Cornea-reflection-based Extrinsic Camera Calibration without a Direct View

Kosuke Takahashi, Dan Mikami, Mariko Isogawa and Akira Kojima

NTT Media Intelligence Laboratories, Nippon Telegraph and Telephone Corporation,
1-1, Hikarinooka Yokosuka-Shi, Kanagawa, Japan

Keywords: Camera Calibration, Cornea Reflection, Spherical Mirror.

Abstract: In this paper, we propose a novel method to extrinsically calibrate a camera to a 3D reference object that is not directly visible from the camera. We use the spherical human cornea as a mirror and calibrate the extrinsic parameters from the reflections of the reference points. The main contribution of this paper is to present a cornea-reflection-based calibration algorithm with minimal configuration; there are five reference points on a single plane and one mirror pose. In this paper, we derive a linear equation and obtain a closed-form solution of extrinsic calibration by introducing two key ideas. The first is to model the cornea as a virtual sphere, which enables us to estimate the center of the cornea sphere from its projection. The second idea is to use basis vectors to represent the position of the reference points, which enables us to deal with 3D information of reference points compactly. Besides, in order to make our method robust to observation noise, we minimize the reprojection error while maintaining the valid 3D geometry of the solution based on the derived linear equation. We demonstrate the advantages of the proposed method with qualitative and quantitative evaluations using synthesized and real data.

1 INTRODUCTION

Determining the geometric relationship between a camera and a 3D reference object is called extrinsic camera calibration, and has been a fundamental research field in computer vision for many years(Hartley and Zisserman, 2004; Zhang, 2000). This technique is widely used as an essential element of various applications, such as 3D shape reconstruction from multi-view images(Matsuyama et al., 2012; Agarwal et al., 2010), and augmented reality(Azuma et al., 2001). Conventional extrinsic calibration techniques have a fundamental assumption: *the camera should observe the 3D reference object directly*.

Display-camera systems such as laptop computers, smart phones, and digital signage have become popular and thus gained much attention as a useful device for many tasks in computer vision. For example, Hirayama *et al.*(Hirayama et al., 2010) estimate the interest of users who are watching a digital signage. They assume that the user's gaze points represent his/her interests in the displayed contents. As another example, Kuster *et al.*(Kuster et al., 2012) propose a gaze correction method with a display-camera setup for home video conferences. For these applications, they have to know the relative posture and position
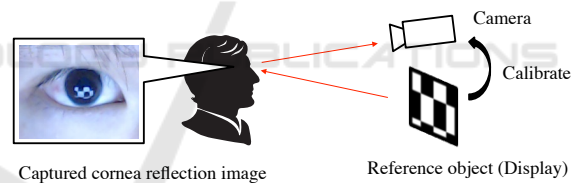

Figure 1: Cornea-reflection-based extrinsic camera calibration. The goal of this paper is to calibrate the camera against the reference object which lies out of the camera's field of view.

of the camera against the display. However, the fundamental assumption of extrinsic camera calibration, *the camera should observe the 3D reference object directly*, does not hold in some cases of display-camera system calibration, as in (Hirayama et al., 2010) and (Kuster et al., 2012). In this paper, we focus on extrinsic camera calibration where the reference object lies out of the camera's field of view.

If the reference object is hidden from the camera, mirrors can be used to offset the occlusion. Some studies on calibration with a mirror have described setups to simplify calibration (Sturm and Bonfort, 2006; Kumar et al., 2008; Rodrigues et al., 2010; Hesch et al., 2010; Nayar, 1997; Takahashi et al., 2012; Agrawal, 2013; Francken et al., 2007; De-

launoy et al., 2014). Techniques include decreasing the number of required reference points or mirror poses, because a simpler setup offers many advantages for more robust calibration and lower computational cost. Takahashi *et al.*(Takahashi et al., 2012) and Hesch *et al.*(Hesch et al., 2010) proposed calibration algorithms with three reference points and three poses of a planar mirror, which is the minimal setup for planar mirrors. To decrease the number of mirror poses, Agrawal(Agrawal, 2013) proposed an algorithm with one pose of a spherical mirror and eight reference points. As a calibration method with no additional hardware, Nitschke *et al.*(Nitschke et al., 2011) used the cornea as a spherical mirror. This method needs three reference points and both cornea spheres, i.e., two spherical mirror poses.

In this paper, we focus on cornea-reflection-based extrinsic camera calibration for occluded reference objects (Figure 1). The contribution of this paper is to present a calibration algorithm with *minimal configuration*, that is five reference points on a single plane and one spherical mirror (cornea sphere) pose. In this paper, we derive a linear equation for estimating extrinsic parameters by introducing two key ideas. The first is to model the cornea as a virtual sphere, which enables us to estimate the center of the cornea sphere from its projection. The second is to represent the position of reference points with basis vector expression, which enables us to treat 3D information of the reference points compactly. By solving this linear equation, we obtain extrinsic parameters under the minimal configuration in a linear manner. Besides, in order to make our method robust to observation noise, we minimize the reprojection error while maintaining the valid 3D geometry of the solution based on the derived linear equation.

