Model-based Gaze Estimation with Transparent Markers on Large Screens

Koshikawa, Koki

Nagamatsu, Takashi

Takemura, Kentaro

(Citation)
Proceedings of the ACM on Human-Computer Interaction,6(ETRA):1-16

(Issue Date)
2022-05-13

(Resource Type)
journal article

(Version)
Accepted Manuscript

(URL)
https://hdl.handle.net/20.500.14094/0100485161

Model-based Gaze Estimation with Transparent Markers on Large Screens

KOKI KOSHIKAWA, Tokai University, Japan
TAKASHI NAGAMATSU, Kobe University, Japan
KENTARO TAKEMURA, Tokai University, Japan

Several technical issues that affect eye-tracking have arisen concomitantly with the steadily increasing sizes of personal displays recently. One such issue is the loss of the illumination reflection of the near-infrared light-emitting diodes used as reference points around the edge of the display. Another issue is that reference identification is required for practical usage. Therefore, this paper proposes gaze estimation with transparent markers for large display environments to solve these problems. The transparent markers can be distributed on the screen, and a unique ID is assigned to each marker using linear polarization angles. The reference is detected using a polarization camera through the reflection on the cornea. The results of experiments conducted using a 50-inch display indicate that the proposed method can estimate the point-of-gaze to within 2.1 degrees of error. We confirmed that on-screen markers in sizable displays could be effectively used as references instead of illumination sources.

CCS Concepts: • **Human-centered computing → Human computer interaction (HCI)**; • **Computing methodologies → Computer vision**.

Additional Key Words and Phrases: Polarization, Eye tracking, Large display

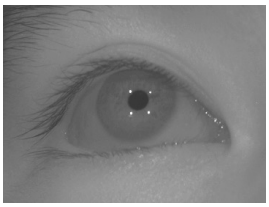
ACM Reference Format:
Koki Koshikawa, Takashi Nagamatsu, and Kentaro Takemura. 2022. Model-based Gaze Estimation with Transparent Markers on Large Screens. *Proc. ACM Hum.-Comput. Interact.* 6, ETRA, Article 147 (May 2022), 16 pages. https://doi.org/10.1145/3530888

1 INTRODUCTION

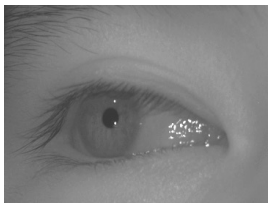
Human–computer interaction with eye-trackers enables non-contact, intuitive, and rapid operation, and eye-based interaction is expected in various situations. Consequently, various gaze estimation methods have been proposed to improve the performance of eye-trackers in aspects such as accuracy and tolerance to head movement [7]. Most methods emit near-infrared (NIR) light to track the corneal reflections (i.e., glints) as references for calculating the point-of-gaze (PoG) on a screen. Model-based methods use glints as a cue, and the number of glints necessary has also been analyzed [21]. Guestrin and Eizenman [5] presented a general theory of model-based methods, and they indicated that two or more glints are required to determine the center of the corneal sphere. Additionally, it has been found that binocular model-based gaze estimation enables high accuracy and implicit user calibration [11, 13]. However, hardware calibration is required, as the geometric relationship between cameras and the display must be calculated in advance. Yoo et al.

Authors' addresses: Koki Koshikawa, Tokai University, Hiratsuka, Kanagawa, 259-1292, Japan, kouki-k@takemura-lab.org; Takashi Nagamatsu, Kobe University, Kobe, Hyogo, 658-0022, Japan, nagamatu@kobe-u.ac.jp; Kentaro Takemura, Tokai University, Hiratsuka, Kanagawa, 259-1292, Japan, takemura@tokai.ac.jp.

2573-0142/2022/5-ART147 $15.00
https://doi.org/10.1145/3530888



(a) Illumination sources are reflected appropriately on the cornea



(b) Several corneal reflections are missing

Fig. 1. Comparison of corneal reflections related to eyeball pose.

[26] proposed cross-ratio (CR)-based gaze estimation with a liquid crystal display (LCD) to which four or more NIR-LEDs were attached. The CR method does not require strict hardware calibration, and the methods derived on this basis have been actively studied. Although both methods can effectively improve the accuracy and robustness by employing multiple glints, new problems, such as glint identification, are introduced. A standard solution for identification is based on the variation in the illumination pattern [14, 15]. However, synchronization of the camera and illumination is required, and the system configuration is complicated. Another solution is to apply a heuristic rule using the glint location, in which the increase of glints incurs the absence of the pupil area. The pupil center is used to calculate the gaze vector; therefore, the accuracy of the gaze vector is influenced by the lack of the pupil area. Lie et al. [10] filled the hole causing the occurrence of glint to estimate the pupil center accurately when glint was located on the pupil. In addition, with the size of personal displays steadily increasing in recent years, conventional methods that employ NIR-LEDs as references have become inadequate. This is because the mounting position of LEDs is limited to the outer frame of the display, and the loss of several glints is inevitable, as shown in Figure 1.

To solve the problems outlined above, we propose a method in which transparent markers are distributed on the screen as references. In our proposed method, partially polarized light with half-wave retarders is emitted from the screen as markers, and the marker ID is detected using the angle of linear polarization reflected on the corneal surface as a marker. Moreover, an eye image without reflections is generated using polarization for pupil tracking. We believe that the transparent markers will be a breakthrough to solve the problems related to glints. In particular, the method exhibits high effectiveness for large personal displays located near users.

The remainder of this paper is organized as follows. Section 2 describes related work and outlines our contributions. Section 3 presents the preliminaries of the polarization processing. Section 4 explains the proposed markers with half-wave retarders and the model-based gaze estimation using these markers. Section 5 presents the results of the evaluation experiments conducted. Section 6 discusses the results obtained and the limitations of the proposed method. Finally, our conclusions and scope for future research are provided in Section 7.

