Extending the Measurement Angle of a Gaze Estimation Method Using an Eye Model Expressed by a Revolution about the Optical Axis of the Eye

Nagamatsu, Takashi
Hiroe, Mamoru
Arai, Hisashi

(Citation)
IEICE Transactions on Information and Systems,E104.D(5):729-740

(Issue Date)
2021-05

(Resource Type)
journal article

(Version)
Version of Record

(URL)
https://hdl.handle.net/20.500.14094/90008326

PAPER

Extending the Measurement Angle of a Gaze Estimation Method Using an Eye Model Expressed by a Revolution about the Optical Axis of the Eye

Takashi NAGAMATSU[†a)], ***Member***, **Mamoru HIROE**[†b)], ***and*** **Hisashi ARAI**[†*], ***Nonmembers***

SUMMARY An eye model expressed by a revolution about the optical axis of the eye is one of the most accurate models for use in a 3D gaze estimation method. The measurement range of the previous gaze estimation method that uses two cameras based on the eye model is limited by the larger of the two angles between the gaze and the optical axes of two cameras. The previous method cannot calculate the gaze when exceeding a certain limit of the rotation angle of the eye. In this paper, we show the characteristics of reflections on the surface of the eye from two light sources, when the eye rotates. Then, we propose a method that extends the rotation angle of the eye for a 3D gaze estimation based on this model. The proposed method uses reflections that were not used in the previous method. We developed an experimental gaze tracking system for a wide projector screen and experimentally validated the proposed method with 20 participants. The result shows that the proposed method can measure the gaze of more number of people with increased accuracy compared with the previous method.

key words: *aspherical eye model, eye tracking, gaze tracking*

1. Introduction

Video-based gaze trackers are widely used in scientific studies, human–computer interactions, etc. [1]–[3]. There are many types of gaze tracking methods [4]. These methods can be roughly categorized into 2D regression-based, 3D model-based, and appearance-based gaze estimation methods. Among these, the 3D model-based gaze estimation method has advantages in terms of accuracy and tolerance of head movement.

3D model-based gaze estimation methods typically use a spherical model of the front surface of the cornea [5]–[10]. The point of gaze (POG) is estimated using the reflections of point light sources on the corneal surface. However, Shih and Liu [5] and Guestrin and Eizenman [6] pointed out that a spherical model may not be suitable for modeling the boundary region of the cornea. In order to describe the corneal region near the sclera boundary more accurately than a spherical model of the cornea can, Nagamatsu et al. [11] proposed a method of gaze estimation in which the whole eye is modeled as a surface of revolution about the optical axis of the eye. This method allows a single point calibration and head movement. They experimentally showed that the method was more accurate than the method based on a spherical model of the cornea [12]. As this model is very complicated, the calculation method for this model has not been improved since the model was initially proposed.

In this paper, we show an improvement of Nagamatsu's method by extending the gaze tracking range of the direction. Hereinafter, we call Nagamatsu et al.'s method [11] "the previous method."

The main contributions of this paper are to show the behavior of the reflections from two light sources when the eye rotates, and to achieve gaze estimation using the reflections from the light sources that were not used in the previous method, to realize a wide-angle gaze estimation.

We denote 3D vectors in boldface without superscripts. The uppercase vector indicates the position vector, and the lowercase vector indicates the direction vector. Of the uppercase vectors, the vector with (') indicates the point on the image sensor, and the vector with (") indicates the point on the surface of the eyeball. Vectors that have a superscript 0 or 1 on the left-hand side are vectors in the image coordinate system of camera 0 or 1, respectively, which are 2D vectors.

2. Related Work

Gaze estimation methods can be categorized into 2D regression-based, 3D model-based gaze estimation, and appearance-based gaze estimation methods.

2D regression-based methods use eye features such as the pupil center and corneal reflection by external light sources for calculating gaze by regression [2]. 2D regression-based methods need several gaze points for calibration. These methods are widely used in commercial eye trackers.

3D model-based methods detect eye features, and then calculate the direction of the eye geometrically. The advantages of the 3D model-based method are a reduction in the number of required calibration points and that the user can move their head. One point calibration is achieved by the 3D model-based approach with a spherical model of the cornea, external light sources, and a calibrated stereo pair of cameras [5]–[9]. While early 3D model-based methods needed high-resolution cameras and infrared light sources, recent approaches only need a single webcam [13]. More recent

Manuscript received April 1, 2020.
Manuscript revised November 14, 2020.
Manuscript publicized February 4, 2021.

[†]The authors are with the Graduate School of Maritime Sciences, Kobe University, Kobe-shi, 658–0022 Japan.
[*]Presently, with Daifuku Co., Ltd.
a) E-mail: nagamatu@kobe-u.ac.jp
b) E-mail: hiroe@stu.kobe-u.ac.jp

DOI: 10.1587/transinf.2020EDP7072

work uses machine learning to improve the accuracy of feature detection [14], [15].

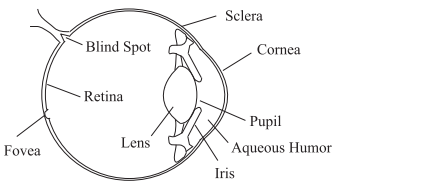
Appearance-based methods directly learn a mapping function from 2D input images to POG using machine learning [16]. Recent works use both large-scale training data and deep learning to improve the gaze estimation accuracy [17]. The disadvantage of appearance-based methods is that their accuracy is mostly lower than for 3D model-based methods with infrared light sources [18]. For a robust performance they need large training image databases [18].

Because we prioritize accuracy and head movement, we focus on 3D model-based methods with light sources. For a 3D model-based approach, the shape of the eye model used is important. The structure of an eye is shown in Fig. 1 (a). Light is focused by the cornea and lens onto the retina. Part of the light is reflected from the smooth surface of the cornea which behaves like a mirror. The edge of the outer surface of the cornea is smoothly connected to the sclera, which does not have a smooth surface. A typical eye model used in a model-based approach [5]–[10] is shown in Fig. 1 (b). It is based on Emsley's reduced eye [19]. The model consists of a combination of large and small spheres; the cornea is modeled as a small sphere. This model models only the front surface of the cornea accurately. Several methods model the corneal surface as an ellipsoid [20], [21] because the shape of a normal adult cornea is much closer to an ellipsoid than to a sphere. However, gaze estimation methods that use a spherical or ellipsoidal model of the front surface of the cornea cannot be used to model the corneal region near the sclera boundary. When a user looks at the peripheral area, a reflection is generated on the peripheral area of the cornea, which is no longer modeled accurately; therefore, the estimation becomes poor.

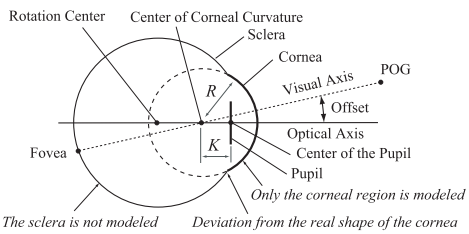
Nagamatsu et al. [11] proposed an eye model for gaze estimation that is more accurate than the spherical or ellipsoidal model of the cornea, as shown in Fig. 1 (c). The shape of their eye model is that of the surface of a revolution about the optical axis of the eye. Because Nagamatsu et al.'s gaze estimation method (the previous method) is based on a model that is a general surface of revolution about the optical axis of the eye, it is not necessary to express the shape of the model in an equation. This is an advantage over all methods that use spherical and ellipsoidal models with mathematical expressions. The model includes various shapes; regardless of the eye shape, the gaze can be estimated. Therefore, the model suits the eyes of more people than the spherical or ellipsoidal models.

Furthermore, Model and Eizenman [22] pointed out that stereo camera systems can estimate the POG accurately. However, their gaze tracking range is limited. The range of the gaze direction is limited by the larger of the two angles between the gaze and the optical axes of two cameras. The larger angle is typically 30°. Therefore, the tracking range of the previous method is limited.

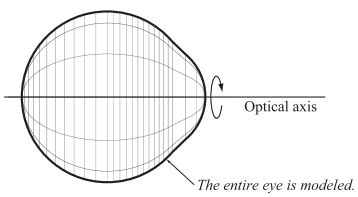
Thus, we study a method that extends the gaze tracking range using the same eye model as proposed by Nagamatsu et al.