The rest of this paper is organized as follows. Section 2 provides a review of conventional techniques that use mirrors for calibration and clarifies the novelty of the proposed method. Section 3 describes a measurement model for calibration first, and then introduces key constraints and the algorithm. Section 4 details evaluations conducted on synthesized data and real data to demonstrate the performance of our method. Section 5 provides the discussions on the effects of noise on the cornea model and the validity of using reprojection error as a criteria for detecting a local minimum. Section 6 concludes this paper.

2 RELATED WORK

This section reviews conventional mirror-based calibration approaches and clarifies the contribution of

this paper. Mirror-based calibration algorithms that use indirect observations of 3D reference objects can be categorized in terms of the mirror shape, the number of minimal reference points and mirror poses (See Table 1). First, we categorize them into two groups in terms of mirror shape: (1) Planar mirrors (Sturm and Bonfort, 2006; Kumar et al., 2008; Rodrigues et al., 2010; Hesch et al., 2010; Nayar, 1997; Takahashi et al., 2012), and (2) Spherical mirrors (Agrawal, 2013).

Planar Mirrors: The conventional methods in this group can be categorized based on whether the mirror duplicates the camera (mirrored camera approach) or the reference points (mirrored point approach). Hesch *et al.*(Hesch et al., 2010) take the mirrored camera approach. They estimate the extrinsic parameters between the mirrored camera and the true reference points (not reflections) by solving the P3P problem(Haralick et al., 1994). They use them for estimating the extrinsic parameters between the camera and the true reference points with the configuration of three reference points and three mirror poses. On the other hand, Takahashi *et al.*(Takahashi et al., 2012) adopt the mirrored point approach. They introduce an orthogonality constraint that should be satisfied by all families of reflections of a single reference point and utilize it to estimate extrinsic parameters with the same configuration. Note that Sturm and Bonfort(Sturm and Bonfort, 2006) revealed that at least three mirror poses are required to uniquely determine the extrinsic parameters if the mirror is planar. Therefore, three reference points and three mirror poses is the minimal configuration for planar mirror based methods.

Spherical Mirrors: Agrawal(Agrawal, 2013) proposed a spherical mirror based calibration method. They obtain an *E* matrix similar to the essential matrix, by using a *coplanarity constraint* with eight point correspondences and retrieve the extrinsic parameters from the matrix.

Nitschke *et al.*(Nitschke et al., 2011) proposed a method for calibrating display-camera setups from the reflections in the user's eyes (corneas) with no additional hardware. They estimate 3D positions of the reference points by finding the intersection of two rays connecting a reference point to the center of the eye ball. Their method needs three reference points and both eyes, i.e., two spherical mirrors.

Our novel calibration method is also based on cornea reflections because eliminating additional hardware for calibration is important for casual display-camera systems, such as webcams and smartphones. In this paper, we propose a calibration algorithm that assumes the minimal configuration of five

Table 1: Configuration for each method: shape of mirrornumber of reference points and number of mirror poses.

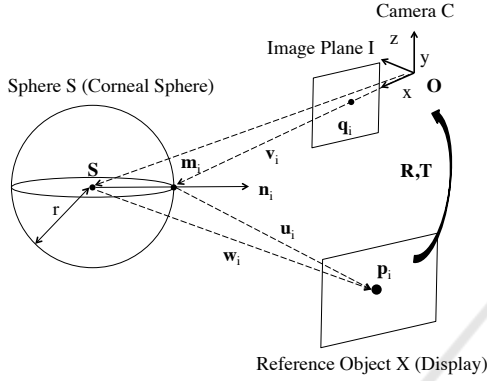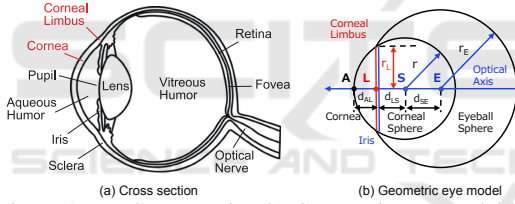|	Shape	Points	Poses
Kumar *et al.*(Kumar et al., 2008)	Plane	5	3
Rodorigues *et al.*(Rodrigues et al., 2010)	Plane	4	3
Hesch *et al.*(Hesch et al., 2010)	Plane	3	3
Takahashi *et al.*(Takahashi et al., 2012)	Plane	3	3
Agrawal (Agrawal, 2013)	Sphere	8	1
Nitschke *et al.*(Nitschke et al., 2011)	(Cornea) Sphere	3	2
Proposed	(Cornea) Sphere	5	1

Figure 2: Reflection model of spherical mirror.

Figure 3: (a) Cross section(b) Geometric eye model based on (Nakazawa and Nitschke, 2012).

reference points on a single plane and one pose of a spherical mirror (cornea sphere) by introducing a cornea sphere model and basis vector expression.

3 EXTRINSIC CAMERA CALIBRATION USING CORNEA REFLECTION

This section introduces our cornea reflection based calibration algorithm; it determines the extrinsic parameters representing the geometric relationship between the camera and an obscured planar reference object.

