Screen corner detection using polarization camera for cross-ratio based gaze estimation

Sasaki, Masato
Nagamatsu, Takashi
Takemura, Kentaro

(Citation)
ETRA '19: Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications:1-9

(Issue Date)
2019-06

(Resource Type)
conference paper

(Version)
Accepted Manuscript

Screen Corner Detection using Polarization Camera for Cross-Ratio Based Gaze Estimation

Masato Sasaki
Tokai University
Hiratsuka, Kanagawa, Japan
masato-s@takemura-lab.org

Takashi Nagamatsu
Kobe University
Kobe, Hyogo, Japan
nagamatu@kobe-u.ac.jp

Kentaro Takemura
Tokai University
Hiratsuka, Kanagawa, Japan
takemura@tokai.ac.jp

ABSTRACT

Eye tracking, which measures line of sight, is expected to advance as an intuitive and rapid input method for user interfaces, and a cross-ratio based method that calculates the point-of-gaze using homography matrices has attracted attention because it does not require hardware calibration to determine the geometric relationship between an eye camera and a screen. However, this method requires near-infrared (NIR) light-emitting diodes (LEDs) attached to the display in order to detect screen corners. Consequently, LEDs must be installed around the display to estimate the point-of-gaze. Without these requirements, cross-ratio based gaze estimation can be distributed smoothly. Therefore, we propose the use of a polarization camera for detecting the screen area reflected on a corneal surface. The reflection area of display light is easily detected by the polarized image because the light radiated from the display is polarized linearly by the internal polarization filter. With the proposed method, the screen corners can be determined without using NIR LEDs, and the point-of-gaze can be estimated using the detected corners on the corneal surface. We investigated the accuracy of the estimated point-of-gaze based on a cross-ratio method under various illumination and display conditions. Cross-ratio based gaze estimation is expected to be utilized widely in commercial products because the proposed method does not require infrared light sources at display corners.

CCS CONCEPTS

• **Human-centered computing** → *Human computer interaction (HCI)*; • **Computing methodologies** → *Computer vision tasks*;

KEYWORDS

Eye gaze estimation; Cross-ratio method; Polarized image

ACM Reference Format:
Masato Sasaki, Takashi Nagamatsu, and Kentaro Takemura. 2019. Screen Corner Detection using Polarization Camera for Cross-Ratio Based Gaze Estimation. In *2019 Symposium on Eye Tracking Research and Applications (ETRA '19), June 25–28, 2019, Denver , CO, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3314111.3319814

ETRA '19, June 25–28, 2019, Denver , CO, USA

ACM ISBN 978-1-4503-6709-7/19/06…$15.00
https://doi.org/10.1145/3314111.3319814

1 INTRODUCTION

Eye tracking methods have been proposed for a wide variety of devices, such as desktop monitors, TVs, head-mounted displays, automobiles, and wearable devices [Kar and Corcoran 2017]. Diverse gaze estimation methods have been developed, such as infrared-based and non-infrared-based approaches. In commercial products, near-infrared (NIR) light-emitting diodes (LEDs) are often used for eye tracking, and the pupil center and the first Purkinje image are tracked as features for achieving highly accurate estimations of the so-called point-of-gaze. These methods are classified into two types: regression-based (or interpolation-based) methods, and three-dimensional (3D) model-based methods [Hansen and Qiang Ji 2010]. Regression-based methods [Ji and Yang 2002; Morimoto and Mimica 2005] estimate the point-of-gaze with a regression formula that uses a pupil-center corneal-reflection (PC-CR) vector. These methods do not require strict hardware calibration, because regression formulas are calculated exclusively using the pupil center and corneal reflection. However, these techniques require user calibration—specifically, that the user stares at several points in advance—and the estimated point-of-gaze is sensitive to head movements. On the other hand, model-based methods [Beymer and Flickner 2003; Guestrin and Eizenman 2006] consider a 3D geometric model of the human eye to compute the optical axis and visual axis, and the point-of-gaze is estimated as the intersection between the vector and an object such as the screen. These methods merely require brief user calibration, and user-calibration-free methods have been reported [Nagamatsu et al. 2009]. Additionally, these methods are tolerant of head movements. However, proper hardware calibration is needed in order to calculate the relationship between the eye camera and the screen when computing the intersection as the point-of-gaze, and users cannot adjust the location of the eye camera and the screen by themselves.

As a method different from the ones using NIR LEDs, Yoo et al. [2002] proposed a cross-ratio (CR) based method, which uses IR-LEDs located at the screen corners. The advantages of the CR method are no hardware calibration requirement and allowing user's head movements. However, at least four IR-LEDs are required to estimate the point-of-gaze for CR-based gaze estimation. In other words, the system configuration includes a specific monitor to which IR-LEDs are attached. Nevertheless, we expect the CR method to be used widely in commercial products, and as such, the eye tracking system should be separated from a specific monitor.

Therefore, we propose CR gaze estimation using a polarization camera to detect the screen corners on the corneal surface precisely without IR-LEDs attached to the screen corners. Most of liquid crystal displays (LCDs) include a polarization filter, and the screen area can be extracted stably using polarized light. We implemented

two approaches to extracting the pupil center or the limbus center for CR gaze estimation using the polarized image. When the limbus center is employed as a feature, IR emission is not required. When an IR-LED is used to detect the pupil, it is expected to estimate the point-of-gaze with high accuracy. If the use of CR-based gaze estimation in various conditions is facilitated, we expect eye tracking to be used widely in user interfaces because of its easy setup. We conducted an experiment to compare these two approaches, and we confirmed the feasibility of using the polarization camera for eye tracking under several conditions.

