3D Model-Based Gaze Tracking Via Iris Features With a Single Camera and a Single Light Source

Jiahui Liu, Jiannan Chi, Wenxue Hu, and Zhiliang Wang

***Abstract*—Traditional 3D gaze estimation methods are usually based on the models of pupil refraction and corneal reflection. These methods typically rely on multiple light sources. The 3D gaze can be estimated using single-camera-single-light-source systems only when certain user-dependent eye parameters are available *a priori*, which is rarely the case. This article proposes a 3D gaze estimation method which works based on iris features using a single camera and a single light source. User-dependent eye parameters involving the iris radius and the cornea radius are user-calibrated. The 3D line-of-sight is estimated from the optical axis and the positional relationship between the optical axis and the visual axis, and then optimized using a binocular stereo vision model. The feasibility and robustness of the proposed method are assessed by simulations and practical experiments. The system configuration required by the method is simpler than that required by the state-of-the-art methods, which shows significant potential value, especially with regard to mobile device applications.**

***Index Terms*—3D gaze estimation, iris radius, kappa angle, single-camera-single-light-source.**

I. Introduction

EYE gaze tracking systems are widely used in human–computer interaction, medical diagnosis, traffic safety, military, human factors analysis, and virtual reality applications. They are relatively noninvasive, simple to operate, and perform well. They work based on features of the human face, eyes, and their movements, which play an important role in understanding the user's desires and needs [1].

Nonintrusive system design is a popular research topic with regard to modern gaze tracking technology. It is mainly based on the appearance or the features of the eyes. Appearance-based methods map the eye appearance to the point of regard (POR) on the screen directly through a regression function, which is constructed from a large set of statistical samples [2]. This does not, however, account for the influence of natural head movements and variable illumination conditions [3], [4]. By contrast, feature-based gaze tracking methods are more accurate and allow for natural head movement [4]. Existing feature-based methods may be 2D mapping-based or 3D model-based [5]. 2D gaze tracking systems typically estimate the POR from a calibrated gaze-mapping function without knowing the 3D direction of the gaze [3], [6]–[14]. 3D gaze tracking systems model the common physical structures of the human eye geometrically to estimate the 3D gaze direction [15], [16], [17], [18]–[27].

The most common approach for 3D gaze estimation involves geometrical models of pupil refraction and one or more corneal reflections [17]–[25]. The 3D positions of the cornea center and the pupil center need to be estimated. The line of sight (LOS) is usually estimated using multiple light sources. Certain user-dependent eye parameters cannot be solved for gaze estimation in the existing single-light-source 3D gaze tracking systems, unless they are already known or set according to the population averages [18]–[20]. Villanueva and Cabeza [20] stated that the determination of the 3D cornea center was necessary to use a single camera and two light sources when the cornea radius was known. Guestrin and Eizenman [18] demonstrated that a single camera and a single light source can estimate a 3D gaze only when the cornea radius, the distance between cornea center and pupil center (DCP), the indices of refraction, and the distance between the eye and the camera are all known *a priori*. Unfortunately, this is not normally the case. Individual differences are also not taken into consideration if they are preset according to the population averages. Ohno *et al.* [19] used various cornea radius and DCP information ($+/-10\%$) using a single camera and a single infrared light-emitting diode (LED) array to find that gaze estimation accuracy is improved by inputting the DCP of individual users. To this effect, the calibration of user-dependent eye parameters can markedly enhance the performance of the single-light-source gaze tracking system.

Most existing methods also rely on pupil features to estimate the pupil center and to reconstruct the optical axis (OA) of the eye for 3D gaze estimation [17]–[25], [28]. The geometric model between the real pupil and the pupil imaging ellipse is expressed as per the pupil refraction on the corneal surface. Although the cornea radius and the DCP can be calibrated, refraction calculation equations are relatively complicated; the pupil radius is the variable parameter that can change the shape of the eye

Manuscript received July 29, 2019; revised February 8, 2020, June 20, 2020, and September 9, 2020; accepted October 24, 2020. Date of publication November 25, 2020; date of current version March 12, 2021. This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFB1002804, in part by the Beijing Municipal Natural Science Foundation under Grant 4172040, in part by the Fundamental Research Funds for the Central Universities under Grant FRF-GF-20-04 A, in part by the Scientific and Technological Innovation Foundation of Shunde Graduate School, USTB, and in part by the Opening Project of Key Laboratory of Operation Safety Technology on Transport Vehicle, Ministry of Transport, China. This article was recommended by Associate Editor H. Zhou. *(Corresponding author: Jiannan Chi.)*

Jiahui Liu, Jiannan Chi, and Wenxue Hu are with the School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China (e-mail: ljh@xs.ustb.edu.cn; ustbjnc@ustb.edu.cn; g20183012@xs.ustb.edu.cn).

Zhiliang Wang is with the School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China (e-mail: wzl@ustb.edu.cn).

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/THMS.2020.3035176

model [1]. The iris, conversely, is less influenced by refraction; its radius is an invariant user-dependent eye parameter that can be easily calibrated [1]. The special structure of the iris, namely, the transition from white-to-dark and dark-to-white, makes it possible to segment the iris from eye region reliably [26]. At present, iris features are mainly used in 2D mapping-based [3], [6]–[8], [29] and appearance-based [2], [30], [31] methods, though some have extended them to 3D gaze estimation methods [15], [16], [26], [32]. Wang and Sung [26] used the iris normal vectors of both eyes to determine the gaze based on a two-camera system rather than the traditional depth-camera-based system. The gaze accuracy is low, however, due to negligence of the deviation angle between the OA and the visual axis (VA).

The method proposed in this article is iris-feature-based and works in the single-camera-single-light-source system. The 3D gaze is estimated after the user-calibration of iris radius and cornea radius. The final POR is determined via binocular combination optimization. The novelty of this method is twofold.

1) Iris features are used to resolve limitations inherent to traditional 3D gaze estimation methods based on the models of pupil refraction and corneal reflections, and realize the calibration of user-dependent eye invariant parameters through a user calibration process in a single-light-source system.

2) A novel binocular combination optimization method is used to optimize the 3D PORs, which significantly improves 3D gaze estimation accuracy over traditional methods.

The rest of this article is organized as follows. Section II discusses previous research on 3D gaze estimation. Section III shows the structure of the eyeball and the eye gaze. Section IV defines the gaze tracking problem in the existing multi-light-source system; the proposed method is described in Section V. Section VI presents the computer simulations and practical experiments conducted to validate the proposed method. Section VII concludes this article.

II. Related Works

Feature-based gaze tracking methods estimate the 2D POR via mapping model or the 3D LOS via geometric model based on features of the eyeball that are visible in the captured image (e.g., the pupil, iris, eye corner, or glint). Extant methods, as mentioned above, may be 2D mapping-based or 3D model-based. The proposed method is a 3D gaze estimation method, so this discussion centers on 3D model-based gaze tracking technology. 3D gaze trackers estimate the real 3D gaze direction through the 3D eyeball structure and eye imaging model. These approaches include depth-sensor-based methods and common-camera-based methods.

A. Depth-Sensor-Based Gaze Tracking Technology

Depth-sensor-based methods usually use Kinect to obtain depth information and convert the POR from the eyeball center, which are simple in configuration and allow for natural head movements. Li *et al.* [15] performed 3D gaze estimation under natural head movements, wherein the eyeball center was computed with the inner eye corner as the anchor point for head pose tracking. The gaze estimation accuracy was 1.38°–2.71°. Although this model is simple, the proposed screen calibration method is overly complex. Zhou *et al.* [16] presented a personal calibration method that only requires one calibration point with a two-eye model; its estimation accuracy can reach 1.99°. Its gaze point accuracy is relatively low, however, as the LOS is approximately presented by the line connecting the iris center and the POR. These methods estimate the POR according to the vector superposition of middle vectors, so the accuracy of the POR is very dependent on the accurate detection and estimation of each parameter. To this effect, these methods have relatively low accuracy on the whole.

Unlike this, Wang and Ji [27] proposed a real-time eye gaze tracking method using a 3D deformable eye–face model. They first constructed the generic 3D deformable eye–face model based on the recovered 3D rigid facial landmarks and the estimated 3D eyeball center. The personal eye parameters and individual 3D eye–face model were then determined during a personal calibration to estimate 3D eye gaze. This method is more robust against head movement compared to other model-based methods; however, more features are needed, such as facial landmarks, and the gaze estimation framework is more sophisticated.

