Extrinsic Camera Calibration of Display-Camera System with Cornea Reflections**

Kosuke TAKAHASHI†(a), Nonmember, Dan MIKAMI(b), Senior Member, Mariko ISOGAWA(c), Nonmember, Akira KOJIMA†(d), and Hideaki KIMATA†(e), Members

SUMMARY 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 a human cornea as a spherical 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 a simple configuration: 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 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 use basis vectors to represent the position of the reference points, which enables us to deal with 3D information of reference points compactly. We demonstrate the performance of the proposed method with qualitative and quantitative evaluations using synthesized and real data.

key words: camera calibration, cornea reflection, spherical mirror

1. Introduction

Display-camera systems have become commonly used in many studies and applications. For example, Hirayama et al. [2] used a display-camera system to estimate people’s interest from their gaze while they were watching contents shown on a display. Kuster et al. [3] proposed a gaze correction method using a display-camera system to reduce miscommunication caused by mismatching gaze direction in video chatting. For these applications, extrinsic calibration, which is for estimating relative pose and posture between a display and a camera, should be done beforehand. Conventionally, to do the calibration, the reference object should be directly observed by the camera [4], [5]. However, in some cases when using display-camera systems [2], [3], the display that is supposed to be the reference object cannot be directly observed from the camera, so the conventional methods do not work in such situations. In this paper, we focus on extrinsic calibration in the situations where the reference object does not lie in the camera’s field of view.

Several mirror-based approaches have been proposed for when the reference object is not observable from the camera. Some approaches aim to calibrate with simpler set-ups [6]–[10], [12]–[17], which mean 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. Takahashi et al. [9] and Hesch et al. [8] proposed planar mirror-based approaches that need three mirror poses and three reference points. Agrawal [10] proposed a spherical mirror-based approach that needs one mirror pose and eight reference points. 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 [3].

The key contribution of this paper is to propose a new extrinsic calibration method without such additional devices and with a simpler setup than conventional methods. In this paper, we focus on utilizing the human cornea as a spherical mirror as described in [11], [12], [27], because the cornea surface reflects light as a mirror and cornea size does not vary widely from person to person [18]. Takahashi et al. [9] also proposed a cornea reflection-based extrinsic camera calibration method with minimal configuration consisting of one mirror (cornea) pose and three reference points. However, there are limits to the mirror position, e.g., the mirror should be set at the same distance from the camera center and the reference points. In this paper, we propose a novel cornea reflection-based calibration method without such limitations on the mirror position. We derive a linear equation for estimating extrinsic parameters by modeling a cornea as a spherical mirror and introduce basis vector representation that enables us to treat 3D information of the reference points compactly.

The rest of this paper is organized as follows. Section 2...
This section provides a review of related work done in mirror-based calibration techniques. Section 3 provides a measurement model and our calibration algorithm. Section 4 describes evaluations conducted on synthesized data and real data to demonstrate the performance of our method. Section 5 discusses details of our method, especially the effects of noise on the cornea model and the validity of using reprojection error as a criterion for detecting a local minimum. Section 6 concludes the paper with a summary of the key points.

2. Related Work

This section provides a review of related work done in mirror-based calibration techniques [6]–[15]. Mirror-based calibration methods can be categorized in terms of configuration (i.e., mirror shape), the required number of mirrors, the required number of reference points, limitations, and the number of additional non-reference points required (see Table 1).

Hesch et al. [8] and Takahashi et al. [9] proposed planar mirror-based calibration methods with the minimal configuration of three mirror poses and three mirror reference points. Hesch et al. [8] estimated the extrinsic parameters between the mirrored camera and the true reference points (not reflections) in P3P scenarios. Takahashi et al. [9] introduced an orthogonality constraint that must be satisfied by the reflections of a single reference point and the intersection of mirror planes and utilized it to estimate extrinsic parameters linearly. Takahashi et al. [13] generalized the planar mirror-based calibration method [9] for configurations of more than three mirror poses and more than three reference points. Sturm and Bonfort [14] revealed that three mirror poses are the minimal configuration for determining the extrinsic parameters in planar mirror cases. On the other hand, Agrawal [10] proposed a spherical mirror-based calibration method that utilizes a matrix $E$ similar to the essential matrix in mirror geometry by using a coplanarity constraint with eight point correspondences. These methods need a planar or spherical mirror as an additional device, but it is sometimes troublesome to utilize such additional devices to calibrate every time in casual scenarios, such as gaze correction in video chatting.