2 RELATED WORKS

Yoo et al. [2002] proposed a pioneer eye tracking method using the relationship of the cross-ratio between the screen and the reflections on the cornea. In recent CR methods, NIR LEDs are attached to screen corners to compute the two homography matrices [Hartley and Zisserman 2000]. Assuming that the reflections of IR lights on the cornea are coplanar, the transformation from the image plane to the screen plane is expressed by a homography transformation. The homography matrix is calculated using the IR-LED's reflections on the cornea. This method has two advantages: hardware calibration is unnecessary, and it is tolerant of head movements. However, CR methods suffer from the following problems: they assume that reflections on the cornea and pupil center are on the same plane, and the optical axis of the eye computed by the pupil center is used as the line of sight. Thus, many researchers have proposed methods to improve accuracy. Yoo and Chung [2005] proposed an improved CR method in which the virtual tangent plane is located on the corneal surface, and the reflection is projected on the virtual tangent plane as the coplanar for reflections before computing the homography transformation. Coutinho and Morimoto [2006] compensated for the difference between the visual axis and the optical axis of the eye with the CR method, and Kang et al. [2007] and Hansen et al. [2010] proposed a homography-based correction, which calculates a correction matrix through user calibration to improve the accuracy. Arar et al. [2015a] proposed a CR method with a few number of calibration points by using regularized least-squares regression. These bias-correction methods are successful, but the accuracy of the point-of-gaze decreases when the user's head moves from the location where the calibration was performed.

Tolerance to head movement has been considered in the latest research, and Coutinho and Morimoto [2013] proposed two solutions(CR-DD and PL-CR) to pronounced head movement: CR-DD considers the distance between the camera and the eye using the quadrilateral size of reflections on the corneal surface to improve depth compensation with the CR method, and PL-CR compensates for head movement using a weak-perspective camera model and an eye model consisting of the cornea center and iris plane. On the other hand, Huang et al. [2014] reported that PL-CR requires a camera with considerable focal length, and that the method has a limited field-of-view. Thus, they proposed a head-movement correction method using an adaptive homography matrix obtained using simulation data without a weak-perspective camera model. Furthermore, Zhang and Cai [2014] proposed a CR binocular fixation method, which calculates homography matrices from both eyes jointly for improved accuracy. Arar et al. [2015b] also proposed

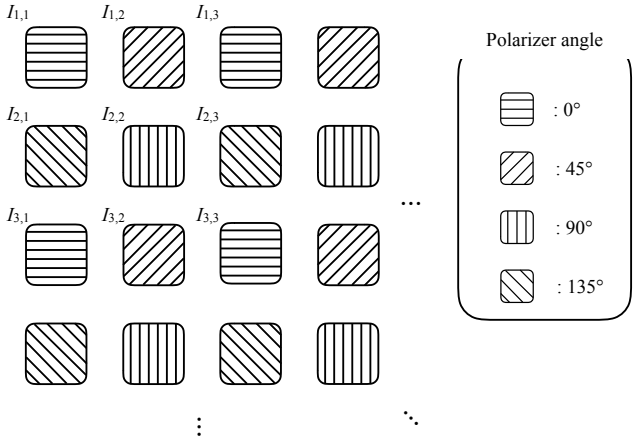

Figure 1: Overview of polarizing image sensor

CR-based gaze estimation with multiple cameras for improvement of head movements. In these head-movement compensation approaches, sometimes reflections of NIR LEDs exist on the sclera as changing the relationship between an eye, a camera, and light sources, and it is difficult to detect the reflections on the sclera in the image. Recent studies [Huang et al. 2014; Zhang and Cai 2014] have used more than NIR LEDs around the display to detect the reflections on the corneal surface robustly for the accuracy of gaze estimation.

Therefore, CR-based gaze estimation has been improved actively, and there is considerable potential for the method. However, CR methods require more than four IR-LEDs to detect screen corners. Therefore, it is necessary to prepare a specific monitor with NIR LEDs. We believe that the eye-tracker should be independent of the monitor for general versatility. Hansen et al. [2010] tried to implement a CR method without IR-LEDs, but the extracted screen area depends on the textures displayed on the monitor. Therefore, we focus on the polarization filter as a solution to extracting the display area without IR emission. Rädle et al. [2018] proposed PolarTrack, which detects the display area in an image using an RGB camera and polarization filter with a motor to rotate the polarization angle of the filter. The light radiated by the LCD is usually linearly polarized because most LCDs have an internal polarization filter. Therefore, we propose a CR method using polarization in optics to extract screen corners without near-infrared LEDs on the display.

3 METHODOLOGY

In CR methods, the reflections of screen corners are detected using IR-LEDs attached to the display and the IR camera. In contrast to the general approach, we have proposed a polarization camera system to detect the screen area without IR emission[Sasaki et al. 2018]. The camera system consists of two cameras with polarization filters, a camera lens, an optical breadboard, and optomechanical components. However, this system is quite large. Consequently, we employed a polarization camera (PHX050SC, Lucid Vision Labs Inc., Richmond, B.C., Canada) to capture the image. This camera includes a specific gray scale image sensor (IMX250MZR, Sony Corporation, Minato-ku, Tokyo, Japan) that has a nanowire grid

polarizer for each pixel. Figure 1 shows an overview of this sensor. A unit of the polarizer consists of four pixels, and each polarizer has a different polarization angle. Linear polarized lights through the polarizer array have different intensities. Hence, the polarization information can be derived from the value of the intensity from each unit. The method for computing the polarization information is described in the following section.

3.1 Calculating polarization images

First, the polarization angle image is computed from the image captured by the polarization camera, as shown in Figure 2(a). It looks like a general monochrome image. The brightness between neighboring pixels is different due to the polarized light in the polarization camera. Figure 2(b) shows the enlarged image of a part of the original polarized image; a pattern that depends on the layout of the polarization filter can be observed. When each pixel value has a different polarizer angle, there are two approaches to generate the polarization angle image: substitution or compensation. According to the substitution approach, the resolution of the polarization angle image is one-quarter of the captured image. On the other hand, the size of the polarization angle image and the captured image are the same with the compensation approach. Huang et al. [2014] reported that the accuracy of CR-based gaze estimation depends on the resolution. Therefore, we employed the compensation approach in this study to compute the polarization angle image. The computation of a polarization angle of zero degrees is provided as an example. The pixel value, which does not have a value filtered by zero degrees, is calculated by interpolating the neighboring pixel value as follows:

$$I_{1,2}(0^\circ) = (I_{1,1}(0^\circ) + I_{1,3}(0^\circ))/2, \quad (1)$$

$$I_{2,1}(0^\circ) = (I_{1,1}(0^\circ) + I_{3,1}(0^\circ))/2, \quad (2)$$

$$I_{2,2}(0^\circ) = (I_{1,1}(0^\circ) + I_{3,3}(0^\circ))/2, \quad (3)$$

where $I_{i,j}(0^\circ)$ is the intensity of pixel (i, j) through the 0° polarization filter. In the case of other polarization angles (45, 90, 135 degrees), the same procedure can be applied to the whole image.