(a) Structure of the eye



(b) Spherical corneal model



(c) Model expressed by surface of revolution about the optical axis

Fig. 1 Structure of eye and the eye models for 3D model-based methods.

3. Characteristics of Reflection from Two Light Sources When the Eye Rotates, and the Limitation of the Previous Method

In this section, we explain the previous method [11], characteristics of reflection from two light sources when the eye rotates, and the limitations of the previous method.

3.1 Calculation of the Previous Method

To calculate the optical axis of the eye using the previous method [11], a special arrangement of cameras and light sources is necessitated (Fig. 2). The position of each light source is intended to be the same as the nodal point of the camera. The two cameras capture the images of the same

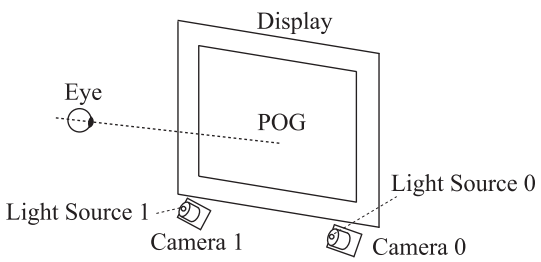


Fig. 2 Arrangement of cameras and light sources for estimating POG.

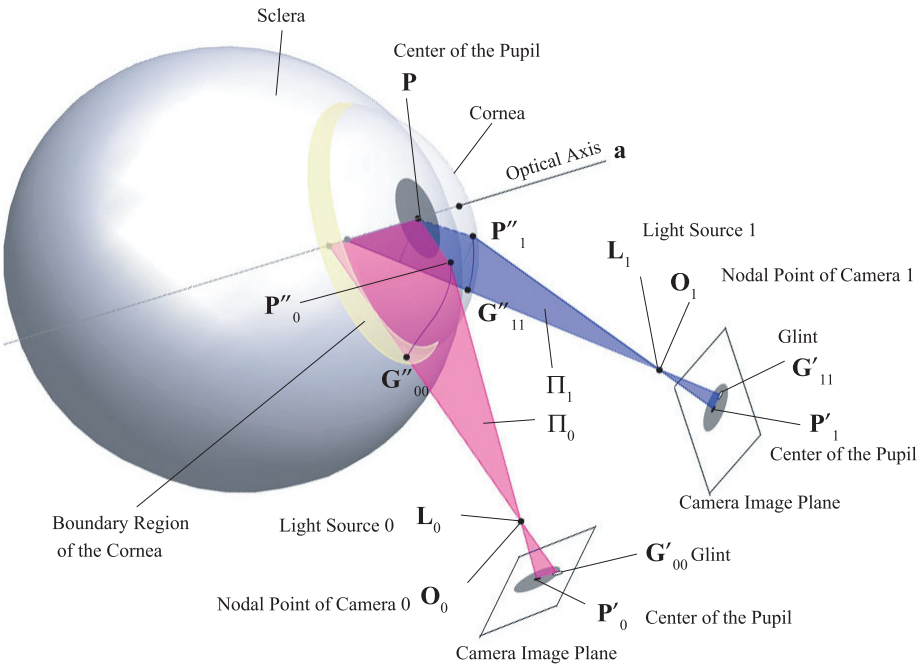


Fig. 3 Ray tracing diagram based on the previous method.

eye from different angles. In this study, we use the same arrangement of hardware, i.e., two calibrated cameras with a light source attached to each one. This special arrangement allows the calculation of the proposed method described in Sect. 4 using this eye model.

Figure 3 shows a ray tracing diagram for the calculation of the previous method. A ray from a light source ($\mathbf{L}_j$; $j = 0, 1$) is reflected back at a point ($\mathbf{G}''_{jj}$) on the corneal surface along its incident path. It passes through the nodal point of camera ($\mathbf{O}_j$) and intersects the camera image plane at a point $\mathbf{G}'_{jj}$. A ray originating from the center of the pupil ($\mathbf{P}$) is refracted at a point $\mathbf{P}''_j$ on the corneal surface, passes through the nodal point of camera ($\mathbf{O}_j$), and intersects the camera image plane at point $\mathbf{P}'_j$. These rays are on the same plane that is expressed in magenta ($j = 0$) or blue ($j = 1$) in Fig. 3. The optical axis of the eye is calculated by the intersection of two planes (magenta and blue). The visual axis of the eye can be calculated after user calibration in which the user gazes at a single point.

3.2 Characteristics of the Reflection from Two Light Sources When the Eye Rotates

Figure 4 shows ray tracing diagrams (upper images) and schematic eye images (bottom images) captured by cameras 0 and 1 when the eye rotates to three positions. The upper images, i.e., the wireframe models, are the top view of the eye, and the blue lines indicate the rays of reflections. The gray ellipses in the bottom images indicate the cornea.

Figure 4 (a) shows the eye directed forward. A ray from light source 0 ($\mathbf{L}_0$) is reflected back at a point $\mathbf{G}''_{00}$ on the corneal surface along its incident path. It passes through the nodal point of camera 0 ($\mathbf{O}_0$) and intersects the camera image plane at a point $\mathbf{G}'_{00}$. In addition, a ray from $\mathbf{L}_0$ is reflected at a point $\mathbf{G}''_{01}$ on the corneal surface and goes to camera 1. It passes through the nodal point of camera 1 ($\mathbf{O}_1$) and intersects the camera image plane at a point $\mathbf{G}'_{10}$. Similarly, a ray from $\mathbf{L}_1$ produces glints $\mathbf{G}'_{11}$ and $\mathbf{G}'_{01}$. Figure 5 shows the images captured by cameras 0 and 1 when the user looks forward. The reflections of two light sources (glints) are clear and appear as small white dots. The glints from light sources 0 and 1 captured by camera 0 are indicated by ${}^0\mathbf{G}_0$ and ${}^0\mathbf{G}_1$, respectively. The superscripts indicate the coordinate system of the camera number. The subscripts indicate the light source number that produces the glint. The glints from light sources 0 and 1 captured by camera 1 are indicated by ${}^1\mathbf{G}_0$ and ${}^1\mathbf{G}_1$, respectively.

Figure 4 (b) shows the eye rotated to the left. As $\mathbf{G}''_{00}$ is on the sclera, the glint ${}^0\mathbf{G}_0$ in the image of camera 0 is scattered. Figure 6 shows the images when the user looks left. On the image of camera 0, the glint from light source 1 (${}^0\mathbf{G}_1$) is shown on the cornea, but the glint from light source 0 (${}^0\mathbf{G}_0$) is shown on the sclera and is scattered. Therefore, the position of ${}^0\mathbf{G}_0$ cannot be detected accurately. On the image of camera 1, the glints from light source 0 (${}^1\mathbf{G}_0$) and light source 1 (${}^1\mathbf{G}_1$) are shown on the cornea.

Observing the produced reflections carefully, the rays that produce glints ${}^0\mathbf{G}_1$ and ${}^1\mathbf{G}_0$ are reflected at the same location ($\mathbf{G}''_{10} = \mathbf{G}''_{01}$); therefore, when the eye rotates further left, both glints become to reflect from the sclera simultaneously. Figure 4 (c) shows the eye rotated further left. The glints ${}^0\mathbf{G}_0$, ${}^0\mathbf{G}_1$, and ${}^1\mathbf{G}_0$ are scattered. Figure 7 shows the images when the user looks further left. On the image of camera 0, the glints from light source 0 (${}^0\mathbf{G}_0$) and light source 1 (${}^0\mathbf{G}_1$) are shown on the sclera, and they are scattered and merged. On the image of camera 1, the glint from

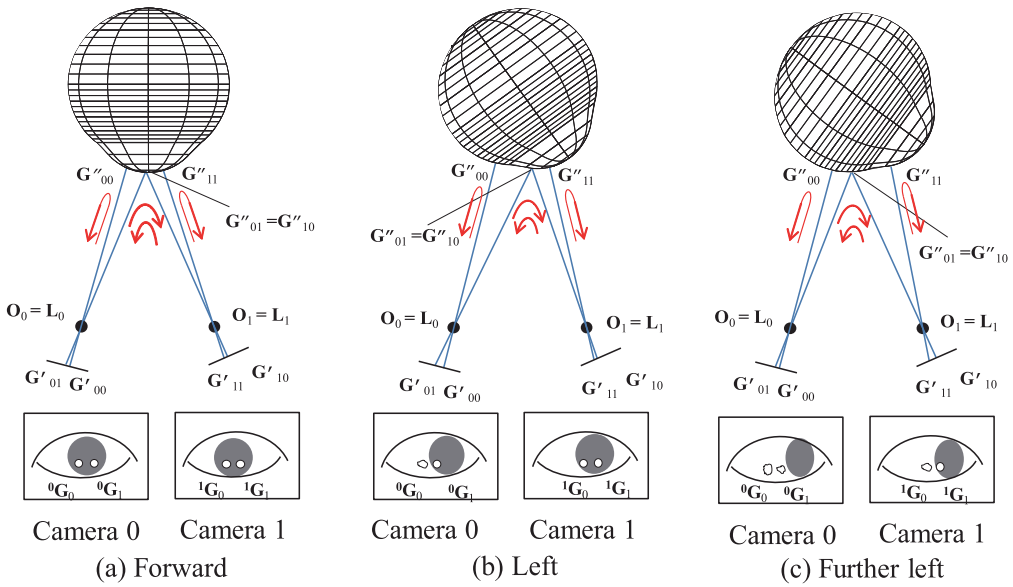


Fig. 4 Ray tracing diagrams for three cases of eye rotations.