To address this problem, Nitschke et al. [11] proposed a method for calibrating display-camera setups from the reflections in the user’s eyes (corneas) with no additional hardware. They estimated 3D positions of the reference points by finding the intersection of two rays connecting a reference point to the center of the eyeball. Their method needs three reference points and both eyes, i.e., two spherical mirrors. Takahashi et al. [12] also proposed cornea reflection-based calibration with minimal configuration consisting of one mirror pose and three reference points. Although this method works with such a minimal configuration, it has 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. Note that such spherical-based methods need at least five additional non-reference points for detecting an ellipse as the projection of the sphere.

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 without limitations on the mirror position. Our algorithm assumes the simple configuration of five 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

Figure 2 illustrates the reflection model with a spherical mirror. Let $X$ be the reference object and $i.e.$ a display. This $X$ lies out of the camera $C$’s field of view and $X$ has $N_p$ reference points $p_i$ $(i = 1, \cdots, N_p)$. The 2D points $q_i$ in image plane $I$ denote the projections of reflections of $p_i$. The rotation matrix $R$ and the translation vector $t$ are the extrinsic parameters, which transform the reference object coordinate system $\{X\}$ into a camera coordinate system $\{C\}$ and satisfy

Table 1 Features of each mirror-based calibration method: mirror shape, number of reference points, number of mirror poses, limitations, and number of additional non-reference (NR) points.

<table>
<thead>
<tr>
<th>Shape</th>
<th>Points</th>
<th>Poses</th>
<th>Limitations</th>
<th>NR points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kumar et al. [6]</td>
<td>Plane</td>
<td>5</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>Roodrigues et al. [7]</td>
<td>Plane</td>
<td>4</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>Hesch et al. [8]</td>
<td>Plane</td>
<td>3</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>Takahashi et al. [9], [13]</td>
<td>Plane</td>
<td>3+</td>
<td>3+</td>
<td>-</td>
</tr>
<tr>
<td>Agrawal [10]</td>
<td>Sphere</td>
<td>8</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Nitschke et al. [11]</td>
<td>(Cornea) Sphere</td>
<td>3</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>Takahashi et al. [12]</td>
<td>(Cornea) Sphere</td>
<td>3</td>
<td>1</td>
<td>mirror position</td>
</tr>
<tr>
<td>Proposed</td>
<td>(Cornea) Sphere</td>
<td>5</td>
<td>1</td>
<td>-</td>
</tr>
</tbody>
</table>
the following equation,
\[ p_i = Rp_i[X] + t, \] 
where \( p_i[X] \) denotes the 3D position of \( p \) in \( \{X\} \). Here, we assume that the camera coordinate system \( \{C\} \) is the world coordinate system and omit this superscript if a vector \( p \) is represented in \( \{C\} \). The goal of extrinsic camera calibration is to estimate these 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 eyeball as a spherical mirror.

As illustrated by Fig. 3, the human eyeball can be modeled as two overlapping spherical mirrors. Here, we focus on the cornea sphere whose center is \( S \) and radius is \( r \). Let \( m_i \) denote the reflection point of reference point \( p_i \) on the cornea sphere. Let us assume the unit vector from \( m_i \) to \( p_i \) expressed as \( u_i \), \( p_i \) is expressed as
\[ p_i = k_i u_i + m_i, \] 
where \( k_i \) is the distance between \( m_i \) and \( p_i \). The laws of reflection provide the following equation regarding \( u_i \),
\[ u_i = v_i + 2(-v_i^T \cdot n_i) n_i, \] 
where \( n_i \) denotes the normal vector of a spherical mirror at \( m_i \). Since \( n_i \) is the unit vector from the center of the cornea sphere \( S \) to \( m_i \), \( n_i \) is expressed as \( n_i = (m_i - S)/||m_i - S|| \), where ||*|| means the Euclidean norm.