2 RELATED WORK

2.1 Roles of illumination in eye-tracking

Illumination is used for three roles in eye-tracking. First, it is used to track the pupil. The pupil area is not observed clearly from brown eyes. Therefore, near-infrared light is emitted to extract the pupil area clearly based on the differences in the absorption rates between the iris and pupil. Second, illumination functions as a reference. When the PoG on the screen is calculated with a

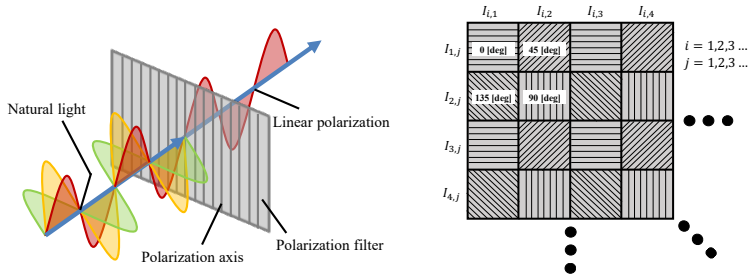
remote eye-tracker, illumination is required as a reference that describes the geometric relationship between user, cameras, illumination sources, and display. The regression-based method, which computes the PoG directly using the pupil center and corneal reflection, has been used as a classical approach in eye-tracking. The regression formula is calculated by performing user calibration, which determines the relationship between the display and the vector of the pupil center-corneal reflection. However, as the head movement tolerance is relatively low, multiple light sources are used to improve the robustness to head movements in the regression-based method [1, 20]. Yoo et al. [25, 26] proposed the CR-based method as an alternative for gaze estimation. In CR-based gaze estimation, the screen corners are used for computing the homography matrices to describe the relationship between the eye image, corneal surface, and screen. Therefore, at least four NIR-LEDs are attached to the display to compute the PoG. Recently, several methods have been proposed for improving the accuracy of gaze estimation and the observation of glints[3, 6, 28]. The regression and CR-based methods utilize the first and second roles of illumination.

The role for which illumination is used in the model-based method differs from the first and second roles used by the regression and CR-based methods. Specifically, the third role is employed in which illumination is used to estimate the center of the corneal sphere. The corneal center can be determined using two glints and a stereo camera pair [5, 21, 24]. The optical axis, which passes through the pupil center and corneal sphere center, is calculated geometrically using the eyeball model. Furthermore, the visual axis, which indicates the line of sight (LoS), is determined by defining the individual offset angle from the optical axis.

As mentioned above, gaze estimation tends to increase the number of illumination sources to improve the robustness of the eyeball pose variations in the above methods. Therefore, it is necessary to discuss the identification of illumination sources.

2.2 Eye-tracking for large displays

In conventional situations, the user is far from the large display, and thus the issue is to capture the eye region within the angle of view. Cho et al. [2] employed two wide-view cameras and a narrow-view camera mounted on a pan-tilt unit to deal with this issue. However, the style of human–computer interaction changes this paradigm, and users are often close to the display, such as an electric blackboard. Additionally, the size of the personal display has also been enlarged to as much as 49 inches. Therefore, the variation of eyeball pose has also increased progressively in association with the enlargement of the display size. Nagamatsu et al. [12] proposed a wide-angle gaze estimation method that extends the gaze-tracking range using the whole eye modeled as a revolution surface about the optical axis of the eye. However, in their method, the identification of glints depends on heuristic rules. Furthermore, eye-tracking that does not require references has emerged for large displays. A smooth-pursuit-based approach, which calculates the correlation between smooth-pursuit eye movements and the motion of the object [23] is highly versatile, calibration-free, and enables interaction with dynamic objects on the screen. Therefore, this approach has been employed effectively in large display environments. Khamis et al. [8] used the smooth-pursuit-based method and integrated an eye-tracker and a one-axis linear motor for a large public display. Their method enabled eye-based interaction on large public displays, primarily by responding to changes in user position. Furthermore, appearance-based methods have been studied [22, 27] to estimate the PoG using a model trained by a large amount of data. Therefore, appearance-based methods can work in many scenarios with public displays. However, high accuracy and handling significant eyeball rotation is required for sizable personal displays.



(a) Polarization of light (b) Layout of polarizer array

Fig. 2. Principle of polarization and layout of polarizers in the CMOS sensor.

2.3 Polarization in eye-tracking

In recent years, polarization has been employed as a novel feature in eye-tracking. Sasaki et al. [17, 19] proposed a method that focused on the polarization of light emitted from the display. In their method, the screen corners were extracted without installing multiple NIR-LEDs in the display and the PoG was estimated using CR-based gaze estimation. Moreover, they extended the method to determine the PoG under multiple display environments [18]. Koshikawa et al. [9] attached a polarizing filter to NIR-LEDs to polarize the emitted light during eye-tracking. When multiple light sources are used, the illumination IDs can be correctly identified using the polarization angle. Their research indicated that polarization has great potential as a breakthrough in eye-tracking. However, these methods extract features that are located outside the screen. Therefore, reference locations should be investigated for enlarged displays.

2.4 Contributions

In the conventional method, NIR-LEDs are required as the reference points. Since the NIR-LEDs must be installed outside the display or on the display frame, the display size is limited by the requirement to stably observe the reflection point on the corneal surface. Additionally, if multiple NIR-LEDs are used, their reflection points could interfere with pupil tracking. Therefore, we propose a novel method for installing the reference points on a screen using half-wave retarders that change the polarization angle. Half-wave retarders can be arranged on the screen as transparent markers, and estimating the PoG even on a large display with a significant eyeball rotation is possible.

The contributions of this study are as follows:

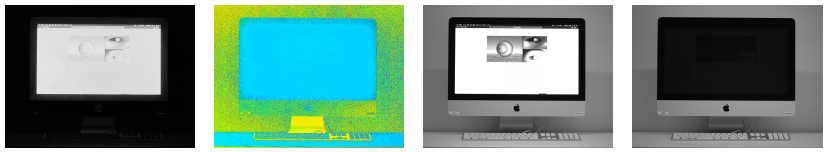
- We propose transparent markers using polarization, which can be arranged on the screen as novel reference points for eye-tracking.
- The use of eye-tracking has previously been limited to small displays to which NIR-LEDs are attached. In contrast, our proposed method can work on large displays.
- With a higher number of illumination sources, the pupil area is partially lacking. However, despite the increased number of markers, the pupil area can be clearly observed using polarization.

3 PRELIMINARIES

Light is a type of electromagnetic radiation that has wave–particle duality. Its characteristics are used in various scientific and engineering fields. In this study, we focus on the polarization of light, which is expected to offer the potential to solve an unresolved issue in the eye-tracking field.

(a) Original (b) $\mathbf{I}(0°)$ (c) $\mathbf{I}(45°)$ (d) $\mathbf{I}(90°)$ (e) $\mathbf{I}(135°)$

Fig. 3. Original image captured by polarization camera and angle images generated by using the pixel on which each polarizer is arranged by interpolation.