Second, the Stokes parameters are calculated from the captured image. The Stokes parameters have four values to describe the polarization properties of the light wave. Sony's CMOS sensor can calculate three Stokes parameters defined for each pixel as follows:

$$S_0(i,j) = I_{i,j}(0^\circ) + I_{i,j}(90^\circ), \quad (4)$$

$$S_1(i,j) = I_{i,j}(0^\circ) - I_{i,j}(90^\circ), \quad (5)$$

$$S_2(i,j) = I_{i,j}(45^\circ) - I_{i,j}(135^\circ). \quad (6)$$

Two kinds of images can be computed from these Stokes parameters: a degree of linear polarization (DoLP) image, and an angle of linear polarization (AoLP) image. DoLP and AoLP images describe the intensity of polarization in a polarized image and the angle of linear polarization of each pixel, respectively. The DoLP and AoLP are calculated as follows:

$$DoLP(i,j) = \frac{\sqrt{S_1(i,j)^2 + S_2(i,j)^2}}{S_0(i,j)}, \quad (7)$$

$$AoLP(i,j) = \frac{1}{2}\arctan\left(\frac{S_2(i,j)}{S_1(i,j)}\right). \quad (8)$$

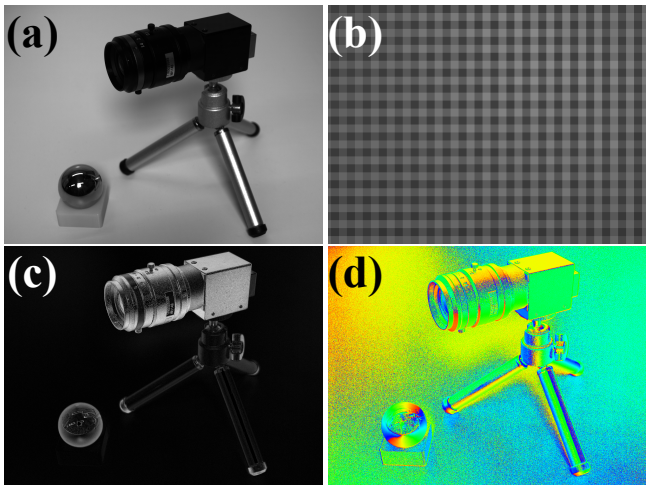

Figure 2: Polarization information: (a) original polarized image, (b) enlarged image of a part of original polarized image, (c) DoLP image, and (d) AoLP image

Figure 2(c) and (d) show the calculation result of the polarized image of a camera and a spherical mirror. The reflected light that reaches the specular surface is polarized depending on an incidence angle also called Brewster's angle. The shape can be confirmed easily by using the DoLP image, as shown in Figure 2(c). On the other hand, the AoLP image shows the angle of polarization that describes 0 to 180°, and it is possible to comprehend the angle of polarization directly as a color. The spherical mirror's surface is curved and its AoLP value changes smoothly because the AoLP value depends on the vector normal to the object's surface. Therefore, the spherical mirror's color in Figure 2(d) appears iridescent.

3.2 Detecting screen corners

CR-based gaze estimation requires at least four points on the screen to compute homography matrices. Hansen et al. [2010] used a standard camera to extract the screen area, and their approach uses a binary image of the visible spectrum image. When a white background is displayed on the screen, the screen area can be extracted using an visible spectrum image, but it is uncertain how well this works in complex backgrounds. Therefore, we employ polarization to detect screen corners. LCDs generally include a linear polarization filter to adjust the emission of screen light, and the light radiated from the display is polarized. Therefore, the intensity of the reflection on the cornea can be controlled by an external polarization filter attached to the camera. The polarization camera detects the area of the screen using the polarization image on the corneal surface. Figure 3 (a) shows an image captured by the polarization camera. The screen area is detected using the DoLP and AoLP images computed by the polarization camera. Figure 3(b) respectively shows DoLP image of the eye when the display is reflected on the cornea. The DoLP value of the display light reflected on the cornea is higher than others because the light beam radiated by the display is polarized linearly. The reflection area on the corneal surface is detected using a binary image of the DoLP image and the limbus

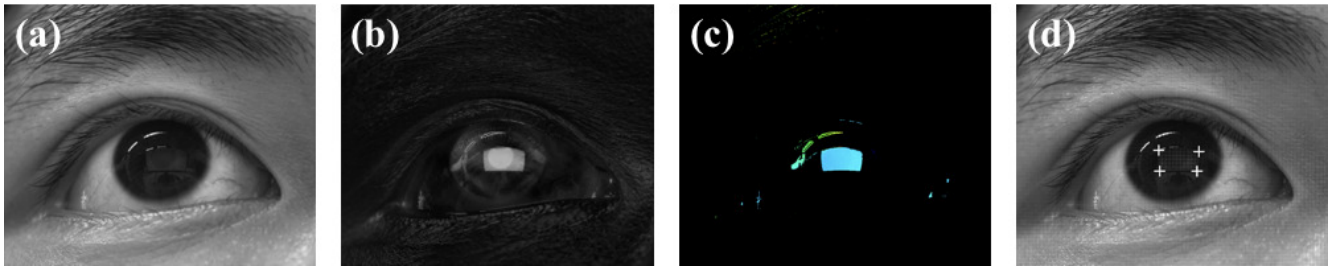

Figure 3: Detecting screen corners: (a) original image, (b) DoLP image, (c) AoLP image masked by binary image of (b), and (d) detection result

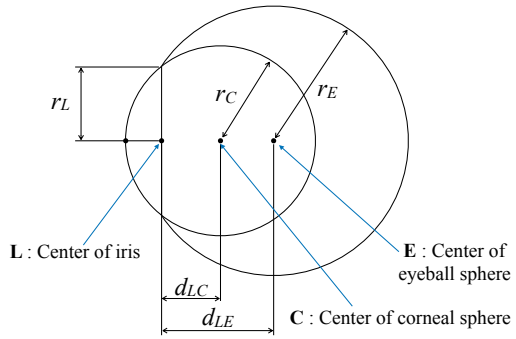

Figure 4: 3D eye model composed of an eyeball sphere and a corneal sphere

(or pupil) center coordinates because the limbus (or pupil) center is close to the display area. Figure 3(c) shows the AoLP image, which is masked by the binary image of the DoLP image. As the light radiated from the display has the same polarization angle due to linear polarization, the computed AoLP values of the screen area on the corneal surface are nearly equal to the polarization angle of light emitted from the display. Therefore, the screen area in Figure 3(c) is shown using almost the same color. Figure 3(d) shows the result of the screen corner detection; the display corners are detected correctly.