B. Common-Camera-Based Gaze Tracking Technology

Common-camera-based gaze tracking methods usually estimate the 3D LOS based on pupil refraction and cornea reflection. This begins with determination of the user's eye-invariant parameters. These eye-invariant parameters are then used to estimate eye variable parameters during the 3D gaze tracking process; the OA is reconstructed and the 3D LOS of the eye is estimated accordingly. Eye invariant parameters are invariant for each user but with pronounced individual differences (e.g. the cornea radius, iris radius, and kappa angle). Eye variable parameters are those which are closely related to the 3D LOS during eye movement (e.g., the cornea center, pupil center, and iris center). Common-camera-based gaze tracking systems can be divided into single-camera-single-light-source, single-camera-multi-light-source, and multi-camera system categories based on their respective complexity.

1) Single-Camera-Single-Light-Source System: A few geometric methods do include the use of a single-camera system to estimate 3D gazes relying on population averages for certain eye parameters. Guestrin and Eizenmen [18] demonstrated that the POR can be estimated in a single-camera-single-light-source system if the distance between the eye and the computer screen is known or the user's head is fixed, provided that the cornea radius, DCP, and indices of refraction are all known. These *a priori* parameters differ between different users, so they must be determined individually rather than using fixed values for all users. This is not possible with traditional single-light-source 3D gaze tracking methods. Ohno *et al.* [19] represented the 3D gaze by the OA based on a network camera and an infrared LED array, wherein the cornea radius, DCP, and refractive index were initially set as fixed values taken from the literature, and then

adjusted between 10% smaller and 10% larger in their experiment. They found that the gaze estimation accuracy improves when the DCP of each user is accurately determined.

2) Single-Camera-Multi-Light-Source System: 3D gaze estimation can be achieved in a single-camera-multi-light-source system [28], [33]. Villanueva and Cabeza [20] demonstrated that the minimal hardware needed for the geometric model, which is based on glint positions and pupil ellipse in the image, is a single camera and two light sources. They presented a typical such method. First, when the cornea radius is known, the cornea center can be solved as it consistently falls into the reflection planes and the distance from each reflection point on the cornea surface to the cornea center is equal; the pupil center can be solved by the refractions of pupil edge points, which satisfies the distance from each pupil edge point to the pupil center is equal. Thus, the OA can be reconstructed by the cornea center and the pupil center to determine the LOS. Representing the 3D positions of the cornea center and the pupil center by a series of nonlinear equations can reduce the calculation speed.

Other researchers have used preset eye parameters to replace complex computational process at the expense of estimation accuracy. Morimoto *et al.* [21], for example, modeled a spherical convex mirror for cornea center estimation; they defined the gaze direction by the vector connecting the cornea center and the pupil center, wherein the cornea radius, cornea index of refraction, and the DCP data were taken from the Gullstrand model. Certain individual differences, however, were not taken into consideration.

3) Multi-Camera System: Most 3D model-based gaze estimation methods are operated in multi-camera systems [17], [18], [22]–[25], [34]. Shih and Liu [17] demonstrated that the 3D LOS computed as per the 3D position of the cornea center requires at least two cameras and at least two light sources, when the eye-specific parameters are unknown. The cornea center can be determined by the intersection of the reflection planes of multiple light sources, and then the OA can be directly denoted by the intersection line of the refraction planes as the cornea center and the pupil center both fall into the pupil refraction plane [18], [22], or reconstructed by the cornea center and the pupil center, in which the pupil center needs to be estimated by the relation of pupil refraction first [17], [23]–[25]. Guestrin and Eizenman [23] presented a POR estimation method that tolerates head movements and requires a one-point calibration procedure with three cameras and multiple light sources. They found that it would be more robust using more calibration points for user calibration. Beymer and Flickner [25] used a wide-angle stereo system to detect the face under free head movement and steered an active narrow FOV stereo system to track the eye at high resolution. 3D gaze estimation based on multiple cameras has a straightforward user calibration process and high accuracy even under natural head movement conditions, but the system calibration process is relatively complex.

In summary, most 3D gaze trackers consist of a single camera and multiple light sources or multiple cameras and multiple light sources. These trackers have workable gaze estimation models and fairly accurate LOS output, but require complicated system calibration processes. Certain eye-invariant parameters (e.g., cornea radius, DCP) must be known in single-light-source 3D gaze tracking systems to accurately estimate the 3D gaze. They are usually predefined for all users according to the population averages due to a lack of standardized equations for solving them. This drives down the estimation accuracy of these systems on the whole. Therefore, it is challenging to calibrate the eye-invariant parameters in a single-light-source system through user calibration to achieve universal, accurate 3D gaze estimation.

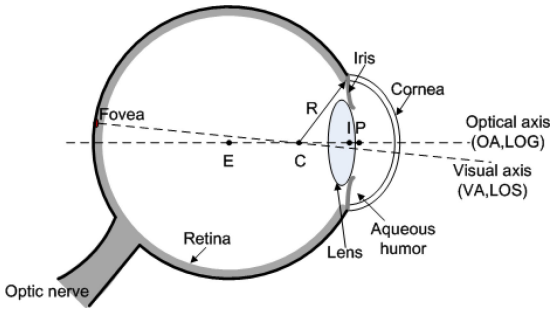


Fig. 1. Eye structure.

III. Eye Model and Eye Gaze

Gaze tracking technology depends on eye parameters obtained from the user. A typical eye model is shown in Fig. 1. Several terms commonly used in gaze tracking systems are listed below.

Optical Axis: This is considered the symmetrical axis of the eyeball [35], wherein the eyeball center (E), cornea center (C), iris center (I), and the pupil center (P) are all in the OA of the eye. The iris and the pupil are the visible part of the eye in a face image, so iris or pupil features are usually used for OA reconstruction.

Visual Axis: This is the line connecting the fovea and the cornea center. The fovea is a small area in the retina with a high density of cones that are responsible for high visual detail discrimination and an individual's central vision [20]. To this effect, the VA is considered to be the real direction of the gaze (namely, the LOS).

Kappa Angle: This is the angular offset between the visual and the optical axes, in which the intersection is the cornea center. The kappa angle is user-dependent and varies among individuals [20]. It is approximately 5° for most.

In summary, the key to effective eye gaze tracking is estimation of the LOS. It is difficult, however, to estimate the LOS directly due to the invisibility of the fovea; the LOS is instead estimated based on the reconstruction of the OA. Once the OA is reconstructed by iris or pupil features, the LOS can be estimated by the kappa angle.

IV. Problem Statement

According to the law of reflection, the incident and reflected light is contained in a plane together with the light source, the cornea center, the camera optical center, and the glint. When multiple light sources are used, the line connecting the cornea

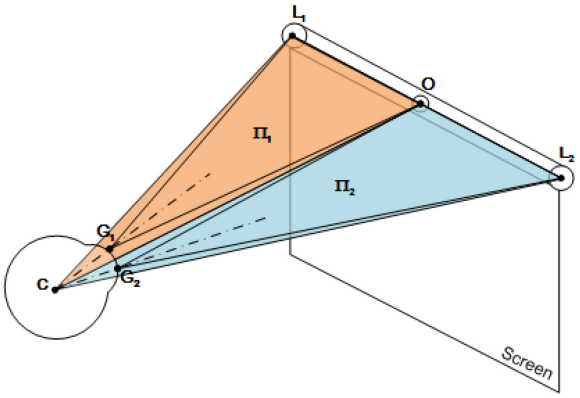


Fig. 2. Reflection model.

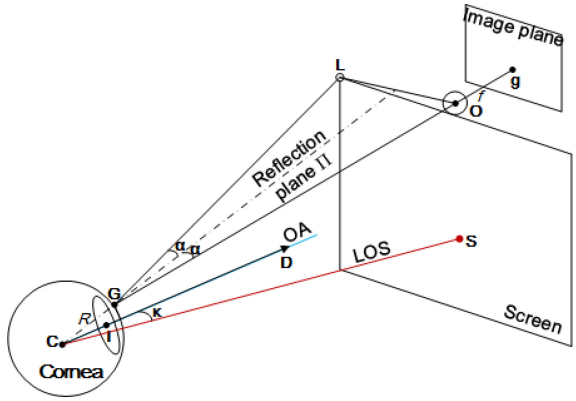


Fig. 3. Geometric relationship between the eye, camera, and the screen when the user looks at S.

center C and the camera optical center O can be determined per the plane intersection to estimate the cornea center [17], [18], [20], [22]–[24]. If the multiple light sources and the camera optical center are collinear, however, the system consisting of multiple light sources is equivalent to a system with a single light source.

