

PAPER

User-Calibration-Free Gaze Estimation Method Using a Binocular 3D Eye Model

Takashi NAGAMATSU^{†a)}, Ryuichi SUGANO^{†b)}, Yukina IWAMOTO^{†*}, *Nonmembers*, Junzo KAMAHARA^{†c)}, and Naoki TANAKA^{†d)}, *Members*

SUMMARY This paper presents a user-calibration-free method for estimating the point of gaze (POG). This method provides a fast and stable solution for realizing user-calibration-free gaze estimation more accurately than the conventional method that uses the optical axis of the eye as an approximation of the visual axis of the eye. The optical axis of the eye can be estimated by using two cameras and two light sources. This estimation is carried out by using a spherical model of the cornea. The point of intersection of the optical axis of the eye with the object that the user gazes at is termed POA. On the basis of an assumption that the visual axes of both eyes intersect on the object, the POG is approximately estimated using the binocular 3D eye model as the midpoint of the line joining the POAs of both eyes. Based on this method, we have developed a prototype system that comprises a 19" display with two pairs of stereo cameras. We evaluated the system experimentally with 20 subjects who were at a distance of 600 mm from the display. The root-mean-square error (RMSE) of measurement of POG in the display screen coordinate system is 1.58°.

key words: gaze tracking, calibration-free, eye movement, eye model

1. Introduction

Gaze tracking technology is used as a human interface [1], [2]. However, the user-calibration phase employed by current systems still requires some effort by the user in that the user must gaze at several points. Because of the requirements of user calibration, current gaze tracking systems find limited applications; for example, it is difficult to develop systems meant for the general public such as applications for digital signage. Therefore, it would be desirable to eliminate user calibration.

Several studies have attempted to reduce the effort required for user calibration using a model-based approach. Figure 1 shows a typical model of an eye that is used in the model-based approach [3]–[6]. There are two important axes that are modeled: one is the optical axis, which is the line passing through the geometric center of the eye, and the other is the visual axis, which is the line of sight connecting the fovea and the point of gaze (POG). In this model, it is approximated that the two axes of the eye intersect at the center of the corneal curvature. The offset between the

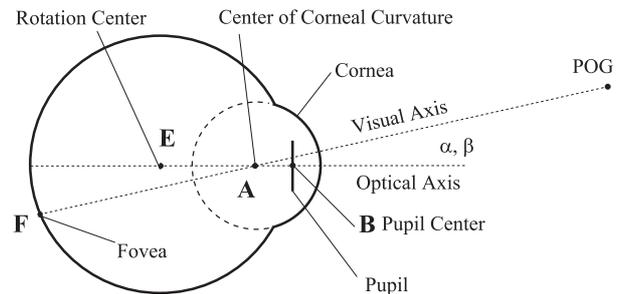


Fig. 1 Eye model.

optical and visual axes of the eye is described by two parameters (α and β). If we use horizontal and vertical angles to express the offset, the horizontal and vertical angles between the optical and visual axes of the eye range from 3.5° to 7.5° (average of 5.5°) and from 0.25° to 3.0° (average of 1.0°), respectively [7]; these angles vary among individuals.

Shih and Liu [3] and Guestrin and Eizenman [8] presented methods for reconstructing the optical axis of the eye using stereo cameras. These methods do not require any information about the characteristics of an individual's eye, such as the radius and the refractive index of the surface of the cornea. However, in order to estimate the visual axis of the eye, these methods employ a one-point user-calibration technique that requires a user to gaze at a single point. The estimation error of these systems is under 1° in view angle.

Several systems claim to be user-calibration-free [9], [10]; however, these systems measure only the optical axis of the eye. The optical axis of the eye is used as an approximation for the visual axis of the eye; therefore, the error of estimating the visual axis of the eye is expected to be approximately 5°.

In order to realize user-calibration-free gaze estimation, it is insufficient to use only a single eye. Because the fovea is unobservable by a camera, the position of the visual axis of the eye is not determined without gazing at a known point. Therefore, using both eyes is required. Recently, Model et al. [11], [12] reported an automatic personal calibration technique using both eyes. Because 1000 widely-distributed POGs are used for convergence calculations, it takes some time to calibrate the system. Although the system is suitable for studies with young children, etc., it may not be suitable for public use such as digital signage (people look at advertisements etc. for a short time).

Manuscript received December 15, 2010.

Manuscript revised May 23, 2011.

[†]The authors are with the Graduate School of Maritime Sciences, Kobe University, Kobe-shi, 658-0022 Japan.

*Presently, with the Mitsubishi Electric Corporation

a) E-mail: nagamatu@kobe-u.ac.jp

b) E-mail: 101w110w@stu.kobe-u.ac.jp

c) E-mail: kamahara@maritime.kobe-u.ac.jp

d) E-mail: ntanaka@maritime.kobe-u.ac.jp

DOI: 10.1587/transinf.E94.D.1817

In this paper, we propose a fast and stable theoretical solution for estimating POG on the basis of a method to estimate the optical axis of the eye [13], [14].

The main contributions of this paper are as follows

- We propose a binocular 3D eye model that includes the optical and visual axes of both eyes. The key point is the fact that the visual axes of both eyes *intersect on the object* that the user gazes at.
- We propose a method for estimating the POG using the midpoint of the optical axes of both eyes. Our method is not just calculation of an average of two optical axes but an estimation of the offsets between the optical and visual axes of both eyes.

The remainder of the paper is organized as follows. In Sect. 2, we explain the estimation of the optical axis of a single eye. In Sect. 3, we propose a binocular 3D model and a theoretical solution for user-calibration-free gaze estimation. In Sect. 4, we present the implementation of the system and the evaluation results. In Sect. 5, we present methods for improving the stability of the user-calibration-free gaze estimation. We discuss related research efforts in Sect. 6 and conclude the paper in Sect. 7.

In this paper, capitalized boldface letters denote position vectors and small boldface letters denote unit direction vectors.

2. Estimation of Optical Axis of Single Eye

The optical axis of a single eye can be estimated by the method described in [3], [4], [15]. This method requires a minimum of two cameras and two point light sources, as shown in Fig. 2. The cameras are modeled as pinhole-type cameras. Prior to the estimation, the positions of the display and the light sources are measured and the intrinsic and extrinsic camera parameters are determined by camera calibration.

Figure 3 shows a ray-tracing diagram for the estimation of the optical axis of the eye. The cornea is modeled as a sphere. In what follows, i and j can be 0 and 1. \mathbf{A} is the position of the center of the corneal curvature. \mathbf{L}_i is the position of the light source i and \mathbf{C}_j is the nodal point of

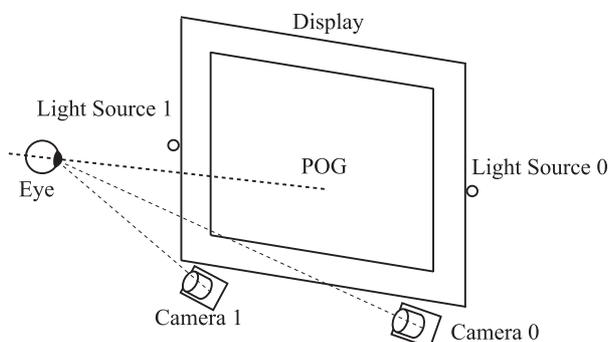


Fig. 2 Gaze tracking system using two cameras and two point light sources.

camera j . The value of \mathbf{L}_i is measured and \mathbf{C}_j is determined beforehand by camera calibration.

A ray from \mathbf{L}_i is reflected on the corneal surface such that it passes through \mathbf{C}_j and intersects the camera image plane at a point \mathbf{P}'_{ji} (first Purkinje image: reflections on the outer surface of the cornea). The plane including \mathbf{C}_j , \mathbf{L}_i , and \mathbf{P}'_{ji} is expressed as

$$\{(\mathbf{P}'_{ji} - \mathbf{C}_j) \times (\mathbf{L}_i - \mathbf{C}_j)\} \cdot (\mathbf{X} - \mathbf{C}_j) = 0, \quad (1)$$

where $\mathbf{X} (= (x, y, z)^T)$ is a point on the plane. When a ray reflected on a spherical surface, a plane passing through the incident and reflection vectors includes the center of the sphere; therefore, the plane expressed by Eq. (1) includes \mathbf{A} .

