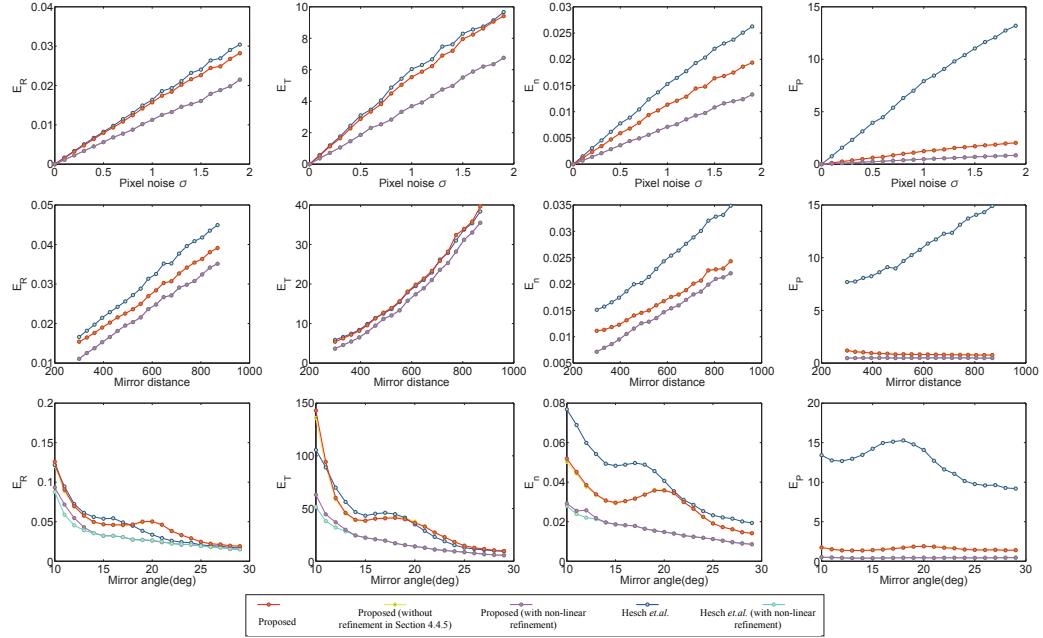


Figure 4.5: Calibration errors against different (a) pixel noise, (b) distance between the mirrors and camera, and (c) mirror poses. Note that (1) all trials of (b) and (c) have Gaussian pixel noise with zero-mean and standard deviation $\sigma = 1.0$, and (2) the plots of red and yellow are almost overlapped. This applies to the magenta and cyan plots as well.

The reprojection error is defined as follows:

$$E_P = \frac{1}{N_\pi \times N_p} \sum_{j=1}^{N_\pi} \left(\sum_{i=1}^{N_p} (\check{\mathbf{q}}_j^i - \mathbf{q}_j^i) \right), \quad (4.24)$$

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

Results Figure 4.5 shows results of quantitative evaluations. The top row (a) shows the performance against pixel noise. We added Gaussian noise with zero-mean and standard deviation $\sigma = 0, \dots, 2$ to the synthesized 2D observation \mathbf{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

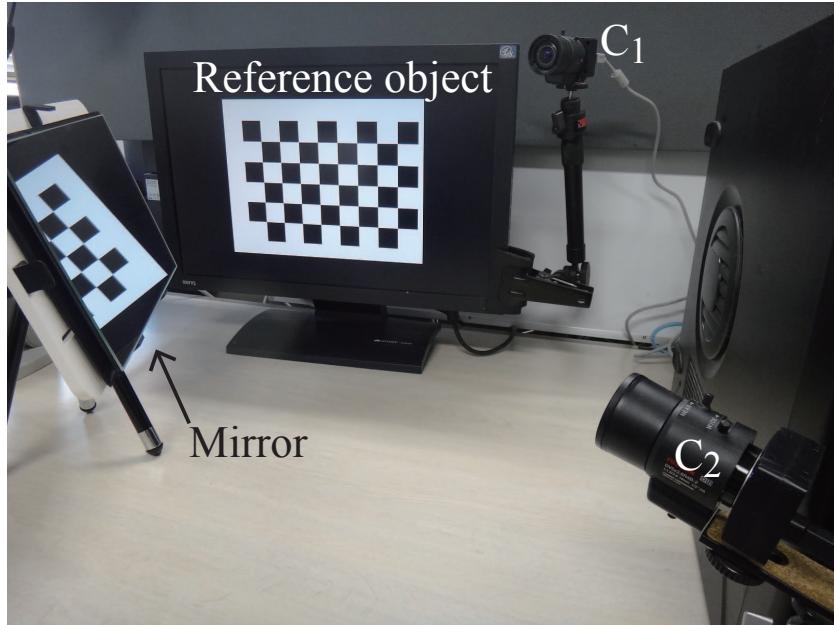


Figure 4.6: Calibration setup

1000 trials for each noise level, distance, and angle. The plots in magenta and cyan show results by the proposed and [HMR09] followed by the non-linear optimization of reprojection errors. These results quantitatively prove that (1) our method outperforms [HMR09] in terms of reprojection errors, and (2) that improvement is mainly provided by the improvement on n . Plus, our result is comparable with the results after non-linear optimization. Even though both non-linear optimizations started from ours and [HMR09] 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 solving the orthogonal Procrustes problem does not contribute much in terms of these estimation accuracies. In this sense, this refinement can be regarded just as a numerical process to ensure R be a rotation matrix, not to improve the result.

4.4.2 Qualitative Evaluations with Real Data

Experiment Environment

Figure 4.6 shows an overview of the setup for calibration. We have two cameras (PointGrey Chameleon CMLN01352C) C_1 and C_2 , a flat-panel display

4.4. Performance Evaluations in $N_p = 3$ case

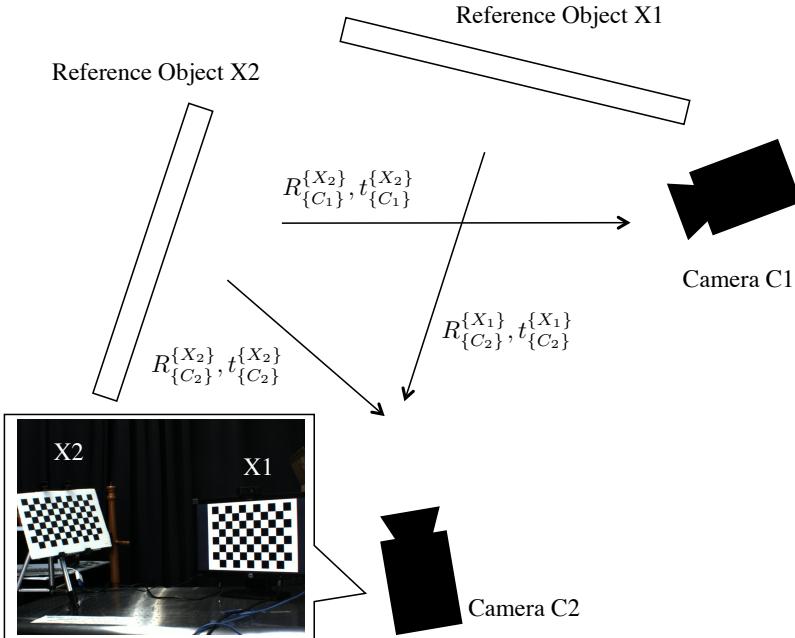


Figure 4.7: Environment for calibration with real data.

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 display. 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.

Baseline calibration

In order to evaluate the performance with real data, we prepared a baseline extrinsic parameter of camera C_1 against X_1 . We locate C_2 where it can observe both X_1 and the mirror directly. Using the observation of X_1 , we can obtain the extrinsic parameter of C_2 against X_1 by Zhang's method[Zha00]. 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 4.7, suppose $R_{\{C_k\}}^{X_l}$, $t_{\{C_k\}}^{X_l}$ denote the rotations and translations from $\{X_l\}$ to $\{C_k\}$ respectively. That is, the 3D position of a 3D point $\mathbf{p}^{i\{X_l\}}$ of X_l in $\{C_k\}$ is given by

$$\mathbf{p}^{i\{C_k\}} = R_{\{C_k\}}^{X_l} \cdot \mathbf{p}^{i\{X_l\}} + \mathbf{t}_{\{C_k\}}^{X_l} \quad (k = 1, 2, l = 1, 2). \quad (4.25)$$

We can calibrate $R_{\{C_1\}}^{X_1}$, $t_{\{C_1\}}^{X_1}$ by the proposed method as well as by [SB06]

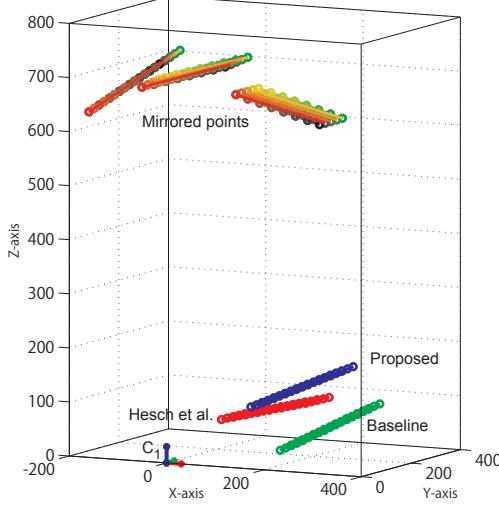


Figure 4.8: Estimated positions of the reference object by the proposed method (blue), by [HMR09] (red), and by baseline (green). Notice that this figure renders estimate positions of ${}^C\mathbf{p}$ and X_1 in $\{C\}$, and therefore C_1 is located at $(0, 0, 0)^\top$.

Table 4.1: Comparison of the proposed method and [HMR09] against baseline

	Hesch <i>et al.</i> [HMR09]	Proposed
E_R	0.239	0.057
E_t	60.57	48.04
E_p	3.653	0.170

and [RBN10]. In addition, by calibrating $R_{\{C_2\}}^{\{X_1\}}$, $\mathbf{t}_{\{C_2\}}^{\{X_1\}}$, $R_{\{C_1\}}^{\{X_2\}}$, $\mathbf{t}_{\{C_1\}}^{\{X_2\}}$, $R_{\{C_2\}}^{\{X_2\}}$ and $\mathbf{t}_{\{C_2\}}^{\{X_2\}}$ by Zhang’s method, these parameters also provide $R_{\{C_1\}}^{\{X_1\}}$, $\mathbf{t}_{\{C_1\}}^{\{X_1\}}$ as a baseline calibration:

$$\begin{aligned} R_{\{C_1\}}^{\{X_1\}} &= R_{\{C_1\}}^{\{X_2\}} R_{\{C_2\}}^{\{X_2\}\top} R_{\{C_2\}}^{\{X_1\}} \\ \mathbf{t}_{\{C_1\}}^{\{X_1\}} &= R_{\{C_1\}}^{\{X_2\}} R_{\{C_2\}}^{\{X_2\}\top} (\mathbf{t}_{\{C_2\}}^{\{X_1\}} - \mathbf{t}_{\{C_2\}}^{\{X_2\}}) + \mathbf{t}_{\{C_1\}}^{\{X_2\}}. \end{aligned} \quad (4.26)$$

4.4.3 Results

Figure 4.8 shows results by the proposed method, by [HMR09], and by baseline. This figure renders estimate positions of ${}^C\mathbf{p}$ and X_1 in $\{C\}$, and therefore C_1 is located at $(0, 0, 0)^\top$. Table 4.1 reports quantitative comparisons of

them. E_R and E_t are the differences against baseline. E_P is the reprojection errors against the original observations. Notice that even though the chess pattern provides more than 3 reference points, we used only three points of them for both our method and [HMR09]. From these results, we can conclude that our method performs better than [HMR09].

4.5 Performance Evaluations in $N_p > 3$ case

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.* [SB06] and by Rodrigues *et al.* [RBN10] 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 (fx, fy, cx, cy) ; fx and fy represents the focal length in pixels, and cx and cy represent the 2D coordinates of the principle point. We set them to $(500, 500, 300, 250)$ in this evaluation respectively.

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

The reference object consists of N_p reference points 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.

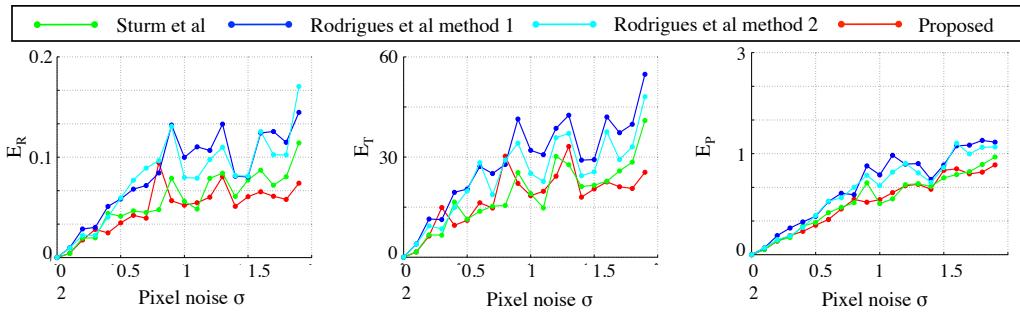
We represent the ground truth of rotation matrix as a product of three elemental rotation matrices, that is $R = R_1(\theta_1)R_2(\theta_2)R_3(\theta_3)$, and we set random values to each angles θ_1, θ_2 and θ_3 within $[-10 : 10]$ respectively. The position \mathbf{t} is generated of each trial by assigning a random value within $[-5 : 5]$ to each x, y and z element of \mathbf{t} .

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

- (a) σ : the standard deviation of Gaussian pixel noise of zero-mean.
- (b) N_p : the number of reference points.

Parameter	Min	Max	Step	Default
σ	0	2	0.1	1
N_p	4	20	1	4
N_π	3	20	1	3

Table 4.2: The range of changing parameters.


 Figure 4.9: Estimation error of each parameter in changing the standard deviation σ of pixel noise added to the input.