The \( m_i \) is expressed as
\[ m_i = k_i v_i, \] 
where \( v_i \) denotes the unit vector connecting \( O \) and \( m_i \) and \( k_i \) denotes the distance between them. This \( v_i \) is expressed as \( v_i = (K^{-1} q'_i)/||K^{-1} q'_i|| \), where \( q_i \) denotes the projection of \( m_i \) and \( K \) consists of intrinsic parameters, which are assumed to be given beforehand. The \( m_i \) lies on the surface of the cornea sphere and satisfies \( ||m_i - S|| = r \). By substituting Eq. (4) for this equation and multiplying it by itself, we have
\[ k_i^2 ||v_i||^2 - 2k_i v_i^T S + ||S||^2 - r^2 = 0. \] 
This equation provides two solutions: \( k'_i = (v_i^T S \pm \sqrt{(v_i^T S)^2 - ||v_i||^2(||S||^2 - r^2)})/||v_i||^2 \). 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:
\[ Rp_i[X] + t = k_i u_i + m_i. \] 
In Eq. (6), both \( u_i \) and \( m_i \) can be expressed with unknown parameters \( S \) and \( r \). 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 cannot solve Eq. (6) and obtain extrinsic parameters by simply increasing the number of reference points. To reduce the unknown parameters, we introduce (1) a geometric model of the cornea sphere and (2) a basis vector expression to represent 3D reference point position.

3.2.1 Geometric Model of Cornea Sphere

This section describes a method to estimate the center of the cornea sphere, \( S \), from limbus projection introducing a geometric model[19]. On the basis of the work reported by Snell [18], we assume that the average radius of the cornea sphere, \( r \), and the average radius of the cornea limbus, \( r_L \), are respectively 7.7 and 5.6 mm.

Figure 4 shows that the limbus is projected as an ellipse with five parameters: the center, \( i_L \), the major and minor radii, \( (r_{max} \) and \( r_{min} \), respectively), and rotation angle \( \phi \). We assumed weak perspective projection since the depth variation of a tilted limbus is much smaller than the distance between the cornea sphere and the camera. 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 \( d = f \cdot r_L / r_{max} \), where \( f \) represents the focal length in pixels. Gaze direction \( g \) is approximated by the optical axis of the eye and is theoretically determined by \( g = [\sin \tau \sin \phi, -\sin \tau \cos \phi, -\cos \tau]^T \).

Fig. 3  (a) Cross section. (b) Geometric eye model based on [19]. (Cited from [12].)  
Fig. 4  Estimating cornea sphere center from limbus projection proposed by Nakazawa et al. [19].
where \( \tau = \pm \arccos(r_{\min}/r_{\text{max}}) \); \( \tau \) corresponds to the tilt of the limbus plane with respect to the image plane. Since the center of the cornea sphere, \( S \), is located at distance \( d_{LS}(= \sqrt{r_{c}^2 - r_{	ext{limb}}^2} \approx 5.3 \text{ mm}) \) from the limbus, which is the radius of the cornea sphere from \( L \), we compute \( S \) as

\[
S = L - d_{LS}g. \tag{7}
\]

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

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, \( p = \sum_{j=0}^{N_{c}} a_{j} e_{j} \), where \( e_{j} \) \((j = 0 \ldots N_{c} - 1)\) denotes the basis vector of \( N_{c} \) 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, \( N_{c} = 3 \). With this basis vector representation, \( p_{i} \) in the reference object coordinate system \( \{X\} \) is expressed as

\[
p_{i}^{[X]} = \sum_{j=0}^{2} a_{j}^{[X]} e_{j}^{[X]}, \tag{8}
\]

where \( a_{j}^{[X]} \) denotes the coordinates of \( p^{[X]} \) with respect to basis \( e_{j}^{[X]} \). By assuming \( p_{i}^{[X]} \) and \( e_{j}^{[X]} \) are given a priori, \( a_{j}^{[X]} \) can be computed. By substituting Eq. (8) into Eq. (1), we have

\[
p_{i} = \sum_{j=0}^{2} a_{j}^{[X]} Re_{j}^{[X]} + t. \tag{9}
\]