Consider two light sources as an example as shown in Fig. 2. The light source L_1 produces a reflection on the corneal surface and the reflection point is G_1. The reflected light passes the camera optical center O and intersects with the image plane at a glint. Thus L_1, C, and O form a plane Π_1. Similarly, the light source L_2 produces a reflection on the corneal surface; the reflection point is G_2. The reflected light passes the camera optical center O and intersects with the image plane at the glint. Thus L_2, C, and O form a plane Π_2. Because L_1, L_2, and O are collinear, Π_1 and Π_2 actually form the same plane. Therefore, when the multiple light sources and the camera optical center are collinear, the system can be simplified to a single-light-source system. 3D gaze estimation in the single-light-source system was conducted in this study based on this assumption.

V. Proposed Method

In this article, a 3D gaze tracking method using a single camera and a single light source is proposed. It mainly includes two steps: user calibration and 3D gaze estimation. The subsections below describe each step in further detail.

A. User Calibration

User calibration is performed to obtain user-dependent eye parameters which replace the traditional population averages or empirical values. This mitigates the differences in individual human eyes. The eye parameters to be calibrated for 3D gaze estimation in this case are the iris radius and the cornea radius.

1) Iris Radius: As shown in Fig. 3, the light source L produces a reflection on the corneal surface. The reflection point is denoted by G. The reflected light passes the camera optical center O and intersects with the image plane at g. The normal line passes through the cornea center C and the reflection point G. According to the law of reflection, both incident and reflected lights are contained in a plane Π together with the light source, cornea center, camera optical center, and glint. The normal vector of Π is $\overrightarrow{n} = (L - O) \times (g - O)$ and the cornea center C satisfies

$$(L - O) \times (g - O) \bullet (C - O) = 0. \tag{1}$$

The OA of the user's eye is always perpendicular to the plane determined by the iris edges and passes through the iris center. Thus, the points on the OA can be expressed by using the iris center I and its normal vector $\overrightarrow{D}$ approximately as

$$P(t) = \begin{bmatrix} I & \overrightarrow{D} \end{bmatrix} \begin{bmatrix} 1 \\ t \end{bmatrix}. \tag{2}$$

where t is the parametric variable of the OA parametric equation. The spatial iris center I and its normal vector $\overrightarrow{D}$ are estimated by the iris features detected from the face image using the spatial circular target reconstruction method [36] since the iris of the user's eye can be regarded as a circular target. Suppose that the equation of iris imaging ellipse in the image coordinate system is $a_u u^2 + b_u v^2 + c_u uv + d_u u + e_u v + f_u = 0$ and the camera focal length is f_c, the equation of the cone constructed with the camera optical center as the apex and the iris imaging ellipse as the bottom can be expressed as $ax^2 + by^2 + cxy + dxz + eyz + fz^2 = 0$, where $a = a_u f_c^2$, $b = b_u f_c^2$, $c = c_u f_c^2$, $d = -d_u f_c$, $e = -e_u f_c$, and $f = f_u$. According to the eigenvalues λ_1, λ_2, λ_3 and corresponding eigenvectors e_1, e_2, e_3 of the real symmetric matrix $\mathbf{P} = \begin{bmatrix} a & c/2 & d/2 \\ c/2 & b & e/2 \\ d/2 & e/2 & f \end{bmatrix}$, the spatial iris center and its normal vector can be expressed as $I = r \begin{bmatrix} e_1 & e_2 & e_3 \end{bmatrix} \left[\pm\sqrt{\frac{|\lambda_3|(|\lambda_1|-|\lambda_2|)}{|\lambda_1|(|\lambda_1|+|\lambda_3|)}} \quad 0 \quad \sqrt{\frac{|\lambda_1|(|\lambda_2|+|\lambda_3|)}{|\lambda_3|(|\lambda_1|+|\lambda_3|)}} \right]^T$ and $\overrightarrow{D} = \begin{bmatrix} e_1 & e_2 & e_3 \end{bmatrix} \left[\pm\sqrt{\frac{(|\lambda_1|-|\lambda_2|)}{(|\lambda_1|+|\lambda_3|)}} \quad 0 \quad -\sqrt{\frac{(|\lambda_2|+|\lambda_3|)}{(|\lambda_1|+|\lambda_3|)}} \right]^T$, in which r is the iris radius. It shows that the iris center I is related to the iris radius r.

The cornea center C is not only on the line of OA but also in the reflection plane, so the parametric variable in (2) can be calculated by solving the equations composed of (1) together

with (2)

$$t_c = -\frac{\overrightarrow{n} \bullet I}{\overrightarrow{n} \bullet \overrightarrow{D}}. \tag{3}$$

In this case, the cornea center can be obtained as $C = P(t_c)$.

When the user looks at different on-screen calibration points $S_i (i = 1, 2, \ldots, n)$, the deviation between the OA and the LOS of one's eye is a constant K [37]. The OA direction is expressed by the iris normal vector $\overrightarrow{D_i}$, and the VA direction is expressed by the vector between the cornea center C_i and the known calibration point S_i, which satisfies

$$\frac{\overrightarrow{D_i} \bullet (C_i - S_i)}{\|\overrightarrow{D_i}\| \|C_i - S_i\|} = K. \tag{4}$$

Therefore, the iris radius r can be estimated based on the invariance of the kappa angle. The use of multiple calibration points allows for a solution to (4) because the unknowns are r and K. When $n = 2$, the iris radius r can be obtained.

We used an iterative method in this study to improve the robustness of the calibration result of r. The initial value r_0 in this case is first set according to the theoretical range of iris radius, then the iris center corresponding to each image is uniquely determined. The OA corresponding to each image is constructed according to the determined iris center and the iris normal vector. The cornea center is calculated by solving (1) and (2). Then the kappa angle for each image that corresponds to the set iris radius is estimated respectively. Deviations between the estimated maximum kappa angle and minimum kappa angle among all the calibration images are stored in an array. The iris radius is determined by its position corresponding to the minimum value of this array.

2) Cornea Radius: Fig. 3 shows the geometric relationship between the eye, the camera, and the screen as the user looks at the on-screen point S. The cornea radius is R. The law of reflection states that the angles of the incident and reflected light relative to the normal line are equal. Therefore, the reflection of the light source L on the corneal surface can be expressed as

$$\frac{(L - G) \bullet (G - C)}{\|L - G\|} = \frac{(G - C) \bullet (O - G)}{\|O - G\|}. \tag{5}$$

The reflection point G is imaged as g on the imaging plane of the camera, so G, O, and g are collinear. G can be represented by multiplying g by a proportional coefficient k. The iris radius r is calibrated at this time as well, so the cornea center C can be calculated by the intersection of the reflection plane of the light source and the constructed OA. According to (5), the only unknown is k. It can be estimated by (5) to obtain the coordinates of G.

The distance between the reflection point G and the cornea center C is the cornea radius R[13], [18]–[20], that is

$$\|C - G\| = R. \tag{6}$$

When multiple calibration points S_i are used, the cornea radius is optimized as follows:

$$R^* = \frac{1}{n} \sum_{i=1}^{n} \|C_i - G_i\|. \tag{7}$$

B. 3D Gaze Estimation

Once user calibration process is complete, the iris radius and cornea radius can be directly applied to estimate the user's 3D gaze. In this process, the OA is reconstructed by the cornea center and the iris center; the LOS is then estimated through the cornea center and the VA direction obtained by the OA and kappa angle transformation matrix.

1) OA Reconstruction: The spatial relationship between the cornea and the iris is fixed in the eyeball model. The distance between the cornea center and iris center is a constant that can be determined by their respective coordinates during the calibration process. Assume that this is denoted by E, thus:

$$\|C - I\| = E. \tag{8}$$

The cornea center can be estimated by (1), (5), (6), and (8), in which the unknowns are the cornea center and the proportional coefficient between the reflection point G and the glint g. In order to eliminate any inaccurate solutions, we used the distance between the cornea centers of the left and right eyes (as- obtained in the calibration process) as a screening measure. The distance between the cornea centers of both eyes corresponding to each solution was estimated and compared with that obtained in the calibration process. The cornea centers of the left and right eyes corresponding to the smallest deviation are reported here as the final results.

Once the 3D cornea center is determined, the OA of the eye can be reconstructed by the cornea center C and the iris center I estimated by the iris radius and iris features.

2) LOS Estimation: The OA and VA have a certain deviation angle, namely the kappa angle, so the kappa angle should be estimated in order to obtain the LOS from the OA. Assume that the unit vector of the OA in any calibration position of all calibration positions is $\overrightarrow{O_{p0}}$ as-obtained during user calibration. The OA direction when the user is in any gaze estimation position can be reconstructed (Section V-B.1), which is denoted by $\overrightarrow{O_{p1}}$, so the rotations of the OA can be expressed as

$$\overrightarrow{O_{p1}} = \mathbf{R}\overrightarrow{O_{p0}}. \tag{9}$$

where $\mathbf{R} = \begin{bmatrix} \cos\beta & 0 & -\sin\beta \\ -\sin\alpha\sin\beta & \cos\alpha & -\sin\alpha\cos\beta \\ \cos\alpha\sin\beta & \sin\alpha & \cos\alpha\cos\beta \end{bmatrix}$, which is represented by $g(\alpha, \beta)$ in this article from here on. α and β are, respectively, the rotation angles around the *x*-axis and the *y*-axis as estimated by (9) in real time.