(c) N_π : the number of mirror poses.

Table 4.2 describes the min, max, increment step and default value of the parameters. We computed the average of the estimation errors of 100 trials for each of combinations. While changing these parameters respectively, the other parameters are set to values in Default column, that is the minimum setup for Sturm *et al.* [SB06] and Rodrigues *et al.* [RBN10].

Throughout this experiment, we utilized the same error functions introduced in Section 4.4 for evaluating extrinsic parameters, that is the Riemanian distance [Moa02] E_R , the root mean square error E_t and the reprojection error E_p .

Results

Figure 4.9 shows results for different standard deviations σ of pixel noise added to the observations. The averages of each estimation error of proposed method are smaller than those of Sturm *et al.* [SB06] and Rodrigues *et al.* [RBN10]. This fact indicates that our method can estimate better initial values from same inputs, because of the larger number of constraints involved in the estimation. Notice that there exist some trials in each of which all the methods result in a same optimal value regardless of the differences between initial values returned by their linear methods. Besides, there exists some

4.5. Performance Evaluations in $N_p > 3$ case

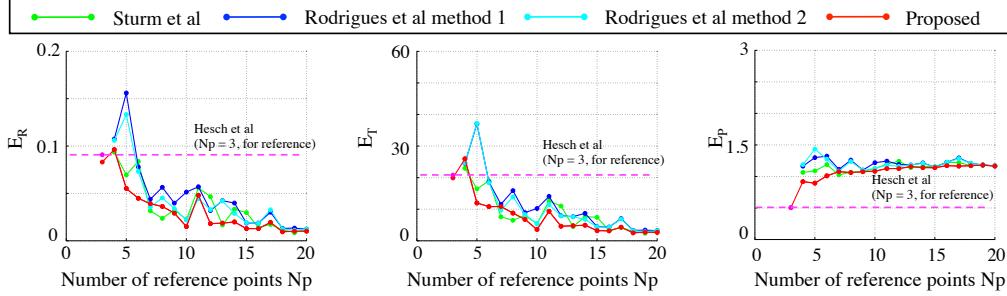


Figure 4.10: Estimation error of each parameter in changing the number of reference points N_p .

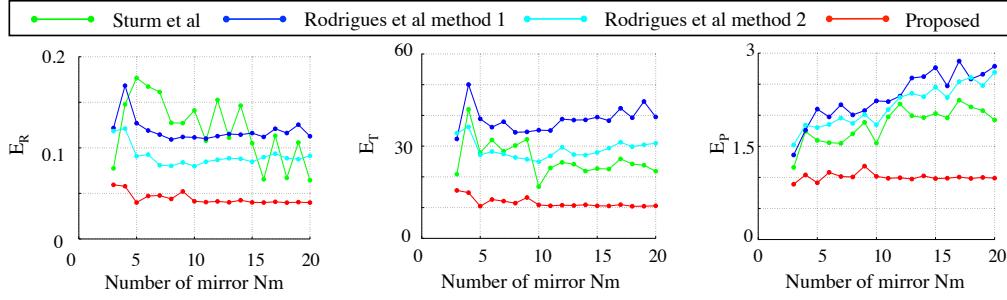


Figure 4.11: Estimation error of each parameter in changing the number of mirror poses N_π .

trials where all the methods fall in local minima. This is the reason for the spikes.

Figure 4.10 shows results for difference number of reference points. We added the results of Hesch *et al.* [HMR09], which is method for $N_p = 3$ scenario, as a reference. From these results, we can observe that the number of reference points affects the performance of each method drastically. These improvements are thought to be due to the improvement of estimation of mirrored reference point by PnP. In fact, Lepetit *et al.* [LMNF09] shows that the result of PnP in increasing number of points follows similar pattern of this experiments in their paper.

Figure 4.11 shows results in changing the number of mirror poses N_π . While Sturm's method does not improve with increasing number of mirror pose, we can see that proposed method and Rodrigues's method improves. This is considered to be due to the scalability of formulation for mirror pose, that is the number of equations for estimating extrinsic parameters in proposed method and Rodrigues's method changes depends on the number of mirror poses.

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

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. We utilized the same configuration introduced in 4.4.2 (Figure 4.7). 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 7×10 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_π UXGA images of different mirror poses $\pi_j (j = 1, \dots, N_\pi)$ for calibration.

Results

Figure 4.12 renders the estimated positions of the reference 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. This precision is acceptable for applications using display-camera system such as gaze detection for digital-signage or gaze correction[KPB⁺12] in a video conference scenario.

Figure 4.13 and 4.14 shows results in changing the number of reference points N_p and the number of mirror poses N_π respectively. Notice that the estimated parameters by each method are almost identical, and therefore we can observe only one line in Figure 4.14.

From these results, we can observe that our method performs better than conventional methods [SB06, RBN10, HMR09] in real situation qualitatively and quantitatively.

4.6 Discussion

4.6.1 Degenerate Case

Our algorithm does not work if it cannot compute enough axis vectors $\mathbf{m}_{jj'}$ for estimating mirror normals. This happens in the following three cases.

4.6. Discussion

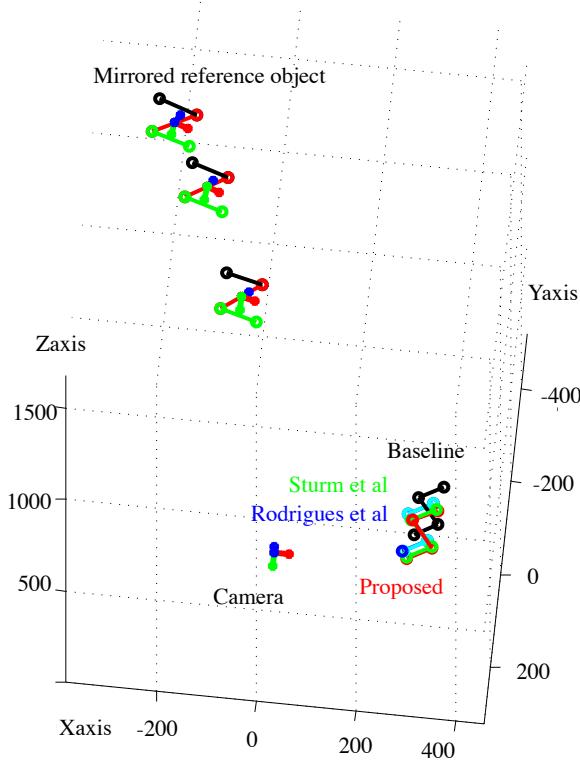


Figure 4.12: Estimated positions of the reference object by the proposed method (red), by [RBN10](blue and cyan), by [SB06](green) and by Eq (4.26) (black). Notice that this figure renders estimate positions of \mathbf{p} and X_1 in $\{C\}$, and therefore C_1 is located at $(0, 0, 0)^\top$.

(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 (4.10). These (1) and (2) cases has been originally observed by Sturm *et al.* [SB06]. (3) If reference points and the intersection of two mirrors π_j and $\pi_{j'}$ are on a same plane, the axis vector $\mathbf{m}_{jj'}$ is not be computable by solving Eq (4.8) though $\mathbf{m}_{jj'}$ does exist physically because of the following reason which makes the two rows of $M'_{jj'}$ corresponding to π_j and $\pi_{j'}$ be linearly dependent.

Proposition 3. *If two reference points \mathbf{p}^i and $\mathbf{p}^{i'}$ and the intersection of two mirrors π_j and $\pi_{j'}$ are on a same plane, the two lines connecting results of different Householder transformations of \mathbf{p}^i and $\mathbf{p}^{i'}$, i.e., the lines connecting \mathbf{p}_j^i to $\mathbf{p}_{j'}^i$ and $\mathbf{p}_j^{i'}$ to $\mathbf{p}_{j'}^{i'}$, are parallel (Figure 4.15).*

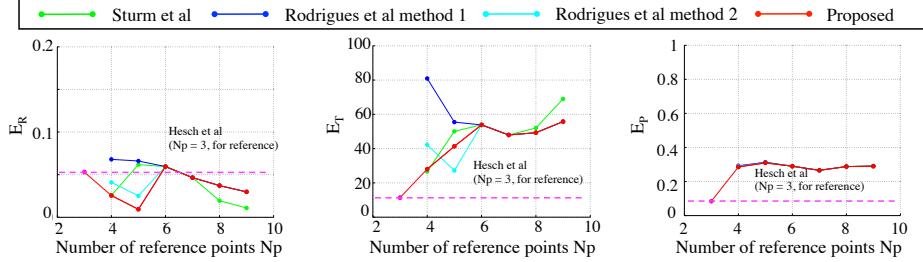


Figure 4.13: Estimation error of each parameter with real data in changing the number of reference points N_p .

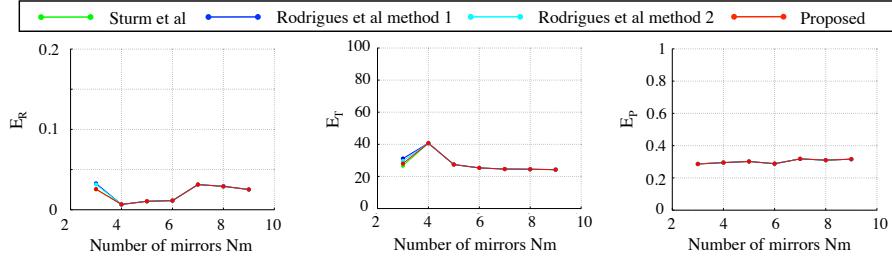


Figure 4.14: Estimation error of each parameter with real data in changing the number of mirror poses N_π .

Proof. Suppose the line connecting \mathbf{p}^i and $\mathbf{p}^{i'}$ intersects with the intersection of the two mirrors at \mathbf{O} as shown in Figure 4.15. By definition of the reflection, the distance from \mathbf{p}^i to \mathbf{O} is equal to the one from \mathbf{p}_j^i to \mathbf{O} . Similarly, the distance from $\mathbf{p}^{i'}$ to \mathbf{O} is equal to the one from $\mathbf{p}_j^{i'}$ to \mathbf{O} . Also, these distances are equal to the ones from $\mathbf{p}_{j'}^i$ to \mathbf{O} , and from $\mathbf{p}_{j'}^{i'}$ to \mathbf{O} respectively. Here $\triangle \mathbf{O}\mathbf{p}_j^i\mathbf{p}_{j'}^i$ and $\triangle \mathbf{O}\mathbf{p}_j^{i'}\mathbf{p}_{j'}^{i'}$ are isosceles triangles sharing the apex $\angle \mathbf{p}_j^i\mathbf{O}\mathbf{p}_{j'}^i$. Therefore, the two lines \mathbf{p}_j^i to $\mathbf{p}_{j'}^i$ and $\mathbf{p}_j^{i'}$ to $\mathbf{p}_{j'}^{i'}$ are parallel. \square

These three degenerate cases can be detected by observing the rank of $M_{jj'}$ in Eq (4.8). If the rank is less than 2, we can discard the mirror pair and try with more mirrored images in practice.

4.6.2 Sufficiency of the Orthogonality Constraint

The orthogonality constraint holds for two reflections of a single reference point and the axis vector, as a necessary condition. Obviously, this does not constrain the position of the mirror (the parameter d of Eq (4.3)). That is a mirror of another distance satisfies Eq (4.4) as long as it has the same intersection direction.

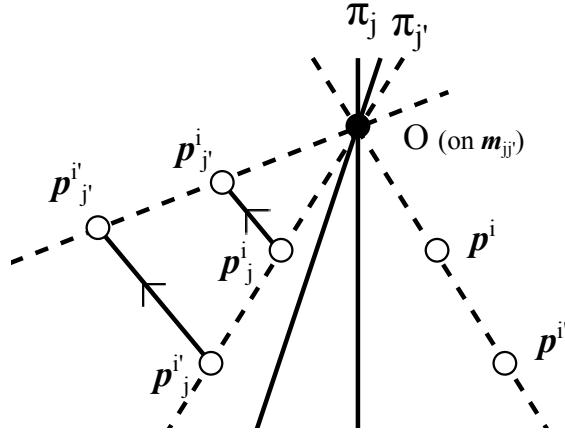


Figure 4.15: Degenerate case

This fact indicates that the orthogonality constraint itself does not serve as the sufficient condition to determine all of the mirror parameters. Instead, given three mirrors, it becomes the sufficient condition to obtain the mirror normals uniquely as described in Sections 4.3.2 and 4.3.3. Using the estimated normals, we can define the linear equations (Eq (4.11)) based on the measurement model (Eq (4.3)).

4.7 Summary

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

In this chapter, the proposed method utilized a single reflection of a reference object. However, in case of calibrating widely scattered cameras or an omnidirectional cameras, the reference object cannot be observable via single reflection. For such cases, the multiple reflections should be considered and the extrinsic camera calibration with them is the future work.

Chapter 5

Extrinsic Camera Calibration using Human Cornea Reflections

This chapter provides novel algorithms of extrinsic calibration using reflections on a human cornea for calibrating between a camera and a reference object located out of the camera's field of view.

For the cases where the reference object is not observable from the camera, some mirror-based approaches have been proposed as introduced in Section 4. Some of them aim to calibrate with simpler setups[SB06, KIFP08, RBN10, HMR09, Nay97, Agr13, FHB07, DLJP14], which means decreasing the number of required reference points or mirror poses, because a simpler setup provides a lot of advantages for more robust calibration and computational cost. However, it is sometimes troublesome to prepare a mirror and to calibrate the camera and the display every time in a casual scenario, *e.g.* gaze correction in a video conference as Kuster *et al.* reported [KPB⁺12].

In this chapter, we focus on cornea-reflection-based extrinsic camera calibration for occluded reference objects (Figure 5.1). The contribution of this work is to present two calibration algorithms with *simpler configuration*: they are three reference points and one spherical mirror pose (cornea sphere) and five reference points and one spherical mirror pose respectively.