3.3 Iris tracking using 3D eye model

CR-based gaze estimation requires tracking the pupil center to estimate the point-of-gaze. In a visible image, the pupil cannot be observed clearly, so we detect the iris center using the 3D eyeball model shown in Figure 4 in place of the pupil center. The model proposed in the previous work [Takemura et al. 2014] is employed, and the eyeball model consists of a corneal sphere and an eyeball sphere. The corneal surface protrudes from the eyeball sphere, and the eyeball parameters are defined as shown in Table 1 using anatomical data [Snell and Lemp 1989].

First, the geometric relationship between the camera and the eye model is computed using the limbus in the image. Because the limbus size is constant in contrast to the pupil, the limbus center **L** in 3D space is expressed as follows:

$$\mathbf{L} = \left[d\frac{i_{Lx} - c_x}{f}, d\frac{i_{Ly} - c_y}{f}, d \right]^T, \tag{9}$$

Table 1: Anatomical parameters of a 3D eye model

Description	Parameter	Size [mm]
Radius of eyeball sphere	r_E	12.1
Radius of corneal sphere	r_C	7.8
Radius of limbus	r_L	5.6
Distance between the center of the limbus and the center of the eyeball sphere	d_{LE}	10.8
Distance between the center of the limbus and the center of the cornea	d_{LC}	5.6

where (i_{Lx}, i_{Ly}) is the computed ellipse center of the limbus,and the focal length f and the image center (c_x, c_y) are acquired during camera calibration in advance. The distance d between the image plane and the 3D limbus center **L** is calculated using the radius of limbus r_L, as follows:

$$d = f\frac{r_L}{r_{max}}, \tag{10}$$

where r_{max} denotes the major axes of the limbus ellipse. The optical axis **g** of the eye model is defined using the angle of the 3D iris plane τ as

$$\mathbf{g} = [\sin(\tau)\sin(\phi), -\sin(\tau)\cos(\phi), -\cos(\tau)]^T, \tag{11}$$

where ϕ is the angle of the limbus ellipse in the image. The angle of the iris plane τ is calculated from the limbus ellipse as follows:

$$\tau = \pm\cos^{-1}\left(\frac{r_{min}}{r_{max}}\right), \tag{12}$$

where r_{min} denotes the minor axes of the limbus ellipse.

The center of the corneal sphere in 3D space is computed by the optical axis **g** and the distance d_{LC} between the 3D limbus center **L** and corneal sphere center **C**, as follows:

$$\mathbf{C} = -d_{LC}\frac{\mathbf{g}}{\|\mathbf{g}\|} + \mathbf{L}. \tag{13}$$

The center of the eyeball sphere **E** is calculated as

$$\mathbf{E} = -(d_{LE} - r_C)\frac{\mathbf{g}}{\|\mathbf{g}\|} + \mathbf{C}, \tag{14}$$

where d_{LE} is the distance between the center of the eyeball sphere **E** and the limbus center **L** defined in Table 1. Figure 5 shows the estimated eyeball pose in the camera coordinate as the initialization.

Second, the iris is tracked using the initialized 3D eyeball model. Because the color of the iris area is darker than the sclera, the inverted binary image **B** of the captured image is computed as shown in Figure 6(a). When it is assumed that the relationship

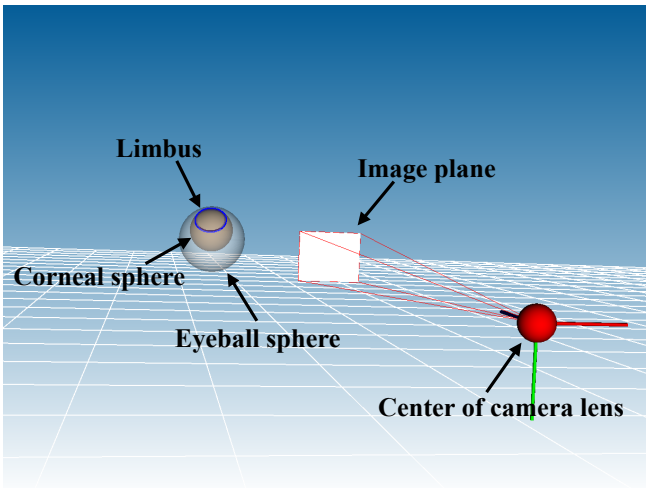

Figure 5: Relationship between the eye and the camera

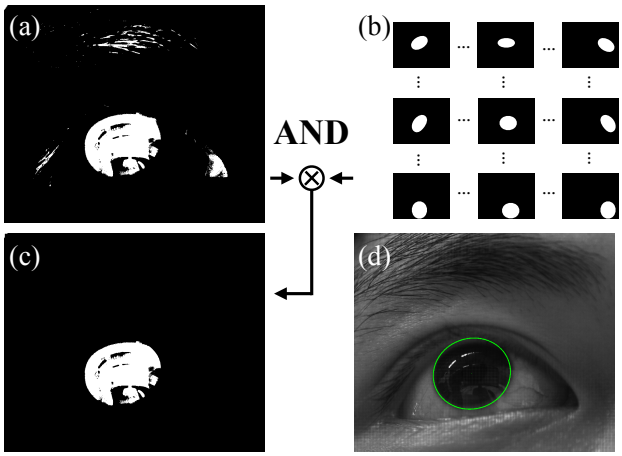

Figure 6: Algorithm for iris tracking using a 3D eye model: (a) inverted binary image of input image, (b) simulated images of the iris area using a 3D eye model, (c) the logical product of these binary images, and (d) the estimated result of iris tracking

between the image plane and the eyeball center is constant, the iris area of any eyeball pose can be projected onto the image plane. The projected iris area $\mathbf{V}$ can be computed as shown in Figure 6(b). When the sum of the logical conjunction between the projected iris area and the inverted binary image reaches its maximum, the pose $(\hat{\theta}, \hat{\psi})$ of the eye is estimated as follows,

$$(\hat{\theta}, \hat{\psi}) = \arg\max_{\theta,\psi} \sum_{j=1}^{m} \sum_{i=1}^{n} \mathbf{B}_{ij} \mathbf{V}_{\theta\psi ij}. \quad (15)$$

Figure 6(c) shows the calculated image of the sum of the logical conjunction. Once the pose of the eye is determined, the limbus of the eye model can be projected onto the image plane as the estimated limbus, as shown in Figure 6(d).