We obtain four planes when we use two cameras and two light sources ($i = 0, 1$ and $j = 0, 1$). Because all planes contain \mathbf{A} , it can be calculated if at least three planes are given.

A ray from \mathbf{B} refracts at \mathbf{B}'_j on the corneal surface, passes through \mathbf{C}_j , and reaches \mathbf{B}''_j (center of the pupil on the image sensor). Therefore, the plane that includes \mathbf{A} , \mathbf{B} , \mathbf{B}''_j , \mathbf{C}_j , and \mathbf{B}'_j also includes the optical axis (line connecting \mathbf{A} and \mathbf{B}). The normal vector of the plane is $(\mathbf{C}_j - \mathbf{B}'_j) \times (\mathbf{A} - \mathbf{C}_j)$. Because we use two cameras, we can calculate the optical axis as the intersection of the two planes ($j = 0, 1$). The direction of the intersection line is the direction of the cross product of the two normal vectors of the planes. Therefore, the unit vector of the optical axis \mathbf{d} is written as

$$\mathbf{d} = \frac{((\mathbf{C}_0 - \mathbf{B}'_0) \times (\mathbf{A} - \mathbf{C}_0)) \times ((\mathbf{C}_1 - \mathbf{B}'_1) \times (\mathbf{A} - \mathbf{C}_1))}{\|((\mathbf{C}_0 - \mathbf{B}'_0) \times (\mathbf{A} - \mathbf{C}_0)) \times ((\mathbf{C}_1 - \mathbf{B}'_1) \times (\mathbf{A} - \mathbf{C}_1))\|}. \quad (2)$$

Thus, the optical axis of the eye is estimated as $\mathbf{X} = \mathbf{A} + t\mathbf{d}$ (where t is a parameter) without user calibration.

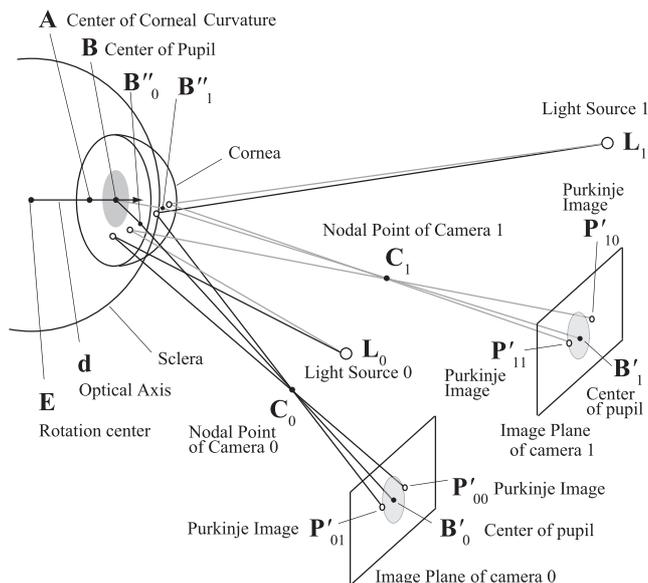


Fig. 3 Ray-tracing diagram for estimation of the optical axis of the eye.

3. User-Calibration-Free Gaze Estimation Method

The research carried out in the field of model-based gaze estimation have conventionally involved the use of only one eye [3], [4], [15], and the optical axis of the eye is estimated accurately using two cameras and multiple light sources. Our new method provides a solution by using both eyes to improve the accuracy of gaze estimation without user calibration.

3.1 Binocular 3D Eye Model

Figure 4 shows a 3D model of the optical and visual axes of both eyes when a user gazes at a certain point on an object (e.g., a display). It is known that the visual axis of the eye is inclined toward the nose, away from the optical axis of the eye [16]. The key point here is the fact that the visual axes of both eyes intersect on the object when the user gazes at a point on the object. The intersection point is called POG and its position vector is denoted by \mathbf{X}_{POG} . The point of intersection of the optical axis of the eye with the object is denoted by POA in this paper. The position vectors of the POAs of the left and right eyes are denoted by \mathbf{X}_{POAL} and \mathbf{X}_{POAR} , respectively.

The offset between the optical and visual axes of one eye is expressed by two parameters, as shown in Fig. 1. Therefore, the offsets of both eyes are expressed by four parameters: $\alpha_L, \beta_L, \alpha_R,$ and β_R , that is, two parameters each for the left eye and the right eye.

We suppose that $\alpha_L, \beta_L, \alpha_R,$ and β_R are user-dependent and constant. Because we can estimate the optical axis of the eye without user calibration by the method described in Sect. 2, the problem to solve here is the estimation of $\alpha_L, \beta_L, \alpha_R,$ and β_R given that the optical axes of both eyes are known and the visual axes of both eyes intersect each other on the object.

3.2 Estimation of POG

In this paper, we propose a method to determine the four pa-

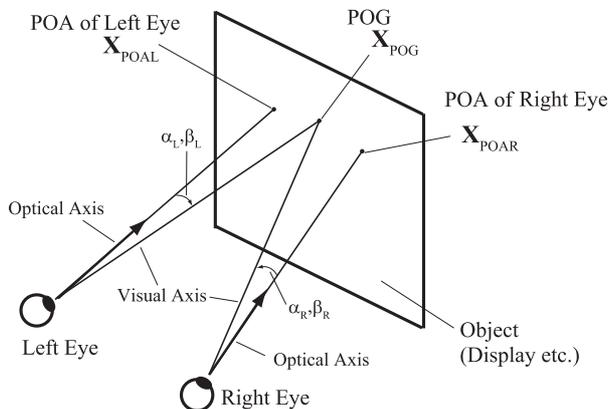


Fig. 4 Binocular 3D eye model.

rameters efficiently. In other words, we determine the four parameters using image data from just one frame. The estimation is carried out in two steps: 1) estimating the POG and 2) estimating the four parameters; the first step is described in this section and the second step is described in Sect. 3.3.

As described in the introduction, the average values of the horizontal and the vertical angle between the optical and visual axes of the eye are 5.5° and 1.0° , respectively. In this study, we focus on the estimation of the horizontal angle between the optical and visual axes of the eye (which makes up the majority of the offset). We ignore the vertical angle because it is smaller than the horizontal angle and its average value is only 1.0° .

Figure 5 shows a detail model of binocular eyes when the vertical angle is approximated by 0° . We can estimate \mathbf{X}_{POAL} and \mathbf{X}_{POAR} by determining the points of intersections between the optical axes of both eyes and the object. We assume the human body to be symmetric about the sagittal plane, and therefore, the horizontal angles between the optical and visual axes of both eyes to be similar. Therefore, the POG is determined such that the angles between the visual and the optical axes of both the eyes are equal. For fast calculation, we determine the midpoint of the POAs, which can serve as a good approximation of the POG; \mathbf{X}_{POG} is described as follows:

$$\mathbf{X}_{\text{POG}} \approx \frac{1}{2}(\mathbf{X}_{\text{POAL}} + \mathbf{X}_{\text{POAR}}). \quad (3)$$

Because our method can be employed using only the real-time data, a system leveraging our gaze tracking

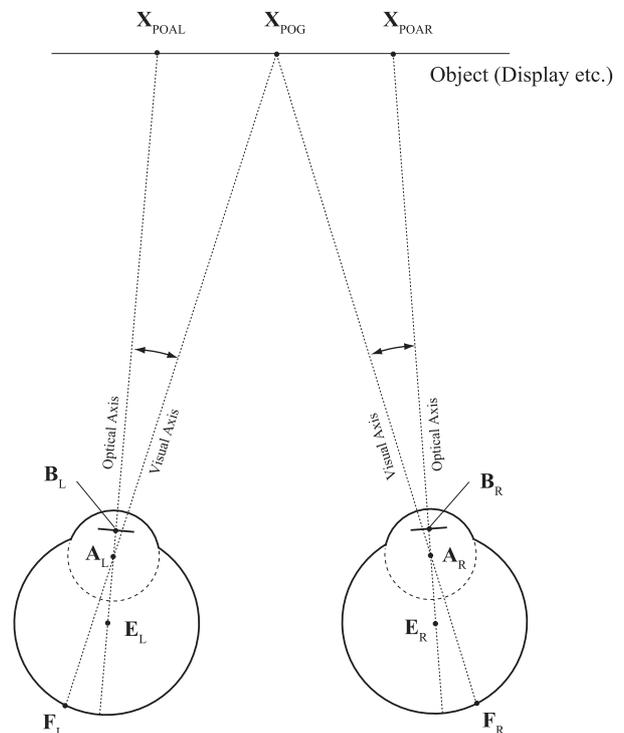


Fig. 5 Binocular eye model in detail.

method can be used from the moment the user comes in front of the object and begins gazing at some point upon it.

