IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 4 November 2022, accepted 28 December 2022, date of publication 4 January 2023, date of current version 9 January 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3234120

RESEARCH ARTICLE

An Improved Cross-Ratio Based Gaze Estimation Method Using Weighted Average and Polynomial Compensation

JIAHUI LIU[1,2], (Member, IEEE), JIE WU[1], HUIJIE YANG[1], AND JIANNAN CHI[1,2,3]
[1]School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China
[2]Beijing Engineering Research Center of Industrial Spectrum Imaging, University of Science and Technology Beijing, Beijing 100083, China
[3]Shunde Innovation School, University of Science and Technology Beijing, Foshan 528399, China

Corresponding author: Jiannan Chi (ustbjnc@ustb.edu.cn)

This work was supported in part by the National Science Foundation for Young Scholars of China under Grant 62206016, in part by the Beijing Municipal Natural Science Foundation under Grant 4212023, in part by the Fundamental Research Funds for the Central Universities under Grant FRF-GF-20-04A, and in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2022A1515140016.

ABSTRACT The cross-ratio (CR)-based method exploits the invariance property of CRs in projective transformation to determine a screen point corresponding to the pupil center. However, this point is essentially the intersection of the eyeball optical axis (OA) and the screen, rather than the actual point-of-regard (POR). In addition, the premise of CR calculation is that the corneal reflection points of four on-screen light sources are coplanar with the 3D pupil center, but they are only assumed to be coplanar. To solve these issues, this paper proposes an improved CR-based gaze estimation method using weighted average and polynomial compensation. Under the configuration of a single camera and two light sources, the 3D corneal center and the normal vector of virtual pupil plane are first estimated using the eyeball imaging model, and then four reference planes parallel to the virtual pupil plane are determined based on the geometric model of pupil and screen corner points. The screen point corresponding to the intersection of each reference plane and the line connecting the camera optical center and the imaging pupil center is calculated using the conventional CR-based method. Thus, the point where the OA intersects the screen is determined by the weighted average of these four points. Finally, a polynomial is learned to compensate it to the POR. The experimental results show that the gaze accuracy can reach 1.33° when the more accurate eye is selected, and it would be improved by 24% by calculating the joint POR, which is competitive with the state-of-the-art methods using more complex systems. On the basis of simplifying the system configuration of CR-based methods, the proposed method avoids the non-coplanarity of 3D pupil center and corneal reflection plane, and improves the gaze estimation performance.

INDEX TERMS Gaze estimation, cross-ratio, corneal reflection, virtual pupil, polynomial compensation.

I. INTRODUCTION

Gaze tracking aims to analyze the user's current gaze direction or gaze point from the user's face or eye features, so as to track the eye gaze. At present, it has been used in many fields such as human-computer interaction, virtual reality, and human factors analysis [1], [2], [3]. The research on gaze tracking is usually divided into appearance-based methods and model-based methods. Appearance-based gaze estimation methods use large training samples to learn the mapping model between face/eye images and 2D/3D gazes, thereby predicting the gaze information of a new image based on the trained model. This method has low requirements on system configuration and strong robustness, but is affected

The associate editor coordinating the review of this manuscript and approving it for publication was Szidonia Lefkovits.

by the limitations of training samples. When there are differences in appearance such as individuals, environments, and head pose, it is difficult to accurately predict the gaze using the trained model, making unconstrained gaze estimation a challenging task [4]. Model-based gaze estimation methods specifically study the relationship between eye features and gaze in the state of eye movement. They use the visual features, such as pupil, iris, and glints, to estimate the 2D POR using the 2D mapping model obtained by user calibration or to estimate the 3D line-of-sight using the geometric imaging model. Model-based methods can effectively deal with the influence of individual differences and head movements, and the gaze accuracy is often between 0.5° and 2° [5].

Model-based gaze estimation methods are generally divided into 2D mapping-based methods and 3D model-based methods. 2D mapping-based methods are based on the invariant features of eye movement, and perform gaze tracking by simulating the 2D mapping function between the variation features of eye movement and the POR. 3D model-based methods are to solve the spatial points such as corneal center and eyeball center according to the eyeball structure and the geometric imaging model, which are used to estimate the OA of the eyeball, and then estimate the 3D line-of-sight combined with the kappa angle. Among the 2D mapping-based methods, the CR-based method is a typical method. It utilizes the projections of four light sources located at the four corners of the screen. Taking the virtual tangent plane determined by the corneal reflection points of light sources as the medium, the screen point corresponding to the 3D pupil center is calculated using the CR invariance, which is regarded as the POR. However, this method has two deficiencies: (1) The CR mapping point on the screen of the pupil center is essentially the intersection of the OA and the screen (POA), rather than the actual POR; (2) The premise of CR calculation is that the corneal reflection points of four light sources are coplanar with the 3D pupil center, but they are only assumed to be coplanar. This deviation is related to parameters such as the distance from 3D corneal center to 3D pupil center and the corneal radius [6].