The OA and the VA of the user's eye have a certain positional relationship in the eyeball coordinate system, so the rotation matrix $\mathbf{R}$ is also applicable to the relationship between the VA directions in these two positions. During user calibration, the VA of the user's eye can be estimated by the known calibration point and the estimated cornea center. Assume that the unit vectors of the VA in the selected calibration position and in this gaze estimation position are $\overrightarrow{V_{p0}}$ and $\overrightarrow{V_{p1}}$, respectively. The VA direction in this gaze estimation position can be estimated by

$$\overrightarrow{V_{p1}} = \mathbf{R}\overrightarrow{V_{p0}}. \tag{10}$$

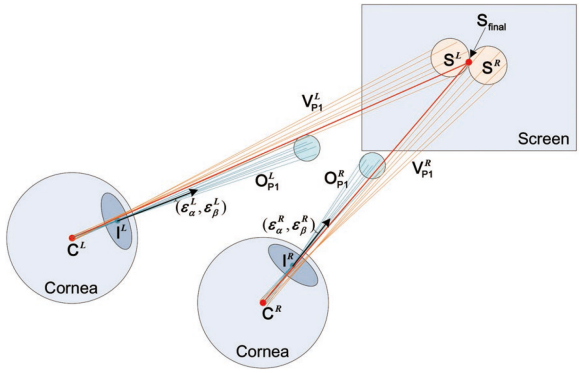


Fig. 4. Relationship between eyeball and LOS as user gazes at any point.

Therefore, the LOS can be determined by $\overrightarrow{V_{p1}}$ and the estimated cornea center C. The POR can also be obtained by the intersection of the LOS and the screen plane.

3) POR Optimization: An optimization method was also established in this study to minimize any errors introduced by insufficient image processing precision or the geometric approximation.

The OA of the user's eye is reconstructed with the cornea center and the iris center. The calculated horizontal and vertical rotation angles corresponding to the OA from the calibration position to this two-eye gaze estimation position are represented by $\alpha_0 = [\alpha_0^L, \alpha_0^R]$ and $\beta_0 = [\beta_0^L, \beta_0^R]$, respectively. The superscripts "L" and "R" represent the parameters of left and right eyes. The real rotation angles are $\alpha = \alpha_0 + \varepsilon_\alpha$ and $\beta = \beta_0 + \varepsilon_\beta$, where $\varepsilon_\alpha = [\varepsilon_\alpha^L, \varepsilon_\alpha^R]$ and $\varepsilon_\beta = [\varepsilon_\beta^L, \varepsilon_\beta^R]$ are the deviations of rotation angles (Fig. 4).

The OA directions of two eyes in any calibration position are $\overrightarrow{O_{p0}} = [\overrightarrow{O_{p0}}^L, \overrightarrow{O_{p0}}^R]$, which are expressed by the corresponding iris normal vector. The OA directions of two eyes in the gaze estimation position are $\overrightarrow{O_{p1}} = [\overrightarrow{O_{p1}}^L, \overrightarrow{O_{p1}}^R]$, asestimated by the method presented in Section V-B.1. Therefore, the real-time rotation matrix $\mathbf{R}$ can be obtained.

Since the VA directions of two eyes in any calibration position are $\overrightarrow{V_{p0}} = [\overrightarrow{V_{p0}}^L, \overrightarrow{V_{p0}}^R]$, which are expressed by the corresponding unit vector between this calibration point and the cornea center, the VA directions of two eyes in the gaze estimation position $\overrightarrow{V_{p1}} = [\overrightarrow{V_{p1}}^L, \overrightarrow{V_{p1}}^R]$ can be estimated by (10). The cornea centers of two eyes in the gaze estimation position are estimated by (1), (5), (6), and (8). The POR $S = [S^L, S^R]$ can be expressed as

$$S = C + \mathbf{M}\overrightarrow{V_{p1}} \tag{11}$$

where $\mathbf{M} = [m^L, m^R]$ can be estimated by the screen equation. The deviation of the PORs of left and right eyes can be derived as

$$B(\varepsilon_\alpha, \varepsilon_\beta) = \|S^L - S^R\|. \tag{12}$$

The gazes of the left and right eyes intersect when the user stares at one point, which satisfies $B(\varepsilon_\alpha, \varepsilon_\beta) = 0$. The equation converts the optimization of the estimated 3D gaze into an estimation of the deviations of rotation angles $\varepsilon_\alpha, \varepsilon_\beta$. The least square method is used to determine these parameters. The solution is

$$\hat{X} = -(A^T A)^{-1} A^T Y \tag{13}$$

where $A = [\frac{\partial S^L}{\partial \alpha^L} \ -\frac{\partial S^R}{\partial \alpha^R} \ \frac{\partial S^L}{\partial \beta^L} \ -\frac{\partial S^R}{\partial \beta^R}], Y = \|S^L - S^R\|$. $\hat{X}$ is used to optimize the POR, as shown in Algorithm 1.

Algorithm 1: POR Optimization Procedure.

Input: The directions of OA and VA in the calibration position $\overrightarrow{O_{p0}}$ and $\overrightarrow{V_{p0}}$; the direction of OA in the gaze estimation position $\overrightarrow{O_{p1}}$; step length $\hat{X}$; upper limit W; coefficient matrix $\mathbf{M}$; given accuracy ε.

Output: The optimal POR S_{final};

$X_0 = g^{-1}(\overrightarrow{O_{p1}}\overrightarrow{O_{p0}}^{-1})$
for $k = 0$ to W **do**
  $X = X_0 + k\hat{X}$;
  $\mathbf{R} = g(X)$;
  $\overrightarrow{V_{p1}} = \mathbf{R}\overrightarrow{V_{p0}}$;
  $S = C + \mathbf{M}\overrightarrow{V_{p1}}$;
  $B = \|S^L - S^R\|$;
  if $\max(B) < \varepsilon$ **then**
    $S_{\text{final}} = \frac{(S^L + S^R)}{2}$;
    Break;
  else
    $k = k + 1$;
  end if
end for
return S_{final};

VI. Simulations and Experiments

Computer simulations and practical experiments were performed to test the feasibility and robustness of the proposed method. We simulated the proposed method using confirmed, accurate data first, and then added different levels of noise to the considered error sources to mimic measurement errors in the real scene. The user calibration and 3D gaze estimation results in the presence of noise characterize the robustness of our method. Subjects were recruited to test the system in a practical experiment as well, where their iris and cornea radii were calibrated and their PORs were estimated using the proposed method. We also compared our method against other iris-feature-based and state-of-the-art gaze tracking methods in terms of gaze estimation performance, as discussed in detail below.

A. Computer Simulations

Computer simulations were carried out using Rhino 5.0 and MATLAB R2015a. The simulation model was developed in Rhino 5.0 as shown in Fig. 5. The eyeball structure was drawn based on the structural parameters in the Gullstrand-Le Grand eye model, wherein the corneal radius is 7.8 mm. The iris radius was set to 5.7 mm and the kappa angle was set to 5°. The light

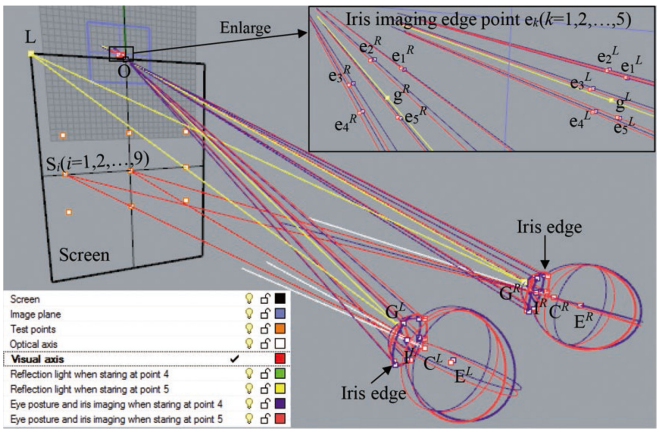


Fig. 5. Simulation model.