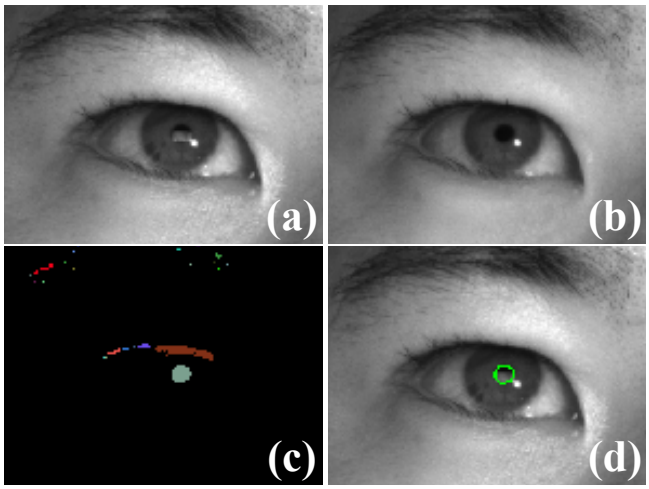

Figure 7: Two polarized images (a) and (b), which have different polarization angles, are generated from an original polarized image. The image (c) is generated by binarization and connected-component labeling from the image without screen reflection. The pupil area (d) is detected by using the degree of circularity.

3.4 Pupil detection using infrared image

Iris tracking is described above, but it faces a problem: occlusion by the eyelid. It is known well that pupil tracking works better than iris tracking. Therefore, we also propose CR-based gaze estimation using screen corner reflection with a NIR LED module for pupil tracking. Sony's sensor can capture the range of NIR light, and thus, pupil tracking can proceed using the same camera. When an IR light source is attached far from the camera, a dark pupil image can be captured as shown in Figure 7(a) and (b). Figure 7(a) and (b) are generated for the different angles of polarization (0 and 90°). These images are computed from the original image captured using the calculated DoLP and AoLP images. Because the polarization angle of the light radiated from the display is 0°, the screen can be observed as a reflection in Figure 7(a). On the other hand, the reflection area of the screen is removed in Figure 7(b), because the image is generated using the pixel that has a polarization angle of 90°. Screen reflection is an outlier in pupil tracking, and therefore we use the polarized image, which has a polarization angle of 90°, to detect the pupil precisely. Each blob area is labeled on the binary image converted from the 90° polarized image as shown in Figure 7(c). The degree of the circularity of all labeled areas is calculated, and the area with the highest degree of circularity is detected as the pupil. Figure 7(d) shows a detected pupil area. The center of the pupil can also be used to estimate the point-of-gaze.

3.5 Cross-ratio based gaze estimation

The CR method was employed to estimate the point-of-gaze on a screen in our study. One of the advantages of the CR method is that hardware calibration is not required. The CR method solves the relationship between the screen and the eye using homography transformation. Thus, it is able to estimate the point-of-gaze using

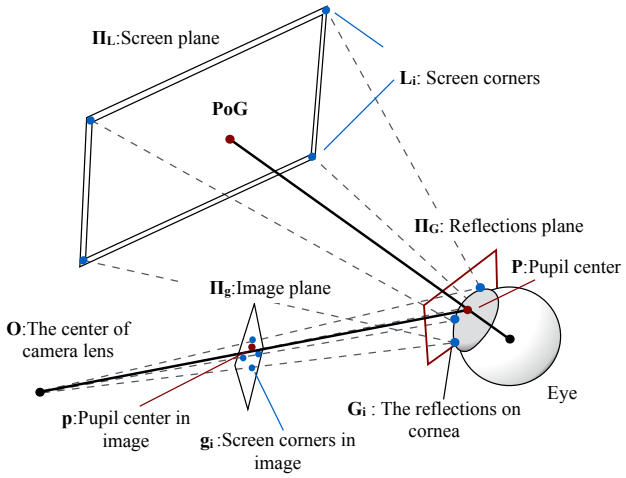

Figure 8: Geometric setup of cross-ratio based gaze estimation

the iris (or the pupil) center and the screen corners in an image. Previous CR methods have used at least four IR-LEDs attached to the screen to detect corners, but screen corners can be extracted without IR-LEDs using our proposed method. Figure 8 illustrates the geometric relationship with CR-based gaze estimation, and the screen plane Π_L, the reflections plane Π_G, and the image plane Π_g are considered when estimating the point-of-gaze **PoG**. Assuming that reflections on the cornea and the pupil center are coplanar, the point-of-gaze is estimated by homography matrices from Π_g to Π_G and Π_G to Π_L. The relationships from the reflected screen corners $\mathbf{g_i}$ on the image plane to the reflected screen corners $\mathbf{G_i}$ on the cornea and the reflected screen corners $\mathbf{G_i}$ on the cornea to the screen corners $\mathbf{L_i}$ are defined as $\mathbf{H_{gG}}$ and $\mathbf{H_{GL}}$ as homography matrices, respectively. For homography transformation, the relationship of the cross-ratio is invariant. The CR method is based on the projective transformation between planes, and thus, the point-of-gaze **PoG** is expressed using these matrices and the iris (or pupil) center $\mathbf{p}$ in the image coordinate:

$$\begin{aligned} \mathbf{PoG} &= \mathbf{H_{GL}}(\mathbf{H_{gG}}(\mathbf{p})) \\ &= CRf(\mathbf{g_i}, \mathbf{p}). \end{aligned} \tag{16}$$