As illustrated by Figure 2, we assume that reference object X (Display) is located out of camera C's field-of-view and there are N_p reference points $p_i(i = 1, \cdots, N_p)$ on X. These reference points p_i are mirrored by the eye ball and projected onto image

plane I as q_i. Extrinsic parameters (rotation matrix R and translation vector T), which transform the reference object coordinate system $\{X\}$ into the camera coordinate system $\{C\}$, satisfy the following equation.

$$p_i = Rp_i^{\{X\}} + T, \qquad (1)$$

where $p^{\{X\}}$ denotes the 3D position of p in $\{X\}$. We assume that $\{C\}$ is the world coordinate system in this paper and omit this superscript if vector p is represented in $\{C\}$. Our goal is to estimate extrinsic parameters R and T from the projections of the reference points.

3.1 Measurement Model based on Cornea Reflection

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

The human eyeball can be modeled as two overlapping spheres as illustrated by Figure 3. Since the reflections of reference points can be seen at the cornea, we utilize the cornea sphere as a spherical mirror whose center is S and radius is r.

As illustrated by Figure 2, m_i denotes the reflection point of reference point p_i on the cornea sphere. Suppose the unit vector from the camera center O to m_i and unit vector from m_i to p_i are expressed as v_i and u_i, respectively, p_i is expressed as follows:

$$p_i = k_i u_i + m_i, \qquad (2)$$

where k_i denotes the distance between m_i and p_i. Based on the laws of reflection, u_i is expressed as,

$$u_i = v_i + 2(-v_i^\top \cdot n_i)n_i, \qquad (3)$$

where n_i denotes the normal vector at m_i. Since the normal vector n_i is the unit vector from the center of cornea sphere S to m_i, n_i is expressed as $n_i = (m_i - S)/|m_i - S|$.

With the unit vector v_i, m_i is expressed as,

$$m_i = k_i' v_i, \qquad (4)$$

where k_i' denotes the distance between O and m_i. By using projection q_i, we obtain $v_i = (K^{-1}q_i^I)/|K^{-1}q_i^I|$,
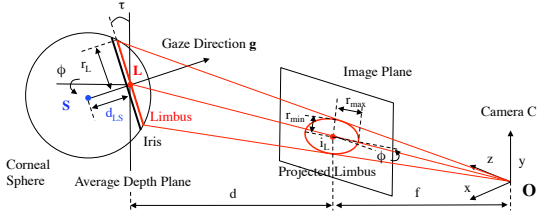
19

Figure 4: Estimating the center of the cornea sphere from limbus projection.

where matrix K denotes the intrinsic parameters and supposed to be given beforehand.

Since m_i is on the cornea sphere, m_i satisfies $|m_i - S| = r$. By substituting Eq (4) for this equation and multiplying it by itself, we have

$$k_i'^2 |v_i|^2 - 2k_i' v_i^\top S + |S|^2 - r^2 = 0, \qquad (5)$$

Solving Eq (5) yields two solutions such as $k_i' = (v_i^\top S \pm \sqrt{(v_i^\top S)^2 - |v_i|^2(|S|^2 - r^2)})/|v_i|^2$. Since m_i is the point closer to the camera among the intersections of v_i and the sphere surface, the smaller k_i' represents the distance between O and m_i.

By substituting Eq (2) into Eq (1), we obtain the following equation:

$$Rp_i^{\{X\}} + T = k_i u_i + m_i. \qquad (6)$$

In this paper, we define Eq (6) as the measurement model.

3.2 Reducing Unknown Parameters in Measurement Model

Since only $p_i^{\{X\}}$ is known in Eq (6), we can not solve Eq (6) and obtain extrinsic parameters by simply increasing the number of reference points. In order to reduce the unknown parameters, we introduce two ideas (1) a geometric model of the cornea sphere and (2) basis vector expression to represent 3D reference point position, and propose an extrinsic calibration method with minimal configuration, that is five reference points and one mirror pose.

3.2.1 The Geometric Model of Cornea Sphere

In this section, we describe a method to estimate the center of the cornea sphere, S, from limbus projection by introducing a geometric model(Nakazawa and Nitschke, 2012). 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(Richard S.Snell, 1997).

As illustrated in Figure 4, the limbus projection is modeled as an ellipse represented by five parameters: the center, i_L, the major and minor radii, r_{max} and r_{min}, respectively, and rotation angle ϕ. Since the depth variation of a tilted limbus is much smaller than the distance between camera and the cornea sphere, we assume the weak perspective projection. Under this assumption, the 3D position of the center of limbus L is expressed as $L = dK^{-1}i_L$, where d denotes the distance between the center of the camera O, and the center of the limbus 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 g is approximated by the optical axis of the eye, and is theoretically determined by $g = [\sin\tau\sin\phi, -\sin\tau\cos\phi, -\cos\tau]^\top$, 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, S, is located at distance $d_{LS}(= \sqrt{r^2 - r_L^2} = \sqrt{7.7^2 - 5.6^2} \approx 5.3mm)$ form the limbus, the radius of the cornea sphere from L, we compute S as follows,

$$S = L - d_{LS}g. \qquad (7)$$

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

From the above, by introducing the geometric model of the cornea sphere, we can obtain unknown parameters r and S in Eq (6).