Here, we comment on the validity of Eq. (3). In a strict sense, the equation is true under the following assumptions:

1. The line through the centers of both eyes is parallel to the object plane.
2. The POG is located on the perpendicular bisector of the line segment between both eyes.

We evaluated the error of approximation caused by Eq. (3) using the diagram shown in Fig. 6. The error is defined as the distance between the point that is determined, such that the angles between the optical and visual axes of both the eyes are equal (\mathbf{X}_{POG}), and the point calculated by Eq. (3) ($(\mathbf{X}_{POAL} + \mathbf{X}_{POAR})/2$).

Figure 7 shows the simulation results of the error in degree when $\psi = 0^\circ$, with changing gaze direction (i.e. ϕ), horizontally to the right, at the distance of 600 mm. In this simulation, the angles between the optical and visual axes of both eyes are set at 7.0° , which is about the largest value possible for a human. According to the simulation, if the user looks at the 19" display (the edge of the display corresponds to $\phi = 17.4^\circ$), the error is only 0.06° ; moreover, if the user rotates their eyes until $\phi = 40^\circ$, the error is 0.36° .

Table 1 shows the simulation results when changing ψ

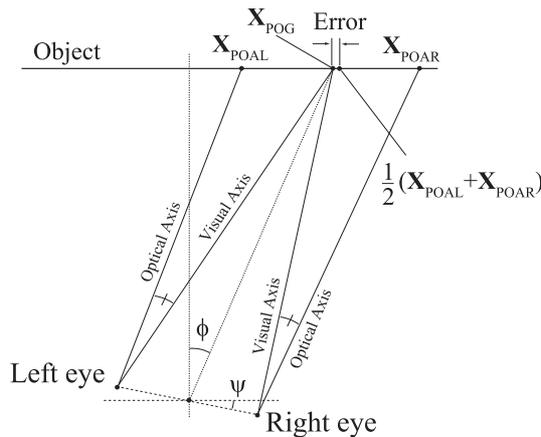


Fig. 6 Diagram for error evaluation for Eq. (3).

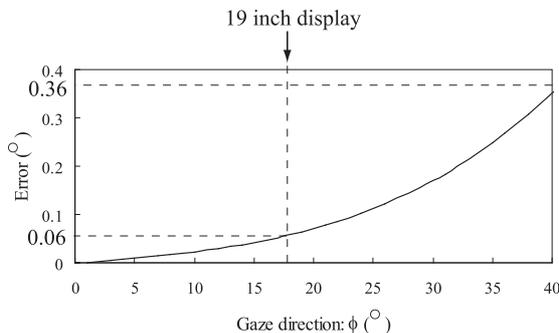


Fig. 7 Simulation of error when gaze direction changes horizontally.

and ϕ . In this simulation, the line connecting both eyes increases in angle from 0° to 40° , i.e., $\psi = 0 \sim 40^\circ$, in each case, and $\phi = -40 \sim 40^\circ$. From the results, the error when the user looks at the 19" display is at most 0.26° .

3.3 Estimation of the Offset of the Optical and Visual Axes of the Eye

By using the calculated POG, the offsets between the visual and the optical axes of both eyes can be calculated using the one-point calibration gaze estimation method proposed by Nagamatsu et al [15], [17]. At the time, the POG estimated in 3.2 is considered a fiducial point that a user gazes at when one-point calibration is carried out.

Most previous works (e.g. [3], [4]) did not accurately consider the eyeball kinematics that are described by Listing's law. We employed Nagamatsu's method, as it provides a more accurate description of the eyeball kinematics.

Listing's law states that there is an eye position called the primary position, from which any eye position can be reached by a single rotation, and that all the rotational axes lie on a plane (Listing's plane) [18], [19].

Figure 8 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. **A₀** and **A** are the center of the corneal curvature at the primary position and the center of the corneal curvature after eye movement, respectively.

In the introduction section, we said that the offset between the optical and visual axes of the eye is described by two parameters (α and β). However, there are various ways

Table 1 Simulation results with changing ψ and ϕ ($^\circ$).

ψ	$\phi = -40^\circ$	$\phi = 17.4^\circ$	$\phi = 0^\circ$	$\phi = 17.4^\circ$	$\phi = 40^\circ$
0°	0.36	0.06	0.00	0.06	0.36
10°	0.35	0.01	0.06	0.11	0.37
20°	0.36	0.03	0.12	0.17	0.40
30°	0.38	0.04	0.16	0.22	0.43
40°	0.41	0.04	0.18	0.26	0.47

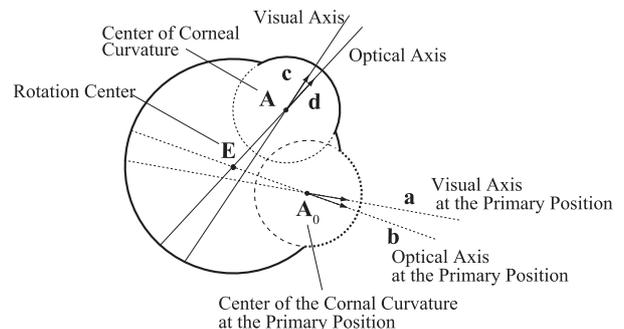


Fig. 8 Eyeball model of rotation.

of describing these two parameters, depending on which coordinate system we use.

Because the direction of **a** can be approximated by the position held when looking straight ahead at an object at eye level (e.g. (0, 0, -1)), estimation of the offset of the optical and visual axes of the eye is the same as the estimation of **b**, which is described by two parameters as $(\alpha, \beta, \sqrt{1 - (\alpha^2 + \beta^2)})$.

The one-point calibration method calculates the rotation axis and the rotation angle of the eyeball from the primary position, on the basis of Listing’s law, thereby allowing for the calculation of the offset of the optical and visual axes of both eyes [15], [17].

In this paper, the unit direction vector of the optical axis of the left eye is expressed by two parameters: α_L and β_L , and the unit direction vector of the optical axis of the right eye is expressed by two parameters: α_R and β_R .

4. Implementation and Experiment

4.1 Implementation of the System

A prototype system for estimation of the POG of a display was developed, as shown in Fig. 9. In the system, we suppose the primary positions of both eyes are supposed to be parallel to the normal vector of the display, which is used in the calculation of the offset between the optical and visual axes of the eye [15].

This system consists of four synchronized monochrome IEEE-1394 digital cameras (Firefly MV, Point Grey Research Inc.), three infrared light sources (LED), a 19" (for the user) and a 22" (for checking the system) LCD, and a Windows-based PC (Windows XP, Intel Core 2 Quad). Each camera uses a 1/3" CMOS image sensor whose resolution is 752 × 480 pixels. A 50-mm lens and an IR filter are attached to each camera. In order to capture high-resolution images of the eyes, we used lenses with a narrow field of view. A pair of cameras was used for capturing the left eye, and another pair was used for the right eye. These cameras were positioned under the display. The cameras were calibrated

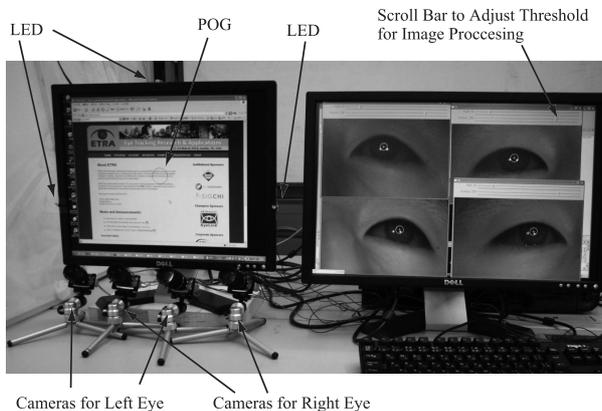


Fig. 9 Prototype system.

beforehand to set the intrinsic and extrinsic camera parameters. The reason for using three LED is to reduce the estimation error of the optical axis of the eye caused by asphericity of the cornea as described in Nagamatsu et al [15].