Homography matrices are calculated using at least four reflections on the image plane. Thus, **PoG** is defined using the CR method function *CRf*. The *CRf* function combines two steps: calculating matrices using the four reflections in the image, and estimating the gaze from these calculated matrices and the coordinates of the pupil center on the image plane. The point-of-gaze is estimated by substituting the reflections $\mathbf{g_i}$ and the pupil center $\mathbf{p}$ for *CRf*. However, the visual axis is not considered, and pupil center and reflections on cornea is not coplanar in the CR-based gaze estimation to estimate the point-of-gaze. Therefore, correction is required from the pupil center as a offset, and homography-based correction [Hansen et al. 2010; Kang et al. 2007] is employed. The point-of-gaze of the CR method with homography-based correction $\mathbf{PoG_{HOM}}$ is expressed as

$$\mathbf{PoG_{HOM}} = \mathbf{H_{LL}}(CRf(\mathbf{g_i}, \mathbf{p})), \tag{17}$$

where $\mathbf{H_{LL}}$ is the correction matrix calculated during the user calibration, in which a user stares at some calibration points. This approach corrects the error of *CRf* based on the user position during calibration using homography matrix. Homography-based correction requires at least four calibration points to calculate the correction matrix.

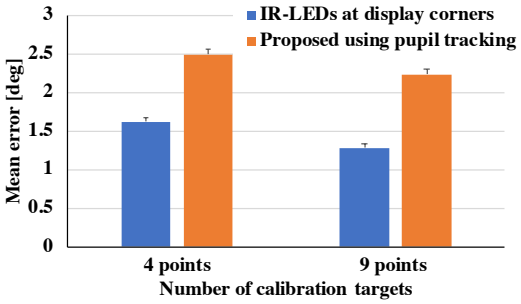

Figure 9: Mean error of gaze estimation with corner detection using IR-LEDs and polarization information

4 EVALUATION

4.1 Evaluation of screen-corner detection

Several experiments were performed to confirm the feasibility of the proposed method for detecting display corners using a polarization camera. First, we compared the proposed method using pupil tracking to a previous method that uses IR-LEDs, in order to evaluate the performance of the screen corner detection. It is difficult to evaluate the accuracy of screen detection directly. Thus, the point-of-gaze on the display was estimated for comparison, because the accuracy of the point-of-gaze depends on the accuracy of screen detection.

A 21.5-inch LCD, on which the pattern of the calibration image was indicated, was located 600 mm in front of the user, and IR-LEDs were attached to the display corners. The calibrated image consisted of nine blue cross markers and a white background. Four participants looked at nine points on the screen for homography-based user calibration, and the point-of-gaze was determined when they looked at 25 subsequent points to evaluate the accuracy. Figure 9 shows the results of this experiment. The proposed method detected the display area exactly, although the root-mean-square error of the estimated point-of-gaze using IR-LEDs was lower than that of polarization. Several problems might explain this, but one is that the display corners were influenced by the corneal surface, and thus, the observed corners were often distorted as a round shape.

4.2 Accuracy of the estimated point-of-gaze under several illumination conditions

Screen corners are detected using the degree of linear polarization in the wavelength of visible light. As such, the accuracy of the estimation might be influenced by the illumination conditions. Therefore, we verified the accuracy of the estimated point-of-gaze

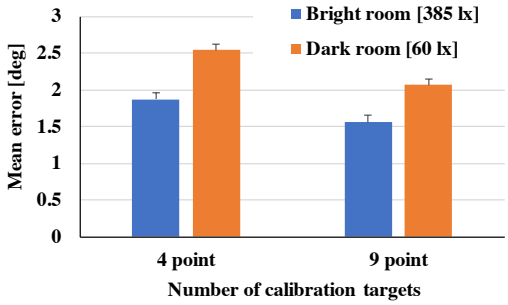

Figure 10: Mean error of gaze estimation in different room conditions

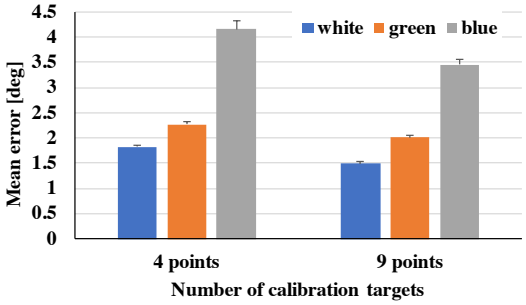

Figure 11: Mean error of gaze estimation with different colors projected on display

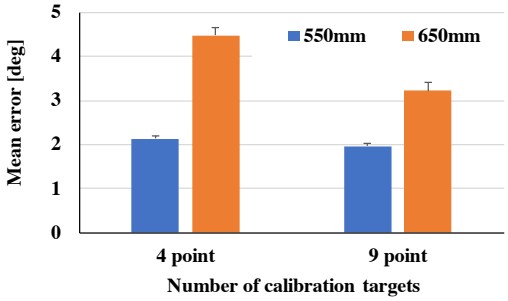

Figure 12: Mean error of gaze estimation from different distances between the display and the eye

under two illumination conditions. The experimental setup was the same as for the previous experiment (described in Section 4.1) and the pupil is detected using NIR LED. Illumination was turned on and off as two illumination conditions, where the illuminance in the room was about 385 and 60 lx, respectively. There were seven participants for this experiment, and Figure 10 shows the root-mean-square error and the standard error of the estimated point-of-gaze. The average error of gaze estimation in the brighter condition was lower than the error in the darker condition. When the illuminance is low, most of illumination input to the camera is radiated by the display, making it difficult to separate the display area reflected on the cornea.

4.3 Accuracy of the estimated point-of-gaze under several displayed colors on the screen

The previous section describes our evaluation of the influence of illumination, yet one more condition might affect the accuracy: the screen color. Therefore, we performed the experiment with different background screen colors. Three kinds of background patterns, white, green, and blue, were used for calibration and the experiment. Four participants stared at nine targets on the display for user calibration, and then looked at 25 points for the evaluation. For both calibration and the evaluation, the background colors were the same. Figure 11 shows the root-mean-square error and the standard error of the gaze direction calculated from the point-of-gaze. When a white background was selected, the mean error of the gaze direction was about 1.49 degrees, and brighter colors worked better than darker colors as a background.