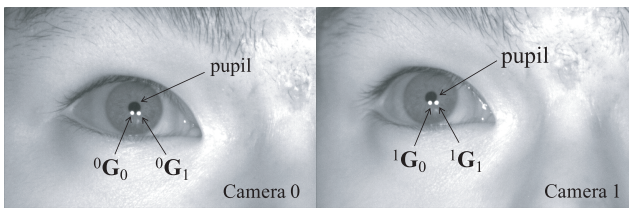


Fig. 5 Case (a): eye image when the gaze direction is toward the front.

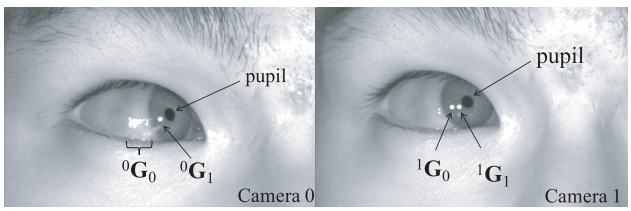


Fig. 6 Case (b): eye image when the gaze direction is to the left.

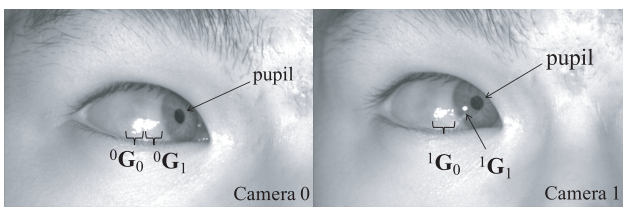


Fig. 7 Case (c): eye image when the gaze direction is further to the left.

light source 1 is shown on the cornea, but the glint from light source 0 ($^1\mathbf{G}_0$) is shown on the sclera and is scattered.

3.3 Limitation of the Previous Method

In the previous method, when the optical axis of the eye is estimated, two glints ($^0\mathbf{G}_0$ and $^1\mathbf{G}_1$) are required. In case (a) as shown in Fig. 5, the previous method is effective. In cases (b) and (c) as shown in Figs. 6 and 7, $^0\mathbf{G}_0$ is unclear; therefore, the previous method may have large errors.

Thus, the problem in the previous method is that when the eye rotates beyond a certain limit, the light becomes reflected on the sclera; therefore, the reflections are scattered. The gaze estimation then becomes inaccurate or impossible.

In this paper, we herein propose a calculation method for case (b). The improved point from [11] is that the proposed method can calculate the gaze for case (b), which realizes the extension of gaze direction. For case (c), where we can use only one clear reflection, we do not have sufficient information for calculating the gazes; therefore, this case will not be discussed herein but will be investigated in future work.

4. Proposed Novel Gaze Estimation Method

4.1 Estimation of the Optical Axis of the Eye

In case (b) as shown in Fig. 6, when the eye rotates to the left, $^0\mathbf{G}_0$ becomes scattered and is difficult to detect accurately. Therefore, we use clear glints $^0\mathbf{G}_1$ and $^1\mathbf{G}_0$ instead of $^0\mathbf{G}_0$ in the calculation, although glints $^0\mathbf{G}_1$ and $^1\mathbf{G}_0$ were considered as disturbances for image processing in the previous method. Meanwhile, when the eye rotates to the right, $^1\mathbf{G}_1$ is difficult to detect accurately. It is therefore treated in the same manner.

Figure 8 shows a ray tracing diagram for calculating gaze using $\mathbf{G}'_{01}$, $\mathbf{G}'_{10}$, and $\mathbf{G}'_{11}$, which are calculated from $^0\mathbf{G}_1$, $^1\mathbf{G}_0$, and $^1\mathbf{G}_1$ using camera parameters, respectively. First, the blue plane (Π_1) is determined by the following equation:

$$\{(\mathbf{P}'_1 - \mathbf{O}_1) \times (\mathbf{G}'_{11} - \mathbf{O}_1)\} \cdot (\mathbf{X} - \mathbf{O}_1) = 0, \quad (1)$$

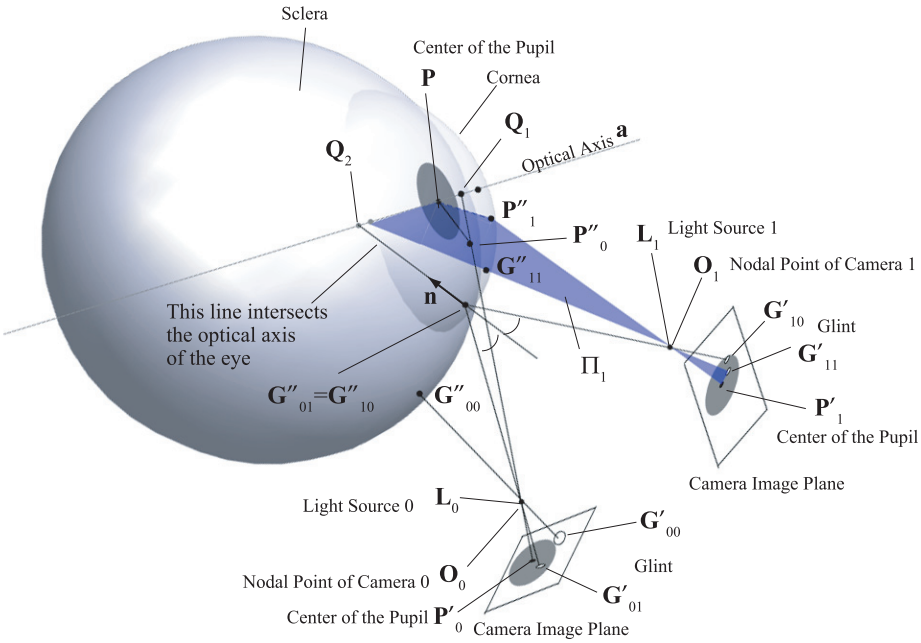


Fig. 8 Ray tracing diagram based on the proposed method.

where $\mathbf{O}_1$ is the nodal point of the camera 1 and is determined by camera calibration, $\mathbf{P}'_1$ is the center of the pupil on the image plane of camera 1, and $\mathbf{G}'_{11}$ is the glint from the light source $\mathbf{L}_1$ on the image plane. The pupil center $\mathbf{P}$ is modeled on the optical axis of the eye. As the blue plane includes $\mathbf{P}$ and the reflection surface is modeled as the surface of revolution about the optical axis of the eye, the blue plane includes the optical axis of the eye.