The software was developed in the C++ language using OpenCV 1.0 [21]. A circle is displayed at the estimated POG, which can be eliminated. The image processing results and system status are displayed on the 22" display, and the thresholds of binarization for the image processing required for the four cameras can be adjusted by using the scroll bars on each window that displays each camera image. The software can perform both the function of estimating POG and that of displaying the fiducial point that the subject gazes at in the evaluation experiment and saving the results of the estimation. The experimenter can control the position of the fiducial point and save the measured result.

4.2 Experiment for Evaluating the Accuracy of Proposed Method

We evaluated the accuracy of the prototype system in a laboratory with 20 adult subjects (15 men and 5 women) who did not wear glasses or contact lenses. The ages of the subjects range from 21 to 40.

The proposed gaze estimation method mathematically allows the user to move; however, the current implementation cannot capture the user’s eyes in a large area. Therefore, in order to measure the performance of the gaze estimation method while avoiding the errors caused by head movement, the head was supported by a chin rest to prevent it from being out of focus/field of view of the cameras in this experiment. Therefore, the objective of this experiment was to confirm that the method can estimate the POG with greater accuracies as compared to the user-calibration-free gaze estimation methods that approximate the visual axis of the eye by the optical axis of the eye.

Figure 10 shows the experimental setup. The eyes were approximately 600 mm from the display. The subjects were asked to fixate on 25 points that appeared one after the other on the display. The procedure of the experiment was as follows:



Fig. 10 Experimental setup.

1. The experimenter displayed one fiducial point using the keyboard.
2. The experimenter asked the subject to fixate on the fiducial point.
3. The experimenter checked the image processing result using the 22" display. If the image processing had a problem, the experimenter adjusted the thresholds of binarization.
4. The experimenter pressed the keyboard to start recording the data. The recorded data were time, the position of the fiducial point, the position of the POA of the left eye, the position of the POA of the right eye, the position of the POG, the position of the center of the corneal curvature of the left eye, the position of the center of the corneal curvature of the right eye, and the error (the difference between the fiducial point and the POG in the view angle). The data were recorded when the optical axes of both eyes were detected. The data recording continued until 10 data points were obtained. If the experimenter found that the subject blinked or looked somewhere else, the experimenter repeated recording of the data.
5. The experimenter displayed another fiducial point using the keyboard.
6. The experiment was finished when data for 25 fiducial points were recorded.

4.3 Results

Figures 11 and 12 show the experimental results of 20 subjects. The black diamond-shaped points on the grids represent the fiducial points that were intentionally gazed at by the subjects. The view angles between grid points are about 7.1° horizontally and 5.7° vertically. The triangle- and square-shaped points indicate the average POAs of the left and the right eye, respectively. The lines connecting a POAL and a POAR indicate that these were measured at the same time. The black circular points indicate the averages of estimated POGs. Before plotting the graph, in cases where the POA of either the left eye or the right eye deviated from the median value by more than 5° , in terms of view angle, the data were removed as an outlier. In such a case, it is assumed that the subject gazed at another target or that the image processing failed to detect the center of the pupil or the first Purkinje images, because of blinking, etc. In the preliminary experiment [13], we found the system had sufficient accuracy to distinguish between neighboring points. Therefore, we think that 5° is large enough to identify outliers. We also used the data, excluding outliers, in the analysis below.

The average of the root-mean-square errors (RMSE) of the system evaluation are listed in Table 2 in view angle. The RMSE of the POAL (POA of left eye), POAR (POA of right eye), and POG denote the RMSE between the POAL and the fiducial point, the POAR and the fiducial point, and the POG and the fiducial point, respectively. The $RMSE_x$ and

$RMSE_y$ indicate the horizontal and vertical RMSE, respectively. The values in the table are the averages of the RMSE for 25 fiducial points, for each subject. The second lowest row (Ave.) and the lowest row (σ) indicate the average and the standard deviation, respectively.

The $RMSE_x$ is calculated by the following equation:

$$RMSE_x = \sqrt{\frac{1}{n} \sum_{i=1}^n (X - x_i)^2} \quad (4)$$

where X and x_i are true and measured values, respectively. The $RMSE_y$ is calculated by the following equation:

$$RMSE_y = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y - y_i)^2} \quad (5)$$

where Y and y_i are true and measured values, respectively.

The RMSE is given by $\sqrt{RMSE_x^2 + RMSE_y^2}$. The $RMSE_x$, $RMSE_y$, and RMSE were calculated for each fiducial point, excluding outliers; at least 7 data points were used for the calculation for each fiducial point. After calculating the averages of the RMSE, in mm, for 25 fiducial points, we converted the value to degrees; we used $\theta = \arctan(d/L)$ as the conversion equation, from d mm to θ° , where L is the distance of the eye from the display.

As shown in Table 2, the average of the RMSE of POG was 1.58° , which is equivalent to 16.55 mm on the display for a distance of 600 mm from the eye. The accuracy of the method that treats the optical axis as an approximation of the visual axis of the eye (optical axis method) is same as the accuracy of estimating POAL or POAR. The averages of RMSE of POAL and POAR were 3.21° and 2.75° , respectively, as shown in Table 2. Therefore, as compared to the optical axis method, our method is more accurate. Furthermore, the standard deviation σ of our method was smaller than the optical axis method; the RMSE of POAL was found to range from 1.38° to 8.27° , the RMSE of POAR was found to range from 0.80° to 5.71° , and the RMSE of POG was found to range from 0.64° to 3.21° .

As shown in Table 2, the average of the $RMSE_x$ and $RMSE_y$ of POG is 0.80° and 1.23° , respectively. While the horizontal error is improved from the optical axis method, the vertical error is similar to the average value of the vertical offset for humans reported in previous literature [7] (Horizontal: $3.5 - 7.5^\circ$ (Ave. 5.5°), Vertical: $0.25 - 3.0^\circ$ (Ave. 1.0°)).

4.4 Discussion of this Experiment

In order to discuss the effect of Listing's law, based on it, we simulated the POAL, POAR, and POG. The position of the display and fiducial points are the same as in the prototype system. The positions of the left and right eyeballs are the same as the positions of the left and right centers of the chess-board patterns for the camera calibration of the prototype system, respectively. In the simulation, the visual axes