TABLE I
ERRORS OF CALIBRATION RESULTS IN SIMULATIONS

Distance/mm	Δr/mm	ΔR/mm
160	-0.0001	0.000288
315	-0.0001	0.0000964
500	0	0.0000311

TABLE II
RMSES OF THE PORS IN SIMULATIONS

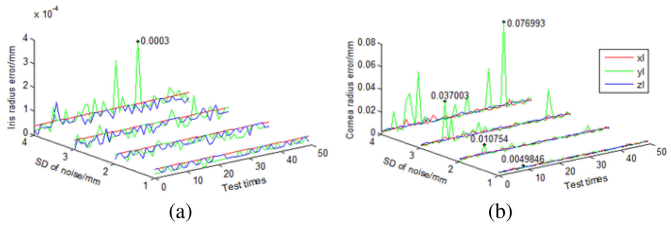
Distance/mm	X/mm	Y/mm
160	0.031	0.0181
315	0.0044	0.2365
500	0.0014	0.1494

source was located at $L = (-80, 0, 0)^T$ in the camera coordinate system. Camera imaging was simulated as the eye was set at three different positions (160 mm/315 mm/500 mm) away from the screen; then data were extracted and input to MATLAB to calibrate the invariant eye parameters and estimate the POR.

1) Feasibility of Proposed Method: The errors in iris radius and cornea radius calibrations are shown in Table I. The deviation between the calibration results and the theoretical values for each eye position is less than 0.0003 mm. It appears that the iris radius and cornea radius calibrations, as per the three positions we tested, are not strongly dependent on the distance from the eye to the screen.

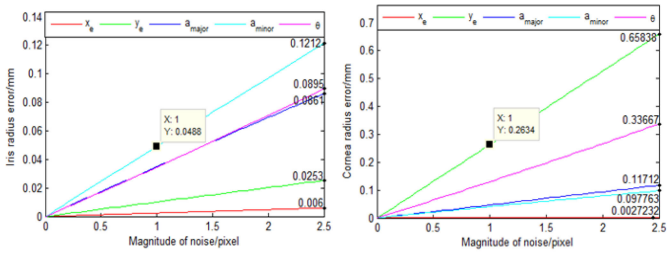
Next, we estimated the PORs based on the calibrated parameters. The root-mean-square errors (RMSEs) of the estimated PORs compared with the predefined calibration points are listed in Table II. Accurate PORs were indeed obtained while using accurate system parameters and image information in the simulations; in other words, the proposed method is feasible. The POR estimations also are not strongly dependent on the distance from the eye to the screen.

2) Analysis of Possible Error Sources: Measurement errors are inevitable in any real-world scene. We simulated noises in the considered error sources, namely, the light source position (x_l, y_l, z_l) and the iris features (involving the major axis of iris imaging ellipse a_{major}, the minor axis a_{minor}, the center of iris imaging ellipse (x_e, y_e), and the inclination angle θ) accordingly.



(a) (b)

Fig. 6. Calibration errors caused by light source with noise. (a) Iris radius. (b) Cornea radius.



(a) (b)

Fig. 7. Calibration errors caused by iris imaging ellipse with noise.

The proposed method depends on the reflection of the light source on the corneal surface, and there may be an error in the position of the light source obtained via system calibration. Independent zero-mean white Gaussian noises [standard deviations (SD) = 1/2/3/4 mm] were added to the coordinates of the light source x_l, y_l, z_l to simulate the influence of this light source position error on the proposed method. The results are shown in Fig. 6. The error of the iris radius is less than 0.0003 mm even when the SD of the Gaussian noise is 4 mm. The position deviation of the light source has a larger effect on the calibration of the cornea radius than that of the iris radius, especially after adding noise in the y direction. When the SD of the Gaussian noise is 4 mm, the maximum error of the cornea radius is about 0.077 mm.

Since the iris features were used to calibrate the user-dependent eye invariant parameters and estimate the 3D gaze, the accuracy of the feature parameters of iris imaging ellipse is very important. When the noises of 0–2.5 pixels were added to the feature parameters of iris imaging ellipse a_{major}, a_{minor}, x_e, y_e, θ in steps of 0.05 pixels, Fig. 7 shows the errors of calibration results caused by the feature parameters of iris imaging ellipse with noise. The error of a_{major}, a_{minor}, θ has a larger effect on the iris radius calibration than the center of iris imaging ellipse. By comparison, y_e has the greatest influence on the cornea radius calibration, followed by θ. Iris and glint detections are based on a subpixel method here, so the error on the image is generally less than 1 pixel. When the noise of each iris feature parameter is 1 pixel, the error of the iris radius is less than 0.05 mm. Error in the iris feature parameters, to this effect, has little effect on the iris radius calibration. The error of the cornea radius reaches 0.2634 mm when the noise added to y_e is 1 pixel, however, which indicates that iris feature parameter error markedly influences the cornea radius calibration.

TABLE III
RMSEs OF THE PORs WITH DIFFERENT LIGHT SOURCE NOISES

SD of noise /mm	x_l		y_l		z_l	
	X/deg	Y/deg	X/deg	Y/deg	X/deg	Y/deg
1	0.119	0.028	0.08	0.041	0.11	0.045
2	0.188	0.042	0.061	0.042	0.127	0.045
3	0.253	0.035	0.15	0.046	0.238	0.046
4	0.292	0.03	0.32	0.136	0.418	0.072

TABLE IV
RMSEs OF THE PORs WITH DIFFERENT IRIS EDGE NOISES

Noise /pixel	x_e		y_e		a_{major}		a_{minor}		θ	
	X/deg	Y/deg	X/deg	Y/deg	X/deg	Y/deg	X/deg	Y/deg	X/deg	Y/deg
0	0.013	0.009	0.019	0.011	0.011	0.012	0.012	0.007	0.012	0.006
0.5	0.091	0.059	0.422	0.400	0.105	0.107	0.109	0.054	0.103	0.041
1	0.162	0.102	0.746	0.812	0.246	0.180	0.233	0.096	0.206	0.081
1.5	0.233	0.145	1.121	1.181	0.373	0.268	0.351	0.144	0.311	0.122
2	0.301	0.186	1.500	1.563	0.560	0.410	0.470	0.192	0.404	0.138
2.5	0.364	0.223	1.755	1.955	0.579	0.434	0.560	0.225	0.507	0.144

Fig. 8. Experimental system.

We next simulated the PORs based on the iris radius and cornea radius calibrated by the light source and iris imaging ellipse with different added noises. The RMSEs of the POR results with different light source noises are listed in Table III, where the RMSE in the X direction is larger than that in the Y direction. When the SD of the Gaussian noise added to the light source is 4 mm, the RMSEs of the PORs are below 0.42°.

The RMSEs of the PORs with different added iris edge noises are shown in Table IV. The RMSEs of the PORs are much larger using y_e with noise than when using other feature parameters of iris imaging ellipse with noise. When the noise added to y_e is 1 pixel, the RMSEs of the PORs are 0.746° and 0.812° in the horizontal and vertical directions, respectively.

In summary, the position error of the light source has less influence on the user calibration and 3D gaze estimation methods proposed in this article. On the contrary, the iris is an important feature of eye movement research. Inaccurate iris feature parameters (especially the ordinate of the center of iris imaging ellipse) will lead the inaccurate calibration of the user-dependent eye invariant parameters and the low accuracy of 3D gaze estimation. Therefore, iris feature parameters, which are the main error source of the proposed method, need to be extracted accurately. Accurate iris detection lays the foundation for achieving high-precision 3D gaze estimation in this article.

(a) (b)

Fig. 9. Experimental user calibration and 3D gaze estimation.

B. Practical Experiments

We conducted experiments using the system shown in Fig. 8 with assistance from seven subjects to further evaluate the proposed method. A single camera and a single light source were used, where the camera is located in the middle and the light source is located on the left side of the camera. The camera is a CMOS camera sensor (model OV7251). Its imaging resolution is 640×480 and the pixel size is 2.2 μm. The focal length of the lens is 3.66 mm. The sclera on the outside of the iris is white and contrasts sharply with the iris, so we used a visible light source rather than an infrared light source in order to detect the features of iris edge more clearly. According to the global calibration method [38], the position of the light source and the screen in the camera coordinate system were calibrated in advance.

Before the experiment began, two calibration points were predefined for the user calibration experiment of each subject; 16 test points distributed on the computer screen were predefined for the 3D gaze estimation experiment of each subject. The screen resolution was 1440×900, and the length and width of the screen was 41 and 25.7 cm, respectively. The subjects were asked to sit in front of the screen at a distance between 350 and 600 mm without glasses and look at the two calibration points comfortably while allowing both eyes to be imaged clearly by the camera, as shown in Fig. 9(a). They were able to slightly move their heads in translation, pitch, and yaw to create slight displacement between the head and the camera. However, head rolling movement was not allowed for the following reason: 3D gaze estimation usually needs to reconstruct the OA direction of the eyeball first, and then convert to the VA direction of the eyeball according to the fixed spatial angle (i.e., kappa angle) between the OA and the VA. If the head rolls, the kappa angle cannot be calculated at this time, and the 3D LOS cannot be estimated. Then they were also asked to naturally cast their gaze at all of the test points, but not to roll their heads, as shown in Fig. 9(b).