Next, we wish to obtain two points ($\mathbf{Q}_1$ and $\mathbf{Q}_2$) on the blue plane that are included in the optical axis of the eye. For the first point ($\mathbf{Q}_1$), the line expressed by Eq. (2) is a line that goes through the virtual pupil and intersects the optical axis of the eye:

$$\mathbf{X} = \mathbf{P}'_0 + s\left(\mathbf{O}_0 - \mathbf{P}'_0\right), \tag{2}$$

where $\mathbf{O}_0$ is the nodal point of camera 0, $\mathbf{P}'_0$ is the center of the pupil on the image sensor of camera 0, and s is a parameter. Subsequently, as the blue plane includes the optical axis of the eye, we determine a point on the optical axis of the eye by the intersection of the blue plane (Eq. (1)) and the line (Eq. (2))

For the other point ($\mathbf{Q}_2$), the calculation method is as follows. First, the reflection point on the cornea ($\mathbf{G}''_{01} = \mathbf{G}''_{10}$) is calculated by the intersection of the two following lines:

$$\mathbf{X} = \mathbf{O}_0 + t_{01}\left(\mathbf{O}_0 - \mathbf{G}'_{01}\right), \tag{3}$$

$$\mathbf{X} = \mathbf{O}_1 + t_{10}\left(\mathbf{O}_1 - \mathbf{G}'_{10}\right), \tag{4}$$

where t_{01} and t_{10} are parameters.

As we use the model of the surface of revolution about the optical axis, the normal vector of the corneal surface at $\mathbf{G}''_{01}$ intersects the optical axis of the eye. The normal vector at $\mathbf{G}''_{01}$ on the corneal surface $\mathbf{n}$ is calculated using

$$\mathbf{n} = \frac{\mathbf{O}_0 - \mathbf{G}'_{01}}{\|\mathbf{O}_0 - \mathbf{G}'_{01}\|} + \frac{\mathbf{O}_1 - \mathbf{G}'_{10}}{\|\mathbf{O}_1 - \mathbf{G}'_{10}\|}. \tag{5}$$

Therefore, the line expressed by Eq. (6) intersects the optical axis of the eye:

$$\mathbf{X} = \mathbf{G}''_{01} + t\mathbf{n}, \tag{6}$$

where t is a parameter. Subsequently, we determine a point on the optical axis of the eye by the intersection of the blue plane (Eq. (1)) and the line (Eq. (6))

Thus, the line connecting these two points is the optical axis of the eye.

4.2 User Calibration and Estimation of POG

The user calibration method is the same as that in the previous method [11], which is a one-point calibration. Here, we briefly describe it.

4.2.1 User Calibration

The following user-dependent parameters are to be estimated: R, the radius of the corneal curvature around the central portion of the outer surface of the cornea; K, the distance between the center of the corneal curvature and the pupil; and the offset between the optical and visual axes of the eye. This estimation is considered as user calibration.

In the user-calibration process, the user is instructed to gaze at a single point (calibration point) at a known position. The position of the calibration point is selected such that the light from the camera is reflected from the central portion of the corneal surface, which is approximated as a sphere.

For a user gazing at a known position, R and K are estimated by the method described in [11], and the offset between the optical and visual axes of the eye is calculated using the method reported by Nagamatsu et al. [23]. The detail of the user-calibration process is described in Appendix A.

4.2.2 Estimation of the Visual Axis of the Eye and POG

After user calibration, the user moves his/her eyes freely. The unit direction vector of the optical axis of the eye can be calculated using the method described in Sect. 4.1. The position of the center of the corneal curvature $\mathbf{C}$ is determined on the optical axis of the eye using R and K. The unit direction vector of the visual axis of the eye is determined from the optical axis of the eye and the offset between the optical and visual axes of the eye using the method reported by Nagamatsu et al. [23]. The intersection point between the visual axis of the eye and the object (e.g., a display) is the POG. The detail of the estimation of the visual axis of the eye and POG after calibration are described in Appendix B.

5. Implementation of Gaze Estimation System

5.1 Producing Dark Pupil Images

The arrangement of cameras and light sources suggests that the system produces bright pupils. A bright pupil is created by the light reflected from the retina when the light source is set to be coaxial with the optical path. However, the intensity of the bright pupil varies between individuals and is affected by changes in head position or pose [4]. Therefore, we decided to implement our system such that it produced dark pupils.

We set infrared (IR) light-emitting diodes (LEDs) sufficiently far from the nodal point of the camera such that they produced dark pupils; eight IR LEDs were attached to each camera such that their centers coincided with the center of the camera, as shown in Fig. 9. The wavelength of IR LED was 830 nm. Because the cornea behaves as a convex mirror, the reflected images of the eight LEDs resembles one small light source at the image resolution of the implementation described in this paper. The glint produced by the eight LEDs looks like a circle. The center of gravity of the glint area is determined as the glint position The position of the center of eight LEDs is the same as the nodal point of the camera to which the LEDs are attached, and is determined by camera calibration. We created the arrangement as shown in Fig. 2 virtually.

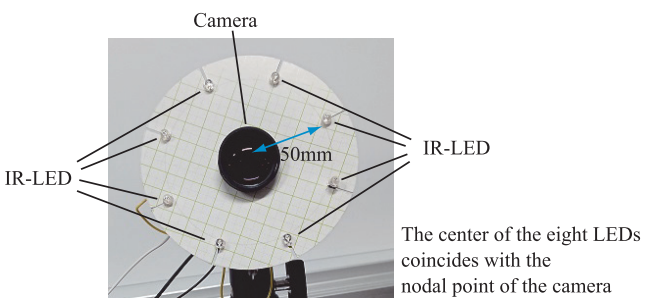


Fig. 9 Camera for producing dark pupils.

5.2 Identifying the Glints

For calculating the optical axis of the eye, we must identify the glints on the cornea that were produced by the light sources. Three glints are required in proposed method for the calculation.

We selected three glints as follows in this implementation stage. When the user looks to the left, as shown in Fig. 10, glints ${}^0\mathbf{G}_1$ and ${}^1\mathbf{G}_0$ are produced by reflecting on the same position of the corneal surface. As the distances between the cameras and the eye are similar, the distance between ${}^0\mathbf{G}_1$ and ${}^0\mathbf{P}$ is similar to the distance between ${}^1\mathbf{G}_0$ and ${}^1\mathbf{P}$. We can obtain three reflections by selecting three reflections near the pupil center (${}^0\mathbf{P}$ or ${}^1\mathbf{P}$). In the case depicted in Fig. 10, we selected ${}^0\mathbf{G}_1$, ${}^1\mathbf{G}_0$, and ${}^1\mathbf{G}_1$

5.3 System Overview

We developed a system for estimating the POG on a projector screen to evaluate a wide range of eye rotation (Fig. 11). This system comprises two monochrome USB2.0 digital

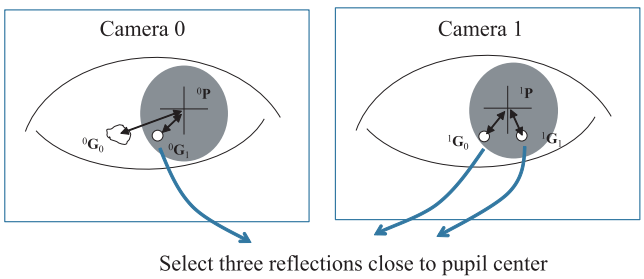


Fig. 10 Identifying the glints.

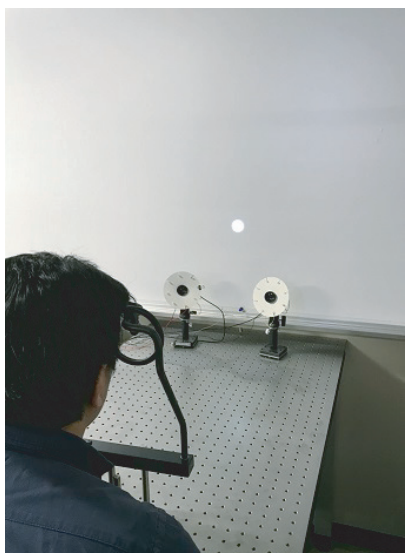


Fig. 11 Developed system.

cameras (Firefly MV, Point Grey Research Inc.) with IR LEDs, a 23″ liquid crystal display for verifying the system status, a digital light processing projector, and a Windows-based personal computer operating Windows 7. The size of the projected area was 1920×1080 mm^2. An ordinary computer display is too small to verify whether the gaze tracking system is capable of wide-angle measurement; therefore, we used a projector to verify the POGs.

The software was developed in the C++ language using OpenCV 2.4 [24]. Each camera is equipped with a 1/3″ CMOS image sensor that has a resolution of 752×480 pixels. A 50 mm lens and a visible light cut filter were attached to each camera in addition to the eight IR LEDs. These cameras were positioned in front of the user. Two cameras captured only the area around the right eye as shown in Fig. 5. The intrinsic and extrinsic parameters of the cameras were determined before the system was set up.