4.4 Comparison of the distance between the display and the eye

The distance between the display point and the user influences the reflection on the cornea. Therefore, the accuracy from two distances was compared. A 21.5-inch monitor was used for this experiment, and the experiment was conducted with five participants. Their heads were fixed using a chin rest at each distance, and the geometric relationship between the camera and display was constant during all trials. Figure 12 shows the root-mean-square error of the estimated point-of-gaze from 550 mm and 650 mm. When the distance between the display and the eye was 550 mm, the mean error was about 2 degrees, and the accuracy from 650 mm was lower than that from 550 mm. We thus confirmed that the polarization information is sensitive to the distance.

4.5 Cross-ratio based gaze estimation using visible light image

We also conducted a comparative experiment for the two proposed approaches: CR-based gaze estimation with IR-based pupil detection, and CR-based gaze estimation without IR emission. The tracking targets differ with each method, and the pupil and the limbus were used with and without IR-based pupil detection, respectively. In this experiment, the distance between the display and the participants was about 600 mm, and the head pose was fixed using the chin rest. Currently, head movement is not supported with model-based iris tracking [Takemura et al. 2014]. The eyes of the five participants were measured with the polarization camera, which covered the infrared wavelength, when the NIR LED was turned on and off. The root-mean-square error and the standard error of the estimated point-of-gaze are shown in Figure 13. When the nine-point homography-based calibration was performed, the mean error of the cross-ratio method with pupil detection was about 1.46 degrees. However, the accuracy of the cross-ratio method without IR emission was lower than with IR emission. The eyelids and eyelashes often influence iris tracking as an occlusion, and thus, it is difficult to compute the center of the iris appropriately.

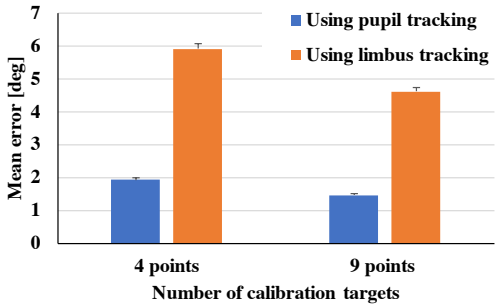

Figure 13: Mean error of gaze estimation using an IR light image and a visible light image

5 DISCUSSION

We confirmed the accuracy and effectiveness of the proposed CR method with the use of a polarization camera through the experiments mentioned above. When nine-point homography-based user calibration was employed, the best accuracy was approximately 1.46° through several experiments. We conducted the experiment under the similar condition between Figure 9 and 13. There is a difference in mean errors for which some reasons are: research participants are different, and the experiment was conducted at different times. A large number of research participants are required for removing these factors, but we can confirm the tendency in each comparative evaluation.

When the differences are compared between the previous method that uses the LED-mounted display and the proposed method that uses one IR-LED for pupil detection, the accuracy of the previous method is still slightly higher than that of the proposed method. One of the reasons for this difference was that edge detection used the detected display area, and the edge of display reflection on the cornea was sensitive to other light sources and noise. It is also likely that the edge of display corners on the corneal surface are rounded by the curvature of the cornea. Therefore, stable edge detection requires more information, such as color, in addition to the polarization information. However, we believe that, for popularization, eliminating IR-LEDs located on the display corners would be effective.

The point-of-gaze could not be estimated when the display was reflected on the sclera with the CR method. Therefore, some CR methods [Huang et al. 2014; Zhang and Cai 2014] use many LEDs for screen detection, and when part of display is missing, the display area is estimated using the observed LEDs. By contrast, the display reflection cannot be divided into a small area with the proposed method. It is thus necessary to reflect the entire area of the display onto the cornea. Therefore, there is a limitation to the size of the display under the current implementation. However, the displayed image has the potential to detect the display area appropriately by considering the missing area.

In previous CR-based researches, light from IR-LEDs is reflected on the cornea, and the part of pupil is often missing in this reflection. Therefore, it is difficult to estimate the center of pupil appropriately, decreasing the accuracy. On the contrary, when the polarization camera was employed, it was possible to cut off the reflected display area, and the center of the pupil could be estimated correctly.

The error of the estimated point-of-gaze in a dark environment was higher than in a bright one. In our consideration, the boundary of the display area is obscured in a dark environment. In the case of low illuminance, the majority of the observed light is polarized because light is emitted from the display. Because the DoLP indicates the ratio of the polarization in each pixel, it is difficult to distinguish the display area in the dark. However, the observed image and the color information might be also used to better detect the display area. In the experiment with multiple background colors, the accuracy of the point-of-gaze was lowest when the background color was darker. We employed a polarization camera that captures images in gray-scale, so we expect that color information will increase the robustness of screen detection.

The proposed method has some limitations: iris (or pupil) tracking is sensitive to eye blinking and droopy eyelids. Furthermore, for model-based iris tracking, it is necessary to maintain the relationship between the camera center and the eye. Therefore, in our method we applied the wearable eye tracker only when we employed model-based iris tracking. However, our method has potential in remote eye tracking when we use pupil tracking. Our contribution is the simplification of the system configuration. CR methods have high potential for head movement tolerance in comparison with PC-CR, and therefore we believe that our method can accelerate the popularization of the CR method because our method enables the use of a standard LCD.

6 CONCLUSION

We proposed a method of detecting the display corners reflected on the cornea for CR-based gaze estimation, and experiments were conducted using the detected display corners. The display area reflected on the corneal surface was extracted using DoLP and AoLP images, and the center of the pupil or the limbus was used for the estimation. An evaluation was conducted under several conditions, such as illumination, background color, the distance between the display and the user, and the combination of IR emission. The accuracy decreases due to several factors, such as illumination and background color on the screen, but we confirmed the feasibility of the proposed method. We proposed two types of eye tracking: model-based iris tracking without IR emission, and pupil tracking with IR emission. Pupil tracking worked better than iris tracking, and thus, IR emission is needed for high accuracy. Although it is difficult to retain IR emissions, the specific display onto which NIR LEDs are installed is unnecessary when using the polarization information to detect display corners for CR-based gaze estimation. Therefore, our eye tracking system can be installed easily by distributing the polarization camera.

In future work, we will attempt to improve the accuracy of detecting display corners by combining polarization and color information, and we will apply CR-based gaze estimation to multiple display conditions.

ACKNOWLEDGMENTS

The work presented in this paper was supported by JSPS KAKENHI, grant number JP18H03279.