1) Eye Feature Detection: As the subject looked at each point on the screen, ten face images are captured by the system camera at a time interval of 100 ms each. Every face image was processed to detect the iris and the glint, as shown in Fig. 10.

Preprocessing: The original color image was converted into a gray image, then the Haar feature was used to recognize the left and right eyes of the user in that image. The Hough transform was used to detect the circular target on the eye images for more accurate iris edge detection. Preliminary screening of the left

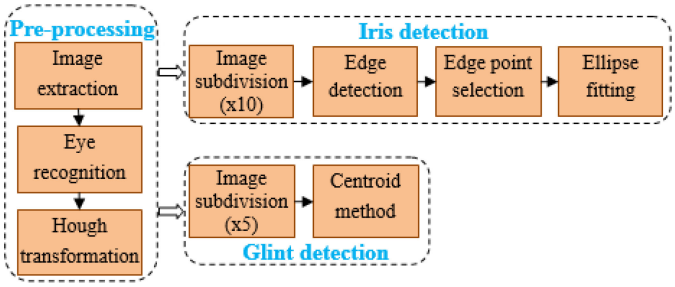


Fig. 10. Eye feature detection procedure.

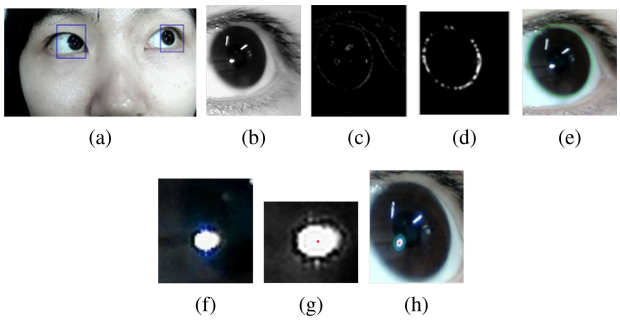


Fig. 11. Eye feature detection results.

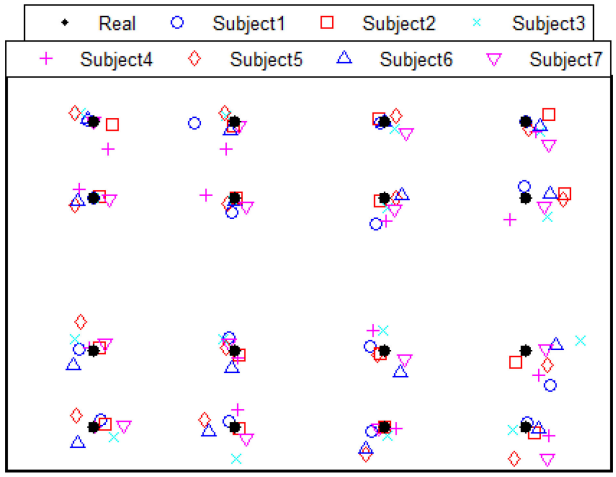


Fig. 12. Deviations between estimated and real PORs of seven subjects.

TABLE V
GAZE ACCURACY OF THE SUBJECTS

Subjects	1	2	3	4	5	6	7	Average
X/deg	0.98	1.02	0.87	1.16	1.28	1.40	1.36	1.15
Y/deg	1.25	0.50	1.01	0.73	1.49	1.29	1.43	1.1

and right eye images were performed to obtain iris and glint images [Fig. 11(a)].

Iris Detection: Left- and right-eye iris images were next subjected to cubic spline interpolation to obtain 10-fold subdivided images [Fig. 11(b)]. The subdivided images were subjected to bilateral filtering to enhance the iris edge. The Canny operator was then used for edge detection on the filtered images to obtain iris edge scatter plots [Fig. 11(c)]. The Hough transform screened the iris edge scatter points further and removed any non-iris edge points introduced by eyelid and eyelash occlusion. Finally, the least squares method was applied for iris edge fitting on the filtered images [Fig. 11(d)] resulting in a set of iris feature parameters. The fitting result is shown in Fig. 11(e). In our experiments, the major axis of the iris in the image was between 22 and 29 pixels, and the minor axis of the iris in the image was between 19 and 27 pixels.

Glint Detection: The cubic spline method was applied to interpolate the glint images to produce fivefold subdivided images [Fig. 11(f)]. The glint image is small and its edge information is easily lost during the process of pixelation, so ellipse fitting is not suitable to determine its center. The centroid method was used to determine its geometric center [Fig. 11(g)]. The result of our experiment is shown in Fig. 11(h).

According to the proposed eye feature detection method, the feature parameters of the iris imaging ellipse and the glint of left and right eyes were extracted. They were then used for subsequent user calibration and 3D gaze estimation.

2) Performance of the Proposed Method: The iris radius and the cornea radius of each subject were user-calibrated by the feature parameters extracted from the face images as the subject looked at two calibration points. The subjects performed no further user calibration afterward. Next, the extracted feature parameters while the subject looked at all of the test points were used to estimate the 3D gazes of each subject. Since ten face images were captured by the system camera when each subject looked at each test point on the screen, ten PORs were obtained for each point. The most accurate one was selected as the final estimated POR corresponding to this test point. Since the test points were predefined and their real coordinates were known, the deviations between the estimated PORs and the real PORs of these subjects on the screen are shown in Fig. 12. The estimated PORs at the bottom of the screen are more scattered than at the top. This is mainly because when the subjects looked at the bottom of the screen, the iris was more seriously occluded by eyelids, eyelashes, and other interference.

By comparing the 16 estimated PORs with the real PORs, the gaze accuracy of each subject as he or she gazed at all test points was calculated using the RMSE, as shown in Table V. The average gaze accuracy of the proposed method is 1.15° in the horizontal direction and 1.1° in the vertical direction. The accuracy of subjects 5–7 was relatively low, but still less than 1.5°. The lower accuracy may be caused by 1) inaccurate iris feature parameters due to eyelid and eyelash occlusion; 2) poor image quality owing to distance selection between the subject and the screen; and 3) subject's inattention during image acquisition. Nevertheless, they are within an acceptable margin of error as the accuracy of gaze tracking in X and Y directions is usually 0.5°–2° [39].

3) Analysis of User Calibration: Ohno *et al.* [19] used cornea radius, DCP, and refractive index data taken from the literature to test their method. Morimoto *et al.* [21] used the fixed values of these parameters from the Gullstrand model. Wang and Sung [26] measured the human eye gaze with a known

TABLE VI
COMPARISON OF GAZE ESTIMATION ACCURACY USING DIFFERENT FIXED PARAMETERS FOR EYE-INVARIANT PARAMETERS

Methods	R/mm	Dcp/mm	r/mm	X/deg	Y/deg
Gullstrand I [21], [40]	7.7	3.6	/	4.54	3.77
Gullstrand-Le Grand [41]	7.8	4.2	/	3.42	3.26
Nararro&Isabel [42], [43]	7.72	4.2	/	3.11	3.09
Liou&Brennan [44]	7.77	4.3	/	3.04	3.07
[19]	7.7	4.5	/	2.34	2.74
[26], [32]	7.8	/	6.3	2.42	2.80
Our	/	/	/	1.15	1.1

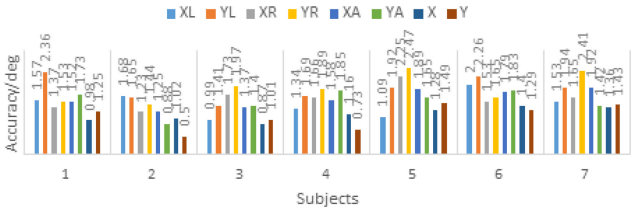


Fig. 13. Comparison of gaze estimation accuracy: single eye versus both eyes; average versus optimize.

TABLE VII
COMPARISON OF GAZE ESTIMATION PERFORMANCE OF IRIS FEATURE-BASED GAZE TRACKING SYSTEM

Categories	Methods	Accuracy/deg	Comments
Appearance-based	[2]	9.95	
	[31]	6.91	
2D mapping-based	[6]	X:0.99;Y:1.33	Fixed head pose
	[8]	2.27	
3D model-based	[15]	1.28-2.71	Kinect
	[26]	1.02	Two cameras system
	Our	X:1.15;Y:1.1	A single camera system

iris radius. Wang *et al.* [32] set the iris radius and the ratio between the eyeball radius and the iris radius as 0.63 cm and 2, respectively, in their experiments. In this study, we estimated user-dependent eye parameters during user calibration based on the single-camera-single-light-source system.