6. Evaluation

6.1 Experiment

6.1.1 Method

The experiment involved 20 adult participants (16 men, 4 women) who did not use glasses. Four participants used soft contact lenses. Each participant was located approximately 930 mm from the screen, as shown in Fig. 12. The experiment was performed for the right eye.

After user calibration, in which the participant gazed at a single point, the participants were asked to stare at 42 fiducial points on the screen, which are shown in Fig. 13. These fiducial points were displayed by a projector one by one. The size of the screen was 1920×1080 mm^2.

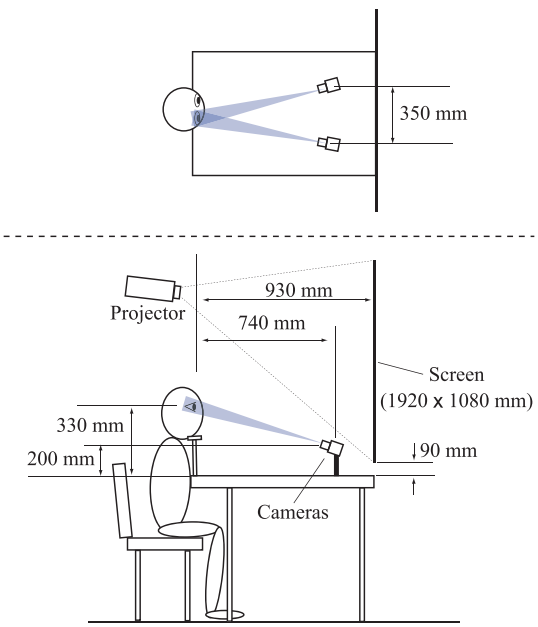


Fig. 12 Experimental setup.

User-dependent parameters (R, K, offset between optical and visual axes of the eye) were saved when the user calibration was completed. After user calibration, the images captured by cameras 0 and 1 while the participants were gazing at 42 points were saved.

The gazes were calculated by three methods: the spherical method [10], the previous method [11], and the proposed method. Here, the spherical method is the method proposed by Chen et al. [10], which is used as a baseline. The previous method is the method proposed by Nagamatsu et al. [11].

6.1.2 Results

Figure 14 shows 42 graphs that are arranged corresponding to the position of the fiducial points shown in Fig. 13, e.g., the graph of No. 1 in Fig. 14 is the graph when the participants were gazing at the fiducial point No. 1 in Fig. 13. Each graph illustrates the accuracy using all three methods (spherical, previous, and proposed) for all 20 participants. The horizontal axis indicates the accuracy in degrees and the vertical axis indicates the cumulative number of participants. The blue and red lines indicate the results using the previous and proposed methods, respectively. The gray-filled area graphs indicate the results using the spherical method. The graphs imply that a method that reaches 20 quickly shows good accuracy.

In Fig. 14, for points 3, 5, 10, 11, 17, 18, 25, and 32, all three methods yielded similar results. This implies that all three methods performed similarly for the central portion of the screen. However, for point 4, the spherical and previous methods performed slightly better than the proposed method. For the other points, the proposed method demonstrated better results than the spherical and previous methods.

Because the central part has a similar accuracy and the number of people measured with high accuracy is large in the peripheral area, the proposed method produced better results than the other two methods.

6.2 Discussion

There are a few studies that deal with an extention of gaze

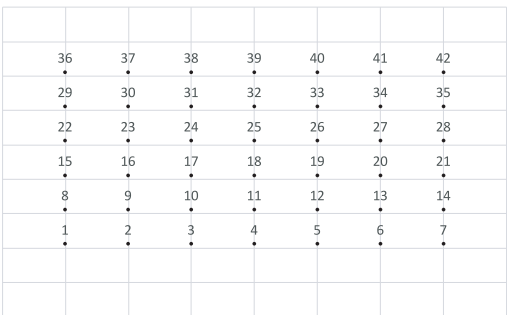


Fig. 13 Positions of fiducial points at which participants intentionally gazed.

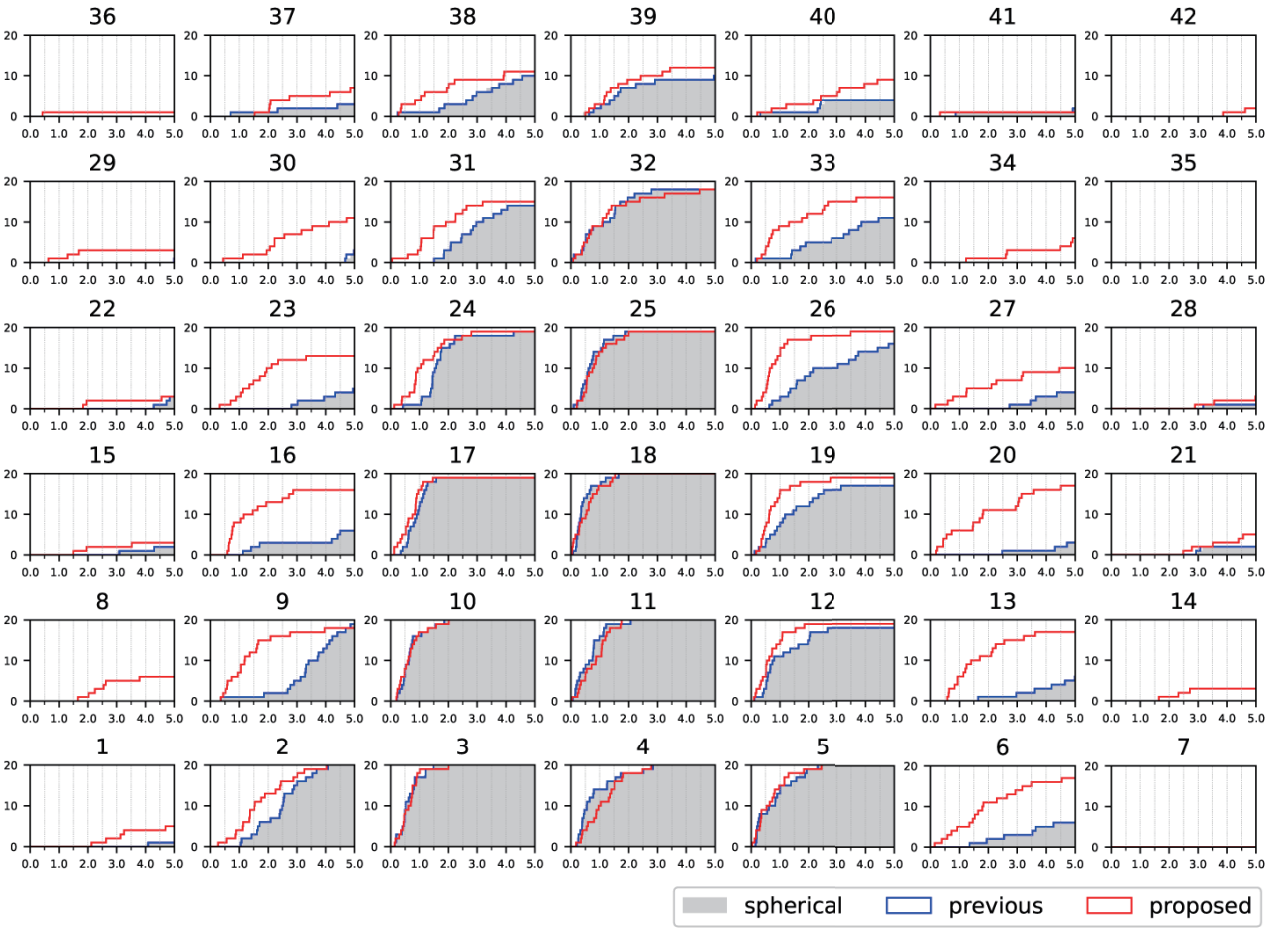


Fig. 14 Results: 42 graphs that are arranged corresponding to the position of the fiducial points shown in Fig. 13. Each graph represents the cumulative number of participants (vertical axis) against the accuracy in degrees (horizontal axis).

measurement range using a 3D model-based gaze estimation method. Model and Eizenman [22] proposed a method for extending the tracking range of 3D gaze tracking using stereo cameras. Their method switches between a one-camera system and a two-camera system. User-dependent parameters must be estimated by the two-camera system for the central tracking area before the one-camera system is used for the peripheral tracking area. Meanwhile, our proposed method can extend the tracking range using a two-camera system.