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

\[
p_{i} = \sum_{j=0}^{2} a_{j}^{[X]} Re_{j}^{[X]} + p_{0}. \tag{10}
\]

### 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 Sects. 3.2.1 and 3.2.2.

By substituting Eq. (10) into Eq. (6) and representing \( p_{0} \) by using Eq. (2), we have

\[
\sum_{j=0}^{2} a_{j}^{[X]} Re_{j}^{[X]} + k_{0}u_{0} + m_{0} = k_{i}u_{i} + m_{i}. \tag{11}
\]

We define each basis vector as \( e_{0}^{[X]} = [1, 0, 0]^T \), \( e_{1}^{[X]} = [0, 1, 0]^T \), \( e_{2}^{[X]} = [0, 0, 1]^T \). From Eq. (11) for the \( N_{p} \) reference points, we can derive the following linear equation:

\[
AX = B, \tag{12}
\]

where

\[
A = \begin{bmatrix}
A_{0}^{1} & A_{1}^{1} & A_{2}^{1} & u_{0} & W_{1} \\
A_{0}^{2} & A_{1}^{2} & A_{2}^{2} & u_{0} & W_{2} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
A_{0}^{N_{p}-1} & A_{1}^{N_{p}-1} & A_{2}^{N_{p}-1} & u_{0} & W_{N_{p}-1} \\
\end{bmatrix},
\]

\[
B = \begin{bmatrix}
m_{1}^{*} & m_{2}^{*} & \cdots & m_{N_{p}-1}^{*} \\
\end{bmatrix} \begin{bmatrix}
u_{0} & u_{1} & \cdots & u_{N_{p}} \end{bmatrix}^{T},
\]

\[
W_{k} = \begin{bmatrix}
W_{k}^{1} \\
\vdots \\
W_{k}^{N_{p}}
\end{bmatrix},
\]

\[
X = \begin{bmatrix}
r_{0}^{T} & r_{1}^{T} & r_{2}^{T} & k_{0} & k_{1} & \cdots & k_{N_{p}-1}
\end{bmatrix}^{T},
\]

\[
m_{0}^{T} = (m_{i} - m_{0})^{T}.
\]

Vectors \( r_{0}, r_{1}, r_{2} \) denote the first, second, and third columns of the rotation matrix \( R = [r_{0} \ r_{1} \ r_{2}] \).

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 \( a_{j} \) are zero. By removing \( r_{2} \), which is the unknown parameter corresponding to \( a_{2}^{[X]} \) in Eq. (12), we have the following linear equation:

\[
A'^{X'} = B', \tag{20}
\]

where

\[
A' = \begin{bmatrix}
A_{0}^{1} & A_{1}^{1} & u_{0} & W_{1} \\
A_{0}^{2} & A_{1}^{2} & u_{0} & W_{2} \\
\vdots & \vdots & \vdots & \vdots \\
A_{0}^{N_{p}-1} & A_{1}^{N_{p}-1} & u_{0} & W_{N_{p}-1} \\
\end{bmatrix},
\]

\[
X' = \begin{bmatrix}
r_{0}^{T} & r_{1}^{T} & k_{0} & k_{1} & \cdots & k_{N_{p}-1}
\end{bmatrix}^{T}. \tag{21}
\]

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 \( X' = A'^{-1}B' \), where \( A'^{+} \) is the pseudo-inverse matrix of \( A' \). \( r_{2} \) is given by the cross product of \( r_{0} \) and \( r_{1}, i.e. r_{2} = r_{0} \times r_{1}. \) Note that if the reference points are on or very close to a line, Eq. (20) is not solvable due to the lack of constraints.