To deal with these issues, Yoo and Chung [7] proposed to use a scale factor α to convert the corneal reflection points to be coplanar with the pupil. Coutinho and Morimoto [8] not only calibrated the scale factor α, but also calculated an offset vector to compensate for the kappa angle. They also proposed a method to convert the pupil center to the corneal reflection plane [9]. They calculated the intersection of the visual axis (VA) and the corneal reflection plane as the adjusted pupil center, and used the adjusted pupil center and glints to estimate the POR using the CR invariance. Hansen et al. [10] predefined a normalization plane, and used two homography matrices from the image plane to the normalization plane and the normalization plane to the screen plane to convert from the imaging pupil center to the POR. The homography matrix from the normalized plane to the screen plane was determined through a user calibration process with no less than four calibration points. Cheng et al. [11] proposed the idea of using a dynamic virtual tangent plane to improve the computational bias introduced by the pupil center not being in the virtual tangent plane, and described the relationship between the reflections of LEDs and their virtual point using a dynamic matrix. These CR-based methods meet the premise of CR calculation, but the methods require four light sources located at the four corners of the screen, even a light source placed on the camera optical axis to determine a plane tangent to its corneal reflection point. To ensure that all light sources are imaged on the cornea, the range of head movement is limited. Although Sasaki et al. [12] proposed to achieve CR-based gaze estimation without installing near-infrared LEDs on the screen by using a polarized camera to detect the screen, this method has high system requirements.

To improve the actualities, this paper proposes a CR-based gaze estimation method using a single-camera-two-light-source (SCTLS) system. First, the 3D corneal center is estimated by the corneal reflection of the two light sources, and the normal vector of the virtual pupil is determined using the 3D reconstruction method of the spatial circular target. Then, four reference planes parallel to the virtual pupil plane are determined using the relationship between the pupil features and the projections of screen corner points. The screen point corresponding to the intersection of each reference plane and the line connecting the camera optical center and the pupil center is calculated using the conventional CR-based method. Thus, the POA is determined by the weighted average of these four points. Finally, the POR is converted from the POA through polynomial compensation.

The primary contributions of this paper are as follows.

(1) The eyeball imaging model is adopted to achieve the CR-based gaze estimation in a SCTLS system, which eliminates the infrared light sources at the four corners of the screen and simplifies the system configuration of the CR-based method.

(2) This paper proposes to use the four corners of the screen to construct four reference planes, and use the CR invariance to calculate the screen point corresponding to the intersection of the reference plane and the line connecting the camera optical center and the pupil center, which avoids the bias introduced by the non-coplanarity of 3D pupil center and corneal reflection plane.

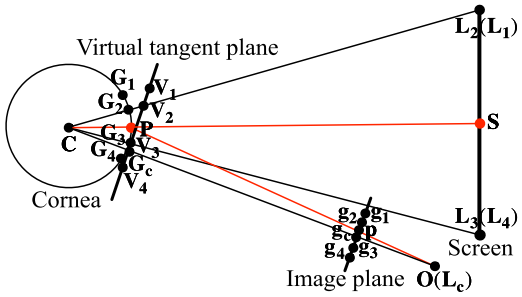
(3) The polynomial compensation not only corrects the POA error obtained by the weighted average, but also considers the kappa angle between the OA and the VA, which improves the gaze estimation performance.

The rest of this paper is organized as follows. Section II introduces the basic principles of the CR-based method and discusses some improved CR-based methods. Section III presents the proposed gaze estimation method in detail. Experimental validation involving computer simulations and practical experiments are analyzed in Section IV. A conclusion is provided in Section V.