This section is structured as follows. In Section 5.1, the measurement model based on a geometric relationship that holds when treating the human eyeball as a spherical mirror is introduced. In Section 5.2 and Section 5.3, the two types of human-cornea based algorithm are proposed and the performances of them are evaluated Section 5.4 and Section 5.5 respectively. Section 5.6 discusses the effects of differences of human cornea among individuals. Section 5.7 summarizes this chapter.

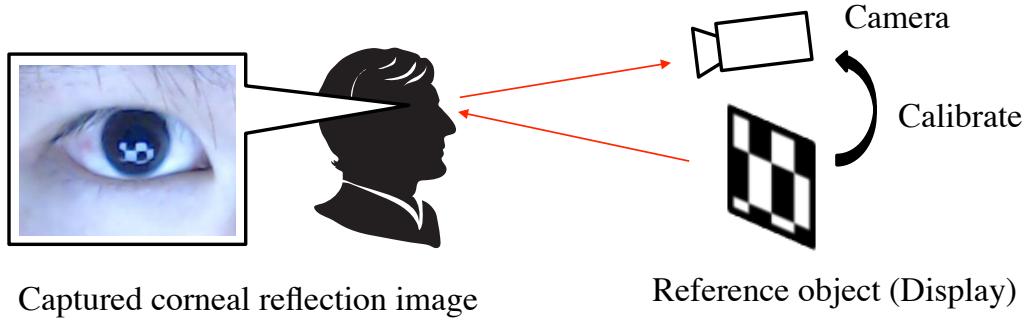


Figure 5.1: Cornea-reflection-based extrinsic camera calibration. The goal of our algorithm is to calibrate the camera and the reference object which lies out of the camera's field of view.

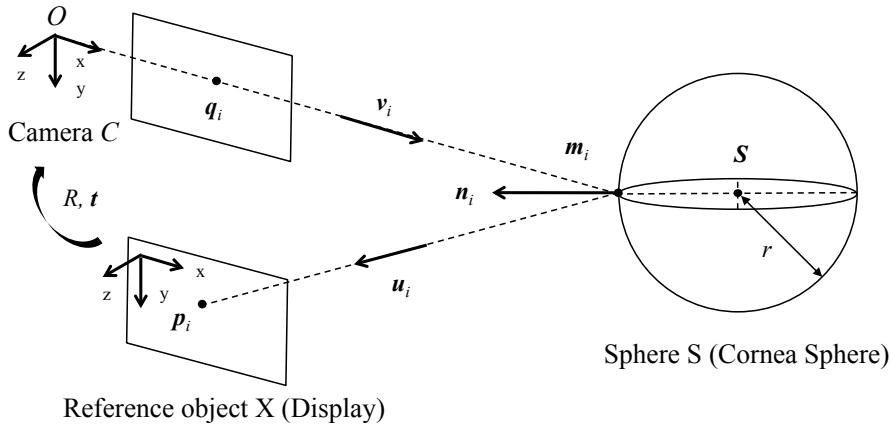


Figure 5.2: Reflection model of spherical mirror.

5.1 Measurement Model

In this section, we define a measurement model based on a geometric relationship that holds when treating the human eyeball as a spherical mirror.

Figure 5.2 illustrates the reflection model with a spherical mirror. Let X be the reference object, *i.e.* a display. This X lies out of the camera C 's field-of-view and X has N_p reference point \mathbf{p}_i ($i = 1, \dots, N_p$). The 2D points \mathbf{q}_i in image plane I denote the projections of reflections of \mathbf{p}_i . A rotation matrix R and a translation vector \mathbf{t} are the extrinsic parameters, which transform the reference object coordinate system $\{X\}$ to camera coordinate system $\{C\}$ and satisfy the following equation,

$$\mathbf{p}_i = R \mathbf{p}_i^{\{X\}} + \mathbf{t}, \quad (5.1)$$

5.1. Measurement Model

where $\mathbf{p}^{\{X\}}$ denotes the 3D position of \mathbf{p} in $\{X\}$. Here, we assume that the camera coordinate system $\{C\}$ is the world coordinate system and omit this superscript if a vector \mathbf{p} is represented in $\{C\}$. The goal of extrinsic camera calibration is to estimate these extrinsic parameters, R and \mathbf{t} , from the projections of the reference points.

As illustrated in Figure 5.2, we denote a center of the spherical mirror, its radius, and a reflection point of a reference point \mathbf{p}_i on the cornea sphere by \mathbf{S} , r and \mathbf{m}_i respectively. Let us assume a unit vector from \mathbf{m}_i to \mathbf{p}_i expressed as \mathbf{u}_i , \mathbf{p}_i is expressed as

$$\mathbf{p}_i = k_i \mathbf{u}_i + \mathbf{m}_i, \quad (5.2)$$

where k_i is the distance between \mathbf{m}_i and \mathbf{p}_i . By substituting Eq.(5.2) into Eq.(5.1), we obtain the following equation:

$$R\mathbf{p}_i^{\{X\}} + \mathbf{t} = k_i \mathbf{u}_i + \mathbf{m}_i. \quad (5.3)$$

This Eq.(5.3) is defined as the measurement model in this configuration.

From the laws of reflection, the unit vector \mathbf{u}_i is expressed as

$$\mathbf{u}_i = \mathbf{v}_i + 2(-\mathbf{v}_i^\top \cdot \mathbf{n}_i)\mathbf{n}_i, \quad (5.4)$$

where \mathbf{n}_i denotes the normal vector of a spherical mirror at \mathbf{m}_i . Since \mathbf{n}_i is the unit vector from the center of the cornea sphere \mathbf{S} to \mathbf{m}_i , \mathbf{n}_i is expressed as

$$\mathbf{n}_i = (\mathbf{m}_i - \mathbf{S}) / \|\mathbf{m}_i - \mathbf{S}\|, \quad (5.5)$$

where $\|\cdot\|$ means the Euclidean norm.

The \mathbf{m}_i is expressed as

$$\mathbf{m}_i = k'_i \mathbf{v}_i, \quad (5.6)$$

where \mathbf{v}_i denotes the unit vector connecting \mathbf{O} and \mathbf{m}_i and k'_i denotes the distance between them. This \mathbf{v}_i is expressed as

$$\mathbf{v}_i = (K^{-1} \mathbf{q}_i^I) / \|K^{-1} \mathbf{q}_i^I\|, \quad (5.7)$$

where \mathbf{q}_i denotes the projection of \mathbf{m}_i and K consists of intrinsic parameters, which are assumed to be given beforehand. The \mathbf{m}_i lies on the surface of the cornea sphere and satisfies $\|\mathbf{m}_i - \mathbf{S}\| = r$. By substituting Eq.(5.6) for this equation and multiplying it by itself, we have

$$k'^2 \|\mathbf{v}_i\|^2 - 2k'_i \mathbf{v}_i^\top \mathbf{S} + \|\mathbf{S}\|^2 - r^2 = 0, \quad (5.8)$$

This equation provides two solutions:

$$k'_i = (\mathbf{v}_i^\top \mathbf{S} \pm \sqrt{(\mathbf{v}_i^\top \mathbf{S})^2 - \|\mathbf{v}_i\|^2(\|\mathbf{S}\|^2 - r^2)}) / \|\mathbf{v}_i\|^2. \quad (5.9)$$

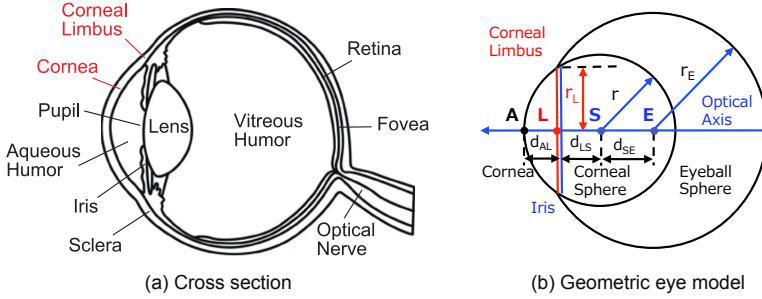


Figure 5.3: A human eyeball can be modeled as two overlapping spherical mirrors. (a) Cross section, (b) Geometric eye model based on [NN12]

Since \mathbf{m}_i is the point closer to the camera among the intersections of \mathbf{v}_i and the sphere surface, the smaller k'_i represents the distance between \mathbf{O} and \mathbf{m}_i .

From the above, the measurement model Eq (5.3) has 10 unknown parameters, that is R , \mathbf{t} , \mathbf{S} and r . Following sections provide two approaches for estimating the unknown parameters by modeling a human cornea as a spherical mirror whose center is \mathbf{S} and its radius is r as illustrated in Figure 5.3.

5.2 Approach 1: 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*. Assume that there exists two corresponding point sets $\mathbf{p}_i (i = 1, \dots, N_p)$ and $\mathbf{p}_i^{\{X\}}$, that satisfy Eq (5.3). By defining,

$$\begin{aligned}\bar{\mathbf{p}} &= \frac{1}{N_p} \sum_{i=1}^{N_p} \mathbf{p}_i & \mathbf{p}_{Ci} &= \mathbf{p}_i - \bar{\mathbf{p}}, \\ \bar{\mathbf{p}}^{\{X\}} &= \frac{1}{N_p} \sum_{i=1}^{N_p} \mathbf{p}_i^{\{X\}} & \mathbf{p}_{Ci}^{\{X\}} &= \mathbf{p}_i^{\{X\}} - \bar{\mathbf{p}}^{\{X\}},\end{aligned}\tag{5.10}$$

where N_p denotes the number of points, the correlation matrix \mathbf{H} is defined by

$$\mathbf{H} = \sum_{i=1}^{N_p} \mathbf{p}_{Ci} \mathbf{p}_{Ci}^{\{X\}\top}.\tag{5.11}$$

5.2. Approach 1: Solving Absolute Orientation Problem

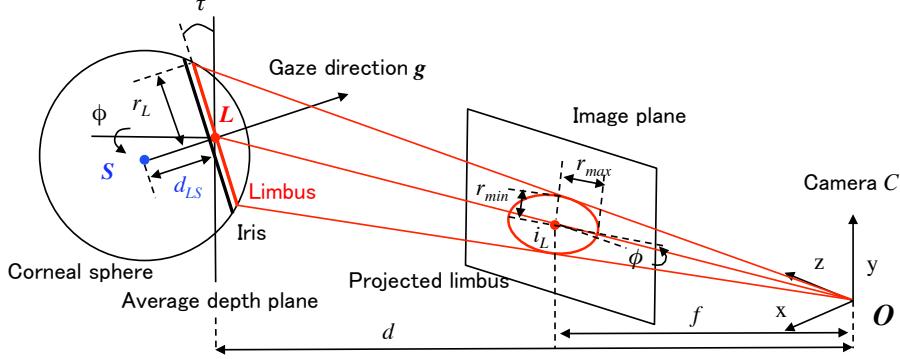


Figure 5.4: Estimating the center of the cornea sphere from limbus projection.

If the singular value decomposition of \mathbf{H} is given by $\mathbf{H} = \mathbf{U}\Lambda\mathbf{V}^\top$, then the optimal R and \mathbf{t} are as follows:

$$\begin{aligned} R &= \mathbf{V}\mathbf{U}^\top \\ \mathbf{t} &= \bar{\mathbf{p}} - R\bar{\mathbf{p}}^{\{X\}}. \end{aligned} \quad (5.12)$$

By using this method, we can obtain the extrinsic parameters from at least three point correspondences. Since the 3D positions of reference points $\mathbf{p}_i^{\{X\}}$ in $\{X\}$ are supposed to be given a priori, we estimate the 3D positions of reference point \mathbf{p}_i in $\{C\}$ by estimating $k_{m_i p_i}$, \mathbf{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 Estimating Parameters of Cornea Sphere Based on its Geometric Model

This section introduces a method to estimate the center of the cornea sphere, \mathbf{S} , from limbus projection by introducing a geometric model[NN12]. The average radius of the cornea sphere, r , and the average radius of the cornea limbus, r_L , are 7.7 mm and 5.6 mm respectively[RS97].

As illustrated in Figure 5.4, the limbus projection is modeled as an ellipse represented by five parameters: the center, \mathbf{i}_L , the major and minor radii, r_{max} and r_{min} , respectively, and rotation angle ϕ . 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 \mathbf{L} is expressed as

$$\mathbf{L} = dK^{-1}\mathbf{i}_L, \quad (5.13)$$

where d denotes the distance between the center of camera, \mathbf{O} , and the center of limbus \mathbf{L} , and is expressed as $d = f \cdot r_L / r_{max}$. f and K represent the focal length in pixels and intrinsic parameters, respectively. Gaze direction \mathbf{g} is approximated by the optical axis of the eye, and is theoretically determined by

$$\mathbf{g} = [\sin \tau \sin \phi, -\sin \tau \cos \phi, -\cos \tau]^\top, \quad (5.14)$$

where $\tau = \pm \arccos(r_{min}/r_{max})$; τ corresponds to the tilt of the limbus plane with respect to the image plane. Since the center of cornea sphere, \mathbf{S} , is located at distance d_{LS} ($= \sqrt{r^2 - r_L^2} = \sqrt{7.7^2 - 5.6^2} \approx 5.3$ mm), the radius of the cornea sphere from \mathbf{L} , we compute \mathbf{S} as follows,

$$\mathbf{S} = \mathbf{L} - d_{LS}\mathbf{g}. \quad (5.15)$$

In this way, we estimate \mathbf{S} from the ellipse parameters of the limbus projected onto the image plane, that is $(\mathbf{i}_L, \phi, r_{max}, r_{min})$.

5.2.2 Equidistance Constraint

To obtain k_i , we introduce the *Equidistance Constraint*. The *Equidistance Constraint* states that the distance from reference point \mathbf{p}_i to the center of cornea sphere S is equal to the distance from the center of the camera, \mathbf{O} , to \mathbf{S} .