3.2.2 Using Basis Vector Representation of 3D Reference Point Position

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

$$p_i^{\{X\}} = \Sigma_{j=0}^2 a_j^{i\{X\}} e_j^{\{X\}}, \qquad (8)$$

where $a_j^{i\{X\}}$ denotes the coordinates of $p^{\{X\}}$ with respect to basis $e_j^{\{X\}}$. By assuming $p_j^{\{X\}}$ and $e_j^{\{X\}}$ are given a priori, $a_j^{i\{X\}}$ can be computed. By substituting Eq (8) into Eq (1), we have

$$p_i = \Sigma_{j=0}^2 a_j^{i\{X\}} R e_j^{\{X\}} + T. \qquad (9)$$

In cases where $p_0^{\{X\}}$ represents the origin of the reference object coordinate system, p_0 can be considered as translation vector T. Therefore p_i can be expressed as follows

$$p_i = \Sigma_{j=0}^2 a_j^{i\{X\}} R e_j^{\{X\}} + p_0. \qquad (10)$$

3.3 Derivation of Linear Equation for Estimating Extrinsic Parameters

In this section, we derive a linear equation for estimating extrinsic parameters by using two ideas introduced in Section 3.2.1 and 3.2.2.

By substituting Eq (10) into Eq (6) and representing p_0 by using Eq (2), we have

$$\Sigma_{j=0}^2 a_j^{\{X\}} R e_j^{\{X\}} + k_0 u_0 + m_0 = k_i u_i + m_i. \qquad (11)$$

We define each basis vector as $e_0^{\{X\}} = [1,0,0]^\top, e_1^{\{X\}} = [0,1,0]^\top, e_2^{\{X\}} = [0,0,1]^\top$. From Eq (11) for the N_p reference points, we can derive the following linear equation:

$$AX = B, \qquad (12)$$

where,

$$A = \begin{bmatrix} A_0^1 & A_1^1 & A_2^1 & u_0 & W_1 \\ A_0^2 & A_1^2 & A_2^2 & u_0 & W_2 \\ & & \vdots & & \\ A_0^{N_p-1} & A_1^{N_p-1} & A_2^{N_p-1} & u_0 & W_{N_p-1} \end{bmatrix}, \qquad (13)$$

$$A_j^i = a_j^{i\{X\}} I_{3\times3}, \qquad (14)$$

$$W_k = \begin{bmatrix} w_1^k & w_2^k & \cdots & w_{N_p-1}^k \end{bmatrix}, \qquad (15)$$

$$w_m^l = \left\{ \begin{array}{ll} -u_l & (l=m) \\ 0_{3\times1} & (otherwise) \end{array} \right\}, \qquad (16)$$

$$X = \begin{bmatrix} r_0^\top & r_1^\top & r_2^\top & k_0 & k_1 & \cdots & k_{N_p-1}, \end{bmatrix}^\top \qquad (17)$$

$$B = \begin{bmatrix} m_1' & m_2' & \cdots & m_{N_p-1}' \end{bmatrix}^\top, \qquad (18)$$

$$m_i' = (m_i - m_0)^\top. \qquad (19)$$

Vectors r_0, r_1 and r_2 denote the first, second and third columns of the rotation matrix $R = [r_0 r_1 r_2]$.

In this paper, we assume that we use a planar display as the reference object, that is the reference points lie on the same planea. 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 r_2, which is the unknown parameter corresponding to a_2^i in Eq (12), we have the following linear equation:

$$A'X' = B, \qquad (20)$$

where

$$A' = \begin{bmatrix} A_0^1 & A_1^1 & u_0 & W_1 \\ A_0^2 & A_1^2 & u_0 & W_2 \\ & \vdots & & \\ A_0^{N_p-1} & A_1^{N_p-1} & u_0 & W_{N_p-1} \end{bmatrix}, \qquad (21)$$

$$X' = \begin{bmatrix} r_0^\top & r_1^\top & k_0 & k_1 & \cdots & k_{N_p-1} \end{bmatrix}^\top. \qquad (22)$$

With N_p reference points, we have $(6 + N_p)$ unknowns (X') and $3(N_p - 1)$ constraints (rows of A' and B) in Eq (20). Hence, when $N_p \geq 5$, we can solve Eq (20) by $X' = A'^*B$, where A'^* is the pseudo-inverse matrix of A'. r_2 is given by the cross product of r_0 and r_1, i.e. $r_2 = r_0 \times r_1$.

In real environment, the rotation matrix $R = [r_0 r_1 r_2]$ obtained by solving Eq (20) is not guaranteed to satisfy the constraints to form a valid rotation matrix ($| r_0 |=| r_1 |=| r_2 |= 1, r_0^\top r_1 = r_1^\top r_2 = r_2^\top r_0 = 0$). In order to enforce these constraints, here we solve the orthogonal Procrustes problem (Golub and van Loan., 1996) as done by Zhang's method (Zhang, 2000).