Next, we compare the accuracy of the proposed method with those of other methods. In order to fairly compare the accuracy, the measurement range of the proposed method should be limited to a computer display size. If it is limited to data for the central portion of the screen, i.e., points 3, 4, 5, 10, 11, 12, 17, 18, and 19, which correspond to a 19-inch monitor that is 600 mm from a display, the accuracies of the spherical, previous, and proposed methods are 1.47, 1,48, and 1.06, respectively.

Table 1 shows an accuracy comparison of the proposed method with other gaze estimation methods by computer display size. The values of accuracy of the top three rows and seventh row are taken directly from each paper. The accuracy of the first row is the value obtained from Blignaut's study [25]. This value is the best result for a 9-point calibration configuration. It can achieve an accuracy of less than one degree but is not robust to some head motions. The 3D model-based method by Guestrin and Eizenman [9] achieved an accuracy of 0.6–1.0 degrees for three participants, as shown in the second row. It achieves good accuracy only when light is reflected on the central portion of the cornea. The method by Beymer and Flickner [21] uses the ellipsoidal model. They achieved an accuracy of 0.6 degrees, as shown in the third row. It needs nine-point calibration. However, the experiment was conducted for only one participant.

The accuracy of the spherical, previous, and proposed method are shown in the fourth, fifth, and sixth row, respectively. The accuracy of the proposed method is comparable to those of other methods in terms of the computer display area. Additionally, the proposed method has the ability of measuring over a wide range, as shown in Fig. 14.

By contrast, state-of-the-art appearance-based methods learn direct mapping from intensity images or extracted eye features to gaze directions. This is potentially applicable to relatively low-resolution images. Palmero [26] achieved an

Table 1 Comparison with other methods by computer display size.

Reference	Set up	Method (model)	Accuracy (degree)	Number of Participants	Remarks
Blignaut [25]	1 LED and 1 camera	2D regression-based	0.87	26	9-point calibration, not robust to head motion
Guestrin and Eizenman [9]	4 LEDs and 2 cameras	3D model-based (spherical model of cornea)	0.6–1.0	3	1-point calibration, peripheral area of cornea is not modeld
Beymer and Flickner [21]	2 LEDs and 4 cameras	3D model-based (ellipsoidal model of cornea)	0.6	1	9-point calibration, peripheral area of cornea is not modeld
Spherical	2 LEDs and 2 cameras	3D model-based (spherical model of cornea)	1.473	20	1 point calibration
Previous	2 LEDs and 2 cameras	3D model-based (model of revolution about optical axis of eye)	1.475	20	1 point calibration, entire eye is modeled,
Proposed	2 LEDs and 2 cameras	3D model-based (model of revolution about optical axis of the eye)	1.06	20	1 point calibration, entire eye is modeled, 3 clear glints are used.
Palmero et al. [26]	1 camera	Appearance-based	3.38	16	using a multimodal recurrent CNN

accuracy of 3.38 degrees using a multimodal recurrent convolutional neural network (CNN). Although an appearance-based method has fewer hardware restrictions, a model-based method is better in terms of accuracy.

In our implementation, we used heuristics to determine the reflected light that corresponded to each LED array. In this study, the heuristics yielded good results; however, if the user moved considerably, the heuristics may not be available. Therefore, in the next implementation, we will use LED flashing to identify the reflected light.

7. Conclusion

We showed the characteristics of the reflections on the surface of the eye from two light sources for when the eye rotates, which had not been considered in detail previously. We proposed a new method that extends the measurement angle of the gaze estimation method and uses an eye model of a revolution about the optical axis of the eye. The proposed method uses reflections that were not used in the previous method. We developed an experimental gaze tracking system for a wide projector screen. Furthermore, we experimentally evaluated the validity of the proposed method with 20 participants. The results demonstrated that the proposed method can measure wider gazes than previous methods.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 16H02860 and 20H04229.

References

[1] R.J.K. Jacob, "The use of eye movements in human-computer interaction techniques: What you look at is what you get," ACM Trans. Information Systems, vol.9, no.2, pp.152–169, 1991.
[2] C.H. Morimoto and M.R.M. Mimica, "Eye gaze tracking techniques for interactive applications," Computer Vision and Image Understanding, vol.98, no.1, pp.4–24, 2005.
[3] A.T. Duchowski, Eye Tracking Methodology: Theory and Practice, 3rd ed., Springer-Verlag, 2017.
[4] D.W. Hansen and Q. Ji, "In the eye of the beholder: A survey of models for eyes and gaze," IEEE Trans. Pattern Anal. Mach. Intell., vol.32, no.3, pp.478–500, 2010.
[5] S.-W. Shih and J. Liu, "A novel approach to 3-D gaze tracking using stereo cameras," IEEE Trans. Syst. Man Cybern. Part B, vol.34, no.1, pp.234–245, 2004.
[6] E.D. Guestrin and M. Eizenman, "General theory of remote gaze estimation using the pupil center and corneal reflections," IEEE Trans. Biomed. Eng., vol.53, no.6, pp.1124–1133, 2006.
[7] A. Villanueva and R. Cabeza, "A novel gaze estimation system with one calibration point," IEEE Trans. Syst. Man Cybern. Part B, vol.38, no.4, pp.1123–1138, 2008.
[8] T. Nagamatsu, J. Kamahara, T. Iko, and N. Tanaka, "One-point calibration gaze tracking based on eyeball kinematics using stereo cameras," Proc. 2008 Symposium on Eye Tracking Research & Applications, pp.95–98, 2008.
[9] E.D. Guestrin and M. Eizenman, "Remote point-of-gaze estimation with free head movements requiring a single-point calibration," Proc. 29th Annual International Conference of the IEEE EMBS, pp.4556–4560, 2007.
[10] J. Chen, Y. Tong, W. Gray, and Q. Ji, "A robust 3D eye gaze tracking system using noise reduction," Proc. 2008 Symposium on Eye Tracking Research & Applications, pp.189–196, 2008.
[11] T. Nagamatsu, Y. Iwamoto, J. Kamahara, N. Tanaka, and M. Yamamoto, "Gaze estimation method based on an aspherical model of the cornea: Surface of revolution about the optical axis of the eye," Proc. 2010 Symposium on Eye-Tracking Research & Applications, pp.255–258, 2010.
[12] T. Nagamatsu, Y. Iwamoto, R. Sugano, J. Kamahara, N. Tanaka, and M. Yamamoto, "Gaze estimation method involving corneal reflection-based modeling of the eye as a general surface of revolution about the optical axis of the eye," IEICE Trans. Inf. & Syst., vol.E95-D, no.6, pp.1656–1667, June 2012.
[13] E. Wood and A. Bulling, "Eyetab: Model-based gaze estimation on unmodified tablet computers," Proc. Symposium on Eye Tracking Research and Applications, pp.207–210, 2014.
[14] T. Baltrusaitis, A. Zadeh, Y.C. Lim, and L.-P. Morency, "Openface 2.0: Facial behavior analysis toolkit," 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), pp.59–66, 2018.
[15] S. Park, X. Zhang, A. Bulling, and O. Hilliges, "Learning to find eye region landmarks for remote gaze estimation in unconstrained settings," Proc. 2018 ACM Symposium on Eye Tracking Research & Applications, article no.21, 2018.
[16] K.-H. Tan, D.J. Kriegman, and N. Ahuja, "Appearance-based eye gaze estimation," Proc. Sixth IEEE Workshop on Applications of Computer Vision, pp.191–195, IEEE Computer Society, 2002.
[17] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, "MPIIGaze: Real-world dataset and deep appearance-based gaze estimation," IEEE Trans. Pattern Anal. Mach. Intell., vol.41, no.1, pp.162–175, 2019.
[18] A. Kar and P. Corcoran, "A review and analysis of eye-gaze estimation systems, algorithms and performance evaluation methods in consumer platforms," IEEE Access, vol.5, pp.16495–16519, 2017.
[19] D.A. Goss and R.W. West, Introduction to the Optics of the Eye, Butterworth-Heinemann, 2001.
[20] K. Nishino and S.K. Nayar, "Eyes for relighting," ACM Trans. Graphics, vol.23, no.3, pp.704–711, 2004.
[21] D. Beymer and M. Flickner, "Eye gaze tracking using an active stereo head," Proc. IEEE Computer Society Conference on Com-

puter Vision and Pattern Recognition, pp.451–458, 2003.
[22] D. Model and M. Eizenman, "A general framework for extension of a tracking range of user-calibration-free remote eye-gaze tracking systems," Proc. Symposium on Eye Tracking Research and Applications, pp.253–256, ACM, 2012.
[23] T. Nagamatsu, J. Kamahara, and N. Tanaka, "3D gaze tracking with easy calibration using stereo cameras for robot and human communication," Proc. IEEE RO-MAN 2008, pp.59–64, 2008.
[24] Intel, "Open source computer vision library." http://code.opencv.org/
[25] P. Blignaut, "Mapping the pupil-glint vector to gaze coordinates in a simple video-based eye tracker," Journal of Eye Movement Research, vol.7, no.1, pp.1–11, 2014.
[26] C. Palmero, J. Selva, M.A. Bagheri, and S. Escalera, "Recurrent CNN for 3D gaze estimation using appearance and shape cues," Proc. British Machine Vision Conference (BMVC) 2018, pp.1–13, 2018.