In real environments, 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 \((\|r_{0}\| = \|r_{1}\| = \|r_{2}\| = 1, \ r_{0}^{T}r_{1} = r_{1}^{T}r_{2} = r_{2}^{T}r_{0} = 0)\) due to the observation noise. To enforce these constraints, here we solve the orthogonal Procrustes problem \([20]\) as done by Zhang’s method \([5]\).
This linear solution estimates the correct extrinsic parameters in noiseless environments. As shown in Fig. 6, extrinsic parameter precision degrades substantially if the input data includes observation noise (we describe the experimental environment in detail in Sect. 4). To overcome this difficulty, we solve the non-linear optimization problem of the objective function derived from Eq. (20).

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. To enforce this constraint on the estimated extrinsic parameters, we introduce the following error term

\[ \text{cost}_{\text{model}}(R, t) = ||A'X'(R, t) - B||, \] (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 is widely done in the calibration [21]:

\[ \text{cost}_{\text{rep}}(R, t) = \sum_{i=0}^{N_p-1} ||q_i - \tilde{q}_i(R, t)||, \] (24)

where \( \tilde{q}_i(R, t) \) denotes \( q_i \) calculated from the estimated \( R \) and \( t \).

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

\[ f = c_{\text{model}} \ast \text{cost}_{\text{model}}(R, t) + c_{\text{rep}} \ast \text{cost}_{\text{rep}}(R, t), \] (25)

where \( c_{\text{model}} \) and \( c_{\text{rep}} \) are respectively the coefficients corresponding to \( \text{cost}_{\text{model}} \) and \( \text{cost}_{\text{rep}} \).

3.4.2 Implementation

We implemented our proposed method together with non-linear optimization as illustrated in Fig. 5. First, we estimated the initial values of extrinsic parameters. In doing so, we used a linear solution of extrinsic parameters estimated by solving Eq. (20) as the initial value. Second, we used the Levenberg-Marquardt algorithm to solve the non-linear optimization problem of Eq. (25). However, Eq. (25) 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 \( \text{cost}_{\text{rep}}(R, t)/N_p \) is larger than a threshold for reprojection error \( t_{\text{rep}} \), 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 \( \text{cost}_{\text{rep}}(R, t)/N_p < t_{\text{rep}} \) 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.

4. Experiment

This section details experiments we conducted on synthesized and real data to evaluate the quantitative and qualitative performance of our method. In this section, “linear solution” denotes extrinsic parameters estimated by solving

![Graphs showing error of rotation matrix, translation vector, and reprojection error for different noise levels.](Fig. 6)
Eq. (20) linearly 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 \((f_x, f_y, cx, cy)\); \( f_x \) and \( f_y \) represent the focal length in pixels, and \( cx \) and \( cy \) represent the 2D coordinates of the principal point. For the experiments, we set \((f_x, f_y, cx, cy)\) 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 \( O = (0, 0, 0) \). The 3D positions of the reference points were defined as \( p^0_t = (0, 0, 0), p^1_t = (-50, 0, 0), p^2_t = (50, 0, 0), p^3_t = (0, -50, 0) \), and \( p^4_t = (50, -50, 0) \). On the basis of the work reported by Snell [18], we set the center of the cornea sphere to \( S = (5, 0, 50) \), the \( r_L \) to 5.6 mm, and radius \( r \) to 7.7 mm.

We set the ground truth of \( R \) to \( I_{3 \times 3} \) and \( t \) to \((0, -50, 0)^\top \). In the optimization process, we set \( \epsilon_{\text{model}} \) and \( \epsilon_{\text{rep}} \) to 1 and \( t_{\text{rep}} \) to 2.

Throughout the experiments, we evaluated the distance between estimated parameters and their ground truth, and reprojection errors as error metrics. Here, parameters marked 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 [22] as follows:

\[
D_R = \frac{1}{\sqrt{2}} \| \log(R^T R_g) \|_F, \tag{26}
\]

\[
\log R' = \begin{bmatrix}
0 \\
\sum_{\theta} (R' - R)^T \\
\theta
\end{bmatrix}, \quad (\theta = 0),
\]

\[
\sum_{\theta} (R' - R)^T, \quad (\theta \neq 0), \tag{27}
\]

where \( \theta = \cos^{-1}(\frac{R_{31}}{\|R_{31}\|}) \). The difference between \( t \) and \( t_g \), \( D_t(t,t_g) \), is defined as root mean square:

\[
D_t = \sqrt{|t - t_g|^2 / 3}. \tag{28}
\]

The reprojection error is defined as \( D_p = cost_{\text{rep}}(R,t) / N_p \).