We also performed 3D gaze estimation using the fixed parameters to observe the effects of estimating the user's own eye invariant parameters. As shown in Table VI, the cornea radius R, DCP, and iris radius r from several eye models and literatures are listed. The iris center and the pupil center are approximately coincident, so we used the cornea radius R and the DCP to estimate the required iris radius r by an approximate right triangle. The 3D gazes were then estimated by these eye-invariant parameters using the proposed 3D gaze estimation method. It can be seen from Table VI that the gaze accuracy obtained by preset eye parameters is lower than that using the proposed user calibration method; the latter is more practical for gaze tracking systems with higher accuracy.

4) Analysis of POR Optimization: We established a binocular combination method for POR optimization in this study. To validate this method, we first used the parameters of a single eye to estimate the 3D LOS of both eyes, and calculated the average POR of left and right eyes, then optimized the POR by the binocular stereo vision model. The gaze estimation accuracy of seven subjects based on a single eye/both eyes data are shown in Fig. 13. "XL" and "YL" represent the gaze accuracy of left eye in the horizontal and in vertical directions, respectively; "XR" and "YR" similarly represent the gaze accuracy of right eye. "XA" and "YA" represent the gaze accuracy that averaged the PORs of left and right eyes in horizontal and in vertical directions while "X" and "Y" represent the gaze accuracy of both eyes after binocular optimization. Using the data for both eyes appears to improve the overall estimation performance significantly compared with using single-eye data, and the gaze accuracy optimized by the binocular stereo vision model is higher than that calculated by averaging the PORs of left and right eyes.

5) Comparison of Iris Feature-Based Methods: The proposed 3D gaze estimation method is based on iris features. Previous researchers have also adopted iris-feature-based methods for gaze estimation. We compared our method against several such previously published methods in terms of gaze estimation performance, as shown in Table VII.

The accuracy of the proposed method is much higher than that of the appearance-based method. Certain model-based methods are comparable in accuracy to the proposed method. The 2D-mapping-based method [6], however, requires the user to keep the head fixed, which is difficult to operate as natural head movements are inevitable. By comparison, the proposed method encourages the user to engage with the device comfortably as long as both eyes can be imaged clearly in the camera. The two-camera system 3D gaze estimation method [26] uses multiple cameras to provide more information for gaze tracking, so it is more accurate than the single-camera system. On the whole, the proposed method achieves accuracy comparable to the multi-camera system with a simpler system configuration.

6) Discussion: The proposed method, based on the single-camera-single-light-source system, allows for user calibration of eye-invariant parameters and optimizes the POR via binocular stereo vision model during 3D gaze estimation. It has a gaze accuracy of 1.15° in the horizontal direction and 1.1° in the vertical direction. The user's own invariant parameters (iris radius, cornea radius) obtained from the proposed user calibration process are more universal and more accurate for gaze estimation than using the fixed parameters. The proposed POR optimization method using the binocular stereo vision model can improve the accuracy of gaze estimation over monocular models. Both the user calibration method and POR optimization method outperformed other existing methods on a series of tests. Comparable accuracy was achieved to state-of-the-art methods while using a simpler gaze tracking system.

The proposed method shows high accuracy and strong robustness in several aspects, and shows room for further improvement in others.

Selection of the Light Source: Our gaze tracking system uses a visible light source, which conveniently exploits the contrast between the iris and the sclera for iris detection. However, the visible light source is a little dazzling, and the use of visible light source is susceptible to the influence of ambient light. If the iris can be accurately detected with an infrared light source,

TABLE VIII
COMPARISON WITH STATE-OF-THE-ART METHODS IN RECENT FIVE YEARS

Categories	Methods	Cameras number	Lights number	Features	Accuracy/deg	Head movement	Specialty
Appearance-based	[30]	1	0	Eye appearance	2.37±1.42	Slight	
	[31]	1	0	Iris appearance	6.91	Fixed	
2D mapping-based	[8]	1	0	Iris,Eye corner	2.27	15.5°×15.5°×5°	
	[29]	1	0	Eye	X:1.56;Y:2.56	Fixed	
	[4]	1	2	Pupil,Glints	1.4	Free	RCNN
	[11]	1	5	Pupil,Glints	0.7	Free	Cross-ratio
	[12]	1	4	Pupil,Glints	1.23±0.54	Free	Homography-normalization
3D model-based	[16]	1(Kinect)	0	Iris,Eye corner	1.99	11.63°×13.44°×9.43°	Depth sensor
	[45]	1(Kinect)	0	Iris,Eye corner	X:3.0;Y:4.5	Free	Depth sensor
	[22]	2	2	Pupil,Glints	1.18	50mm×25mm×50mm	
	[28]	1	4	Pupil,Glints	1.3	Free	Without personal calibration
	[46]	4	4	Pupil,Glints	1.5	Free	Head-mounted
	Our	1	1	Iris,Glint	X:1.15;Y:1.1	Free	

the method may provide a better user experience and have better stability.

Selection of the Camera: The camera image resolution used in this study was 640×480, which makes for a highly demanding imaging process. If a high-resolution camera is used to obtain a clearer eye image, the iris detection process would be less cumbersome.

Iris Detection: As computer simulations show, the main source of error in the proposed method is the iris feature parameters. Eyelid and eyelash occlusions complicate the iris ellipse fitting process during iris detection. If these effects could be removed, the estimated PORs when the user regards the bottom of the screen would be more accurate and the overall 3D gaze estimation accuracy would improve.

C. Comparison With State-of-The-Art Methods

State-of-the-art methods with simple, straightforward hardware configurations include appearance-based methods, 2D pupil (iris)-corner technique-based methods, and 3D depth sensor-based methods at present. State-of-the-art 2D methods with high accuracy include cross-ratio-based and homography-normalization-based methods; while 3D methods are based on cornea reflection and pupil refraction using multiple cameras. We made a comparison of several such methods to further assess the proposed method, as shown in Table VIII. Our goal was to achieve high-precision 3D gaze estimation with a simple, straightforward hardware system. The proposed method provides competitive gaze accuracy compared to the state-of-the-art methods we investigated while requiring only a single camera and a single light source.

VII. CONCLUSION

A 3D model-based gaze tracking method based on a single-camera-single-light-source system was developed in this study. The iris features are used in our system, rather than the pupil features used in most of traditional model-based 3D gaze tracking systems. The proposed method works by first calibrating user-dependent eye invariant iris radius and cornea radius, and then estimating the 3D gaze of the user's eyes accordingly. The OA is reconstructed based on the cornea center and the iris center; the LOS is optimized from the OA using a certain positional relationship between the OA and the VA of the eye. We validated our method using computer simulations, practical experiments, and comparison against other state-of-the-art methods. We found that our method estimates PORs with an RMSE of less than 1.2°.

The proposed method uses a single camera and a single light source, which is simpler than the system configuration required by the state-of-the-art methods. This simplified hardware system has great application value, especially with regard to devices such as mobile phones and laptops. The proposed method can be used to effectively user-calibrate the eye-invariant parameters in a single-camera-single-light-source system, and to optimize the LOS using binocular combination during the 3D gaze estimation process.

Currently, we are continuing to work with a single-camera-single-light-source system to achieve automatic user calibration of the iris radius and the cornea radius. Once this work is finished, we will be able to enhance the operability and improve the user experience of the proposed method.