(a) **DoLP** image (b) **AoLP** image (c) $\mathbf{I}_{\max}$ image (d) $\mathbf{I}_{\min}$ image

Fig. 4. Four characteristic images were generated using Stokes parameters. The **DoLP** expresses the degree of linear polarization using the intensity, and the **AoLP** expresses the angle of linear polarization using the color. Subsequently, the $\mathbf{I}_{\max}$ and $\mathbf{I}_{\min}$ were generated using the **DoLP** and the **AoLP** images. The $\mathbf{I}_{\max}$ image emphasizes the polarization component, whereas the $\mathbf{I}_{\min}$ image does not contain the polarization information.

Although natural light oscillates in all directions, the geometry of the oscillations can be specified with a polarization filter, which is an optical element, as illustrated in Figure 2(a). A CMOS sensor (IMX250MZR, Sony, Tokyo, JP) on which polarizers are formed on-chip has recently been developed; subsequently, a polarization camera with this CMOS sensor was released. The polarization camera can capture a monochrome image; however, the image includes information of different angles because the CMOS sensor has a layout of polarizers on the pixels, as indicated in Figure 2(b). Therefore, each angle image is computed using the light that passes through each polarizer. First, the polarizer of each angle is arranged sparsely on the pixels, and thus each angle image $\mathbf{I}(\theta)$ is generated by interpolating the adjacent pixel values similar to the demosaicing process of color cameras. An example formula for computing the absence pixel of the $\mathbf{I}(0°)$ image with the interpolation approach is as follows:

$$
\begin{aligned}
I_{1,2}(0°) &= \frac{I_{1,1}(0°) + I_{1,3}(0°)}{2}, \\
I_{2,1}(0°) &= \frac{I_{1,1}(0°) + I_{3,1}(0°)}{2}, \\
I_{2,2}(0°) &= \frac{I_{1,1}(0°) + I_{3,3}(0°)}{2},
\end{aligned}
\tag{1}
$$

where $I_{i,j}(\theta)$ is the intensity of pixel (i, j). Figure 3(a) is the original image captured by the polarization camera, whereas Figures 3(b), 3(c), 3(d), and 3(e) are generated as angle images $\mathbf{I}(\theta)$, $\theta \in \{0°, 45°, 90°, 135°\}$, respectively. The emitted light from the display is polarized, and the displayed content can be photographed without loss when the polarizer angle is parallel to the polarization angle of the light emitted from the display, as shown in Figure 3(b). By contrast, Figure 3(d) indicates that the loss is increased when the polarizers are orthogonal to one another.

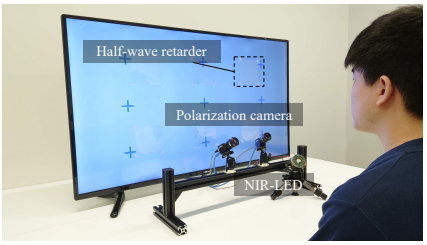


Fig. 5. Display with transparent markers, which consist of some half-wave retarders. The half-wave retarder is a square with a side of 12 cm cut out from the original sheet and attached to the screen surface with transparent tape (base: acetate film, adhesive: acrylic).

Subsequently, the Stokes parameters, which express the polarization state of the light[4], are calculated using the generated angle images. The Stokes parameters consist of four components, which express the polarization state. In this study, three Stokes parameters, which are related to linear polarization, are computed; the other Stokes parameter that expresses the circular polarization state is not required. The three Stokes parameters $\mathbf{S}_k$, $k \in \{0 \dots 2\}$ are calculated for each pixel as follows:

$$S_0(i,j) = I_{i,j}(0^\circ) + I_{i,j}(90^\circ), \tag{2}$$
$$S_1(i,j) = I_{i,j}(0^\circ) - I_{i,j}(90^\circ), \tag{3}$$
$$S_2(i,j) = I_{i,j}(45^\circ) - I_{i,j}(135^\circ). \tag{4}$$

Thereafter, four characteristic images can be generated using the Stokes parameters. The degree of linear polarization (DoLP) and angle of linear polarization (AoLP) are calculated as follows:

$$DoLP(i,j) = \frac{\sqrt{S_1(i,j)^2 + S_2(i,j)^2}}{S_0(i,j)}, \tag{5}$$
$$AoLP(i,j) = \frac{1}{2}\arctan\left(\frac{S_2(i,j)}{S_1(i,j)}\right). \tag{6}$$

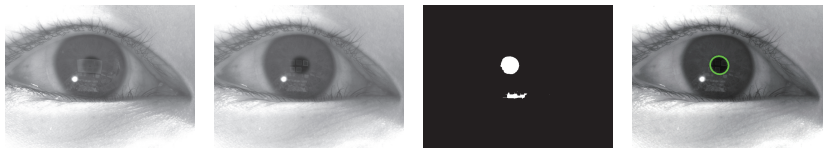
Figures 4(a) and 4(b) present the generated **DoLP** and **AoLP** images, respectively. Moreover, the $\mathbf{I}_{\max}$ and $\mathbf{I}_{\min}$ images are calculated using the Stokes parameter $\mathbf{S}_0$ and the computed **DoLP** image, as follows:

$$I_{\max}(i,j) = \frac{S_0(i,j)}{2}(1 + DoLP(i,j)), \tag{7}$$
$$I_{\min}(i,j) = \frac{S_0(i,j)}{2}(1 - DoLP(i,j)). \tag{8}$$

Figures 4(c) and 4(d) depict the $\mathbf{I}_{\max}$ and $\mathbf{I}_{\min}$ images, respectively. The polarization component is emphasized in the $\mathbf{I}_{\max}$ image. In contrast, the $\mathbf{I}_{\min}$ image does not contain polarization information. Note that the range of pixel values is 0 to 255 ($I_{\max}(i,j) \in 0$–255). These four generated images are used in the proposed method, as described in the following section.

4 METHODOLOGY

As mentioned in Section 2, the glints reflected on the cornea play several roles. In the regression and CR-based methods, they are used to describe the geometric relationship, such as the display location, whereas in the model-based method, they are used as cues to determine the center of the corneal sphere. Therefore, the identification of glints is required when multiple illumination sources are used. In general, NIR-LEDs are attached to the display as illumination sources; however,



(a) Original eye image (b) $\mathbf{I}_{\min}$ image (c) Binary image (d) Pupil area

Fig. 6. Pupil detection using $\mathbf{I}_{\min}$ image.

several glints are often lost owing to the significant variation in the eyeball pose. In particular, this issue frequently occurs when a large display is employed because the locations of the NIR-LEDs are limited to the outside of the display.