If this equidistance constraint is satisfied, triangle $\triangle \mathbf{OSp}_i$ is an isosceles triangle that satisfies $|\mathbf{O} - \mathbf{S}| = |\mathbf{p}_i - \mathbf{S}|$. Let the unit vector representing the bisector of $\angle \mathbf{OSp}_i$ be denoted as \mathbf{l}_i and a point on this bisector be denoted as \mathbf{a}_i . Triangle $\triangle \mathbf{Oa}_i \mathbf{p}_i$ is also an isosceles triangle that satisfies $|\mathbf{O} - \mathbf{a}_i| = |\mathbf{p}_i - \mathbf{a}_i|$. Additionally, among the point set \mathbf{O} , \mathbf{a}_i and \mathbf{p}_i , the laws of reflection can be established at \mathbf{a}_i where \mathbf{l}_i is used as the normal vector. Therefore, when \mathbf{a}_i is the intersection of the bisector and the surface of cornea sphere, \mathbf{a}_i is equal to the \mathbf{m}_i of reference point \mathbf{p}_i .

Therefore, when user sets his/her center of cornea sphere such that the equidistance constraint does hold, triangle $\triangle \mathbf{Om}_i \mathbf{p}_i$ should be an isosceles triangle that satisfies $|\mathbf{p}_i - \mathbf{m}_i| = |\mathbf{O} - \mathbf{m}_i|$, that is $k_i = k'_i$.

From the above, we compute 3D positions of reference point \mathbf{p}_i from Eq (5.2) by introducing two constraints, and obtain extrinsic parameters R and \mathbf{t} by solving the Absolute Orientation Problem (See Algorithm 3). This algorithm works with the minimal configuration where three reference points and one mirror pose, however, note that there is a limitation on the mirror position that the center of the mirror, *i.e.* the center of the human eye, should be located at the same distance from the camera center and each reference point.

5.3. Approach 2:
Direct Derivation Approach

Algorithm 3 Algorithm of approach 1.

Require: Image I , $\mathbf{q}_i(i = 1, 2, 3)$, $\mathbf{p}_i^{\{X\}}$, K

Ensure: R, t

Compute ellipse parameters $(\mathbf{i}_L, \phi, r_{max}, r_{min})$ from projection of limbus in image I .

Compute the center of corneal sphere \mathbf{S} from ellipse parameters $(\mathbf{i}_L, \phi, r_{max}, r_{min})$.

for all $\mathbf{q}_i(i = 1, 2, 3)$ **do**

 Compute an unit vector \mathbf{v}_i from Eq (5.7).

 Compute a reflection point \mathbf{m}_i of a reference point \mathbf{p}_i from Eq (5.6)

 Compute a normal vector \mathbf{n}_i at reflection point \mathbf{m}_i from Eq (5.5).

 Compute an unit vector \mathbf{u}_i from Eq (5.4).

 Compute 3D position of reference point \mathbf{p}_i from Eq (5.2).

end for

Solve Absolute Orientation Problem with $\mathbf{p}_i^{\{X\}}, \mathbf{p}_i$ and obtain extrinsic parameters R and t .

5.3 Approach 2: Direct Derivation Approach

Since only $\mathbf{p}_i^{\{X\}}$ is known in Eq.(5.3), we cannot solve Eq.(5.3) and obtain extrinsic parameters by simply increasing the number of reference points. In order to reduce the unknown parameters, we introduce a geometric model of the cornea sphere as described in Section 5.2 and a basis vector expression to represent 3D reference point position. We also propose an extrinsic calibration method with a simple configuration, *i.e.* five reference points and one mirror pose, without a limitation on the mirror position as introduced in Section 5.2.

5.3.1 Using Basis Vector Representation of 3D Reference Point Position

In this paper, basis vector representation means representing vector \mathbf{p} as the linear combination of basis vectors, that is, $\mathbf{p} = \sum_{j=0}^{N_e-1} a_j \mathbf{e}_j$, where $\mathbf{e}_j(j = 0 \dots, N_e - 1)$ denotes the basis vector of N_e dimensional vector space and is independent linearly, and a_j is the coordinate of \mathbf{p} with respect to the basis \mathbf{e}_j . Here, we assume a three dimensional vector space, that is, $N_e = 3$. With this basis vector representation, \mathbf{p}_i in the reference object coordinate system

$\{X\}$ is expressed as,

$$\mathbf{p}_i^{\{X\}} = \sum_{j=0}^2 a_j^{i\{X\}} \mathbf{e}_j^{\{X\}}, \quad (5.16)$$

where $a_j^{i\{X\}}$ denotes the coordinates of $\mathbf{p}^{\{X\}}$ with respect to basis $\mathbf{e}_j^{\{X\}}$. By assuming $\mathbf{p}_j^{\{X\}}$ and $\mathbf{e}_j^{\{X\}}$ are given a priori, $a_j^{i\{X\}}$ can be computed. By substituting Eq.(5.16) into Eq.(5.1), we have

$$\mathbf{p}_i = \sum_{j=0}^2 a_j^{i\{X\}} R \mathbf{e}_j^{\{X\}} + \mathbf{t}. \quad (5.17)$$

In cases where $\mathbf{p}_0^{\{X\}}$ represents the origin of the reference object coordinate system, \mathbf{p}_0 can be considered as translation vector \mathbf{t} . Therefore \mathbf{p}_i can be expressed as

$$\mathbf{p}_i = \sum_{j=0}^2 a_j^{i\{X\}} R \mathbf{e}_j^{\{X\}} + \mathbf{p}_0. \quad (5.18)$$

5.3.2 Derivation of Linear Equation for Estimating Extrinsic Parameters

By substituting Eq.(5.18) into Eq.(5.3) and representing \mathbf{p}_0 by using Eq.(5.2), we have

$$\sum_{j=0}^2 a_j^{i\{X\}} R \mathbf{e}_j^{\{X\}} + k_0 \mathbf{u}_0 + \mathbf{m}_0 = k_i \mathbf{u}_i + \mathbf{m}_i. \quad (5.19)$$

We define each basis vector as $\mathbf{e}_0^{\{X\}} = [1, 0, 0]^\top$, $\mathbf{e}_1^{\{X\}} = [0, 1, 0]^\top$, $\mathbf{e}_2^{\{X\}} = [0, 0, 1]^\top$. From Eq.(5.19) for the N_p reference points, we can derive the following linear equation:

$$\mathbf{A}\mathbf{X} = \mathbf{B}, \quad (5.20)$$

where

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_0^1 & \mathbf{A}_1^1 & \mathbf{A}_2^1 & \mathbf{u}_0 & \mathbf{W}_1 \\ \mathbf{A}_0^2 & \mathbf{A}_1^2 & \mathbf{A}_2^2 & \mathbf{u}_0 & \mathbf{W}_2 \\ & & \vdots & & \\ \mathbf{A}_0^{N_p-1} & \mathbf{A}_1^{N_p-1} & \mathbf{A}_2^{N_p-1} & \mathbf{u}_0 & \mathbf{W}_{N_p-1} \end{bmatrix}, \quad (5.21)$$

$$\mathbf{A}_j^i = a_j^{i\{X\}} \mathbf{I}_{3 \times 3}, \quad (5.22)$$

$$\mathbf{W}_k = [\mathbf{w}_1^k \ \mathbf{w}_2^k \ \cdots \ \mathbf{w}_{N_p-1}^k], \quad (5.23)$$

$$\mathbf{w}_m^l = \begin{cases} -\mathbf{u}_l & (l = m) \\ \mathbf{0}_{3 \times 1} & (otherwise) \end{cases}, \quad (5.24)$$

$$\mathbf{X} = [\mathbf{r}_0^\top \ \mathbf{r}_1^\top \ \mathbf{r}_2^\top \ k_0 \ k_1 \ \cdots \ k_{N_p-1}]^\top, \quad (5.25)$$

$$\mathbf{B} = [\mathbf{m}'_1 \ \mathbf{m}'_2 \ \cdots \ \mathbf{m}'_{N_p-1}]^\top, \quad (5.26)$$

$$\mathbf{m}'_i = (\mathbf{m}_i - \mathbf{m}_0)^\top. \quad (5.27)$$

Vectors \mathbf{r}_0 , \mathbf{r}_1 and \mathbf{r}_2 denote the first, second and third columns of the rotation matrix $R = [\mathbf{r}_0 \ \mathbf{r}_1 \ \mathbf{r}_2]$.

We assume that we use a planar display as the reference object, that is, the reference points lie on the same plane. In this case, a reference point in the reference object coordinate system can be expressed as $p_i^{\{X\}} = (x_i, y_i, 0)$ and all a_2^i are zero. By removing \mathbf{r}_2 , which is the unknown parameter corresponding to a_2^i in Eq.(5.20), we have the following linear equation:

$$\mathbf{A}'\mathbf{X}' = \mathbf{B}, \quad (5.28)$$

where

$$\mathbf{A}' = \begin{bmatrix} \mathbf{A}_0^1 & \mathbf{A}_1^1 & \mathbf{u}_0 & \mathbf{W}_1 \\ \mathbf{A}_0^2 & \mathbf{A}_1^2 & \mathbf{u}_0 & \mathbf{W}_2 \\ \vdots & & & \\ \mathbf{A}_0^{N_p-1} & \mathbf{A}_1^{N_p-1} & \mathbf{u}_0 & \mathbf{W}_{N_p-1} \end{bmatrix}, \quad (5.29)$$

$$\mathbf{X}' = [\mathbf{r}_0^\top \ \mathbf{r}_1^\top \ k_0 \ k_1 \ \cdots \ k_{N_p-1}]^\top. \quad (5.30)$$

With N_p reference points, we have $(6 + N_p)$ unknowns (\mathbf{X}') and $3(N_p - 1)$ constraints (rows of \mathbf{A}' and \mathbf{B}) in Eq.(5.28). Hence, when $N_p \geq 5$, we can solve Eq.(5.28) by $\mathbf{X}' = \mathbf{A}'^*\mathbf{B}$, where \mathbf{A}'^* is the pseudo-inverse matrix of \mathbf{A}' . \mathbf{r}_2 is given by the cross product of \mathbf{r}_0 and \mathbf{r}_1 , i.e. $\mathbf{r}_2 = \mathbf{r}_0 \times \mathbf{r}_1$.

In real environments, the rotation matrix $R = [\mathbf{r}_0 \mathbf{r}_1 \mathbf{r}_2]$ obtained by solving Eq.(5.28) is not guaranteed to satisfy the constraints to form a valid rotation matrix ($|\mathbf{r}_0| = |\mathbf{r}_1| = |\mathbf{r}_2| = 1, \mathbf{r}_0^\top \mathbf{r}_1 = \mathbf{r}_1^\top \mathbf{r}_2 = \mathbf{r}_2^\top \mathbf{r}_0 = 0$) due to the observation noise. In order to enforce these constraints, here we solve the orthogonal Procrustes problem [GVL96] as done by Zhang's method [Zha00].

This linear solution estimates the correct extrinsic parameters in noiseless environments. However, extrinsic parameter precision degrades substantially if the input data includes observation noise (We describe the experimental environment in detail in Section 5.5). To overcome this difficulty, we solve the non-linear optimization problem of the objective function derived from Eq.(5.28), which is robust to noise.

5.3.3 Solving Non-linear Optimization Problem

Objective Function

We define an objective function for non-linear optimization with two error terms. First, we introduce an error term for the measurement model. Ideal

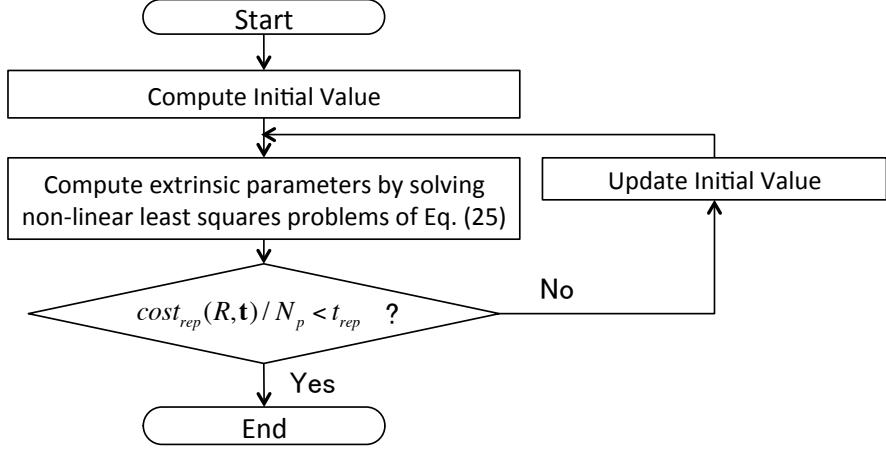


Figure 5.5: Implementation strategy.

extrinsic parameters should satisfy the linear equation of Eq.(5.28), which is derived from the measurement model. In order to enforce this constraint on the estimated extrinsic parameters, we introduce the following error term

$$cost_{model}(R, \mathbf{t}) = \|\mathbf{A}'\mathbf{X}'(R, \mathbf{t}) - \mathbf{B}\|, \quad (5.31)$$

where $\mathbf{X}'(R, \mathbf{t})$ denotes \mathbf{X}' computed from the estimated R and \mathbf{t} .

Second, we introduce an error term to minimize the reprojection error as is widely done in the calibration [TMHF00]:

$$cost_{rep}(R, \mathbf{t}) = \sum_{i=0}^{N_p-1} \|\mathbf{q}_i - \check{\mathbf{q}}_i(R, \mathbf{t})\|, \quad (5.32)$$

where $\check{\mathbf{q}}_i(R, \mathbf{t})$ denotes \mathbf{q}_i calculated from the estimated R and \mathbf{t} .

By introducing these error terms, we define the following objective function f

$$\begin{aligned} f &= c_{model} * cost_{model}(R, \mathbf{t}) \\ &+ c_{rep} * cost_{rep}(R, \mathbf{t}), \end{aligned} \quad (5.33)$$

where c_{model} and c_{rep} are respectively the coefficients corresponding to $cost_{model}$ and $cost_{rep}$.