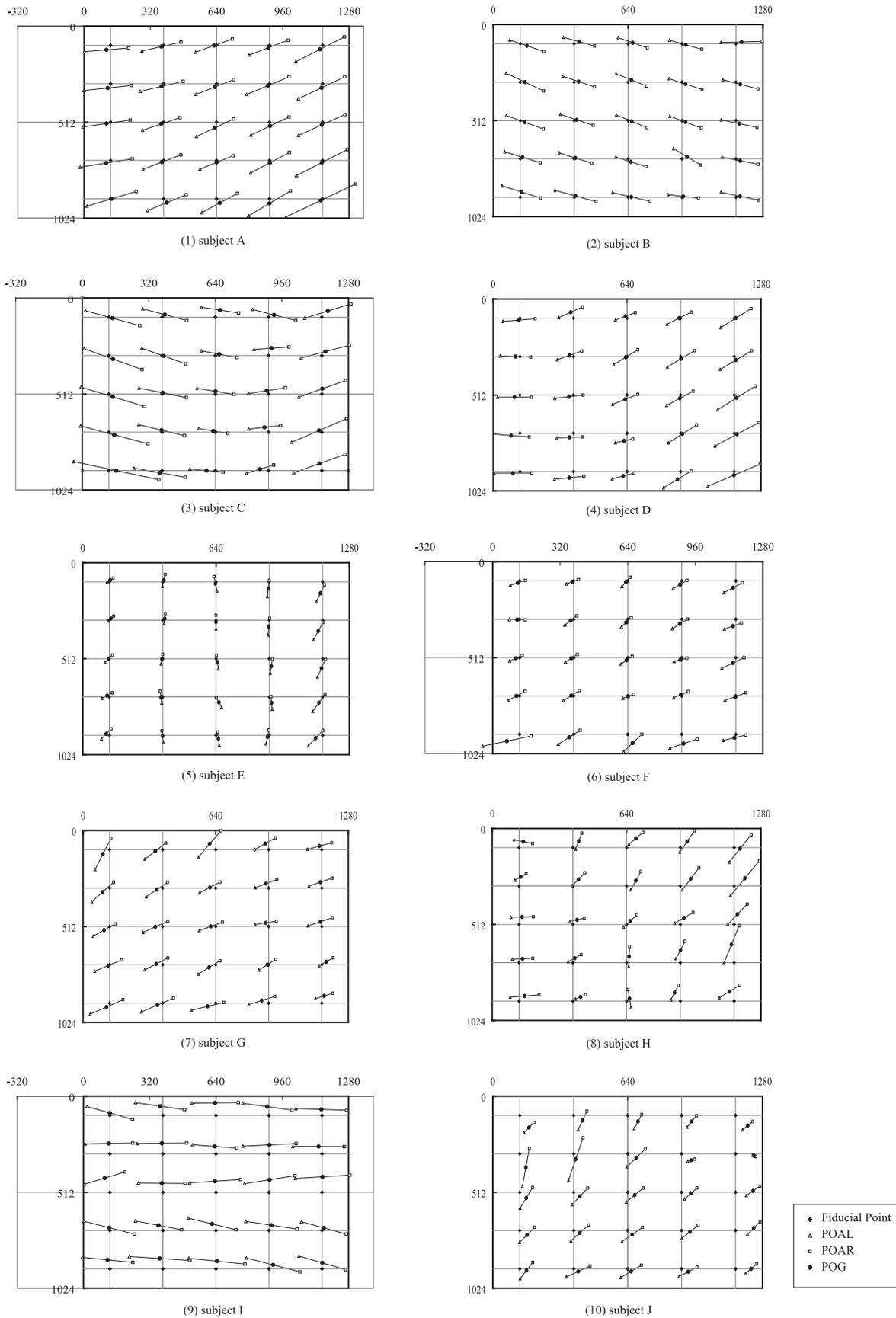


Fig. 11 Experimental results for estimation of the POG, POAL, and POAR (1).

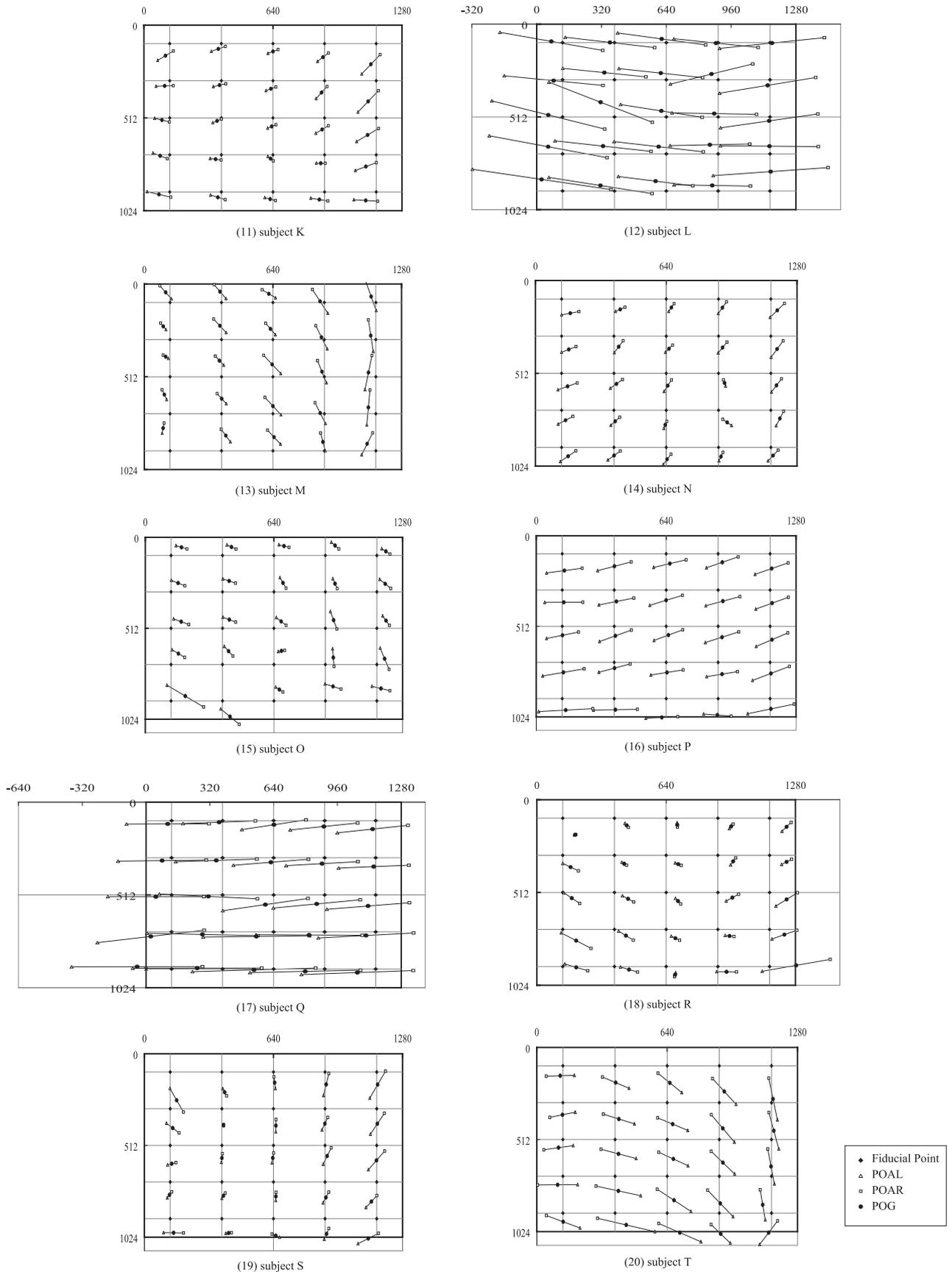


Fig. 12 Experimental results for estimation of the POG, POAL, and POAR (2).

Table 2 The average of the root-mean-square errors (RMSE) for the 20 subjects (°).

Subject	Ave. of RMSE _x of POAL	Ave. of RMSE _y of POAL	Ave. of RMSE of POAL	Ave. of RMSE _x of POAR	Ave. of RMSE _y of POAR	Ave. of RMSE of POAR	Ave. of RMSE _x of POG	Ave. of RMSE _y of POG	Ave. of RMSE of POG
A	3.09	1.66	3.55	2.76	0.94	2.93	0.35	0.51	0.64
B	1.80	0.86	2.03	2.76	0.90	2.92	0.54	0.36	0.67
C	2.99	0.91	3.15	3.24	1.20	3.50	0.52	0.55	0.80
D	2.30	1.23	2.68	1.75	1.01	2.05	0.50	0.63	0.83
E	0.72	1.59	1.80	0.37	0.69	0.80	0.39	0.73	0.86
F	1.51	1.24	2.00	1.01	0.61	1.21	0.54	0.66	0.89
G	2.18	1.08	2.48	1.07	1.24	1.71	0.76	0.68	1.06
H	0.89	0.96	1.38	1.61	2.21	2.86	0.78	1.18	1.46
I	3.51	1.70	3.91	3.30	1.59	3.74	0.54	1.43	1.55
J	0.82	1.81	2.12	2.13	1.05	2.43	1.24	0.84	1.56
K	1.71	1.78	2.61	0.52	1.12	1.27	0.74	1.38	1.65
L	7.58	2.04	7.88	5.56	1.10	5.71	1.21	1.15	1.74
M	0.80	1.60	1.85	1.15	3.02	3.27	0.64	1.73	1.89
N	0.60	2.46	2.57	1.33	1.09	1.78	0.73	1.73	1.90
O	0.68	2.13	2.26	1.97	1.29	2.41	1.19	1.54	1.99
P	2.24	2.51	3.39	2.72	1.48	3.15	0.48	1.92	1.99
Q	8.12	1.30	8.27	4.91	0.67	4.96	1.71	0.85	2.06
R	1.13	1.23	1.71	2.81	1.48	3.28	1.85	1.22	2.28
S	0.69	3.10	3.20	0.92	2.09	2.34	0.53	2.50	2.56
T	2.33	4.41	5.19	1.71	1.85	2.61	0.79	3.06	3.21
Ave.	2.29	1.78	3.21	2.18	1.33	2.75	0.80	1.23	1.58
σ	2.12	0.85	1.90	1.37	0.59	1.20	0.42	0.70	0.69