This linear solution estimates the correct extrinsic parameters in noiseless environments. As shown in Figure 6, we can see that extrinsic parameter precision degrades remarkably if the input data includes observation noise (We describe the experimental environment in detail in Section 4). To overcome this difficulty, we solve the non-linear optimization problem of the objective function derived from Eq (20), which is robust to noise.

3.4 Solving Non-linear Optimization Problem

3.4.1 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 extrinsic parameters should satisfy the linear equation of Eq (20), which is derived from the measurement model. In order to enforce this constraint on the estimated extrinsic parameters, we introduce the following error term,

$$cost_{model}(R,T) = |A'X'(R,T) - B|, \qquad (23)$$

where $X'(R,T)$ denotes X' computed from the estimated R and T.

Second, we introduce an error term to minimize the reprojection error as widely done in the calibration (Triggs et al., 2000):
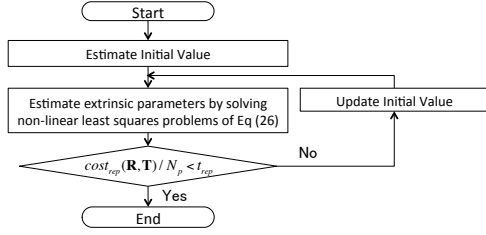
Figure 5: Implementation strategy.

$$cost_{rep}(R,T) = \sum_{i=0}^{N_p-1} |q_i - \check{q}_i(R,T)|, \qquad (24)$$

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

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

$$\begin{aligned} f = \ & c_{model} * cost_{model}(R,T) \\ & + \ c_{rep} * cost_{rep}(R,T), \end{aligned} \qquad (25)$$

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

3.4.2 Implementation

We implement our proposed method together with non-linear optimization as illustrated in Figure 5. First, we estimate the initial values of extrinsic parameters. In this paper, we use a linear solution of extrinsic parameters estimated by solving Eq (20) as the initial value. Second, we use the Levenberg-Marquardt algorithm to solve the non-linear optimization problem of Eq (25). However, Eq (25) is not a convex function and it can converge to a local minimum. Against this problem, we use the reprojection error as the criteria indicating whether the estimated solution is a local minimum or not. When the average reprojection error $cost_{rep}(R,T)/N_p$ is larger than a threshold t_{rep}, that is the estimated solution is a local minimum, we update the initial value of extrinsic parameters by adding random values and resolve the non-linear optimization problem until $cost_{rep}(R,T)/N_p < t_{rep}$ is satisfied.

4 EXPERIMENT

This section details the experiments conducted on synthesized and real data in order to evaluate the quantitative and qualitative performance of our method. In the following, "linear solution" denotes extrinsic parameters estimated by solving Eq (20) in linear manner and "non-linear solution" denotes those estimated by solving the non-linear optimization problem of Eq (25).

4.1 Synthesized Data

4.1.1 Experiment 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. 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 $O = (0,0,0)$. The 3D positions of the reference points are defined as $p_0^{\{X\}} = (0,0,0), p_1^{\{X\}} = (-50,50,0), p_2^{\{X\}} = (50,50,0), p_3^{\{X\}} = (-50,-50,0)$ and $p_4^{\{X\}} = (50,-50,0)$. The center of the cornea sphere is set to $S = (0,45,50)$, the d_{LS} is set to $5.6mm$ and radius r is set to $7.7mm$ on the basis of (Richard S.Snell, 1997).

We represent the ground truth of R as the product of three elemental matrices, one for each axis, that is $R = R_x(\theta_x)R_y(\theta_y)R_z(\theta_z)$, and we set $(\theta_x, \theta_y, \theta_z)$ to $(-0.1, 0.2, 0)[\text{rad}]$. The ground truth of T is set to $(0, 90, 0)$. In the optimization process, c_{model}, c_{rep} and c_{dist} are set to 1 and t_{rep} is set to 2.

Throughout this experiment, we evaluate the distance between estimated parameter and its ground truth, and reprojection error as error metrics. Here, parameters with subscript g indicate ground truth data. The distance between R and R_g, $D_R(R, R_g)$, is defined as the Riemannian distance (Moakher, 2002):

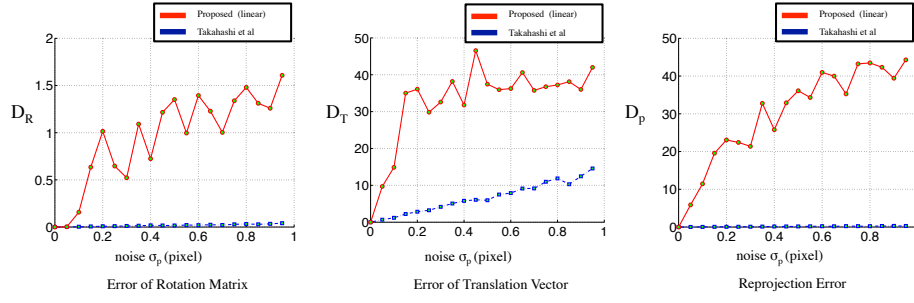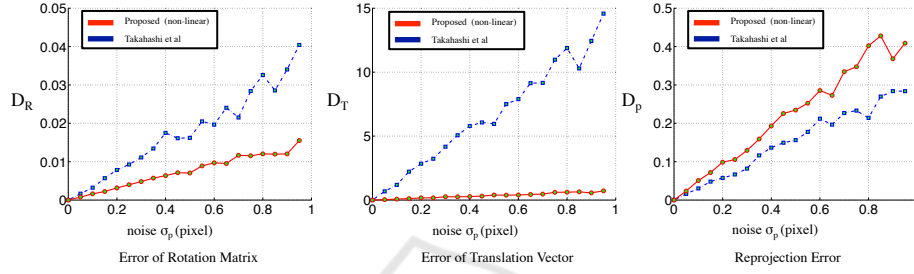$$D_R = \frac{1}{\sqrt{2}} \| \text{Log}(R^{\top}R_g) \|_F, \qquad (26)$$