REFERENCES

Nuri Murat Arar, Hua Gao, and Jean-Philippe Thiran. 2015a. Towards Convenient Calibration for Cross-Ratio Based Gaze Estimation. In *Proceedings of 2015 IEEE Winter Conference on Applications of Computer Vision*. 642–648. https://doi.org/10.1109/WACV.2015.91

Nuri Murat Arar, Hua Gao, and Jean-Philippe Thiran. 2015b. Robust gaze estimation based on adaptive fusion of multiple cameras. In *Proceedings of 2015 11th IEEE International Conference Workshops on Automatic Face and Gesture Recognition*. 1–7. https://doi.org/10.1109/FG.2015.7163121

David Beymer and Myron Flickner. 2003. Eye gaze tracking using an active stereo head. In *in Proceedings of 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2. II–451–8. https://doi.org/10.1109/CVPR.2003.1211502

Flavio Coutinho and Carlos Morimoto. 2006. Free head motion eye gaze tracking using a single camera and multiple light sources. In *Proceedings of 2006 19th Brazilian Symposium Computer Graphics and Image*. IEEE, 171–178. https://doi.org/10.1109/SIBGRAPI.2006.21

Flavio L. Coutinho and Carlos H. Morimoto. 2013. Improving head movement tolerance of cross-ratio based eye trackers. *International Journal of Computer Vision* 101, 3 (2013), 459–481.

Elias Daniel Guestrin and Moshe Eizenman. 2006. General theory of remote gaze estimation using the pupil center and corneal reflections. *IEEE Transactions Biomedical Engineering* 53, 6 (2006), 1124–1133. https://doi.org/10.1109/TBME.2005.863952

Dan Witzner Hansen, Javier San Agustin, and Arantxa Villanueva. 2010. Homography Normalization for Robust Gaze Estimation in Uncalibrated Setups. In *Proceedings 2010 Symposium Eye-Tracking Research & Applications - ETRA '10*, Vol. 1. 13–20. https://doi.org/10.1145/1743666.1743670

Dan Witzner Hansen and Qiang Ji. 2010. In the Eye of the Beholder: A Survey of Models for Eyes and Gaze. *IEEE Transactions Pattern Analysis and Machine Intelligence* 32, 3 (2010), 478–500. https://doi.org/10.1109/TPAMI.2009.30

Richard Hartley and Andrew Zisserman. 2000. *Multiple View Geometry in Computer Vision*. Cambridge: Cambridge University Press.

Jia-bin Huang, Qin Cai, Zicheng Liu, Narendra Ahuja, and Zhengyou Zhang. 2014. Towards Accurate and Robust Cross-Ratio based Gaze Trackers Through Learning From Simulation. In *Proceedings of Symposium Eye-Tracking Research & Applications - ETRA '14*. 75–82. https://doi.org/10.1145/2578153.2578162

Qiang Ji and Xiaojie Yang. 2002. Real-Time Eye, Gaze, and Face Pose Tracking for Monitoring Driver Vigilance. *Real-Time Imaging* 8, 5 (2002), 357–377. https://doi.org/10.1006/rtim.2002.0279

Jeffrey J. Kang, Elias D. Guestrin, W. James Maclean, and Moshe Eizenman. 2007. Simplifying the Cross-Ratios Method of Point-of-Gaze Estimation. In *Proceedings of 30th Canadian medical and biological Engineering Conference (CMBEC30)*. 1–4.

Anuradha Kar and Peter Corcoran. 2017. A Review and Analysis of Eye-Gaze Estimation Systems, Algorithms and Performance Evaluation Methods in Consumer Platforms. *IEEE Access* 5, c (2017), 16495–16519. https://doi.org/10.1109/ACCESS.2017.2735633

Carlos H. Morimoto and Marcio R.M. Mimica. 2005. Eye gaze tracking techniques for interactive applications. *Computer Vision and Image Understanding* 98, 1 (2005), 4–24. https://doi.org/10.1016/j.cviu.2004.07.010

Takashi Nagamatsu, Junzo Kamahara, and Naoki Tanaka. 2009. Calibration-free Gaze Tracking Using a Binocular 3D Eye Model. In *Proceedings of CHI '09 Extended Abstracts on Human Factors in Computing Systems (CHI EA '09)*. 3613–3618.

Roman Rädle, Hans-Christian Jetter, Jonathan Fischer, Inti Gabriel, Clemens Nylandsted Klokmose, Harald Reiterer, and Christian Holz. 2018. PolarTrack: Optical Outside-In Device Tracking that Exploits Display Polarization. In *Proceedings of 2018 CHI Conference Human Factors in Computing Systems - CHI '18*. 1–9. https://doi.org/10.1145/3173574.3174071

Masato Sasaki, Takashi Nagamatsu, and Kentaro Takemura. 2018. Cross-Ratio Based Gaze Estimation using Polarization Camera System. In *Proceedings of 2018 ACM International Conference Interactive Surfaces and Spaces - ISS '18*. 333–338. https://doi.org/10.1145/3279778.3279909

Richard S. Snell and Michael A. Lemp. 1989. *Clinical Anatomy of the Eye*. Blackwell Scientific Publications.

Kentaro Takemura, Tomohisa Yamakawa, Jun Takamatsu, and Tsukasa Ogasawara. 2014. Estimation of a focused object using a corneal surface image for eye-based interaction. *2014 Journal of Eye Movement Research* 7, 3 (2014), 1–9.

Dong Hyun Yoo and Myung Jin Chung. 2005. A novel non-intrusive eye gaze estimation using cross-ratio under large head motion. *Computer Vision and Image Understanding* 98, 1 (2005), 25–51. https://doi.org/10.1016/j.cviu.2004.07.011

Dong Hyun Yoo, Jae Heon Kim, Bang Rae Lee, and Myoung Jin Chung. 2002. Non-contact eye gaze tracking system by mapping of corneal reflections. In *Proceedings of 5th IEEE International Conference on Automatic Face Gesture Recognition*. 101–106. https://doi.org/10.1109/AFGR.2002.1004139

Zhengyou Zhang and Qin Cai. 2014. Improving cross-ratio-based eye tracking techniques by leveraging the binocular fixation constraint. In *Proceedings of Symposium Eye-Tracking Research & Applications - ETRA '14*. 267–270. https://doi.org/10.1145/2578153.2578202