Implementation

We implemented our proposed method together with non-linear optimization as illustrated in Figure 5.5. First, we estimated the initial values of extrinsic parameters. In doing so, we used a linear solution of extrinsic parameters

5.4. Performance Evaluations of Approach 1

estimated by solving Eq.(5.28) as the initial value. Second, we used the Levenberg-Marquardt algorithm to solve the non-linear optimization problem of Eq.(5.33). However, Eq.(5.33) is not a convex function and can converge to a local minimum. To address this problem, we used the reprojection error as the criteria indicating whether or not the estimated solution is a local minimum. When the average reprojection error $cost_{rep}(R, \mathbf{t})/N_p$ is larger than a threshold for reprojection error t_{rep} , that is, the estimated solution was a local minimum, we updated the initial value of the extrinsic parameters by adding random values and solved the non-linear optimization problem until $cost_{rep}(R, \mathbf{t})/N_p < t_{rep}$ was satisfied. Note that in the experiments described in the next section we added zero-mean Gaussian noise to the rotation vector, which is the vector representation of the rotation matrix, and the translation vector with standard deviation $\sigma_R = 0.1$ and $\sigma_t = 10$ as random values at each iteration.

5.4 Performance Evaluations of Approach 1

This section provides evaluations of the performance of our proposed algorithms with synthesized and real data.

5.4.1 Quantitative Evaluations with Synthesized Data Experimental Environment

In order to synthesize the data, we use the following configurations. The matrix of intrinsic parameters, K , consists of (fx, fy, cx, cy) ; fx and fy represent the focal length in pixels, and cx and cy represent the 2D coordinates of the principal point. We set them to $(1400, 1400, 960, 540)$ in this evaluation respectively.

We set the camera coordinate system as the world coordinate system and set the center of camera to $\mathbf{O} = (0, 0, 0)$. The 3D positions of the reference point are defined as $\mathbf{p}_1 = (0, -250, 0)$, $\mathbf{p}_2 = (125, -125, 0)$, $\mathbf{p}_3 = (-125, -125, 0)$. The center of the cornea sphere is set to $\mathbf{S} = (0, -125, 50)$, which satisfies the equidistance constraint, and its radius is set to 5.6 mm on the basis of [RS97]. The distances from \mathbf{S} to \mathbf{O} and to each reference point, $\mathbf{p}_i (i = 1, 2, 3)$, are $\sqrt{125^2 + 50^2}$. We added the Gaussian noise with zero mean and standard deviation $\sigma_p (0 \leq \sigma_p \leq 1)$ to projections \mathbf{q}_i as observation noise. The ground truth of rotation matrix R and translation vector are set to $R = \mathbf{I}$ and $\mathbf{t} = (0, -250, 0)^\top$, respectively. Throughout this experiment, we utilized the same error functions introduced in Section 4.4 for evaluating

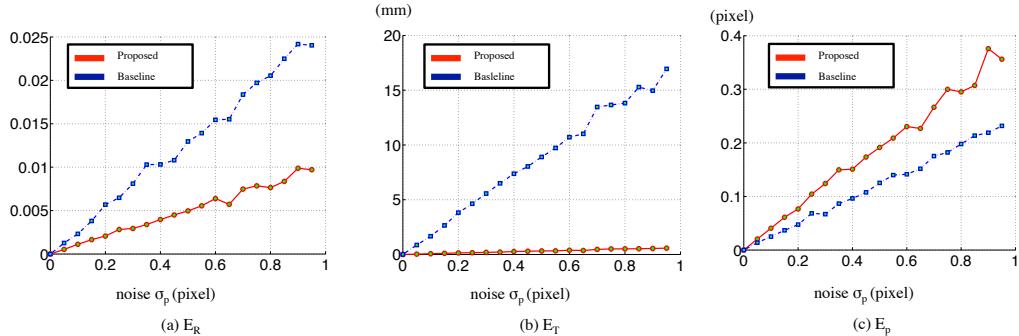


Figure 5.6: Estimation error under Gaussian noise for \mathbf{q}_i with standard deviation σ_p .

extrinsic parameters, that is the Riemannian distance [Moa02] E_R , the root mean square error E_t and the reprojection error E_p .

In this simulation, we compared our method against the state-of-the-art of planar mirror based method proposed in Chapter 4 as a *baseline* method. For fair comparison, we made sure that the projections of reference points using either spherical or planar mirrors occupy a similar pixel area in the image in the same way as in [Agr13].

Results

Figure 5.6 shows the evaluation results of R , \mathbf{t} and reprojection error, from left to right. In each figure, the vertical axis shows the average value over 100 trials and the horizontal axis denotes standard deviation of noise. Our estimation errors, E_R and E_t , are significantly smaller than those of the baseline (59% and 96%, respectively) and these results quantitatively prove that our method outperforms the baseline. However, our reprojection error almost matches that of the baseline. We explain this by noting that the baseline employs non-linear refinement for estimating the extrinsic parameters, which optimizes the reprojection error, while our method doesn't. For further analysis and discussion of this optimization see Section 5.6.3.

5.4.2 Real Data

Experimental Environment

Figure 5.7 overviews the configuration. We used a Logicoal HD Pro Webcam C920t and captured frames had the resolution of 1920×1080 . As illustrated in Figure 5.7, we used a chessboard pattern on display as the reference points

5.4. Performance Evaluations of Approach 1

Table 5.1: Comparison of each function and reprojection error against the baseline.

	Person1			Person2		
	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3
E_R	0.1278	0.1907	0.2117	0.4116	0.3240	0.2674
E_t (mm)	27.4989	28.5416	24.4317	44.7782	37.6088	35.3061
E_p (pixel)	0.4118	0.5230	0.3561	0.3239	1.0900	0.90069

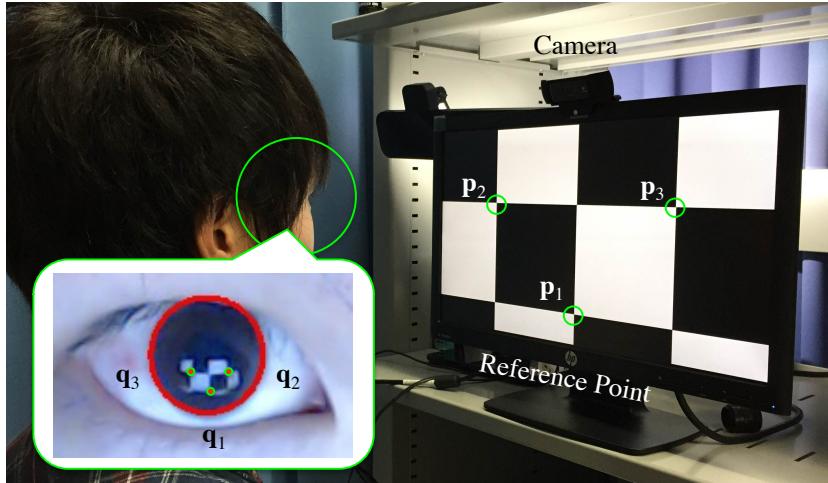


Figure 5.7: Configuration for experiments with real data. The equidistance constraint is satisfied in a casual configuration as discussed in Section 5.6.2. Notice that we used only three points $p_i(i = 1, 2, 3)$ of chessboard pattern as the reference points for calibration.

$p_i(i = 1, 2, 3)$. The length of one side of this chessboard pattern is 125 mm. In order to satisfy our newly proposed *equidistance constraint*, we set the camera, reference points and user's cornea center as in Figure 5.7. Whether this constraint is satisfied or not was verified by visual judgment. The distance between the user's cornea center and the display is about 300 mm. The intrinsic parameter was estimated beforehand by [Zha00].

In order to estimate ellipse parameters $(\mathbf{i}_L, \phi, r_{max}, r_{min})$ from limbus projection, we binarize the input image, apply the Canny detector, and fit an ellipse [FF⁺96] as shown in Figure 5.8. Additionally, we confirmed that we can detect the boundaries of the projected limbus and estimate ellipse parameters under various illumination environments expected where the proposed method will be used. The environments had illumination intensities of 25 lux (dark, only display light used), 600 lux (middle, under distant fluores-

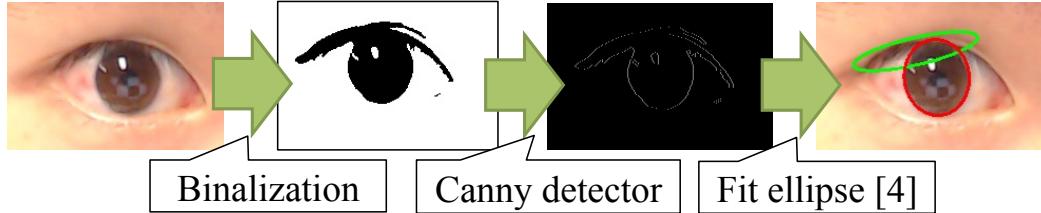


Figure 5.8: A flow of estimating ellipse parameters ($i_L, \phi, r_{max}, r_{min}$) from projection of limbus.

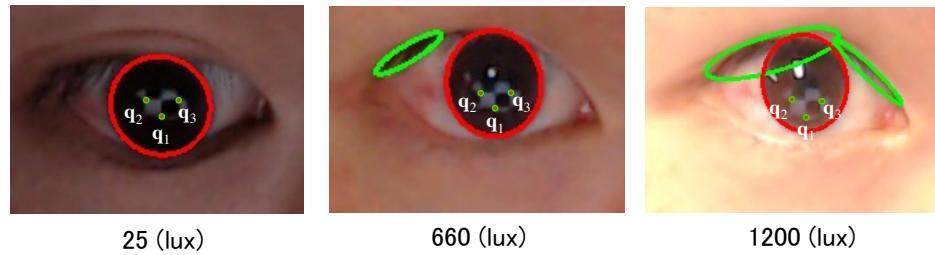


Figure 5.9: Images of cornea reflection in three illumination environments. The distances between each q_i are about $13 \sim 20$ pixels.

cent light) and 1200 lux (bright, near to fluorescent light). We show images of cornea reflection captured in these environments in Figure 5.9.

Results

Figure 5.10 (a) illustrates the geometric relation of camera and reference points. The camera is set at the top of display which has three reference points. Figure 5.10 (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 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 5.10.

Table 5.1 quantitatively compares the parameters estimated by the proposed method and baseline. We evaluated three trials for each of two users. From these results, we confirmed that our method works properly in real environments since the differences in the extrinsic parameters estimated by each method are small. While this precision is not enough for eye gaze tracking, it is acceptable for applications that do not need high precision, such as

5.5. Performance Evaluations of Approach 2

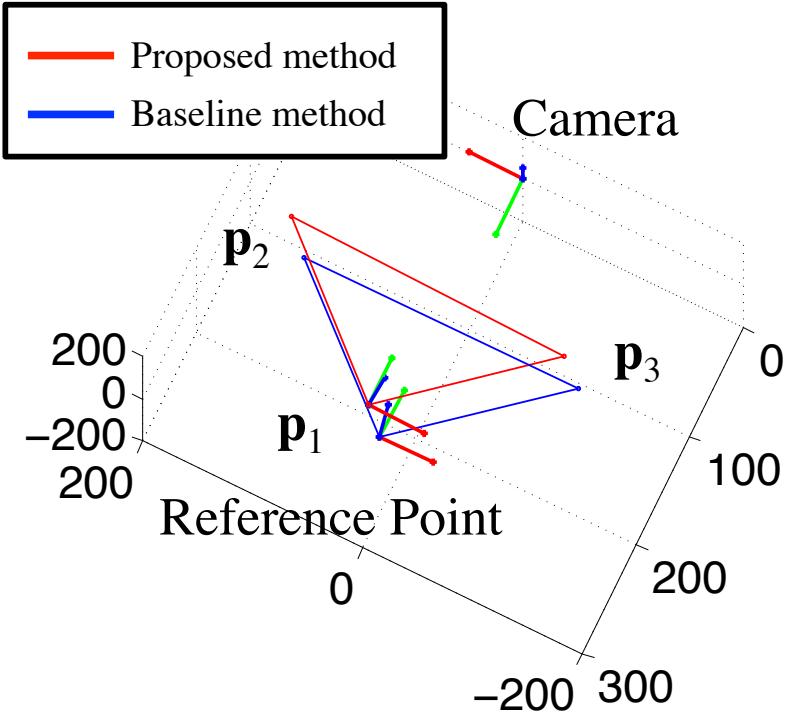


Figure 5.10: (a):View of configuration. (b):Estimated positions of reference points by proposed method (red), by baseline method (blue).

gaze correction [KPB⁺12] using a display and attached web camera system.

Additionally, these results indicate that the estimation results estimated by the proposed method depend on the user. We investigate this dependency in the Section 5.6.

5.5 Performance Evaluations of Approach 2

This section reports the performance of the approach 2 of our proposed method. In this section, “linear solution” denotes extrinsic parameters estimated by solving Eq.(5.28) in a linear manner and “non-linear solution” denotes those estimated by solving the non-linear optimization problem of Eq.(5.33).

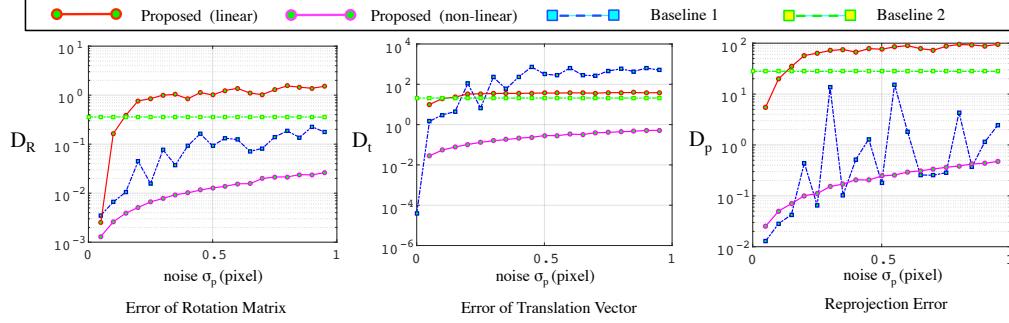


Figure 5.11: Estimation errors of linear solution under Gaussian noise for \mathbf{q}_i with standard deviation σ_p .