$$\text{Log}R' = \begin{cases} 0 & (\theta = 0), \\ \frac{\theta}{2\sin\theta}(R' - R'^{\top}) & (\theta \neq 0), \end{cases} \qquad (27)$$

where $\theta = \cos^{-1}(\frac{trR'-1}{2})$. The difference between T and T_g, $D_T(T, T_g)$, is defined as RMS:

$$D_T = \sqrt{| T - T_g |^2 / 3}. \qquad (28)$$

The reprojection error is defined as $D_p = cost_{rep}(R,T)/N_p$.

In this simulation, we computed linear and non-linear solutions from the projection of reference point q_i with zero-mean Gaussian noise whose standard deviation $\sigma_p(0 \leq \sigma_p \leq 1)$. We compared our method against the state-of-the-art of planar mirror based method proposed by Takahashi(Takahashi et al., 2012). For fair comparison, the projections of reference points using either spherical or planar mirrors are assured to occupy a comparable pixel area in the image as done in (Agrawal, 2013).

Figure 6: Estimation errors of linear solution under Gaussian noise for q_i with standard deviation σ_p.

Figure 7: Estimation errors of non-linear solution under Gaussian noise for q_i with standard deviation σ_p.

4.1.2 Results with Synthesized Data

Figure 6 and Figure 7 show D_R, D_T and D_p of the linear solution and the non-linear solution respectively. In each figure, the vertical axis shows the average value over 50 trials and the horizontal axis denotes the standard deviation of noise.

Linear Solution. From Figure 6, we can observe that D_R, D_T and D_p are zero at $\sigma = 0$, which means that the minimal configuration of our method, that is five reference points and one mirror pose, is sufficient. However, when $\sigma_p > 0$, D_R, D_T and D_p increase remarkably. Additionally, most estimated T values, that is p_0, are located around the surface of the cornea sphere. This is explained as follows: In the proposed algorithm, we estimate the linear solutions of R and T as the parameters X' that minimizes $||A'X' - B||^2$. Since the cornea has a very small radius, unit vector u_i used in A' changes significantly with even trivial observation noise. If u_i is wrong, $||A'X' - B||^2$ increases with the distance between reference points and their reflection points on the surface of the spherical mirror, that is k_i in X'. Therefore, in cases where the input data includes observation noise, it is considered that $||A'X' - B||^2$ is minimized with small k_i, which means T is located around the surface of the cornea sphere.

Non-linear Solution. From Figure 7, we can observe that estimation errors D_R and D_T are significantly smaller than those of Takahashi *et al.*(Takahashi et al., 2012) (57.5%, 94.7%,respectively). These results quantitatively prove that our method outperforms Takahashi *et al.*(Takahashi et al., 2012) and works robustly even if the input data includes observation noise.

4.2 Real Data

4.2.1 Configuration

Figure 8 overviews the configuration. We used a Logicool HD Pro Webcam C920t and captured frames had the resolution of 1920×1080. As illustrated in Figure 8, we projected a chessboard pattern on the display and captured the cornea as the reference points $p_i(i = 0, \cdots, 4)$. The size of each chess block was 125×125mm. The distance between the user's cornea center and the display was about 300 mm. The intrinsic parameter was estimated beforehand by (Zhang, 2000). In order to estimate ellipse parameters $(i_L, \phi, r_{max}, r_{min})$ from limbus projection, we binarize the input image, apply the Canny detector and fit an ellipse (Fitzgibbon and Fisher, 1995) as shown in Figure 9.

Since the ground truth of extrinsic parameters is not available in any real configuration, we used (Takahashi et al., 2012) as the reference parameters.
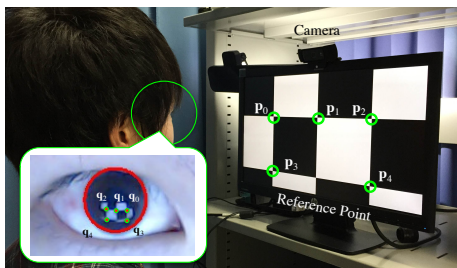
Figure 8: Configuration for experiments with real data. Notice that we use only five points $p_i (i = 0, \cdots, 4)$ of the chessboard pattern as the reference points for calibration. Each q_i is separated by about $10 \sim 13$ pixels in captured image.
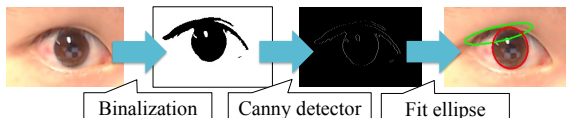