Therefore, we propose a novel gaze estimation method that uses transparent markers that can be arranged on the screen. The proposed method employs the model-based method to estimate the PoG, and thus the implemented system consists of two polarization cameras (BFS-U3-51S5P-C, FLIR Systems Inc., Wilsonville, OR, USA) and half-wave retarders, which have a function to control the angle of linear polarization from the LCD. In addition, an NIR-LED is used only for extracting the pupil because it is not a reference point. The transparent markers consist of half-wave retarders, which are sparsely attached to the screen to translate the polarization angle into any angle. For evaluation, we constructed a setup using a 50-inch display, as illustrated in Figure 5. The markers were temporarily attached to the large display with transparent tape because the marker variation had to be changed during the evaluation. In addition, by using the same adhesives used for the polarization film and liquid crystal panel in LCD, it is possible to construct transparent markers that are almost invisible.

4.1 Pupil detection

Corneal reflections are used in some methods as references to estimate the PoG on a screen; however, pupil detection is adversely affected by the glints and the screen reflection on the cornea. Koshikawa et al. [9] developed a method that reduces the influence of glints using polarizers. Therefore, we focused on the screen reflections, which should be removed for the accurate estimation of the pupil center. We implemented the screen reflection removal using polarization without complex calculations or high computational costs. Figure 6(a) presents the original image captured with the polarization camera. The reflections that appeared on the cornea were observed. The display has a built-in polarizing filter, and the light emitted from the display is polarized. Therefore, the screen reflection on the cornea contains polarization information as features. The $\mathbf{I}_{\min}$ image does not contain polarization information, as described in Section 3. Figure 6(b) shows the computed $\mathbf{I}_{\min}$ image. It can be observed that the screen reflection is reduced; however, the reflections of the half-wave retarders remain on the cornea. Therefore, opening and closing processes were applied to the inverted binary image, as illustrated in Figure 6(c). Finally, the circularity of each blob was computed using the following equation:

$$C_i = \frac{4\pi S_i}{L_i^2}, \tag{9}$$

where C_i is the circularity of blob i. Circularity is an index that expresses quantitative roundness, where a closer value to one indicates greater similarity to a perfect circle. Moreover, S is the number of pixels in the blob detected by connected-component labeling, and L is the perimeter of the blob calculated from the boundary extracted with Moore boundary tracing algorithm. If there is a boundary pixel in 4-neighbors or the diagonal neighbors, one or $\sqrt{2}$ are added to the boundary

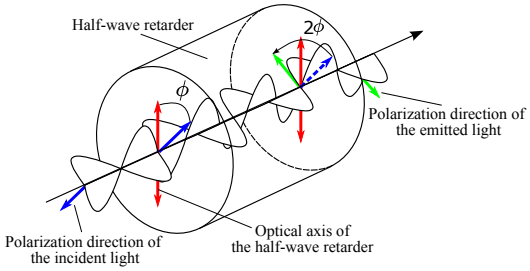
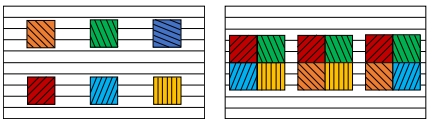


Fig. 7. Principle of half-wave retarder



(a) Single marker (b) Combination marker

Fig. 8. Variations of transparent markers

(a) Original eye image (b) **DoLP** image (c) Extracted screen area (d) Detected marker areas (e) Identified marker ID

Fig. 9. Identification using polarization for single markers reflected on the cornea.

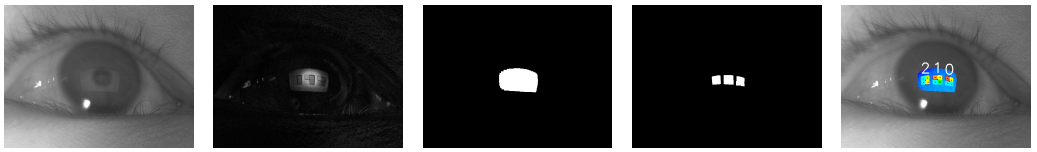
length, respectively. The blob with the highest circularity was detected as the pupil region and elliptical fitting was applied to detect the pupil center, as indicated in Figure 6(d).

4.2 Transparent markers

A wave retarder works to shift two components of vertical and horizontal light waves. The retardation (i.e., phase difference) is determined by its thickness. The wave retarder used in our work is designed to add a retardation π to the waves and is called a half-wave retarder. Thus, the polarization angle of the incident light can be changed to any angle by rotating the half-wave retarder. The blue and green arrows indicate the polarization directions of the incident light and the emitted light, respectively, and the red arrows represent the optical axes of the retarder, as shown in Figure 7. In this case, the angle of the emitted light was twice the angle ϕ formed in these two directions of the blue and red arrows. Therefore, the transparent markers consisted of a half-wave retarder with a controlled optical axis, and they were allocated on the screen by taking advantage of this property. Figures 8(a) and 8(b) depict the proposed markers, with two variations. The first marker had a single half-wave retarder, and there were seven variations based on the AoLP value divided into 22.5-degree intervals. The second marker consisted of four combinations selected from seven patterns; thus, up to $\binom{7}{4} = 35$ variations of marker patterns existed, which was sufficient for attachment to the large screen. There are two reasons it is a combination rather than a permutation. One reason is that even when the display is rotated, the PoG is performed with the same setup. The other reason is to prevent false positives in the identification.

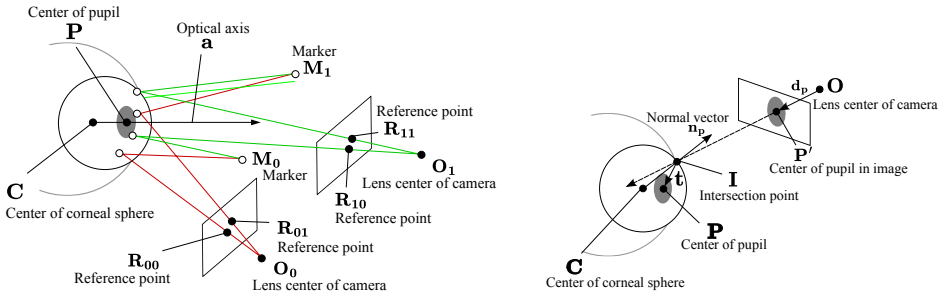
4.3 Detection and identification of transparent markers

As mentioned in Section 4.2, two variations of the proposed markers existed: single and combination. The AoLP value was used for the identification, and most processes were common in these identifications. Figures 9 and 10 present the identification processes; the steps are as follows.