4.1.2 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 yielded by the proposed method with solutions estimated by a planar mirror-based approach [9], [13] (Baseline 1) and by a cornea-based method [12] (Baseline 2). The planar mirror-based approach of Takahashi et al. [9] designed for the \( N_m = 3, N_p = 3 \) configuration is widely used as a baseline method [10], [12], [24]–[26] for calibrating a camera against a reference object located outside its FOV; Takahashi et al. [13] generalized the planar mirror-based approach [9] for \( N_m > 3, N_p > 3 \). To ensure fairness in the comparison of this evaluation, we made sure that the projections of the reference points with either spherical or planar mirrors are occupied comparable pixel areas, which is about 80 \( \times \) 80 pixels in the image, as Agrawal did [10]. Note that we utilized three points \( p_i, p_j \), and \( p_k \) in following the calibration procedure of Baseline 2 [12]. The three points were the minimal configuration for this method [12], but they did not satisfy the equidistance constraint in this evaluation.

Figure 6 shows \( D_R, D_t, \) and \( D_p \) of solutions estimated by each method. In each figure, the vertical axis shows the standard deviation of noise. From the figure, we can observe that \( D_R, D_t, \) and \( D_p \) are almost zero in noiseless environments, \( i.e., \) when \( \sigma_p = 0 \). 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, whereas the precision of the linear solution degrades substantially, the precision of the non-linear solution is comparable with that of Baseline 1 [9].

The effects of noise on estimation precision seem to be smaller for Baseline 2 [12] than for the other methods. We found that the estimation errors of Baseline 2 [12] converge to certain values regardless of the noise level. In addition, Baseline 2 [12] does not estimate valid solutions even if the observation noise is zero. On the other hand, the other methods can estimate them with reasonable precisions. We consider that the effects produced by not satisfying the equidistance constraint overwhelm the effects of the observation noise.

From these results, we can see that

- Our method (linear and non-linear) estimates correct extrinsic parameters in noiseless environment,
- Our method (linear and non-linear) works even if the equidistance constraint required in Baseline 2 [12] does not hold,
- The precisions of non-linear solutions by our method are comparable with those of Baseline 1 even if the input data includes noise.

4.2 Real Data

4.2.1 Configuration

Figure 7 overviews the configuration. We used a Logicool HD Pro Webcam C920t, and captured frames had the resolution of 1920 \( \times \) 1080. As illustrated in Fig. 7, we projected a chessboard pattern on the display and captured the cornea as the reference points \( p_i, (i = 0, \cdots, 4) \). Each chess square was 125 \( \times \) 125 mm. The distance between the user’s cornea center and the display was about 300 mm. The intrinsic parameter was estimated previously by Zhang [5]. Fig-

\(^1\)The code can be downloaded from URL ‘https://compu-

tervision.github.io/takahashi2012cvpr’
Fig. 7  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 \( q_j \) is separated by about 10 - 13 pixels in the captured image.

Fig. 8  Flow in estimating ellipse parameters \( (i_L, \phi, r_{max}, r_{min}) \) from projection of limbus.

Table 2  Error metrics computed by using planar-mirror based method [9] as the ground truth.

<table>
<thead>
<tr>
<th>Method</th>
<th>( D_p )</th>
<th>( D_1 )</th>
<th>( D_p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear solution</td>
<td>0.553</td>
<td>178.896</td>
<td>14.689</td>
</tr>
<tr>
<td>Non linear solution</td>
<td>0.164</td>
<td>33.617</td>
<td>0.260</td>
</tr>
</tbody>
</table>

Fig. 9  Reference point positions estimated by proposed method (red), compared with planar-mirror based method [9], [13] (blue).

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 criterion for detecting local minimum.