Figure 9: A flow of estimating ellipse parameters $(i_L, \phi, r_{max}, r_{min})$ from projection of limbus.
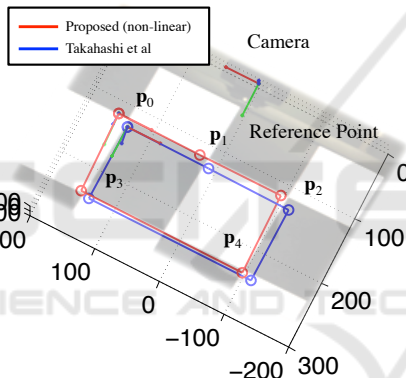

Figure 10: Positions of the reference points as estimated by proposed method (red), by (Takahashi et al., 2012) (blue).

4.2.2 Results with Real Data

Table 2 quantitatively compares the parameters estimated by the proposed method (linear solution and non-linear solution) with (Takahashi et al., 2012).

We can see that the distance functions yielded by the linear solution output large differences. It is considered that some observation noise is present because the estimated T is close to surface of the cornea sphere.

On the other hand, the non-linear solution yields small differences. This point can be verified by visualizing the results as shown in Figure 10. It shows that the positions estimated by the proposed method are almost identical to those of (Takahashi et al., 2012). This confirms that our method works properly in real environments. While this precision may not by enough for eye gaze tracking, it is acceptable

Table 2: Error metrics computed by using (Takahashi et al., 2012) as the ground truth.

	D_R	D_T	D_p
Linear solution	0.553	178.896	14.689
Non linear solution	0.164	33.617	0.260

for applications that do not need high precision, such as gaze correction (Kuster et al., 2012) using a display and attached web camera system.
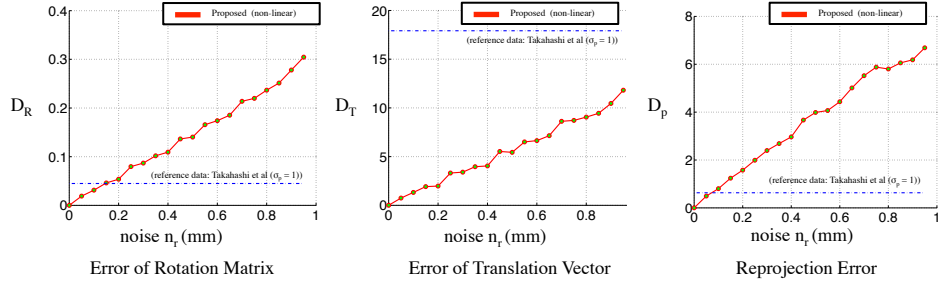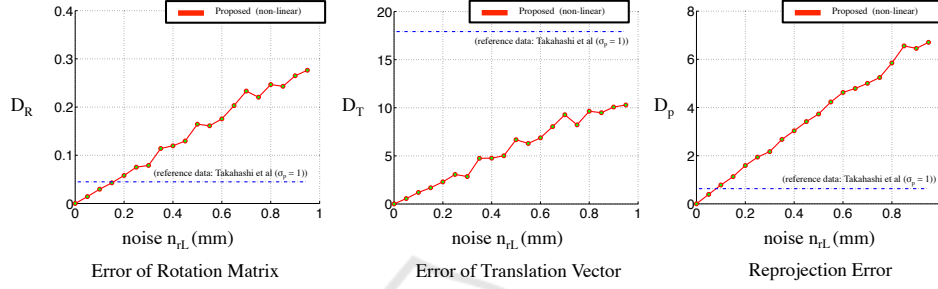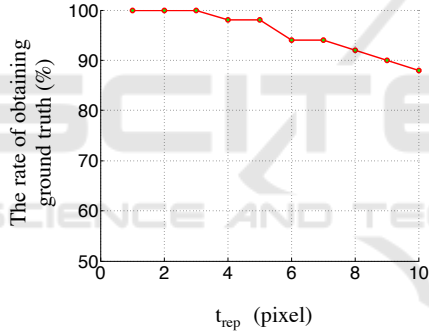
5 DISCUSSION

In this section, we discuss the effects of noise on the cornea model used in the proposed method and the validity of using the reprojection error as the criteria for detecting local minimum.

5.1 Effects of Differences Among Individuals

In our porposed method, we have 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 $r = 7.7$mm (Richard S.Snell, 1997), 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 $r_L = 5.6$mm, but in practice the model parameters can be tailored to suit the individual. To more closely examine the effects of these assumptions, we investigated the effects of noise on these two radii with synthesized data. We used the same configuration as in Section 4.1 and set t_{rep} to 10. We added random noise with uniform distribution n_r and n_{r_L} to r and r_L, respectively, $(0 \leq |n_r| \leq 1, 0 \leq |n_{r_L}| \leq 1)$.