(a) Original eye image (b) **DoLP** image (c) Extracted screen area (d) Detected marker areas (e) Identified marker ID

Fig. 10. Identification using polarization for combination markers reflected on the cornea.



(a) Geometric model for model-based gaze estimation. (b) 3D pupil center detection considering refraction.

Fig. 11. Geometry for estimating visual axis based on the eyeball model.

(1) Generate the **DoLP** images (Figures 9(b) and 10(b)) from the original images (Figures 9(a) and 10(a)) that are captured by the polarization camera.
(2) Binarize the generated **DoLP** images using a threshold because the DoLP value of the screen reflection area is high. The opening (i.e., erosion and dilation) was applied to the binary image to remove noise, and the blobs were detected with the connected-component labeling.
(3) Detect the screen reflection using the blob size and the distance from the pupil center to a blob, as illustrated in Figures 9(c) and 10(c). The screen regions are reflected close to the pupil center and the screen reflection areas are greater than a certain area.
(4) Extract the marker areas by inverting the detected screen area because the AoLP values of the marker areas are different from that of the normal screen area, as illustrated in Figures 9(d) and 10(d). This step also conducted the opening and the connected-component labeling conducted in step 2.
(5) Multiply the **AoLP** image and inverted binary image to extract the AoLP values of the half-retarders.
(6) Identify the marker ID based on the AoLP value detected using the most frequent value in the blob, as indicated in Figures 9(e) and 10(e).

When the single marker was used, the marker ID was determined using the extracted AoLP value directly. In contrast, an additional process was required when using the combination marker; the ID was identified using the combination of the detected half-wave retarders. Following identification, the centers of the detected half-wave retarders were used in addition to the center of the marker as reference points in the model-based gaze estimation.

4.4 Model-based gaze estimation

The model-based method computes the optical axis geometrically using the 3D eyeball model. In the typical model-based method, camera poses, the position of the illumination sources, and the 3D position of the glints and the pupil on the image plane are generally used. Therefore, the position of the illumination sources and camera poses are determined through hardware calibration in advance. In contrast, we employ transparent markers on the screen instead of illumination sources as references. Therefore, the marker positions are easily obtained without strict hardware calibration because the guide can be displayed on the screen for marker arrangement. Figure 11(a) depicts the relationship between the corneal sphere, reference points, and cameras. First, the center of the cornea $\mathbf{C}$ is calculated to determine the optical axis. As the observed reference points $\mathbf{R}_{ji}$, lens centers of the cameras $\mathbf{O}_j$, and centers of the markers $\mathbf{M}_i$ are located on the same plane, they satisfy the following equation:

$$(\mathbf{R}_{ji} - \mathbf{O}_j) \times (\mathbf{M}_i - \mathbf{O}_j) \cdot (\mathbf{C} - \mathbf{O}_j) = 0, \tag{10}$$

where i ($i \in \{0, 1\}$) and j ($j \in \{0, 1\}$) are the index numbers of the markers and cameras, respectively. Therefore, the center of the cornea $\mathbf{C}$ is calculated by using the intersection of the three planes.

Subsequently, the optical axis $\mathbf{a}$ is obtained. The optical axis is a vector between the center of the cornea $\mathbf{C}$ and the pupil center $\mathbf{P}$. However, the observed pupil center is influenced by refraction; thus, it is not observed in the actual position, as shown in Figure 11(b). Therefore, it is necessary to determine the actual 3D pupil position considering refraction [16]. The refracted vector $\mathbf{t}$ can be expressed based on Snell's law, as follows:

$$\mathbf{t} = \frac{n_a}{n_b}\left[\mathbf{d}_\mathbf{p} - \left((\mathbf{d}_\mathbf{p} \cdot \mathbf{n}_\mathbf{p}) + \sqrt{\left(\frac{n_b}{n_a}\right)^2 - 1 + (\mathbf{d}_\mathbf{p} \cdot \mathbf{n}_\mathbf{p})^2}\right)\mathbf{n}_\mathbf{p}\right], \tag{11}$$

where $\mathbf{d}_\mathbf{p}$ is the unit vector from the lens center of the camera to the pupil center in the obtained eye image, and $\mathbf{n}_\mathbf{p}$ is the unit normal vector at the intersection point $\mathbf{I}$. Additionally, n_a and n_b are the refractive indices of air and aqueous humor ($n_a = 1.000$ and $n_b = 1.336$), respectively. Two refracted vectors are computed using the stereo camera pair; thus, the actual pupil position $\mathbf{P}$ is calculated as the intersection of these vectors. The optical axis $\mathbf{a}$ is determined as a vector that passes through $\mathbf{C}$ and $\mathbf{P}$.

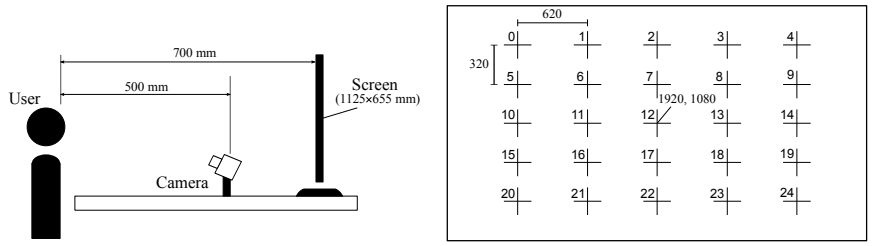
Although the optical axis is a vector between the centers of the cornea and the pupil, it does not match the LoS. Therefore, the visual axis, which indicates the LoS, is required to estimate the PoG. The visual axis is determined using the offset angle (i.e., κ-angle) from the optical axis. The offset angle is generally calculated through one-point calibration (i.e., user calibration). Thereafter, the visual axis is corrected by offset angle. The visual axis is the LoS connecting $\mathbf{C}$ and the PoG and is determined using the optical axis and the offset angle measured by user calibration. Finally, the PoG is calculated as the intersection between the visual axis and the screen.

5 EVALUATION EXPERIMENTS

We conducted experiments to evaluate the effectiveness of the polarization markers and the feasibility of the gaze estimation. This study was approved by the research ethics committee of Tokai university (permission number:20083).