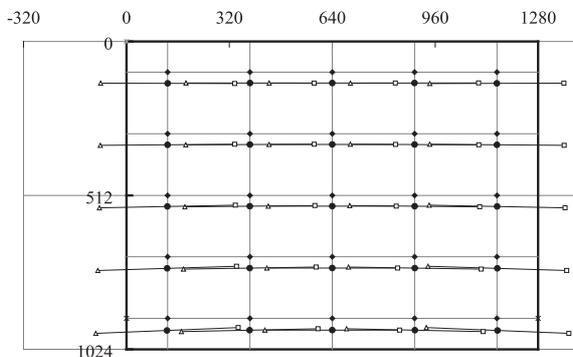


Fig. 13 Simulation of the POAL, POAR, and POG based on Listing's law.

of both eyes incline toward the nose, horizontally, by 5.5° and upward, vertically, by 1.0°, relative to the optical axis of the eye. Figure 13 shows the simulation result. The lines connecting a POAL and a POAR indicate that these were the POAs when a user gazes at the same fiducial point. The lines are almost horizontal. As for the experimental results in Figs. 11 and 12, the graphs of subjects A, B, C, D, F, G, I, K, L, P, and Q show similar trends; the lines are almost horizontal. According to the simulation based on Listing's law, in Fig. 13, the lines in the bottom left area are slightly inclined counter-clockwise, and the lines in the bottom right are slightly inclined clockwise. However, the experimental results do not explicitly express such trends.

In the case of subjects M and T, the POALs and POARs are plotted on the right and left, respectively. In other words, the optical axis for the left eye intersects the screen further to the right than the optical axis for the right eye. Some of the probable causes for this effect are listed as follows: (a) It

could be an image processing problem caused by eyelashes etc. (b) In these subjects, the horizontal angle between the optical and visual axes may be slightly negative. According to Atchison and Smith [16], the horizontal angle is rarely negative. (c) They may suffer from strabismus. Our method is based on the assumption that the visual axes of both eyes intersect on the display; this assumption is valid for the vast majority of subjects, but there can be exceptions such as patients suffering from strabismus.

5. Using One-Eye Model to Keep the Accuracy and Stability of Estimating POG when the Optical Axis of One of the Eyes Is Not Detected

5.1 Algorithm for Stable Operation

Our method is not just a calculation of an average of two optical axes of both eyes but an estimation of the offsets between the optical and visual axes of both eyes. Although there are several studies that use wide angle lenses, (capture multiple people,) and use both eyes to estimate the gaze direction by average of the optical axes of both eyes (e.g. [20]), our method has an advantage in accurate estimation of the visual axis of the eye when the system is unable to detect one eye due to image processing problems etc.

After estimating the offsets between the optical and visual axes of both eyes, when the system is unable to detect one eye, the POG can be calculated from the data of the other eye by using one-eye model.

Figure 14 shows a flowchart to realize this accurate and stable operation. First, the system checks whether the optical axes of both eyes are detected. If they are detected, then POG and both the offsets between the optical and visual axes of the eye are calculated by the method described

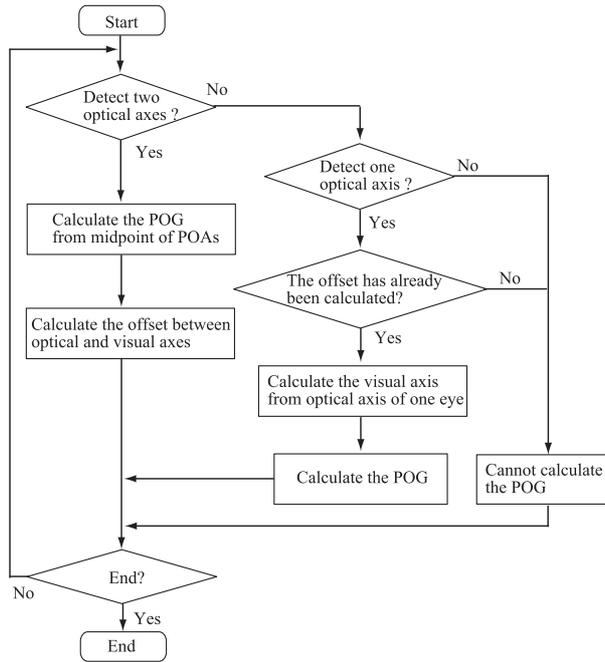


Fig. 14 Flowchart realizing accurate and stable operation.

in Sect. 3. Otherwise, in the case where the system can detect only one eye because of problems in image processing etc., if the system has simultaneously detected the optical axes of both eyes before, the system can accurately calculate the visual axis from the optical axis of the eye by one-eye model [15]. The system cannot calculate POG only in the case where the optical axis of either of the eyes cannot be detected or the optical axes of the eyes have never been detected simultaneously. Thus, the system can calculate the POG in most cases.

5.2 Decision of Detection of the POA

We used three light sources, as shown in Fig. 15. Figure 16 shows the image taken by camera j in the camera image coordinate system. The cross ${}^j\mathbf{B}$ is the center of the pupil of the observed image, which is detected by ellipse fitting. The three small circles are the first Purkinje images. In order to reduce the effect of asphericity of the cornea, we use two light sources whose reflections (first Purkinje images) are closest to the center of the pupil in the camera image, when the POG is calculated. The two selected first Purkinje images are denoted by ${}^j\mathbf{P}_0$ and ${}^j\mathbf{P}_1$. After selecting first Purkinje images, we determine which correspond to each light source by the method described in [15].

Calculation of the POAR necessitates detection of the center of the pupil (${}^0\mathbf{B}$) and Purkinje images (${}^0\mathbf{P}_0$ and ${}^0\mathbf{P}_1$) in the eye image of camera 0, and detection of the center of the pupil (${}^1\mathbf{B}$) and Purkinje images (${}^1\mathbf{P}_0$ and ${}^1\mathbf{P}_1$) in the eye image of camera 1. When the system detects all of 6 feature points, the system concludes that the POAR is detected. Similarly, the system concludes that the POAL is detected when all of ${}^2\mathbf{B}$, ${}^2\mathbf{P}_0$, ${}^2\mathbf{P}_1$, ${}^3\mathbf{B}$, ${}^3\mathbf{P}_0$, ${}^3\mathbf{P}_1$ are detected.

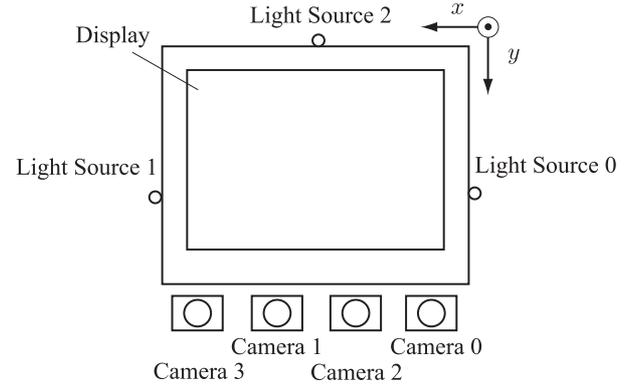


Fig. 15 Arrangement of the three light sources and the four cameras.

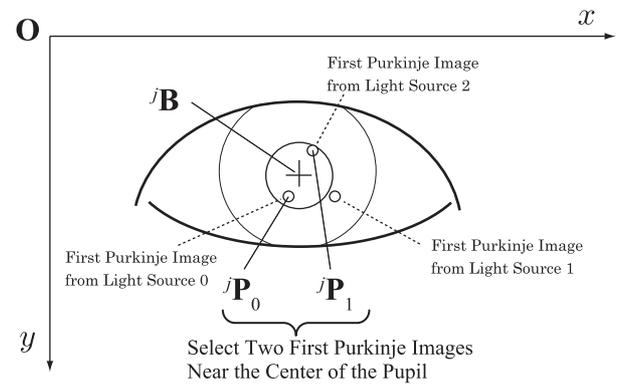


Fig. 16 Eye image with three first Purkinje images in the camera j coordinate system.