Figure 11 and Figure 12 show the results of the averages of each distance function and reprojection error. From Figure 11 and Figure 12, we can see that r and r_L have strong and similar impact to the estiamtion error of extrinsic parameters and reprojection error. This because adding noise to r and r_L affects the precisions of S estimation based on Eq (7) 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 S and r. To solve this problem, it is useful to calibrate the user's eye parameters beforehand.

Figure 11: Estimation errors under random noise for radius of cornea sphere r with uniform distribution.

Figure 12: Estimation errors under random noise for radius of cornea limbus r_L with uniform distribution.

Figure 13: The rate of matching ground truth for each t_{rep}.

5.2 Validity of using Reprojection Error as the Criteria for Detecting Local Minimum

In Section 3.4.2, we use the reprojection error as the metric indicating whether the estimated solution is a local minimum or not. Here we address the validity of this usage by referring to simulation data. In the simulation, we investigate the rate with 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 T as matching to the ground truth if $D_R < t_{D_R}$ and $D_T < t_{D_T}$, which are set to $t_{D_R} = 0.02$ and $t_{D_T} = 6$ respectively based on the result of (Takahashi et al., 2012) with $\sigma_p = 0.5$. We use the same configuration as in Section 4.1. We added Gaussian noise with zero mean and standard

deviation $\sigma_p = 0.5$ to q_i.

Figure 13 shows the rate of matching the ground truth for each t_{rep} over 50 trials. From Figure 13, we can observe that all the estimated solutions convergence 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. Based on this result, we define $t_{rep} = 2$ in Section 4. However, the relationship between σ_p and t_{rep} is not proven theoretically. This is a part of future work of this study.

6 CONCLUSION

In this paper, we proposed a new algorithm that calibrates a camera to a 3D reference object via cornea reflection with the minimal configuration. The key ideas of our method are to introduce a geometric cornea model and to use basis vector expression to represent the 3D positions of reference points. Based on these ideas, we derived a linear equation and obtained a closed-form solution. Additionally, based on the linear equation, we obtained a non-linear solution that is robust to observation noise. In evaluations, our method outperformed a state-of-the-art of planar mirror based method with both synthesized and real data.

REFERENCES

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

Agrawal, A. (2013). Extrinsic camera calibration without a direct view using spherical mirror. In *Proc. of ICCV*.

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

Delaunoy, A., Li, J., Jacquet, B., and Pollefeys, M. (2014). Two cameras and a screen: How to calibrate mobile devices? In *3D Vision (3DV), 2014 2nd International Conference on*, volume 1, pages 123–130. IEEE.

Fitzgibbon, A. and Fisher, R. B. (1995). A buyer's guide to conic fitting.

Francken, Y., Hermans, C., and Bekaert, P. (2007). Screen-camera calibration using a spherical mirror. In *4th Canadian Conference on Computer and Robot Vision*.

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

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

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

Hesch, J., Mourikis, A., and Roumeliotis, S. (2010). Mirror-based extrinsic camera calibration. In *Algorithmic Foundation of Robotics VIII*, volume 57, pages 285–299.

Hirayama, T., Dodane, J.-B., Kawashima, H., and Matsuyama, T. (2010). Estimates of user interest using timing structures between proactive content-display updates and eye movements. *IEICE Trans. Information and Systems*, 93(6):1470–1478.

Kumar, R., Ilie, A., Frahm, J.-M., and Pollefeys, M. (2008). Simple calibration of non-overlapping cameras with a mirror. In *Proc. of CVPR*.

Kuster, C., Popa, T., Bazin, J.-C., Gotsman, C., and Gross, M. (2012). Gaze correction for home video conferencing. *ACM Trans. Graph.*, 31(6):174:1–174:6.

Matsuyama, T., Nobuhara, S., Takai, T., and Tung, T. (2012). *3D Video and Its Applications*. Springer Publishing Company, Incorporated.

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

Nakazawa, A. and Nitschke, C. (2012). Point of gaze estimation through corneal surface reflection in an active illumination environment. In *Proc. of ECCV*.

Nayar, S. (1997). Catadioptric omnidirectional camera. In *Proc. of CVPR*.

Nitschke, C., Nakazawa, A., and Takemura, H. (2011). Display-camera calibration using eye reflections and geometry constraints. *CVIU*, 115(6):835 – 853.

Richard S.Snell, M. A. L. (1997). *Clinical Anatomy of the Eye*. Wiley-Blackwell, second edition.

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

Sturm, P. and Bonfort, T. (2006). How to compute the pose of an object without a direct view. In *Proc. of ACCV*.

Takahashi, K., Nobuhara, S., and Matsuyama, T. (2012). A new mirror-based extrinsic camera calibration using an orthogonality constraint. In *Proc. of CVPR*.

Triggs, B., McLauchlan, P., Hartley, R., and Fitzgibbon, A. (2000). Bundle adjustment a modern synthesis. In Triggs, B., Zisserman, A., and Szeliski, R., editors, *Vision Algorithms: Theory and Practice*, volume 1883 of *Lecture Notes in Computer Science*, pages 298–372. Springer.

Zhang, Z. (2000). A flexible new technique for camera calibration. *TPAMI*, pages 1330–1334.