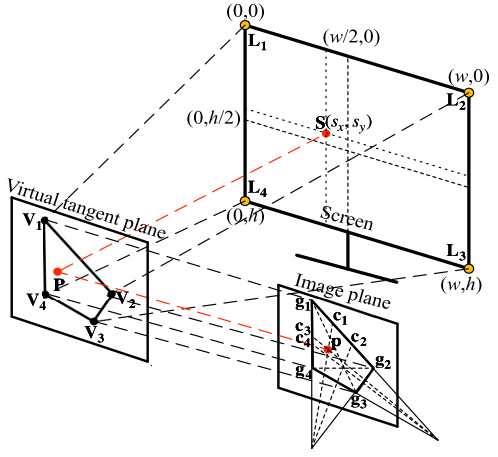
II. RELATED WORKS

A. CONVENTIONAL CR-BASED METHOD

The CR-based method is first proposed by Yoo et al. [13]. It exploits the CR invariance in projective transformation, that is, for a plane polygon, the CR of an edge on the projected image is equal to the CR of the corresponding edge on the plane polygon before geometric transformation. As Fig.1(a) shows, a gaze tracking system using the CR-based method generally consists of five light sources, one of which ($\mathbf{L_c}$) is placed at the camera optical center to determine a virtual tangent plane tangent to its corneal reflection point, and four ($\mathbf{L_i}(i = 1, 2, 3, 4)$) are respectively attached to the four corners of the screen. They have virtual projection points on the virtual tangent plane, and the projection of the virtual projection points on the camera image plane are represented by the glints, as shown in Fig. 1(b). According to the CR invariance, the CR on a certain side of the polygon formed by $\mathbf{L_i}$ is equal to the CR of the points on the corresponding side of the polygon formed by $\mathbf{g_i}$ on the basis of taking the virtual tangent plane as the medium. The 3D pupil center $\mathbf{P}$ is imaged as $\mathbf{p}$, the screen point $\mathbf{S}$ corresponding to $\mathbf{P}$ can be determined by $\mathbf{p}$.



(a) Corneal reflection and virtual projection



(b) CR invariance

FIGURE 1. Basic principle of CR-based method.

If the width of the screen is w and the height is h, and the screen point is $\mathbf{S} = (s_x, s_y)$, the CR of the points on the quadrilateral formed by the screen corner points is

$$\begin{cases} cr_x^{\text{scr}} = \dfrac{s_x}{w - s_x} \\ cr_y^{\text{scr}} = \dfrac{s_y}{h - s_y} \end{cases} . \tag{1}$$

On the image plane, the auxiliary points $\mathbf{c_i} = (x_i^{\text{c}}, y_i^{\text{c}})$ $(i = 1, 2, 3, 4)$ can be determined using the imaging pupil center $\mathbf{p}$ and the diagonal intersection of the quadrilateral formed by $\mathbf{g_i}$. If the glints are expressed by $\mathbf{g_i} = (x_i^{\text{g}}, y_i^{\text{g}})$, then the CR of the points on the corresponding side of the quadrilateral formed by $\mathbf{g_i}$ is

$$\begin{cases} cr_x^{\text{img}} = \dfrac{(x_1^{\text{g}}y_1^{\text{c}} - x_1^{\text{c}}y_1^{\text{g}})(x_2^{\text{c}}y_2^{\text{g}} - x_2^{\text{g}}y_2^{\text{c}})}{(x_1^{\text{g}}y_2^{\text{c}} - x_2^{\text{c}}y_1^{\text{g}})(x_1^{\text{c}}y_2^{\text{g}} - x_2^{\text{g}}y_1^{\text{c}})} \\ cr_y^{\text{img}} = \dfrac{(x_1^{\text{g}}y_3^{\text{c}} - x_3^{\text{c}}y_1^{\text{g}})(x_4^{\text{c}}y_4^{\text{g}} - x_4^{\text{g}}y_4^{\text{c}})}{(x_1^{\text{g}}y_4^{\text{c}} - x_4^{\text{c}}y_1^{\text{g}})(x_3^{\text{c}}y_4^{\text{g}} - x_4^{\text{g}}y_3^{\text{c}})} \end{cases} . \tag{2}$$

According to the correspondingly equal CR between the image plane and the screen, s_x and s_y can be solved, as Eq. (3) shows, which represent the gaze point on the screen.

$$\begin{cases} s_x = \dfrac{w * cr_x^{\text{img}}}{1 + cr_x^{\text{img}}} \\ s_y = \dfrac{h * cr_y^{\text{img}}}{1 + cr_y^{\text{img}}} \end{cases} . \tag{3}$$

B. IMPROVED CR-BASED METHODS

The CR-based method has two major sources of estimation bias: (1) non-coplanarity of 3D pupil center and corneal reflection plane and (2) the angular offset between the OA and the VA. To deal with these biases, there are currently several improved methods, mainly including: CR with feature planarization, CR with kappa angle compensation, homography-normalization (HN)-based method, and CR with dynamic plane projection.