5.1 Design

We evaluated the accuracy (degree) and precision (standard error) of the estimated visual axis with a large display. Although there are still few examples in eye-tracking research, Nagamatsu et al. [12] prepared many focus targets distributed over a wide area for evaluating the extended measurement



(a) Experimental setup (b) Targets used to focus the gaze of the user (pixel)

Fig. 12. System used for the experiment and targets used to focus the gaze of the user

Table 1. Center position of the transparent markers (pixel)

Marker type \ Marker No.	1	2	3	4	5	6
Single marker	(960, 580)	(1920, 580)	(2880,580)	(960,1580)	(1920,1580)	(2880,1580)
Combination marker	(820,1080)	(1920,1080)	(3020,1080)	-	-	-

angle of a gaze. The accuracy was analyzed for each target in relation to the participants. As their study had a similar objective to ours regarding large screens, we designed the experiment by referring to their work.

5.2 Setup and Apparatus

This study estimated the PoG via the model-based approach, and thus, two monocular polarization cameras (BFS-U3-51S5P-C, FLIR Systems Inc., Wilsonville, OR, USA) were used. NIR-LEDs were not required as references; however, an NIR-LED module was set to track the pupil. The focus targets were displayed on a 50-inch display (DME-4K50D, DMM.com LLC., Tokyo, Japan) with the transparent markers attached, and the distance between the user's head and the screen was 700 mm, as shown in Figure 12(a). The half-wave retarder was a 12 cm square; the single and combination markers were 12 and 24 cm squares, respectively. The arrangements of the transparent markers are given in Table 1. The marker can be attached to the specified location using guides displayed on the screen. The user's head was fixed with a chin rest to induce large eye movements. As polarization tends to be affected by illumination conditions, the experiments were conducted under low illumination, with illuminance at approximately 5.40 lx.

5.3 Participants

We recruited 16 healthy research participants (male: 14, female: 2) with average age 22.3 years (min: 20, max: 27). The eye color of all research participants in the experiments was brown, and there was a mixture of people with naked eyes and people wearing contact lenses.

5.4 Procedure

A total of 25 focus targets were allocated every 620 pixels horizontally and every 320 pixels vertically on the screen, as shown in Figure 12(b). These targets were displayed as blue crosses on white background. This study proposed two kinds of transparent markers: single and combination. Therefore, every research participant looked at focus targets under two conditions: single marker and combination marker.

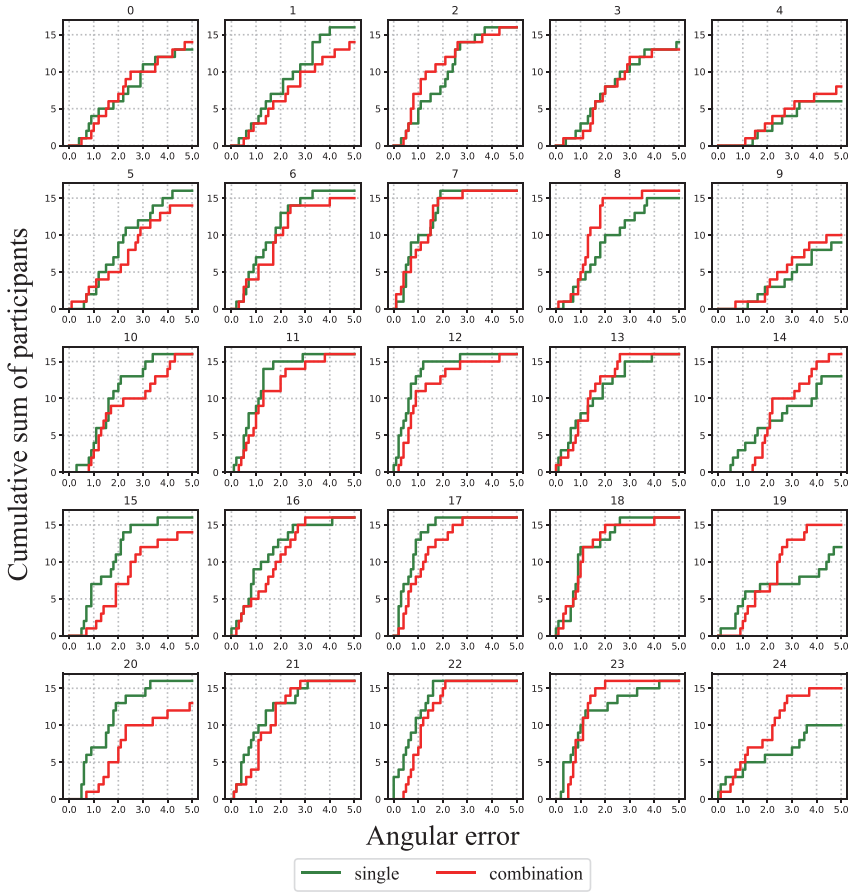


Fig. 13. Accuracy at each focus target displayed on the 50-inch display. X-axis: angular error computed using focus target and estimated visual axis; Y-axis: cumulative sum of participants from the aspect of accuracy at each focus target. The accuracy and precision of using the single and combination markers were 2.09° ± 0.39 and 2.01° ± 0.28, respectively.

Comparative experiments between the proposed and conventional methods are ordinarily required as an evaluation of the proposed method. However, we could not obtain a quantitative result using the conventional method with NIR-LEDs because the display was too sizable and too close to the participant.

5.5 Result

Figure 13 presents the cumulative sum of the participants from the aspect of accuracy at each focus target when the single and combination markers were used. The horizontal axis represents the angular error computed using the focus target and the estimated visual axis, and the vertical axis indicates the cumulative sum of the participants. Although the gaze estimation accuracy is slightly lower in the four corners, the overall gaze estimation achieved high accuracy. The accuracy and precision of using the single and combination markers were 2.09° ± 0.39 and 2.01° ± 0.28, respectively. Additionally, we analyzed the marker detection and identification in detail. The detection rates of the single and combination markers were 99.3% and 99.8%, respectively, and the identification

rates were 97.3% and 99.8%, respectively. The eyeball rotates when staring at the targets on a large display; thus, glints are often lost when markers are placed at display corners only. In contrast, the transparent markers were allocated on the screen instead of at the display corners. Therefore, the marker could be identified stably.