5.5.1 Quantitative Evaluations with Synthesized Data Experimental Environment

The synthesized data was generated as follows. The matrix of the intrinsic parameters, K , consists of (fx, fy, cx, cy) ; fx and fy represent the focal length in pixels, and cx and cy represent the 2D coordinates of the principal point. For the experiments we set them respectively to $(1400, 1400, 960, 540)$.

We set the camera coordinate system as the world coordinate system and set the center of the camera to $\mathbf{O} = (0, 0, 0)$. The 3D positions of the reference points were defined as $\mathbf{p}_0^{\{X\}} = (0, 0, 0)$, $\mathbf{p}_1^{\{X\}} = (-50, 0, 0)$, $\mathbf{p}_2^{\{X\}} = (50, 0, 0)$, $\mathbf{p}_3^{\{X\}} = (0, -50, 0)$ and $\mathbf{p}_4^{\{X\}} = (50, -50, 0)$. On the basis of the work reported by Snell [RS97], we set the center of the cornea sphere to $\mathbf{S} = (5, 0, 50)$, the d_{LS} to 5.6 mm, and radius r to 7.7 mm.

Throughout this experiment, we utilized the same error functions introduced in Section 4.4 for evaluating extrinsic parameters, that is the Riemannian distance [Moa02] E_R , the root mean square error E_t and the reprojection error E_p .

Results

In these simulation experiments, we evaluated the performance of the proposed method under observation noise. We added zero-mean Gaussian noise whose standard deviation was $\sigma_p (0 \leq \sigma_p \leq 1)$. We compared linear and non-linear solutions estimated by the proposed method with solutions estimated by planar mirror-based method introduced in Chapter 4 (Baseline 1) and an approach 1 of proposed method introduced in Section 5.2 (Baseline 2). To ensure fairness in the comparison, we made sure the projections of the reference points, using either spherical or planar mirrors, occupied a comparable

5.5. Performance Evaluations of Approach 2

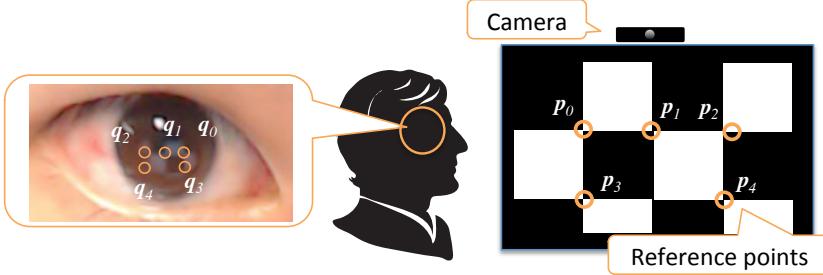


Figure 5.12: Configuration for experiments with real data. Notice that we use only five points $\mathbf{p}_i(i = 0, \dots, 4)$ of the chessboard pattern as the reference points for calibration. Each \mathbf{q}_i is separated by about $10 \sim 13$ pixels in the captured image.

pixel area in the image as Agrawal did [Agr13]. Note that we utilized three points \mathbf{p}_1 , \mathbf{p}_2 and \mathbf{p}_3 in calibrating by Baseline 2. They are the minimal configuration for Baseline 2 but do not necessarily satisfy the equidistance constraint in this evaluation.

Figure 5.11 shows E_R , E_t and E_p of solutions estimated by each method. In each figure, the vertical axis shows the average value over 100 trials and the horizontal axis denotes the standard deviation of noise. From the figure, we can observe that E_R , E_t and E_p are almost zero in noiseless environments, *i.e.* at $\sigma_p = 0$. This means that our method works properly with the configuration of five reference points and one mirror pose, in noiseless environments. In the $\sigma_p > 0$ case, while the precision of the linear solution degrades substantially, we can observe that the precision of the non-linear solution are competitive with the Baseline 1. With regard to Baseline 2, the effects of noise for estimation precision seem to be smaller than those for the other methods. We consider this is because the effects produced by not satisfying the equidistance constraint absorb those of observation noise.

These results show that our method works robustly even if the projections \mathbf{q}_i include observation noise and even if the equidistance constraint required in Baseline 2 does not hold.

5.5.2 Real Data

Configuration

Figure 5.12 overviews the configuration. We used a Logicool HD Pro Webcam C920t and captured frames had the resolution of 1920×1080 . As illustrated in Figure 5.12, we projected a chessboard pattern on the display and captured the cornea as the reference points $\mathbf{p}_i(i = 0, \dots, 4)$. The size

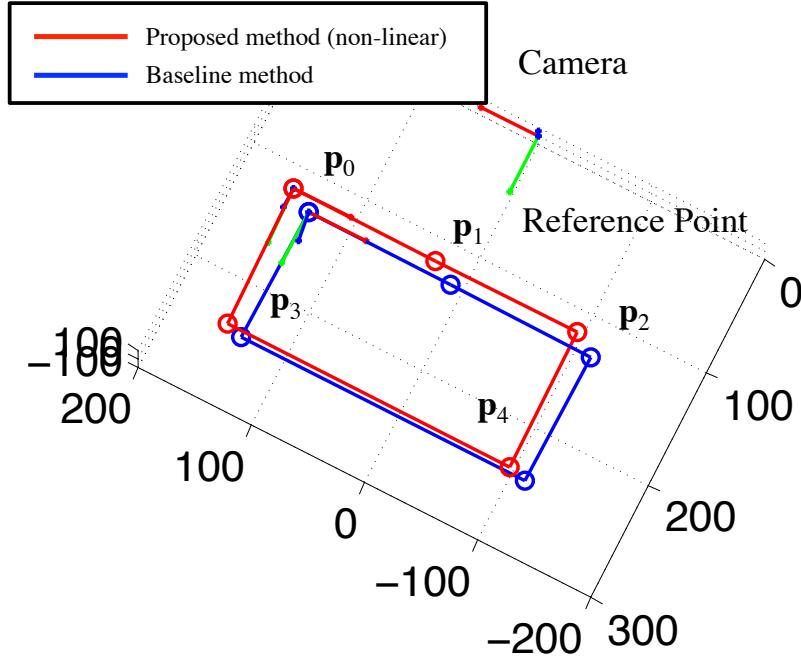


Figure 5.13: Reference point positions estimated by proposed method (red), compared to those estimated by Baseline 1 (blue).

Table 5.2: Error metrics computed by using Baseline 1 as the ground truth.

	E_R	E_t	E_p
Linear solution	0.553	178.896	14.689
Non linear solution	0.164	33.617	0.260

of each chess block was 125×125 mm. The distance between the user's cornea center and the display was about 300 mm. The intrinsic parameter was estimated beforehand by Zhang[Zha00]. We estimated the ellipse parameters $(\mathbf{i}_L, \phi, r_{max}, r_{min})$ from limbus projection as illustrated in 5.8. Since the ground truth of the extrinsic parameters is not available in any real configuration, we used extrinsic parameters estimated by Baseline 1 method as the reference parameters.

Results

Table 5.2 quantitatively compares the parameters estimated by the proposed method (linear solution and non-linear solution) with those estimated by Baseline 1. We can see that the distance functions yielded by the linear solution output large differences due to some observation noise.

5.6. Discussion

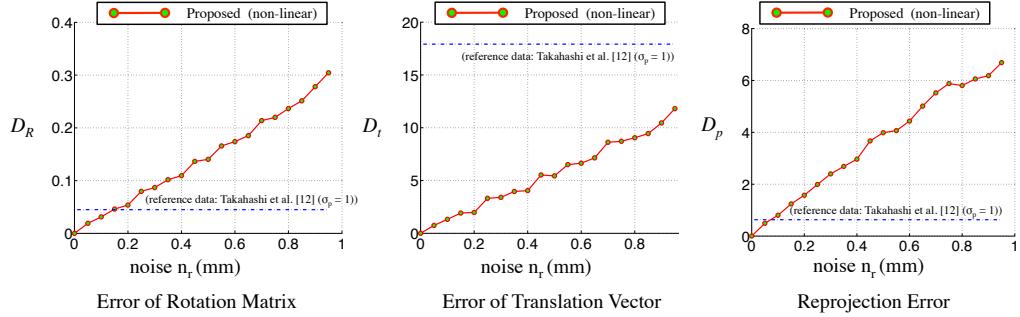


Figure 5.14: Estimation errors under random noise for cornea sphere radius r with uniform distribution.

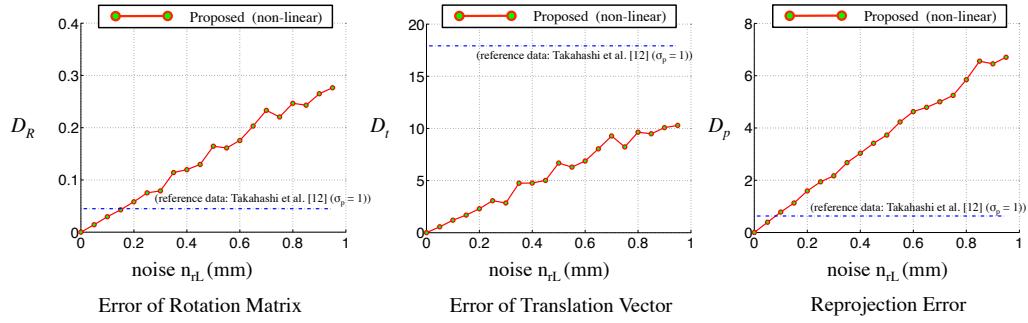


Figure 5.15: Estimation errors under random noise for cornea limbus radius r_L with uniform distribution.

On the other hand, the non-linear solution yields small differences. This point can be verified by visualizing the results as shown in Figure 5.13. It shows that the positions estimated by the proposed method are almost identical to those estimated by Baseline 1. This confirms that our method works properly in real environments. While this precision may not be enough for eye gaze tracking, it is acceptable for applications that do not need high precision, such as gaze correction [KPB⁺12] using a display and attached web camera system.

5.6 Discussion

In this section, we discuss the effects of noise on the constraints of our proposed method, and the method for optimization of extrinsic parameters.

5.6.1 Effects of Differences among Individuals

Both proposed approach 1 and approach 2 make two assumptions about the cornea model. The first assumption is the radius of cornea sphere r . While we use the average radius of the cornea sphere, that is $r = 7.7\text{mm}$ [RS97], it can vary with the individual. The second one is the radius of cornea limbus r_L . In this paper we use the average size $r_L = 5.6\text{mm}$, but in practice the model parameters can be tailored to suit the individual. To more closely examine the effects of these assumptions, we created synthesized data and investigated the effects of noise on these two radii. We used the same configuration as in Section 5.4.1. We added random noise with uniform distribution n_r and n_{r_L} to r and r_L , respectively, ($0 \leq |n_r| \leq 2, 0 \leq |n_{r_L}| \leq 1$) based on [RS97] and [IO12].

Figure 5.14 and Figure 5.15 show the results of the averages of each distance function and reprojection error. From Figure 5.14 and Figure 5.15, we can see that r and r_L strongly impact the estimation error of extrinsic parameters and reprojection error. This because adding noise to r and r_L affects the precision of \mathbf{S} estimates based on Eq (5.15) and $d_{LS} = \sqrt{r^2 - r_L^2}$, and the direction and location of the reflection on the cornea sphere changes significantly depending on \mathbf{S} and r . As to the results in Section 4.2, we can say that these variations do impact estimation performance. To solve this problem, it is useful to calibrate the user's eye parameters beforehand.

5.6.2 Equidistance constraint in a real situation

Our proposed method need to satisfy the equidistance constraint. Although, it is difficult to satisfy the proposed equidistance constraint strictly in a real situation. However, this constraint is not such a strong assumption for calibration because it can easily be satisfied roughly even in a real situation. For example, as illustrated in Fig 5.16, if the reference points \mathbf{p}_i and the camera center are on the same circle \mathbf{O} , and the sphere center \mathbf{S} is on a vector \mathbf{l}_{equi} that is orthogonal to the circle plane and has its footpoint on the center of the circle, the equidistance constraint is satisfied. If one assumes a display-camera system, such as a smartphone or a web-camera attached on the display (Fig 5.12), the reference points and camera center are on almost the same plane and locations of the reference points can be controlled by taking advantage of using "Display" as the reference object. Moreover, we can easily set our eyes on \mathbf{l}_{equi} by displaying the center of the circle as a guide.

We evaluate how well non-experts could satisfy this constraint. We set the camera and reference points in the above manner. Since it is difficult

5.6. Discussion

Table 5.3: The standard deviation of distances from the object to each reference point and the camera σ_{equi} , the average of distances D_{equi} and the mean of placement error e_{equi} for 10 non-experts.

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
σ_{equi} (mm)	8.49	5.45	6.54	3.27	2.96	4.06	3.03	2.18	0.71	3.84
D_{equi} (mm)	184.0	361.3	409.5	377.3	364.5	391.8	358.3	430.5	403.0	421.5
e_{equi} (mm)	8.00	4.38	5.50	2.75	2.50	3.38	2.75	1.75	0.50	3.50

to measure the distances from the center of subject’s cornea sphere to each reference point and the camera by a ruler, we asked them to set an object on \mathbf{l}_{equi} in 3D space with a tripod where it satisfied the equidistance constraint and measured the distances from the object to each reference point and the camera by a ruler. Notice that the ability of placing one’s eye so that it satisfies the equidistance constraint may not be same as the ability of placing object to satisfy the constraint. However we think that if they can set an object, such as their finger, so that it satisfies the constraint, they can achieve to set their eye to satisfy the constraint by using the object as a guide.