1) CR WITH FEATURE PLANARIZATION

To address the non-coplanarity of 3D pupil center and corneal reflection plane, Yoo and Chung [7] used a scale factor α to convert the glints. It is equivalent to performing a scale of corneal reflection points relative to the corneal center, so that the corneal reflection points and the 3D pupil center are coplanar, and then projecting these converted points into the image plane. The scale factor α can be obtained by multi-point calibration since the gaze point during calibration is known, and it is used to compensate the user's subsequent gaze points. However, this method does not consider the kappa angle. That is, the gaze point estimated by this method is actually the POA, rather than the actual POR defined as the intersection of the VA and the screen. It leads to the fact that the scale factor α cannot accurately make the 3D pupil center and the corneal reflection points coplanar, which limits the performance improvement of the CR-based method.

In addition to correcting the corneal reflection points to be coplanar with the 3D pupil center, Coutinho and Morimoto also proposed to convert the 3D pupil center to the

corneal reflection plane [9]. They calibrated three parameters to represent the 3D corneal center and the intersection of the VA and the iris plane in the eyeball coordinate system. Through the coordinate transformation, a corneal reflection plane in the reference coordinate system was determined according to the minimum sum of the distances from the four corneal reflection points to the corneal reflection plane. The intersection of the VA and the corneal reflection plane was regarded as the 3D pupil center to calculate the gaze point based on CR invariance. However, the iris radius represented by the length of the semimajor axis of the iris imaging ellipse in this method is inaccurate under different head pose, and the average value of the corneal radius will also introduce errors due to the presetting of eye invariant parameters.

2) CR WITH KAPPA ANGLE COMPENSATION

To consider the kappa angle, Coutinho and Morimoto [8] introduced a displacement vector to compensate the kappa angle, and they used α to correct all virtual projections. During user calibration, the set of estimated gaze points for a given α candidate was obtained, so the set of displacement vectors was represented by the difference between calibration points and estimated points. The optimal α should make the displacement vector equal to the average displacement vector when looking at each point. According to this criterion, α was calculated, and then the displacement vector was represented by the average vector of the displacement vectors when looking at all the points. Arar et al. [14] used the conventional CR-based method to estimate the initial gaze point, and introduced a personal calibration method using regularized least squares regression to compensate the kappa angle and ensure good accuracy even when there are few calibration points. And they used an adaptive fusion scheme to perform a weighted average of the gaze points of both eyes, which compensates for the error caused by the low-resolution eye data and increases the range of allowed head motion. They also proposed a weighted regression-based calibration method for a more convenient and user-friendly calibration process [15]. In these methods, five light sources are needed.

3) HN-BASED METHOD

Kang et al. [16] proposed to correct the error of the CR-based method through homography mapping, which does not require a light source placed at the camera optical center. The HN-based method calculates two homography matrices of the two projections from the image plane to the corneal reflection plane and from the corneal reflection plane to the screen plane. When estimating the gaze point, the imaging pupil center is first converted to a point in the corneal reflection plane using the homography matrix from the image plane to the corneal reflection plane, and then the point is converted to an on-screen point using the homography matrix from the corneal reflection plane to the screen plane, which is the gaze point. Hansen et al. [10] used a normalized plane to replace the unknown corneal reflection plane. The matrix from the image plane to the normalized plane was calculated by the four glints and the four corners of predefined normalized plane, and the matrix from the normalized plane to the screen plane was determined by user calibration with no less than four calibration points. The HN-based method does not require an extra light source except for the light sources attached to four screen corners, nor does it need the known screen size owing to the introduced normalized space [17]. Moreover, the homography matrices normalize the head motion before estimating the gaze point, and compensate for the kappa angle, so this method is less sensitive to head pose changes [18]. Huang et al. [19] introduced an adaptive homography mapping model based on learning to simultaneously compensate both spatially-varying gaze errors and head pose dependent errors in a unified framework. Zhang and Cai [20] used the binocular fixation constraint to jointly estimate the homography matrices of both eyes. The joint method achieved over 20% with real data compared with the best single eye, and showed an improvement over 7% compared with the typical average method. In addition, there are several methods to deal with the fact that all four corneal reflection points cannot be detected [18], [21].