6 DISCUSSION AND LIMITATIONS

As mentioned above, the effectiveness and feasibility of the proposed method were confirmed through the experiments described in Section 5. In the conventional method, the reflections of NIR-LEDs placed at the edge of the display are used as features. This results in the estimation being inaccurate when glints are lost by rotating the eyeball significantly. This issue needs to be addressed to facilitate eye-tracking with large displays; however, there has been no definitive solution. In contrast, in our proposed method, markers are arranged on the screen, and thus it may achieve a breakthrough in large display environments. However, there are still several limitations depending on its specific characteristics, which we discuss below.

6.1 Single vs. combination marker

We expected that the accuracy of combination markers would be higher than that of single markers because the number of references is high. However, the experimental results showed that the accuracy and precision of the combination marker and single marker are almost the same in this study. We consider that the same size of the half-wave retarders in the single and combination markers may be the reason for this. Additionally, with the arrangement of the single makers presented in Table 1, sufficient references are reflected on the cornea. If the user moved significantly closer to the screen, a difference might be observed.

Single markers use a single half-wave retarder, and thus their setup and identification were easier than those of the combination markers; these are their advantages. In contrast, combination markers consist of four half-wave retarders. The combination markers in this study were designed with 35 simple variations. The scalability of marker variation is an advantage because a unique code can be provided for each marker in extremely large displays. However, each component to which a polarization angle is assigned requires a certain area for detection, and therefore there is a trade-off relationship between marker variation and marker size, which is a limitation.

6.2 Influence of corneal curvature

We confirmed that the accuracy decreased with eye movement outward to the edge of the screen in Section 5. When the user looked outward to the edge of the large display, the marker was reflected around the edge of the cornea owing to the significant rotation of the eyeball. Consequently, the extracted reference points differed from the actual centers of the half-wave retarders because the cornea's curvature is not constant (i.e., the actual cornea is not a sphere as defined in the eyeball model). Although we used a general eyeball model that consists of two spheres, an improved eyeball model has been proposed for the model-based approach [12]. Therefore, one of the solutions might be to employ the improved eyeball model.

6.3 Marker installation

The transparent markers were pasted onto the screen using transparent tape to enable the application of various pattern variations. Currently, transparent markers are additional parts, and the marker installation process should be improved in the aspect of the setup time. However, an electrical system is not required for reference points, unlike with NIR-LEDs; thus, we believe that further improvement is possible toward use with a personal device. Additionally, the half-wave retarders can be installed without the transparent tape using the optical adhesive used in LCDs. We

have tried to attach the markers to a 27-inch display using an optical adhesive, and the PoG was estimated using our proposed method. We confirmed that the gaze estimation can be performed without the influence of artifacts even when an optical adhesive was used.

6.4 Background color and illumination conditions

The background color influences the DoLP value in the LCD, and thus screen reflection might not be extracted correctly when the background color changes. This is another limitation in the current implementation. The countermeasure for this issue is clear: the threshold used in screen extraction should change dynamically in response to the background color. Additionally, the experiments were conducted under restricted and controlled illumination because polarization tends to be affected by the illumination conditions. In future work, we aim to extend our method so that the transparent marker can be used under any illumination conditions.

6.5 Measurable range

Eye-tracking methods that can perform in a wide measurable range have been proposed previously. Cho et al. [2] employed the pan-tilt and focus unit, which controlled the field of view of the camera, and Khamis et al. [8] used a slider to change the location of the eye-tracker. Therefore, these methods have a wide measurable range for location and high flexibility of user location. By contrast, our method is proposed to respond to a sizable personal display, and the measurement target is a user sitting in front of the display. In other words, our proposed method improved measurability width of eyeball rotation. Recently, large displays have begun to be used not only for public but also for personal use. Additionally, a user is often located quite near a display used as a digital blackboard. Therefore, we believe that our proposed method is a breakthrough in eye-tracking for close displays. Specifications for commercial eye-tracking recommend 27-inch and 30-inch displays, and we confirmed that our proposed method could work far beyond these display sizes. If both wide user location and close display condition are required as measurable range, it is also possible to integrate camera control and the proposed transparent markers.

7 CONCLUSIONS

We proposed a method in which transparent markers, which comprise half-wave retarders, are used for eye-tracking, and the PoG is estimated on large displays by allocating these markers on the screen. In recent years, the sizes of personal displays have been steadily increasing; however, NIR-LEDs were attached to the edge of the display. Thus, the references were often lost when the eyeball was rotated significantly. In contrast, the transparent marker can be allocated on the screen, and it changes the light emitted from the display to any angle of polarization. Therefore, the transparent marker works as an alternative reference for extremely large displays. Furthermore, it is necessary to distribute several references in a large display environment, and the identification of references is required. Whereas illumination patterns and heuristic rules are used for identification in conventional approaches, identification is achieved using the angle of linear polarization in the proposed method. The half-wave retarders can be combined, and thus the marker has scalability. Therefore, unique IDs can be assigned to each marker, and the PoG can be estimated by obtaining a snapshot of the eye, which includes marker reflection on the cornea.

We verified the feasibility of the transparent markers in this study; however, there are certain limitations, such as screen extraction and illumination conditions. Therefore, we will improve the method using a color polarization camera and deep learning to adapt to the background variations and various illumination conditions in future work.

REFERENCES

[1] Juan J. Cerrolaza, Arantxa Villanueva, and Rafael Cabeza. 2008. Taxonomic Study of Polynomial Regressions Applied to the Calibration of Video-Oculographic Systems. In *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications* (Savannah, Georgia) *(ETRA '08)*. Association for Computing Machinery, New York, NY, USA, 259–266. https://doi.org/10.1145/1344471.1344530

[2] Dong-Chan Cho and Whoi-Yul Kim. 2013. Long-Range Gaze Tracking System for Large Movements. *IEEE Transactions on Biomedical Engineering* 60, 12 (2013), 3432–3440. https://doi.org/10.1109/TBME.2013.2266413

[3] 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.