Table 5.3 shows the standard deviation of distances from the object to each reference point and the camera σ_{equi} , the average of distances D_{equi} and the mean of placement error e_{equi} for 10 subjects. Notice that $D_{equi} = (\sum_{i=1}^4 |\mathbf{p}_o - \mathbf{p}_i|)/4$, where \mathbf{p}_o represents the 3D position the object and \mathbf{p}_i represents the 3D position of each reference point and camera, and $e_{equi} = (\sum_{i=1}^4 ||\mathbf{p}_o - \mathbf{p}_i| - D_{equi}|)/4$. From Table 3, we can see that non-experts can satisfy the equidistance constraint with $\sigma_{equi} < 10$ mm.

Additionally, we investigate the effects of noise on the equidistance constraint quantitatively. We evaluate the performance of the proposed method by generating synthesized data that includes Gaussian noise with zero mean and standard deviation σ_m ($0 \leq \sigma_m \leq 10$) to \mathbf{p}_i ($i = 1, 2, 3$). Figure 5.17 shows the results of the averages of each distance function and reprojection error. From these results, the equidistance constraint does not have any severely negative effects on extrinsic parameter estimation, even when it is not strictly satisfied. For example, the estimation error of the translation vector is 3 mm at most. We consider that such precision is sufficient for display-camera system applications.

5.6.3 Optimization of Approach 1

In general, estimated extrinsic parameters can be refined by non-linear optimization[TMHF00]. Here we discuss the non-linear optimization for extrinsic parameters esti-

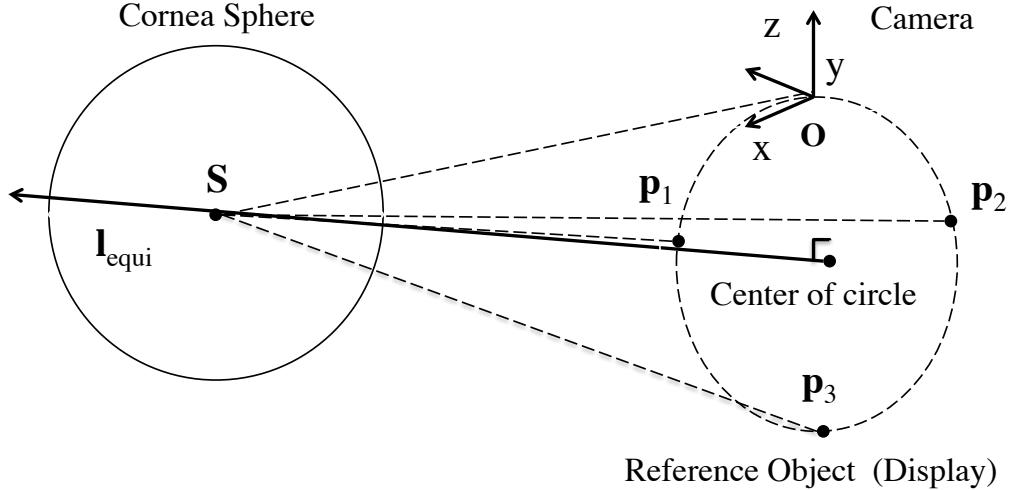


Figure 5.16: A case example where the equidistance constraint holds.

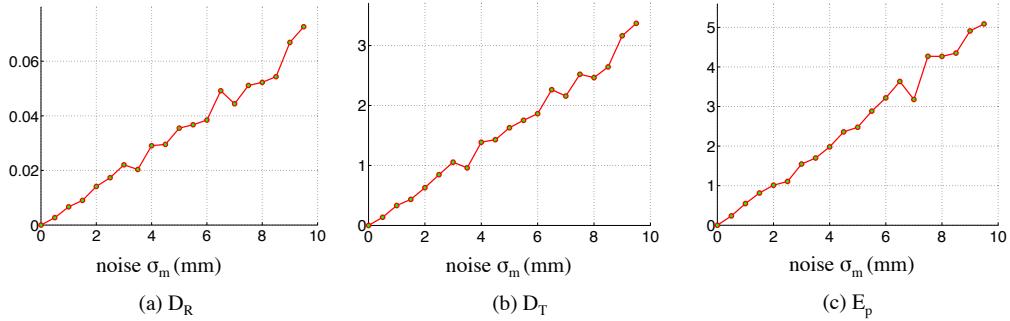


Figure 5.17: Estimation error under Gaussian noise for reference points $p_i (i = 1, 2, 3)$ with standard deviation σ_m .

mated by Approach 1.

In order to optimize the estimated extrinsic parameters R and t , we minimize $E(R, t, S)$ which has reprojection error term and equidistance constraint term as follows,

$$E(R, t, S) = \sum_{i=1}^3 \{|\mathbf{q}_i - \check{\mathbf{q}}_i(R, t, S)| + D_i(R, t, S)\}. \quad (5.34)$$

5.6. Discussion

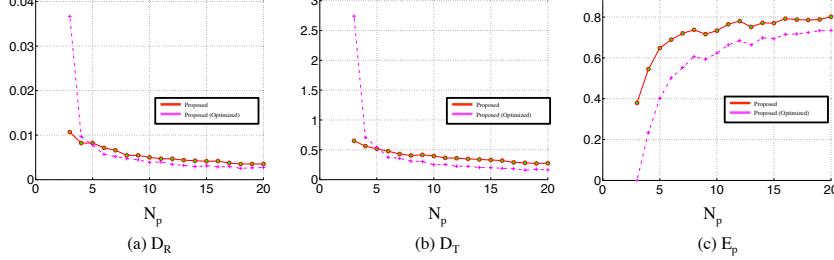


Figure 5.18: Estimation error of optimized extrinsic parameters using N_p reference points under Gaussian noise for \mathbf{q}_i with standard deviation $\sigma_p = 1$.

The term of $D_i(R, \mathbf{t}, \mathbf{S})$ is expressed as follows,

$$\begin{aligned} D_i(R, \mathbf{t}, \mathbf{S}) &= ||\check{\mathbf{p}}_i(R, \mathbf{t}) - \mathbf{S}| - |\mathbf{O} - \mathbf{S}|| \\ &\quad + ||\check{\mathbf{p}}_i(R, \mathbf{t}) - \mathbf{m}_i(R, \mathbf{t}, \mathbf{S})| \\ &\quad - |\mathbf{O} - \mathbf{m}_i(R, \mathbf{t}, \mathbf{S})||, \end{aligned} \quad (5.35)$$

where $\check{\mathbf{p}}_i(R, \mathbf{t})$ and $\mathbf{m}_i(R, \mathbf{t}, \mathbf{S})$ denote the 3D position of reference point \mathbf{p}_i and the reflection point \mathbf{m}_i computed with estimated parameters respectively. We evaluated the performance of this optimization using synthesized data that includes Gaussian noise with standard deviation $\sigma_p = 1$ as in Section 5.4.1.

Figure 5.18 plots the optimization performance under various numbers of reference points. While the distance functions E_R and E_t decrease after non-linear optimization if more than five points are used, they increase with the minimal configuration. This is considered to be due to the lack of redundancy in our method. Future work includes optimization of extrinsic parameters for the minimal configuration.

5.6.4 Validity of Using Reprojection Error as Criteria for Detecting Local Minimum

In Section 5.3.3, we used the reprojection error as the metric indicating whether or not the estimated solution is a local minimum. Here we address the validity of this usage by referring to simulation data. In the simulation experiments, we investigated the rate at which we can match the ground truth in cases where the reprojection error is smaller than t_{rep} ($1 \leq t_{rep} \leq 10$). Note that we regarded the estimated R and \mathbf{t} as matching the ground truth if $E_R < t_{E_R}$ and $E_t < t_{E_t}$, which we respectively set to $t_{E_R} = 0.02$ and $t_{E_t} = 6$ on the basis of the results in Chapter 4 with $\sigma_p = 0.5$. We used the same

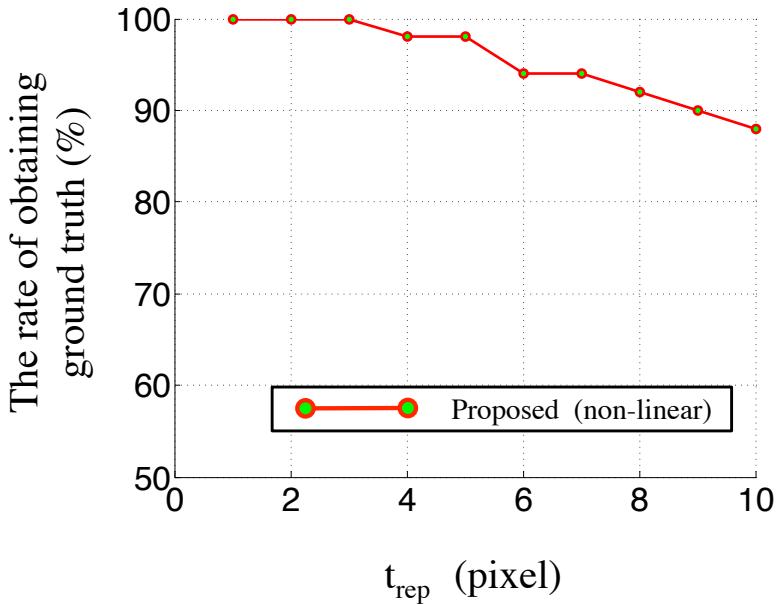


Figure 5.19: Rate of matching ground truth for each t_{rep} .

configuration as we did in Section 5.5.1, while adding Gaussian noise with zero mean and standard deviation $\sigma_p = 0.5$ to \mathbf{q}_i .

Figure 5.19 shows the rate of matching the ground truth for each t_{rep} over 50 trials. From the figure, we can observe that all the estimated solutions converge to the ground truth when $t_{rep} \leq 3$. These simulation results confirm that using the reprojection error as the metric for detecting the local minimum is valid in practice. On the basis of these results, we define $t_{rep} = 2$ in Section 5.5.1. However, the relationship between σ_p and t_{rep} has not been proven theoretically. This is a subject for our future work in this study.

5.7 Summary

In this chapter, we proposed a new algorithm that calibrates a camera to a 3D reference object via cornea reflection with minimal configuration. The key features of our method are its introduction of two constraints: cornea reflection model and equidistance constraint. In evaluations, our method outperformed a state-of-the-art of plane mirror based method with both synthesized and real data.

The proposed method introduced in this chapter mainly focuses on the

5.7. Summary

geometric relationship between the 3D reference point and its 2D projection. While the human eye reflects ray and can be utilized as a spherical mirror, the human eye also reflects a cue of mental condition of humans, such as the gaze direction[HDKM10]. These cues are sometimes available for analyzing the relationship of each parts of a imaging system. The fusion of reflections of a ray and a mental condition is one of the future works.

Chapter 6

Conclusion

6.1 Summary

The geometric camera calibration algorithms presented in this dissertation are based on mirror reflection. 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 Chapter 3. Realization of this intrinsic camera calibration algorithm requires that three problems be solved: chamber assignment, mirror parameter estimation, and intrinsic parameter estimation. The proposed method uses a novel geometric constraint, the *kaleidoscopic projection constraint*, to solve these problems. Qualitative and quantitative evaluations using synthesized and real data demonstrated that the algorithm has performance comparable to that of a one using a known reference object.

For cases in which the camera cannot directly observe a known reference object for extrinsic camera calibration due to a physical constraint on the imaging system, planar-mirror-based and human-cornea-based extrinsic camera calibration algorithms were introduced.

The planar-mirror-based algorithm is based on the use of an *orthogonality constraint*, which is satisfied among reflections by multiple mirror poses, and derives closed-form solutions of the extrinsic parameters by using the

constraint. This algorithm was formulated using the minimal configuration of a planar-mirror-based system, that is, three reference points and three mirror poses. In evaluations with synthesized and real data, it outperformed previous planar-mirror-based algorithms.

The human-cornea-based algorithm was developed for calibrating the camera using an unobservable reference object in casual scenarios. It is based on the fact that the surface of the human eye reflects light like a mirror and can thus be modeled as a spherical mirror. This idea was realized by introducing a geometric model of the human cornea. In evaluations with synthesized and real data, it was robust against observation noise, and the precision of estimating extrinsic parameters was comparable to that of the planar-mirror-based algorithm.

6.2 Future Works

This dissertation explored intrinsic and extrinsic camera calibration using mirror reflection. Future work will focus on improving the feasibility of the proposed mirror-based algorithms in various configurations and scenarios.

Intrinsic Camera Calibration with Various Camera-Lens Model
The proposed algorithm introduced in Chapter 3 is focused mainly on the perspective camera model with simple lens distortion, as described in Eq. (2.3). Along with the development of imaging systems, a wide variety of camera models [NNM⁺15, GN05, KB06] and lens models [BL95, SN00] have been proposed. Adaptation to these models should be investigated.

Extrinsic Camera Calibration via Multiple Reflections The proposed algorithm introduced in Chapter 4 uses a 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. For such cases, multiple reflections are needed, so extrinsic camera calibration via multiple reflections should be investigated.

Extrinsic Camera Calibration with Geometric and Non-geometric Information The proposed algorithm introduced in Chapter 5 is focused mainly on the geometric relationship between the 3D reference point and its 2D projection. While the human eye reflects rays and thus can be modeled as a spherical mirror, the human eye also provides cues, such as the gaze direction, reflecting the person's thinking [HDKM10]. These cues may prove useful in analyzing the relationships among the components of an imaging

6.2. Future Works

system. Therefore, the fusion of ray reflections and a person's thinking should be investigated.