4) CR WITH DYNAMIC PLANE PROJECTION

Cheng et al. [11] pointed out that the error caused by the non-coplanarity of the 3D pupil center and the virtual tangent plane is smaller when the gaze point is closer to the light source, since the 3D pupil center is nearer to the fixed virtual tangent plane. Therefore, to improve the gaze estimation performance, they replaced the fixed virtual plane with a dynamic virtual plane. On the basis of calculating the global projection matrix through the off-line learning process, they calculated the local projection matrix when looking at each calibration point. In the online gaze estimation process, the rough gaze point was first calculated using the conventional CR-based method, and the local projection matrices of multiple calibration points were used to calculate the dynamic projection matrix through interpolation. Then the gaze point was calculated based on CR invariance. This method reduces the error caused by the non-coplanarity of the 3D pupil center and the corneal reflection plane, and improves the gaze estimation accuracy, but it increases the number of calibration points.

III. PROPOSED METHOD

The conventional CR-based method introduces a POR error since the 3D pupil center is not in the corneal reflection plane and the kappa angle is not considered. To eliminate this error, this paper proposes an improved CR-based gaze estimation method, which can accurately estimate the POR using a single camera and two light sources. The basic procedure is shown in Fig. 2. First, the 3D corneal center is estimated by the corneal reflection of the two light sources, and the normal vector of the virtual pupil is determined using the 3D reconstruction method of the spatial circular target. Then, four reference planes parallel to the virtual pupil plane are determined according to the relationship between the

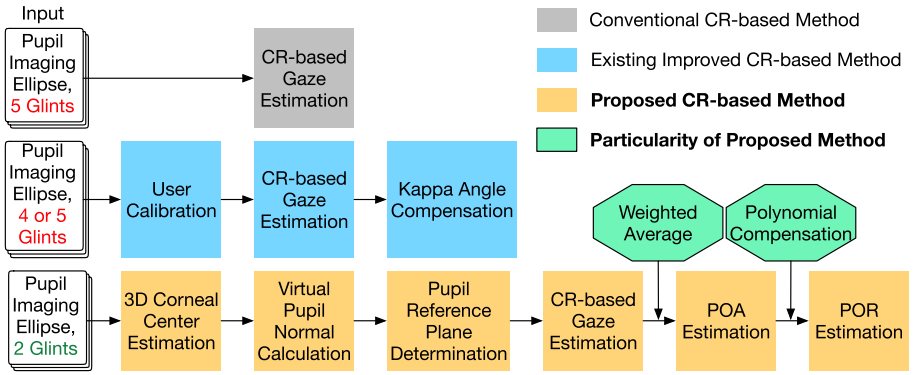


FIGURE 2. Procedure of proposed method.

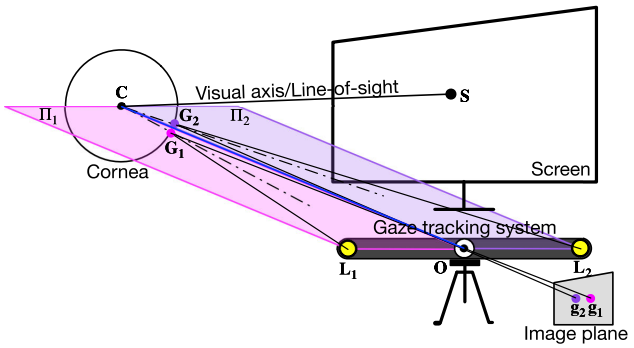


FIGURE 3. Eyeball imaging model for a SCTLS system.

pupil features and the projections of screen corner points. Based on CR invariance, the screen point corresponding to the intersection of each reference plane and the line connecting the camera optical center and the pupil center is calculated respectively, thereby determining the POA by the weighted average. Finally, the POR is converted from the POA through polynomial compensation. The specific methods of each part are discussed in detail below.

A. 3D CORNEAL CENTER ESTIMATION

In a SCTLS gaze tracking system, when the user looks at a point $\mathbf{S}$ on the screen, the user's 3D corneal center can be estimated according to the corneal reflection of the two light sources [22].