5.3 Experiment of Our Algorithm

We conducted an experiment to prove the feasibility of our algorithm for one subject, who did not wear glasses or contact lenses. The subject was asked to fixate on the fiducial point that was displayed at the center of the display. In addition to the recorded data described in 4.2, we recorded whether the POAs were detected or not.

In the experiment, we hid cameras 3 and 1, in that order, by an occluder. The result is shown in Fig. 17. The data is plotted continuously, but the experiment was conducted incorporating a pause every 30 frames; each frame needs about 200 ms. The horizontal axes in Fig. 17 (a) and (b) are the number of frames from the beginning of the experiment. The vertical axes in Fig. 17 (a) and (b) are the x and y axes, respectively, in the display coordinate system. If the POAL was not detected, the POAL at that time was not plotted in the figure; the same goes to the POAR and POG. According to Fig. 17, the POAL was not detected from frames 39 to 56; at that time, camera 3 was hidden by the occluder. According to Fig. 17, the POAR was not detected from frames 68 to 82; at that time, camera 1 was hidden by the occluder. We can see that the POG was always around $(x, y) = (640, 512)$, whereas one of the POAs was not detected. This result shows that the effectiveness of

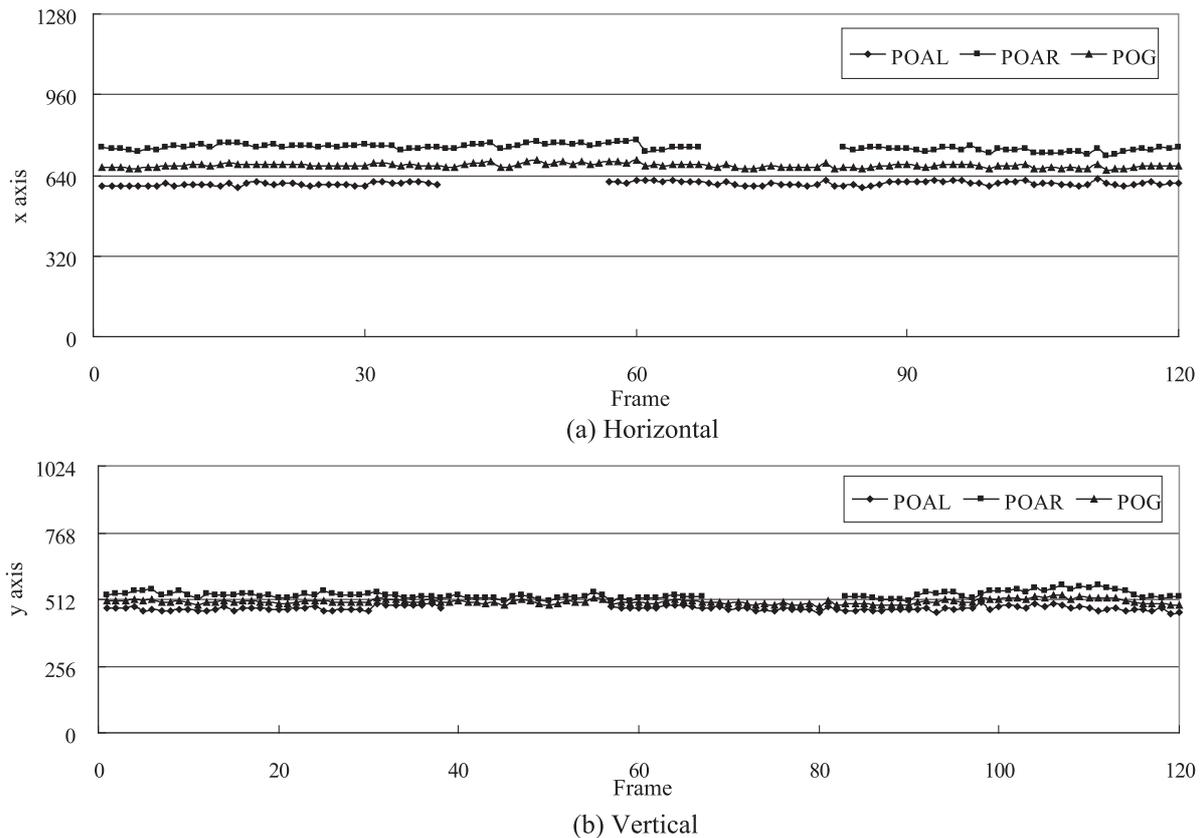


Fig. 17 Estimation of the POG when POSs were partially undetected.

the proposed method.

6. Related Works and Discussion

In this paper, we describe our user-calibration-free gaze tracking method. From the results of the experiment described in Sect. 4, our method had greater accuracy as compared to the optical axis method, while keeping the calibration time down to zero, i.e., the calculation can be performed from the moment the system begins working.

Recently, Model et al. proposed an automatic personal calibration procedure [11], [12] that does not require active user participation. The RMSE of their system was 1.3° on average for four subjects. However, 1000 POGs are used for convergence calculations to estimate the four parameters. If the system processes 30 frames per second to get POGs, this automatic calibration requires 33 seconds to get calibration results. Furthermore, Model's method seems to be difficult to get good estimation results without instructions for the user to gaze widely-distributed direction, at this stage. Comparing our method to the Model's method, we find that while our method is less accurate, it has the advantages of speed and stability.

Although there are many gaze trackers whose accuracy is less than 1.0° [3]–[6], they require active user participation to calibrate the system. User-calibration-free gaze tracker is completely different in a scene of application from

the existing gaze trackers which require user-calibration. Because our system is user-calibration-free and the accuracy is quite good (1.58°), our method has an advantage in applications meant for the general public, i.e., applications in which the user glances at or gazes at for a very short time, such as those used for marketing analysis, for recommendations in a shop window, for providing descriptions of museum displays, for guideboards at tourist spots, etc.

Our current prototype system has a narrow range in which the user can move. Therefore, the system can now be applied in scenarios where the user is almost immobile. In order to discuss issues associated with the practical use of the proposed method, we conducted a brief experiment to evaluate our method subject to changes in the resolution of the eye image. For that purpose, we converted the four camera images to a mosaic image; the mosaic size is $n \times n$ ($n = 1 \sim 10$). Figure 18 shows the result of the experiment for one subject. The horizontal and vertical axes of Fig. 18 (a) and (b) are the same as those of Fig. 17. Every 30 data points correspond to a mosaic size of 1×1 (original image), 2×2 , ..., and 10×10 . The data are plotted continuously, but the experiment was conducted separately with brief pauses. The diameter of the pupil in each camera was about 60 pixels. We can see that, as the resolution of the camera images becomes lower, the deviation becomes larger. From Fig. 18, the limit within which the system can operate stably seems to be constrained to a mosaic size of