Appendix A: User Calibration (Estimation of User-dependent Parameters)

In the user-calibration process, the user is instructed to gaze at a single point (calibration point), the position of which is known. The position of the calibration point is selected such that the light from the camera is reflected from the central portion of the corneal surface that is approximated as a sphere. Furthermore, it is assumed that the pupil can be observed through the portion of the corneal surface that is approximated as a sphere because the pupil is observed sufficiently inside the corneal edge even when the pupil is enlarged. Therefore, the refraction at the corneal surface can be determined on the basis of the spherical model of the cornea.

A.1 Estimation of the Radius of Corneal Curvature (R) [11]

When a user gazes at an object near the camera in the user-calibration process, light is reflected from the spherical corneal surface. Hence, we can use the spherical model of the cornea in the calibration process.

We now estimate the position of the center of corneal curvature $\mathbf{C}$. Figure A· 1 shows a cross section of the cornea including $\mathbf{C}$, the center of corneal curvature; $\mathbf{L}_0$, the position of the light source 0; $\mathbf{L}_1$, the position of the light source 1; $\mathbf{O}_0$, the nodal point of camera 0; and $\mathbf{O}_1$, the nodal point of the camera 1. The positions of $\mathbf{O}_0$ (= $\mathbf{L}_0$) and $\mathbf{O}_1$ (= $\mathbf{L}_1$) are known from the camera calibration. A ray from $\mathbf{L}_0$ reflected from the corneal surface returns to $\mathbf{O}_0$ and reaches $\mathbf{G}'_{00}$. The extension of the path of the ray includes $\mathbf{C}$ because the corneal surface is supposed to be a sphere. Similarly, the line connecting $\mathbf{O}_1$ and $\mathbf{G}'_{11}$ includes $\mathbf{C}$. Therefore, $\mathbf{C}$ can be estimated from the intersection of two lines as follows:

$$\mathbf{X} = \mathbf{O}_0 + t_{00}\left(\mathbf{O}_0 - \mathbf{G}'_{00}\right), \tag{A· 1}$$

$$\mathbf{X} = \mathbf{O}_1 + t_{11}\left(\mathbf{O}_1 - \mathbf{G}'_{11}\right), \tag{A· 2}$$

where t_{00} and t_{11} are parameters.

A ray from $\mathbf{L}_0$ is reflected at a point $\mathbf{G}''_{10}$ on the corneal surface such that the reflected ray passes through $\mathbf{O}_1$ and intersects the camera image plane at a point $\mathbf{G}'_{10}$. Similarly, a ray from $\mathbf{L}_1$ is reflected at a point $\mathbf{G}''_{01}$ on the corneal surface such that the reflected ray passes through $\mathbf{O}_0$ and intersects the camera image plane at a point $\mathbf{G}'_{01}$. To estimate the radius of the cornea, we estimate the reflection point $\mathbf{G}''_{10}$ (= $\mathbf{G}''_{01}$), that is, the intersection of the lines as follows:

$$\mathbf{X} = \mathbf{O}_0 + t_{01}\left(\mathbf{O}_0 - \mathbf{G}'_{01}\right), \tag{A· 3}$$

$$\mathbf{X} = \mathbf{O}_1 + t_{10}\left(\mathbf{O}_1 - \mathbf{G}'_{10}\right), \tag{A· 4}$$

where t_{01} and t_{10} are parameters. Therefore, the radius of corneal curvature R is determined as $R = \|\mathbf{G}''_{10} - \mathbf{C}\|$.

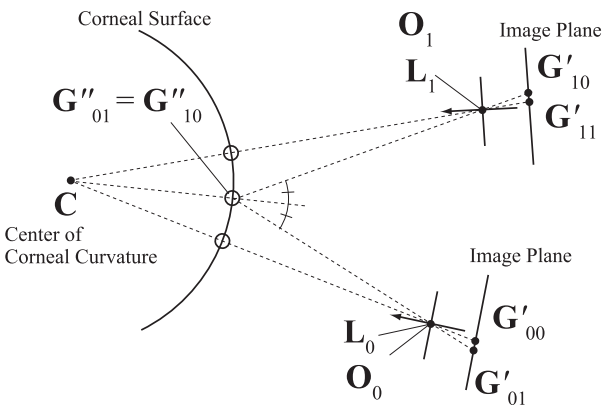


Fig. A· 1 Cross section of the cornea, including the center of corneal curvature; positions of light sources; and nodal points of cameras.

A.2 Estimation of the Distance between the Center of Corneal Curvature and the Center of the Pupil (K) [11]

As shown in Fig. A· 2, a ray originating from the center of the pupil $\mathbf{P}$ gets refracted at a point $\mathbf{P}''_j$, passes through the nodal point $\mathbf{O}_j$ of camera j, and intersects the camera image plane at point $\mathbf{P}'_j$. $\mathbf{P}''_j$ can be determined by solving the equations given below:

$$\mathbf{X} = \mathbf{O}_j + t_j\left(\mathbf{O}_j - \mathbf{P}'_j\right), \tag{A· 5}$$

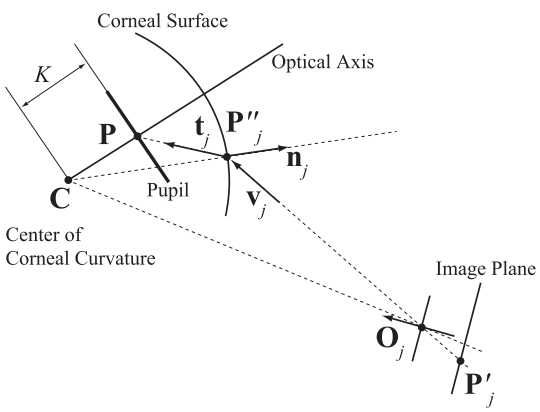


Fig. A· 2 Refraction at the corneal surface.

$$R = \|\mathbf{X} - \mathbf{C}\|. \quad (A\cdot 6)$$

These equations may have two solutions; we select the one closer to $\mathbf{O}_j$.

The vector $\mathbf{t}_j$ (the refracted vector at $\mathbf{P}''_j$, shown in Fig. A· 2) can be obtained by using Snell's law as follows:

$$\mathbf{t}_j = \left(-\rho\mathbf{n}_j \cdot \mathbf{v}_j - \sqrt{1-\rho^2\left(1-(\mathbf{n}_j\cdot\mathbf{v}_j)^2\right)}\right)\mathbf{n}_j + \rho\mathbf{v}_j, \quad (A\cdot 7)$$

where the incident vector $\mathbf{v}_j = (\mathbf{O}_j - \mathbf{P}'_j)/\|\mathbf{O}_j - \mathbf{P}'_j\|$; the normal vector at the point of refraction, $\mathbf{n}_j = (\mathbf{P}''_j - \mathbf{C})/\|\mathbf{P}''_j - \mathbf{C}\|$; and $\rho = n_1/n_2$ (n_1, refractive index of air ≈ 1; n_2, effective refractive index ≈ 1.3375 [6]).

The center of the pupil $\mathbf{P}$ can be determined from the intersection of two rays from the two cameras, as follows:

$$\mathbf{X} = \mathbf{P}''_j + s_j\mathbf{t}_j \quad (j = 0, 1), \quad (A\cdot 8)$$

where s_j is a parameter. Therefore, the distance K between the centers of the corneal curvature and the pupil is calculated as $K = \|\mathbf{P} - \mathbf{C}\|$.