REFERENCES

[1] W. H. Dan and J. Qiang, "In the eye of the beholder: A survey of models for eyes and gaze," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 478–500, Mar. 2010.
[2] E. Wood, T. Baltrušaitis, L. P. Morency, P. Robinson, and A. Bulling, "Learning an appearance-based gaze estimator from one million synthesised images," in *Proc. Biennial ACM Symp. Eye Tracking Res. Appl.*, pp. 131–138, 2016.
[3] J. Sigut and S. A. Sidha, "Iris center corneal reflection method for gaze tracking using visible light," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 2, pp. 411–419, Feb. 2011.
[4] W. Kang, W. Shen, and J. Qiang, "Deep eye fixation map learning for calibration-free eye gaze tracking," in *Proc. Biennial ACM Symp. Eye Tracking Res. Appl.*, pp. 47–55, 2016.
[5] A. Larrazabal, C. G. Cena, and C. Martnez, "Video-oculography eye tracking towards clinical applications: A review," *Comput. Biol. Med.*, vol. 108, pp. 57–66, May 2019.
[6] M. Yu, Y. Lin, X. Tang, D. Schmidt, X. Wang, and Y. Guo, "An easy iris center detection method for eye gaze tracking system," *J. Eye Movement Res.*, vol. 8, no. 3, pp. 1–20, 2015.
[7] A. George and A. Routray, "Fast and accurate algorithm for eye localisation for gaze tracking in low-resolution images," *IET Comput. Vis.*, vol. 10, no. 7, pp. 660–669, Oct. 2016.
[8] Y. M. Cheung and Q. Peng, "Eye gaze tracking with a web camera in a desktop environment," *IEEE Trans. Human-Mach. Syst.*, vol. 45, no. 4, pp. 419–430, Aug. 2015.
[9] X. Li, B. Sheng, W. Wen, L. Ma, and L. Ping, "Accurate gaze tracking from single camera using Gabor corner detector," *Multimedia Tools Appl.*, vol. 75, no. 1, pp. 221–239, Jan. 2016.
[10] P. Blignaut, "Mapping the pupil-glint vector to gaze coordinates in a simple video-based eye tracker," *J. Eye Movement Res.*, vol. 7, no. 1, pp. 1–11, Jan. 2014.

[11] C. Hong *et al.*, "Gazing point dependent eye gaze estimation," *Pattern Recognit.*, vol. 71, pp. 36–44, Apr. 2017.
[12] C. Ma, S. J. Baek, K. A. Choi, and S. J. Ko, "Improved remote gaze estimation using corneal reflection-adaptive geometric transforms," *Opt. Eng.*, vol. 53, no. 5, 2014, Art. no. 053112.
[13] Y. G. Shin, K. A. Choi, S. T. Kim, and S. J. Ko, "A novel single IR light based gaze estimation method using virtual glints," *IEEE Trans. Consum. Electron.*, vol. 61, no. 2, pp. 254–260, May 2015.
[14] M. X. Huang, J. Li, G. Ngai, and V. L. Hong, "Screenglint: Practical, in-situ gaze estimation on smartphones," in *Proc. Conf. Human Factors Comput. Syst.*, pp. 2546–2557, 2017.
[15] S. Li, Z. Liu, and M. T. Sun, "Real time gaze estimation with a consumer depth camera," *Inf. Sci.*, vol. 320, pp. 346–360, Feb. 2015.
[16] X. Zhou, H. Cai, Y. Li, and H. Liu, "Two-eye model-based gaze estimation from a kinect sensor," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2017, pp. 1646–1653.
[17] S. W. Shih and J. Liu, "A novel approach to 3-d gaze tracking using stereo cameras," *IEEE Trans. Syst. Man Cybern. Part B*, vol. 34, no. 1, pp. 234–245, Feb. 2004.
[18] E. D. Guestrin and M. Eizenman, "General theory of remote gaze estimation using the pupil center and corneal reflections," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 6, pp. 1124–1133, Jun. 2006.
[19] T. Ohno, N. Mukawa, and A. Yoshikawa, "FreeGaze: A gaze tracking system for everyday gaze interaction," in *Proc. Eye Tracking Res. Appl. Symp.*, New Orleans, LO, USA, March, 2002, pp. 125–132.
[20] A. Villanueva and R. Cabeza, "A novel gaze estimation system with one calibration point," *IEEE Trans. Syst. Man Cybern. Part B Cybern.*, vol. 38, no. 4, pp. 1123–1138, Aug. 2008.
[21] C. H. Morimoto, A. Amir, and M. Flickner, "Detecting eye position and gaze from a single camera and 2 light sources," in *Proc. Int. Conf. Pattern Recognit.*, 2002, pp. 314–317.
[22] C. C. Lai, S. W. Shih, and Y. P. Hung, "Hybrid method for 3-d gaze tracking using glint and contour features," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 1, pp. 24–37, Jan. 2015.
[23] E. D. Guestrin and M. Eizenman, "Remote point-of-gaze estimation with free head movements requiring a single-point calibration," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2007, pp. 4556–4560.
[24] T. Nagamatsu, R. Sugano, Y. Iwamoto, J. Kamahara, and N. Tanaka, "User-calibration-free gaze tracking with estimation of the horizontal angles between the visual and the optical axes of both eyes," in *Proc. Symp. Eye-Tracking Res. Appl.*, pp. 251–254, 2010.
[25] D. Beymer and M. Flickner, "Eye gaze tracking using an active stereo head," in *Proc. Comput. Vis. Pattern Recognit., Proc. IEEE Comput. Soc. Conf.*, 2003, pp. II–451–458.
[26] J. G. Wang and E. Sung, "Gaze determination via images of irises," *Image Vis. Comput.*, vol. 19, no. 12, pp. 891–911, Oct. 2001.
[27] K. Wang and Q. Ji, "Real time eye gaze tracking with 3d deformable eye-face model," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 1003–1011.
[28] W. Kang and J. Qiang, "3D gaze estimation without explicit personal calibration," *Pattern Recognit.*, vol. 79, pp. 216–227, 2018.
[29] A. Wojciechowski and K. Fornalczyk, "Single web camera robust interactive eye-gaze tracking method," *Bull. Polish Acad. Sci. Tech. Sci.*, vol. 63, no. 4, pp. 879–886, Dec. 2015.
[30] F. Lu, Y. Sugano, T. Okabe, and Y. Sato, "Adaptive linear regression for appearance-based gaze estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 10, pp. 2033–2046, Oct. 2014.
[31] F. Lu, X. Chen, and Y. Sato, "Appearance-based gaze estimation via uncalibrated gaze pattern recovery," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1543–1553, Apr. 2017.
[32] Wang Sung and R. Venkateswarlu, "Eye gaze estimation from a single image of one eye," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2003, pp. 136–143.
[33] C. Hennessey, B. Noureddin, and P. Lawrence, "A single camera eye-gaze tracking system with free head motion," in *Proc. Eye Tracking Res. Appl. Symp.*, pp. 87–94, 2006.
[34] K. Zhang, X. Zhao, Z. Ma, and Y. Man, "A simplified 3d gaze tracking technology with stereo vision," in *Proc. Int. Conf. Optoelectronics Image Process.*, 2011, pp. 131–134.
[35] A. L. Yarbus and B. Haigh, "Eye Movements and Vision," *Quart. Rev. Biol.*, vol. 43, no. 3, p. 360, Sep. 1968.
[36] Y. C. Shiu and S. Ahmad, "3D location of circular and spherical features by monocular model-based vision," in *Proc. IEEE Int. Conf. Syst. Man Cybern.*, 1989, pp. 576–581.
[37] M. F. Land, "The human eye: Structure and function," *Nat. Medicine*, vol. 5, no. 11, pp. 1229–1229, 1999.
[38] J. Chi, Z. Yang, G. Zhang, T. Liu, and Z. Wang, "A novel multi-camera global calibration method for gaze tracking system," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 5, pp. 2093–2104, May 2020.
[39] W. Minfeng, "Computational modeling of gaze tracking," School of Electron. Sci. and Eng., Ph.D. dissertation, Southeast University, 2015.
[40] P. G. Gobbi, F. Carones, and R. Brancato, "Optical eye model for photorefractive surgery evaluation," *Proc. SPIE*, vol. 3591, pp. 10–21, 1999.
[41] A. Villanueva and R. Cabeza, "Evaluation of corneal refraction in a model of a gaze tracking system," *IEEE Trans. Bio-Medical Eng.*, vol. 55, no. 12, pp. 2812–2822, Dec. 2008.
[42] R. S. Navarro, J. Santamara, and J. Bescs, "Accommodation-dependent model of the human eye with aspherics," *J. Opt. Soc. Amer. A Opt. Image Sci.*, vol. 2, no. 8, p. 1273, Sep. 1985.
[43] I. Escuderosanz and R. Navarro, "Off-axis aberrations of a wide-angle schematic eye model," *J. Opt. Soc. Amer. A Opt. Image Sci. Vis.*, vol. 16, no. 8, pp. 1881–91, Sep. 1999.
[44] H. L. Liou and N. A. Brennan, "Anatomically accurate, finite model eye for optical modeling," *J. Opt. Soc. Amer. A Opt. Image Sci. Vis.*, vol. 14, no. 8, pp. 1684–95, Sep. 1997.
[45] J. Li and S. Li, "Gaze estimation from color image based on the eye model with known head pose," *IEEE Trans. Human-Mach. Syst.*, vol. 46, no. 3, pp. 414–423, Jun. 2016.
[46] M. Lidegaard, D. W. Hansen, and N. Krger, "Head mounted device for point-of-gaze estimation in three dimensions," in *Proc. Symp. Eye Tracking Res. Appl.*, pp. 83–86, 2014.