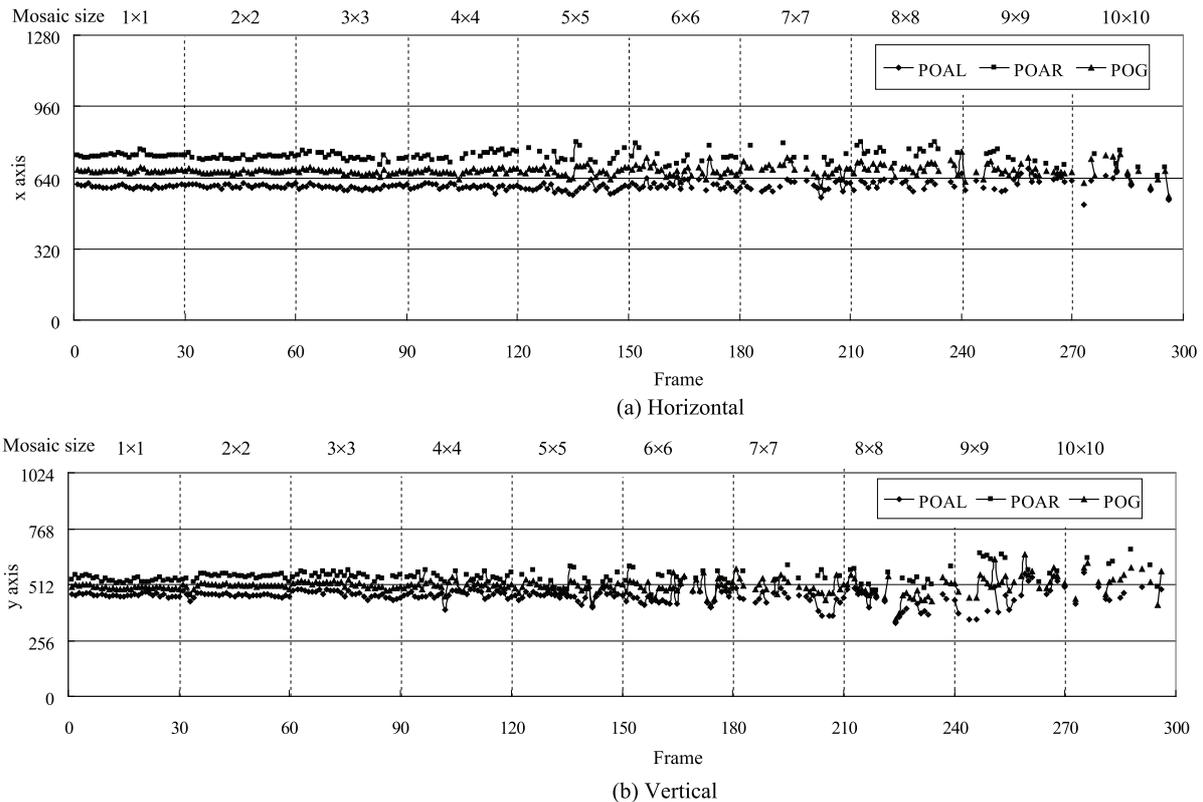


Fig. 18 Estimation of the POG, POAL, and POAR when the mosaic size (resolution) is changed.

4×4 . Since the diameter of the pupil in each camera image is about 60 pixels for our prototype system, we think that the resolution range in which the method can work with acceptable errors is with a pupil diameter at least 15 pixels. If we use a wide field of view lens, the user can move over a range that is 4 times the size of that supported by the current system. There are some approaches that can allow the user to move over a wider area, under the condition where the pupil size is 15 pixels. These include: (1) using a wide field view lens and a high resolution camera, (2) using a pan-tilt unit for tracking eyes, and (3) using many cameras to capture the entire area where the user moves.

7. Conclusion

In this paper, we have proposed a method to enable user-calibration-free gaze tracking using a binocular eye model. This method provides a fast and stable theoretical solution. The POG is estimated as the midpoint of the line joining the POAs of both eyes. From the POG, we calculate the offsets for the optical and visual axes of both eyes. Our method can continue to estimate the POG while the system detects only the optical axis of one eye.

We have developed a prototype system on the basis of the method. The system uses two pairs of stereo cameras. The optical axes of the left eye and the right eye are estimated using two cameras each. We evaluated it experimentally with 20 subjects. The average of the RMSE of the POG

was 16.55 mm for a distance of 600 mm from the eye, which is equivalent to less than 1.58° in terms of the view angle. Our method was better than the optical axis methods.

Acknowledgments

This work was partially supported by KAKENHI (23300047).

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] A.T. Duchowski, *Eye Tracking Methodology: Theory and Practice*, 2nd ed., Springer-Verlag, 2007.
- [3] S.W. Shih and J. Liu, "A novel approach to 3-D gaze tracking using stereo cameras," *IEEE Trans. Syst. Man Cybern. B, Cybern.* vol.34, no.1, pp.234–245, 2004.
- [4] 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.
- [5] 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.
- [6] A. Villanueva and R. Cabeza, "A novel gaze estimation system with one calibration point," *IEEE Trans. Syst. Man Cybern. B, Cybern.* vol.38, no.4, pp.1123–1138, 2008.
- [7] R. Osaka, *Experimental Psychology of Eye Movements (in Japanese)*, The University of Nagoya Press, Nagoya, Japan, 1993.

- [8] 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.
- [9] S.W. Shih, Y.T. Wu, and J. Liu, "A calibration-free gaze tracking technique," Proc. International Conference on Pattern Recognition, pp.201–204, 2000.
- [10] S. Kohlbecher, S. Bardinst, K. Bartl, E. Schneider, T. Poitschke, and M. Ablassemeier, "Calibration-free eye tracking by reconstruction of the pupil ellipse in 3D space," Proc. 2008 Symposium on Eye Tracking Research & Applications, pp.135–138, 2008.
- [11] D. Model and M. Eizenman, "An automatic personal calibration procedure for advanced gaze estimation systems," IEEE Trans. Biomed. Eng., vol.57, no.5, pp.1031–1039, 2010.
- [12] D. Model and M. Eizenman, "User-calibration-free remote gaze estimation system," Proc. 2010 Symposium on Eye-Tracking Research & Applications, 1743672 29-36, Austin, Texas, ACM, 2010.
- [13] T. Nagamatsu, J. Kamahara, and N. Tanaka, "Calibration-free gaze tracking using a binocular 3D eye model," Proc. 27th International Conference Extended Abstracts on Human Factors in Computing Systems, pp.3613–3618, 2009.
- [14] T. Nagamatsu, R. Sugano, Y. Iwamoto, J. Kamahara, and N. Tanaka, "User-calibration-free gaze tracking with estimation of the horizontal angles between the visual and the optical axes of both eyes," Proc. 2010 Symposium on Eye-Tracking Research & Applications, pp.251–254, 2010.
- [15] 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.
- [16] D. Atchison and G. Smith, Optics of the Human Eye, Butterworth-Heinemann, 2000.
- [17] T. Nagamatsu, J. Kamahara, and N. Tanaka, "A model of rotation of the visual and optical axes of the eyeball on the basis of Listing's law, and gaze-estimation and evaluation method using the model," IEICE Trans. Inf. & Syst. (Japanese Edition), vol.J93-D, no.4, pp.511–521, April 2010.
- [18] D. Tweed, "Visual-motor optimization in binocular control," Vision Research, vol.37, no.14, pp.1939–1951, 1997.
- [19] A.M.F. Wong, "Listing's law: Clinical significance and implications for neural control," Survey of Ophthalmology, vol.49, no.6, pp.563–575, 2004.
- [20] Y. Matsumoto, T. Ogasawara, and A. Zelinsky, "Behavior recognition based on head pose and gaze direction measurement," Proc. 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp.2127–2132, 2000.
- [21] Intel, "Open source computer vision library," <http://sourceforge.net/projects/opencvlibrary/>



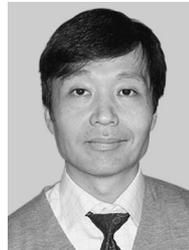
Ryuichi Sugano received the Bachelor degree of Maritime Sciences from Kobe University, Japan, in 2010. He is currently a master course student in Graduate School of Maritime Sciences, Kobe University. He is interested in gaze interaction, and so on.



Yukina Iwamoto received the Bachelor degree of Maritime Sciences from Kobe University, Japan, in 2010. She joined Mitsubishi Electric Corporation in 2010.



Junzo Kamahara received the B.E. and M.E. degrees from Osaka University in 1992 and 1994, respectively and Ph.D. degree in Engineering from that university in 1999. He joined Kobe University of Mercantile Marine in 1996 as an assistant and became a lecturer and an associate professor at that university in 2000 and 2003, respectively. He was a visiting associate professor at Minnesota University during 2003 and 2004. He is presently an associate professor at Kobe University. He is involved in research on multimedia technology, information retrieval, and so on.



Naoki Tanaka received his B.S. degree in communication engineering from Osaka University in 1981 and D.Eng. degree from that university in 1986. He joined Kinki University in 1986 as an assistant and became an associate professor at Kobe University of Mercantile Marine in 1990. He was a visiting associate professor at Washington University during 1992 and 1993. He is presently a professor at Kobe University. He is involved in research on pattern recognition, image processing, and so on.



Takashi Nagamatsu received the B.E. and M.E. degrees in electrical engineering from Kyoto University, Japan, in 1994 and 1996, respectively. In 1996 he joined Mitsubishi Heavy Industries Ltd. In 1999 he joined Kyoto University as a research associate. In 2000 he moved to Kobe University of Mercantile Marine and is presently an assistant professor at Kobe University. He received the Ph.D. degree in energy science from Kyoto University, Japan, in 2004. He is involved in research on human-computer in-

teraction, and so on.