5.1 Effects of Differences among Individuals

In our proposed method, we make two assumptions about the cornea model. The first is the radius of the cornea sphere \( r \). Although it has been reported that the cornea radius \( r \) does not vary greatly among individuals [18], \( r \) should not always satisfy \( r = 7.75 \text{ mm} \) strictly. The second assumption is that the average size of \( r_L \) is 5.6 mm. As with the first assumption, this assumption takes individual differences into account. To examine the effects on the performance of the proposed method, we conducted evaluations with synthesized data including noise added to them. We used the same configuration as in Sect. 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 \( \leq n_{r_L} \leq 1 \)). Note that the \( d_{LS} \) is calculated by \( d_{LS} = \sqrt{r^2 - r_L^2} \) with noised \( r \) and \( r_L \) in these evaluations.

Figures 10 and 11 show the results we obtained in trials (conducted 50 times on average) at each noise level. From these figures, we can see that the noise on these radii affect the performance of the proposed method. The tendencies are similar to those we previously reported [12]. We consider that this is because the estimation precision of \( S \) strongly affects the performance of the proposed method and that noise on \( r \) and \( r_L \) directly affects it based on Eq. (7) and \( d_{LS} = \sqrt{r^2 - r_L^2} \) as we reported previously [12].

From these discussions and the simulation results in Sect. 4, we conclude it is difficult for our proposed method to strictly estimate precise camera parameters due to such differences among individuals. However, our method has a strong advantage in its configuration, i.e., five reference points and one cornea pose without additional devices, and is suitable for casual use in which high precision is not needed.
5.2 Validity of Using Reprojection Error as Criterion for Detecting Local Minimum

In Sect. 3.4.2, 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 \( t_{\text{rep}} \) (\( 1 \leq t_{\text{rep}} \leq 10 \)). Note that we regarded the estimated \( R \) and \( t \) as matching the ground truth if \( D_R < t_{D_R} \) and \( D_t < t_{D_t} \), where \( t_{D_R} = 0.02 \) and \( t_{D_t} = 6 \). These parameters are based on the report [9] of \( \sigma_p = 0.5 \). We used the same configuration as we did in Sect. 4.1, while adding Gaussian noise with zero mean and standard deviation \( \sigma_p = 0.5 \) to \( q_i \).

Figure 12 shows the rate of matching the ground truth for each \( t_{\text{rep}} \) over 50 trials. From the figure, we can observe that all the estimated solutions converge to the ground truth when \( t_{\text{rep}} \leq 3 \). These simulation results demonstrate the validity of using the reprojection error as the metric for detecting the local minimum in practice. On the basis of these results, we define \( t_{\text{rep}} = 2 \) in Sect. 4. However, the relationship between \( \sigma_p \) and \( t_{\text{rep}} \) has not been proven theoretically. This is a subject for our future work in this study.

6. Conclusion

In this paper, we proposed a new method that calibrates a camera to a 3D reference object via cornea reflection with a simpler configuration than previous methods. 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. On the basis of these ideas, we derived a linear equation and obtained a closed-form solution. In evaluations, we showed that our method worked properly with both synthesized and real data with the simpler configuration.

References


Akira Kojima    Manager, Media Innovation Business Department, NTT TechnoCross Corporation. He received the B.E. and M.E. degree in Mathematical Engineering and Information Physics from the University of Tokyo in 1988 and 1990, respectively. Since joining NTT Group in 1990, he has been engaged in research and development on video database, digital library, multimedia information retrieval, video surveillance and high-reality visual communication. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE), the Institute of Image Electronics Engineers of Japan (IIEEJ) and ACM.

Hideaki Kimata    received the B.E. and M.E. degrees in applied physics in 1993 and 1995 and Ph.D. degree in electrical engineering in 2006 respectively from Nagoya University, Nagoya, Japan. He joined Nippon Telegraph and Telephone Corporation (NTT) in 1995 and has been engaged in R&D of video processing of coding, realistic communication, computer vision, and recognition based on machine learning (deep learning). He is currently a Senior Research Engineer, Supervisor, in NTT Media Intelligence Laboratories. He is a member of the Institute of Electronics, Information and Communication Engineers of Japan (IEICE), and a chief examiner of special interest group on audio visual and multimedia information processing of Information Processing Society of Japan (IPSJ).