[4] Dennis H. Goldstein. 2011. *Polarized Light (3rd ed.)*. CRC Press. https://doi.org/10.1201/b10436

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

[6] Dan Witzner Hansen, Javier San Agustin, and Arantxa Villanueva. 2010. Homography Normalization for Robust Gaze Estimation in Uncalibrated Setups. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (Austin, Texas) *(ETRA '10)*. Association for Computing Machinery, New York, NY, USA, 13–20. https://doi.org/10.1145/1743666.1743670

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

[8] Mohamed Khamis, Axel Hoesl, Alexander Klimczak, Martin Reiss, Florian Alt, and Andreas Bulling. 2017. EyeScout: Active Eye Tracking for Position and Movement Independent Gaze Interaction with Large Public Displays. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology* (Québec City, QC, Canada) *(UIST '17)*. Association for Computing Machinery, New York, NY, USA, 155–166. https://doi.org/10.1145/3126594.3126630

[9] Koki Koshikawa, Masato Sasaki, Takamasa Utsu, and Kentaro Takemura. 2020. Polarized Near-Infrared Light Emission for Eye Gaze Estimation. In *ACM Symposium on Eye Tracking Research and Applications* (Stuttgart, Germany) *(ETRA '20 Short Papers)*. Association for Computing Machinery, New York, NY, USA, Article 2, 4 pages. https://doi.org/10.1145/3379156.3391342

[10] Dongheng Li, D. Winfield, and D.J. Parkhurst. 2005. Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*. IEEE, San Diego, CA, USA, 79–79. https://doi.org/10.1109/CVPR.2005.531

[11] Dmitri Model and Moshe Eizenman. 2010. User-Calibration-Free Remote Gaze Estimation System. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (Austin, Texas) *(ETRA '10)*. Association for Computing Machinery, New York, NY, USA, 29–36. https://doi.org/10.1145/1743666.1743672

[12] Takashi NAGAMATSU, Mamoru HIROE, and Hisashi ARAI. 2021. Extending the Measurement Angle of a Gaze Estimation Method Using an Eye Model Expressed by a Revolution about the Optical Axis of the Eye. *IEICE Transactions on Information and Systems* E104.D, 5 (2021), 729–740. https://doi.org/10.1587/transinf.2020EDP7072

[13] Takashi Nagamatsu, Junzo Kamahara, and Naoki Tanaka. 2009. *Calibration-Free Gaze Tracking Using a Binocular 3D Eye Model*. Association for Computing Machinery, New York, NY, USA, 3613–3618. https://doi.org/10.1145/1520340.1520543

[14] Atsushi Nakazawa and Christian Nitschke. 2012. Point of Gaze Estimation through Corneal Surface Reflection in an Active Illumination Environment. In *Computer Vision – ECCV 2012*, Andrew Fitzgibbon, Svetlana Lazebnik, Pietro Perona, Yoichi Sato, and Cordelia Schmid (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 159–172. https://doi.org/10.1007/978-3-642-33709-3_12

[15] Christian Nitschke, Atsushi Nakazawa, and Haruo Takemura. 2011. Practical display-camera calibration from eye reflections using coded illumination. In *The First Asian Conference on Pattern Recognition*. IEEE, Beijing, China, 550–554. https://doi.org/10.1109/ACPR.2011.6166661

[16] Takehiko Ohno, Naoki Mukawa, and Atsushi Yoshikawa. 2002. FreeGaze: A Gaze Tracking System for Everyday Gaze Interaction. In *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications* (New Orleans, Louisiana) *(ETRA '02)*. Association for Computing Machinery, New York, NY, USA, 125–132. https://doi.org/10.1145/507072.507098

[17] Masato Sasaki, Takashi Nagamatsu, and Kentaro Takemura. 2018. Cross-Ratio Based Gaze Estimation Using Polarization Camera System. In *Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces* (Tokyo, Japan) *(ISS '18)*. Association for Computing Machinery, New York, NY, USA, 333–338. https://doi.org/10.1145/3279778.3279909

[18] Masato Sasaki, Takashi Nagamatsu, and Kentaro Takemura. 2019. Cross-Ratio Based Gaze Estimation for Multiple Displays Using a Polarization Camera. In *The Adjunct Publication of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 1–3. https://doi.org/10.1145/3332167.3357095

[19] Masato Sasaki, Takashi Nagamatsu, and Kentaro Takemura. 2019. Screen Corner Detection Using Polarization Camera for Cross-Ratio Based Gaze Estimation. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research &*

Applications (Denver, Colorado) *(ETRA '19)*. Association for Computing Machinery, New York, NY, USA, Article 24, 9 pages. https://doi.org/10.1145/3314111.3319814

[20] Laura Sesma-Sanchez, Arantxa Villanueva, and Rafael Cabeza. 2012. Gaze Estimation Interpolation Methods Based on Binocular Data. *IEEE Transactions on Biomedical Engineering* 59, 8 (2012), 2235–2243. https://doi.org/10.1109/TBME.2012.2201716

[21] Sheng-Wen Shih and Jin Liu. 2004. A novel approach to 3-D gaze tracking using stereo cameras. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 34, 1 (2004), 234–245. https://doi.org/10.1109/TSMCB.2003.811128

[22] Yusuke Sugano, Yasuyuki Matsushita, and Yoichi Sato. 2014. Learning-by-Synthesis for Appearance-Based 3D Gaze Estimation. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Columbus, OH, USA, 1821–1828. https://doi.org/10.1109/CVPR.2014.235

[23] Mélodie Vidal, Andreas Bulling, and Hans Gellersen. 2013. Pursuits: Spontaneous Interaction with Displays Based on Smooth Pursuit Eye Movement and Moving Targets. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Zurich, Switzerland) *(UbiComp '13)*. Association for Computing Machinery, New York, NY, USA, 439–448. https://doi.org/10.1145/2493432.2493477

[24] Arantxa Villanueva and Rafael Cabeza. 2008. A Novel Gaze Estimation System With One Calibration Point. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 38, 4 (2008), 1123–1138. https://doi.org/10.1109/TSMCB.2008.926606

[25] 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

[26] 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 Fifth IEEE International Conference on Automatic Face Gesture Recognition*. IEEE, Washington, DC, USA, 101–106. https://doi.org/10.1109/AFGR.2002.1004139

[27] Xucong Zhang, Yusuke Sugano, Mario Fritz, and Andreas Bulling. 2019. MPIIGaze: Real-World Dataset and Deep Appearance-Based Gaze Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41, 1 (2019), 162–175. https://doi.org/10.1109/TPAMI.2017.2778103

[28] Zhengyou Zhang and Qin Cai. 2014. Improving Cross-Ratio-Based Eye Tracking Techniques by Leveraging the Binocular Fixation Constraint. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (Safety Harbor, Florida) *(ETRA '14)*. Association for Computing Machinery, New York, NY, USA, 267–270. https://doi.org/10.1145/2578153.2578202

Received November 2021; revised January 2022; accepted April 2022