Bibliography

- [ABB⁺01] Ronald Azuma, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair MacIntyre. Recent advances in augmented reality. *IEEE computer graphics and applications*, 21(6):34–47, 2001.
- [AFS⁺10] Sameer Agarwal, Yasutaka Furukawa, Noah Snavely, Brian Curless, Steven M. Seitz, and Rick Szeliski. Reconstructing rome. *Computer*, 43(6):40–47, June 2010.
- [AFS⁺11] Sameer Agarwal, Yasutaka Furukawa, Noah Snavely, Ian Simon, Brian Curless, Steven M Seitz, and Richard Szeliski. Building rome in a day. *Communications of the ACM*, 54(10):105–112, 2011.
- [Agr13] Amit Agrawal. Extrinsic camera calibration without a direct view using spherical mirror. In *Proceedings of International Conference on Computer Vision*, 2013.
- [BAKPW16] Gil Ben-Artzi, Yoni Kasten, Shmuel Peleg, and Michael Weraman. Camera calibration from dynamic silhouettes using motion barcodes. In *Proceedings of IEEE Computer Vision and Pattern Recognition*, pages 4095–4103, 2016.
- [BL95] Anup Basu and Sergio Licardie. Alternative models for fish-eye lenses. *Pattern recognition letters*, 16(4):433–441, 1995.
- [CG00] Roberto Cipolla and Peter Giblin. *Visual motion of curves and surfaces*. Cambridge University Press, 2000.
- [CM05] Ondrej Chum and Jiri Matas. Matching with prosac - progressive sample consensus. In *Proceedings of IEEE Computer Vision and Pattern Recognition*, volume 1, pages 220–226, June 2005.

BIBLIOGRAPHY

- [DK02] Guilherme N. DeSouza and Avinash C. Kak. Vision for mobile robot navigation: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(2):237–267, feb 2002.
- [DLJP14] Amael Delaunoy, Jia Li, Bastien Jacquet, and Marc Pollefeys. Two cameras and a screen: how to calibrate mobile devices? In *Proceedings of 3D Vision*, volume 1, pages 123–130. IEEE, 2014.
- [FF⁺96] Andrew W Fitzgibbon, Robert B Fisher, et al. A buyer’s guide to conic fitting. *DAI Research paper*, 1996.
- [FHB07] Y. Francken, C. Hermans, and P. Bekaert. Screen-camera calibration using a spherical mirror. In *4th Canadian Conference on Computer and Robot Vision*, 2007.
- [FLM92] Olivier D Faugeras, Q-T Luong, and Stephen J Maybank. Camera self-calibration: Theory and experiments. In *Proceedings of European Conference on Computer Vision*, pages 321–334. Springer, 1992.
- [FP09] Yasutaka Furukawa and Jean Ponce. Accurate camera calibration from multi-view stereo and bundle adjustment. *International Journal of Computer Vision*, 84(3):257–268, 2009.
- [FSP10] Friedrich Fraundorfer, Davide Scaramuzza, and Marc Pollefeys. A constricted bundle adjustment parameterization for relative scale estimation in visual odometry. In *Proceedings of IEEE International Conference on Robotics and Automation*, pages 1899–1904, may 2010.
- [FY07] Nathan Funk and Yee-Hong Yang. Using a raster display for photometric stereo. In *Computer and Robot Vision, 2007. CRV’07. Fourth Canadian Conference on*, pages 201–207. IEEE, 2007.
- [GHTC03] Xiao-Shan Gao, Xiao-Rong Hou, Jianliang Tang, and Hang-Fei Cheng. Complete solution classification for the perspective-three-point problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8):930–943, 2003.
- [GN01] Joshua Gluckman and Shree K. Nayar. Catadioptric stereo using planar mirrors. *International Journal of Computer Vision*, 44(1):65–79, 2001.

BIBLIOGRAPHY

- [GN05] Michael D Grossberg and Shree K Nayar. The raxel imaging model and ray-based calibration. *International Journal of Computer Vision*, 61(2):119–137, 2005.
- [GVL96] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, Maryland, third edition, 1996.
- [HDKM10] Takatsugu Hirayama, Jean-Baptiste Dodane, Hiroaki Kawashima, and Takashi Matsuyama. Estimates of user interest using timing structures between proactive content-display updates and eye movements. *IEICE TRANSACTIONS on Information and Systems*, 93(6):1470–1478, 2010.
- [HK94] Glenn E Healey and Raghava Kondepudy. Radiometric ccd camera calibration and noise estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(3):267–276, 1994.
- [HMR09] Joel A Hesch, Anastasios I Mourikis, and Stergios I Roumeliotis. Mirror-based extrinsic camera calibration. In *Algorithmic Foundation of Robotics VIII*, pages 285–299. Springer, 2009.
- [HZ00] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.
- [IO12] Eghosasere Iyamu and Ebi Osuobeni. Age, gender, corneal diameter, corneal curvature and central corneal thickness in nigerians with normal intra ocular pressure. *Journal of optometry*, 5(2):87–97, 2012.
- [IRM⁺12] Ivo Ihrke, Ilya Reshetouski, Alkhazur Manakov, Art Tevs, Michael Wand, and Hans-Peter Seidel. A kaleidoscopic approach to surround geometry and reflectance acquisition. In *CVPR Workshop on Computational Cameras and Displays*, pages 29–36, 2012.
- [ITM⁺13] Chika Inoshita, Seiichi Tagawa, Md. Abdul Mannan, Yasuhiro Mukaigawa, and Yasushi Yagi. Full-dimensional sampling and analysis of bssrdf. *IPSJ Transactions on Computer Vision and Applications*, 5:119–123, 2013.
- [KB06] Juho Kannala and Sami S Brandt. A generic camera model and calibration method for conventional, wide-angle, and fish-eye

BIBLIOGRAPHY

- lenses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(8):1335–1340, 2006.
- [KIFP08] Ram Krishan Kumar, Adrian Ilie, Jan Michael Frahm, and Marc Pollefeys. Simple calibration of non-overlapping cameras with a mirror. In *Proceedings of IEEE Computer Vision and Pattern Recognition*, pages 1–7, 2008.
- [KPB⁺12] Claudia Kuster, Tiberiu Popa, Jean-Charles Bazin, Craig Gotsman, and Markus Gross. Gaze correction for home video conferencing. *ACM Transactionson Graphics (In Proceedings of ACM SIGGRAPH ASIA)*, 31(6):to appear, 2012.
- [KRN97] Takeo Kanade, Peter Rander, and P. J. Narayanan. Virtualized reality: constructing virtual worlds from real scenes. *Multimedia, IEEE*, 4(1):34 –47, jan-mar 1997.
- [LMNF09] Vincent Lepetit, Francesc Moreno-Noguer, and Pascal Fua. Epnp: An accurate $O(n)$ solution to the pnp problem. *International Journal of Computer Vision*, 81(2):155–166, 2009.
- [MF92] Stephen J Maybank and Olivier D Faugeras. A theory of self-calibration of a moving camera. *International Journal of Computer Vision*, 8(2):123–151, 1992.
- [MN99] Tomoo Mitsunaga and Shree K Nayar. Radiometric self calibration. In *Proceedings of IEEE Computer Vision and Pattern Recognition*, volume 1. IEEE, 1999.
- [Moa02] Maher Moakher. Means and averaging in the group of rotations. *SIAM J. Matrix Anal. Appl.*, 24:1–16, January 2002.
- [MTK⁺11] Yasuhiro Mukaigawa, Seiichi Tagawa, Jaewon Kim, Ramesh Raskar, Yasuyuki Matsushita, and Yasushi Yagi. Hemispherical confocal imaging using turtleback reflector. In *Proceedings of Asian Conference on Computer Vision, ACCV'10*, pages 336–349, Berlin, Heidelberg, 2011. Springer-Verlag.
- [Nay97] Shree K. Nayar. Catadioptric omnidirectional camera. In *Proceedings of IEEE Computer Vision and Pattern Recognition*, pages 482–488, 1997.
- [NN98] Sameer A Nene and Shree K Nayar. Stereo with mirrors. In *Proceedings of International Conference on Computer Vision*, pages 1087–1094, 1998.

BIBLIOGRAPHY

- [NN12] Atsushi Nakazawa and Christian Nitschke. Point of gaze estimation through corneal surface reflection in an active illumination environment. In *Proceedings of European Conference on Computer Vision*, 2012.
- [NNM⁺15] Mai Nishimura, Shohei Nobuhara, Takashi Matsuyama, Shinya Shimizu, and Kensaku Fujii. A linear generalized camera calibration from three intersecting reference planes. In *Proceedings of International Conference on Computer Vision*, pages 2354–2362, Dec 2015.
- [NNT11] Christian Nitschke, Atsushi Nakazawa, and Haruo Takemura. Display-camera calibration using eye reflections and geometry constraints. *Computer Vision and Image Understanding*, 115(6):835–853, 2011.
- [PBZP14] Jens Puwein, Luca Ballan, Remo Ziegler, and Marc Pollefeys. Joint camera pose estimation and 3d human pose estimation in a multi-camera setup. In *Asian Conference on Computer Vision*, pages 473–487. Springer, 2014.
- [RBN10] Rui Rodrigues, P Barreto, and Urbano Nunes. Camera pose estimation using images of planar mirror reflections. In *Proceedings of European Conference on Computer Vision*, pages 382–395, 2010.
- [RI13] Ilya Reshetouski and Ivo Ihrke. Mirrors in computer graphics, computer vision and time-of-flight imaging. In *Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications*, pages 77–104. Springer, 2013.
- [RiABR⁺13] Ilya Reshetouski, Alkhazur Manakov and Ayush Bhandari, Ramesh Raskar, Hans-Peter Seidel, and Ivo Ihrke. Discovering the structure of a planar mirror system from multiple observations of a single point. In *Proceedings of IEEE Computer Vision and Pattern Recognition*, pages 89–96, 2013.
- [RMSI11] Ilya Reshetouski, Alkhazur Manakov, Hans-Peter Seidel, and Ivo Ihrke. Three-dimensional kaleidoscopic imaging. In *Proceedings of IEEE Computer Vision and Pattern Recognition*, pages 353–360, 2011.
- [RS97] Michael A. Lemp Richard S.Snell. *Clinical Anatomy of the Eye*. Wiley-Blackwell, second edition, 1997.

BIBLIOGRAPHY

- [SB06] Peter Sturm and Thomas Bonfort. How to compute the pose of an object without a direct view. In *Proceedings of Asian Conference on Computer Vision*, pages 21–31, 2006.
- [SMS06] Davide Scaramuzza, Agostino Martinelli, and Roland Siegwart. A flexible technique for accurate omnidirectional camera calibration and structure from motion. In *Computer Vision Systems, 2006 ICVS'06. IEEE International Conference on*, pages 45–45. IEEE, 2006.
- [SN00] Rahul Swaminathan and Shree K Nayar. Nonmetric calibration of wide-angle lenses and polycameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1172–1178, 2000.
- [SP10] Sudipta N Sinha and Marc Pollefeys. Camera network calibration and synchronization from silhouettes in archived video. *International Journal of Computer Vision*, 87(3):266–283, 2010.
- [TLP02] Dejan Tomažević, Boštjan Likar, and Franjo Pernuš. Comparative evaluation of retrospective shading correction methods. *Journal of Microscopy*, 208(3):212–223, 2002.
- [TMHF00] Bill Triggs, PhilipF. McLauchlan, RichardI. Hartley, and AndrewW. Fitzgibbon. Bundle adjustment — a modern synthesis. In Bill Triggs, Andrew Zisserman, and Richard Szeliski, editors, *Vision Algorithms: Theory and Practice*, volume 1883 of *Lecture Notes in Computer Science*, pages 298–372. Springer Berlin Heidelberg, 2000.
- [TNM12] Kosuke Takahashi, Shohei Nobuhara, and Takashi Matsuyama. A new mirror-based extrinsic camera calibration using an orthogonality constraint. In *Proceedings of IEEE Computer Vision and Pattern Recognition*, pages 1051–1058, 2012.
- [Tsa86] Roger Y. Tsai. An efficient and accurate camera calibration technique for 3d machine vision. In *Proceedings of IEEE Computer Vision and Pattern Recognition*, volume 1. IEEE, 1986.
- [YPH⁺13] Xianghua Ying, Kun Peng, Yongbo Hou, Sheng Guan, Jing Kong, and Hongbin Zha. Self-calibration of catadioptric camera with two planar mirrors from silhouettes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(5):1206–1220, 2013.

BIBLIOGRAPHY

- [Zha00] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.

List of Publications

Journal Papers

1. Kosuke Takahashi, Dan Mikami, Mariko Isogawa, Akira Kojima, "Extrinsic Camera Calibration with Minimal Configuration using Cornea Model and Equidistance Constraint", IPSJ Transactions on Computer Vision and Applications, Vol.8, 2016, pp.20-28
2. Kosuke Takahashi, Shohei Nobuhara, Takashi Matsuyama, "Mirror-based Camera Pose Estimation Using an Orthogonality Constraint", IPSJ Transactions on Computer Vision and Applications, Vol.8, 2016, pp.11-19

International Conference (reviewed)

1. Kosuke Takahashi, Akihiro Miyata, Shohei Nobuhara, Takashi Matsuyama, "A Linear Extrinsic Calibration of Kaleidoscopic Imaging System from Single 3D Point", IEEE, CVPR, June, 2017
2. Kosuke Takahashi, Dan Mikami, Mariko Isogawa, Akira Kojima, "Cornea-Reflection-Based Extrinsic Camera Calibration Without a Direct View", INSTICC, VISAPP, Feb, 2016
3. Kosuke Takahashi, Shohei Nobuhara, Takashi Matsuyama, "A New Mirror-based Extrinsic Camera Calibration Using an Orthogonality Constraint", IEEE, CVPR, June, 2012

Presentation (in Japanese)

1. Kosuke Takahashi, Shohei Nobuhara, Takashi Matsuyama, "A Linear Extrinsic Calibration of Kaleidoscopic Imaging System from Single 3D Point", Technical Report of IPSJ, CVIM, Jan, 2017

BIBLIOGRAPHY

2. Kosuke Takahashi, Dan Mikami, Mariko Isogawa, Akira Kojima, "Extrinsic Camera Calibration of Display-Camera System using Cornea Reflection", Technical Report of IEICE, MVE, Oct, 2015
3. Kosuke Takahashi, Dan Mikami, Akira Kojima, "Extrinsic Calibration of Display-Camera System for AR/VR Applications", Technical Report of IEICE, MVE, Jan, 2015
4. Kosuke Takahashi, Shohei Nobuhara, Takashi Matsuyama, "A New Linear Extrinsic Camera Calibration using Mirrored Images", Technical Report of IPSJ, CVIM, Jan, 2012