A.3 Estimation of the Offset between the Optical and Visual Axes of the Eye [23]

In the case of a user gazing at a known position, the offset between the optical and visual axes of the eye is calculated using the method reported by Nagamatsu et al. [23]. The offset is determined as a relative position of the unit direction vectors of the optical and visual axes of the eye when the eye is at the primary position. The calculation method by Nagamatsu et al. is as follows.

Figure A· 3 shows a model illustrating eyeball rotation. **a** and **b** are the unit direction vectors of the visual and optical axes, respectively, of the eye, when the eye is at the primary position. The primary position is the position of the visual axis relative to the head and is approximately the position when looking straight ahead at an object at eye level. **c** and **d** are unit direction vectors of the visual and optical axes, respectively, of the eye, after the eye movement. **E** is the rotation center of the eyeball. $\mathbf{C}_0$ and **C** are the center of the corneal curvature at the primary position and the center of the corneal curvature after eye movement, respectively.

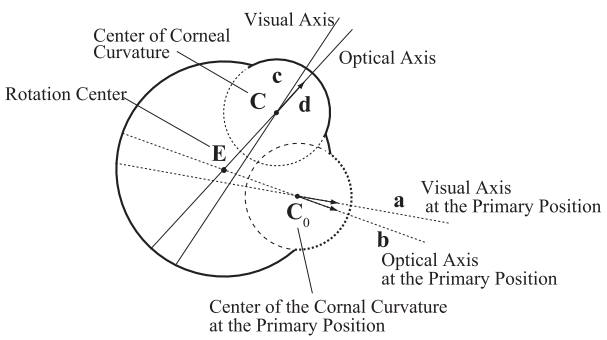


Fig. A· 3 Eyeball model of rotation.

The calibration of our method is a process to estimate the unit direction vector of the opticsl axis at the primary position **b**, from the known **a**, **c**, and **d**.

When the calibration is performed, a user gazes at a point **D**, on the display; hence, **c** is written as $(\mathbf{D} - \mathbf{C})/\|\mathbf{D} - \mathbf{C}\|$. **d** is a unit direction vector of the optical axis and is given as $(\mathbf{P} - \mathbf{C})/\|\mathbf{P} - \mathbf{C}\|$. **a** is known because the coordinate system is determined on the basis of **a**.

The rotation axis **l** and the angle ψ can be calculated on the basis of Listing's law as

$$\mathbf{l} = \frac{\mathbf{a}\times\mathbf{c}}{|\mathbf{a}\times\mathbf{c}|}, \quad (A\cdot 9)$$

$$\psi = \arccos\left(\frac{\mathbf{a}\cdot\mathbf{c}}{|\mathbf{a}||\mathbf{c}|}\right). \quad (A\cdot 10)$$

Then, **b** can be calculated by rotating **d** around **l** by an angle $-\psi$.

Appendix B: Estimation of the Visual Axis of the Eye and POG after User Calibration

After the user calibration, the user moves his/her eyes freely. The optical axis of the eye can be calculated using the method described in Sect. 4. R, K, and the offset between the optical and visual axes of the eye are known from the user calibration. Then, the positions of **C** (the center of corneal curvature) and **c** (the unit direction vector along the visual axis of the eye) are required for calculating the visual axis of the eye ($\mathbf{X} = \mathbf{C} + t\mathbf{c}$).

B.1 Estimation of the Center of Corneal Curvature around the Central Portion of the Outer Surface of the Cornea [11]

Because the pupil is observed sufficiently inside the edge of the cornea, we assume that the corneal surface where the pupil is observed can be approximated as a spherical surface. Therefore, the refraction at the surface of the cornea can be calculated using a spherical model.

The algorithm for searching for the position of **C** is as follows:

1. Set the initial position of **C** on the optical axis of the eye, and select the position that is nearest to the intersection of the two lines, i.e., $\mathbf{X} = \mathbf{O}_j + t_j(\mathbf{O}_j - \mathbf{P}'_j)$ $(j = 0, 1)$.
2. Calculate $\mathbf{P}''_j$ and $\mathbf{t}_j$ by using Eqs. (A· 5), (A· 6), and (A· 7), where R is known from the user calibration.
3. Calculate **P** from the intersection of the two lines described by using Eq. (A· 8).
4. Calculate the distance between **P** and **C**, and compare it to K that was estimated during the user calibration.
5. Shift the position of **C** toward the rotation center of the eye along the optical axis of the eye, and repeat steps 2–5 to determine the accurate position of **C**. The search is finished when $\|\mathbf{P} - \mathbf{C}\| = K$. It is sufficient to search

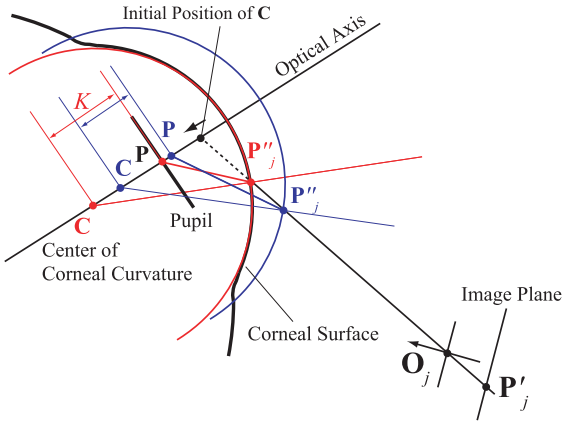


Fig. A·4 Searching the position of **C**.

for the position of **C** for a length of 10 mm because the average radius of the cornea is approximately 7.8 mm.

B.2 Estimation of the Visual Axis of the Eye and POG [23]

The unit direction vector of the visual axis **c**, is calculated from the known **a**, **b**, and **d**. **b** is estimated by the calibration process. **d** is a unit direction vector of the optical axis, which is given as $(\mathbf{P}-\mathbf{C})/\|\mathbf{P}-\mathbf{C}\|$.

The plane including **l** is perpendicular to **a**; hence, the plane is expressed as $\mathbf{a}\cdot\mathbf{x}=0$. Since **b** rotates to **d** around **l**, **l** is also contained in the plane, $(\mathbf{d}-\mathbf{b})\cdot\mathbf{x}=0$. Therefore, **l** is given as

$$\mathbf{l}=\frac{\mathbf{a}\times(\mathbf{d}-\mathbf{b})}{\|\mathbf{a}\times(\mathbf{d}-\mathbf{b})\|}. \tag{A·11}$$

The rotation angle of the eyeball is estimated as

$$\psi=\arccos\left(\frac{((\mathbf{l}\cdot\mathbf{b})\mathbf{l}-\mathbf{b})\cdot((\mathbf{l}\cdot\mathbf{b})\mathbf{l}-\mathbf{d})}{|((\mathbf{l}\cdot\mathbf{b})\mathbf{l}-\mathbf{b})||((\mathbf{l}\cdot\mathbf{b})\mathbf{l}-\mathbf{d})|}\right). \tag{A·12}$$

The direction vector of the visual axis **c** can be calculated by rotating **a** around **l** by an angle ψ.

The intersection point between the visual axis of the eye ($\mathbf{X}=\mathbf{C}+t\mathbf{c}$) and the object (e.g., a display) is the POG.

Takashi Nagamatsu received his B.S. and M.S. degrees in engineering from Kyoto University in 1994 and 1996, respectively, and his Ph.D. degree in energy science from Kyoto University in 2003. He joined Mitsubishi Heavy Industries, Ltd. in 1996. He worked for Kyoto University as a research associate in 1999. He joined Kobe University of Mercantile Marine in 2000. He is currently an associate professor at the Graduate School of Maritime Sciences, Kobe University. His research interests include human-computer interaction, especially gaze-based interaction

Mamoru Hiroe received his B.S. and M.S. degrees in maritime sciences from Kobe University in 2016 and 2019, respectively. He is currently a Ph.D. student of the Graduate School of Maritime Sciences, Kobe University. His research interests include machine learning and gaze tracking.

Hisashi Arai received his B.S. and M.S. degrees in maritime sciences from Kobe University in 2016 and 2018, respectively. He joined Daifuku Co., Ltd. in 2018. His research interests include gaze tracking.