As Fig. 3 shows, the light source $\mathbf{L_i}(i = 1, 2)$ reflects on the corneal surface, and the reflection point is $\mathbf{G_i}$. The reflected light passes through the camera optical center $\mathbf{O}$ and intersects the camera image plane at the glint $\mathbf{g_i}$. The normal line connects the 3D corneal center $\mathbf{C}$ and the reflection point $\mathbf{G_i}$. $\mathbf{G_i}$ is located on the corneal surface, so its distance from the corneal center is equal to the corneal radius. If the corneal radius is R, there is:

$$\|\mathbf{G_i} - \mathbf{C}\| = R. \tag{4}$$

Since $\mathbf{G_i}$, $\mathbf{O}$ and $\mathbf{g_i}$ are collinear, $\mathbf{G_i}$ can be expressed as $\mathbf{G_i} = \mathbf{O} + u_i(\mathbf{g_i} - \mathbf{O})$, where u_i is the proportional coefficient. It can be expressed by the 3D corneal center $\mathbf{C}$ and the corneal radius R as $u_i = g(R, \mathbf{C}) = \frac{(\mathbf{g_i}-\mathbf{O})\bullet(\mathbf{C}-\mathbf{O}) \pm \sqrt{((\mathbf{g_i}-\mathbf{O})\bullet(\mathbf{C}-\mathbf{O}))^2 - (\mathbf{g_i}-\mathbf{O})^2((\mathbf{C}-\mathbf{O})^2 - R^2)}}{(\mathbf{g_i}-\mathbf{O})^2}$.

According to the reflection theorem, the incident light, reflected light and normal are coplanar, that is, the light source $\mathbf{L_i}$, reflection point $\mathbf{G_i}$, 3D corneal center $\mathbf{C}$, camera optical center $\mathbf{O}$, and glint $\mathbf{g_i}$ are all in the reflection plane Π_i. The normal vector of the reflection plane Π_i is

$$\mathbf{n_i^{\Pi}} = (\mathbf{L_i} - \mathbf{O}) \times (\mathbf{g_i} - \mathbf{O}). \tag{5}$$

Since the corneal reflection plane of each light source includes the 3D corneal center $\mathbf{C}$ and the camera optical center $\mathbf{O}$, the unit direction vector connecting $\mathbf{C}$ and $\mathbf{O}$ can be expressed as

$$\mathbf{n_{OC}} = \frac{\mathbf{n_1^{\Pi}} \times \mathbf{n_2^{\Pi}}}{\|\mathbf{n_1^{\Pi}} \times \mathbf{n_2^{\Pi}}\|}. \tag{6}$$

Then the 3D corneal center $\mathbf{C}$ can be expressed as $\mathbf{C} = \mathbf{O} + t * \mathbf{n_{OC}}$, where t is the proportional coefficient.

According to the angles of incidence and reflection are equal, it exists:

$$\frac{(\mathbf{L_i} - \mathbf{G_i}) \bullet (\mathbf{G_i} - \mathbf{C})}{\|\mathbf{L_i} - \mathbf{G_i}\|} = \frac{(\mathbf{O} - \mathbf{G_i}) \bullet (\mathbf{G_i} - \mathbf{C})}{\|\mathbf{O} - \mathbf{G_i}\|}. \tag{7}$$

A nonlinear equation system can be obtained by Eq.(4) and Eq.(7), and there are four unknowns, namely: u_i, t and R. According to **Algorithm 1** shown below, the 3D corneal center $\mathbf{C}$ can be solved.

B. VIRTUAL PUPIL NORMAL CALCULATION

The spatial pupil can be regarded as a circular target, which is imaged as an ellipse in the camera image plane. The parameters of pupil imaging ellipse include: the length of semi-major axis a_{major}, the length of semi-minor axis a_{minor}, center (x_e, y_e), and the angle between the major axis and the horizontal direction θ. As shown in Fig. 4, the projection point of the real pupil center $\mathbf{P_r}$ is not the center of the pupil imaging ellipse $\mathbf{p}$, but the point $\mathbf{p_r}$. The spatial circle with the same radius as the pupil radius reconstructed from the

Algorithm 1 3D Corneal Center Estimation

Input: Light sources $\mathbf{L_i}(i = 1, 2)$, glints $\mathbf{g_i}$, initial parameter matrix $[R_0, t_0]$;

Output: 3D corneal center $\mathbf{C}$;

for $i = 1$ to 2 **do**
  $R = R_0, t = t_0$;
  $\mathbf{n_i^{\Pi}} = (\mathbf{L_i} - \mathbf{O}) \times (\mathbf{g_i} - \mathbf{O})$;
end for
$\mathbf{C} = \mathbf{O} + t * \frac{\mathbf{n_1^{\Pi}} \times \mathbf{n_2^{\Pi}}}{\|\mathbf{n_1^{\Pi}} \times \mathbf{n_2^{\Pi}}\|}$;
for $i = 1$ to 2 **do**
  $u_i = g(R, \mathbf{C})$;
  $\mathbf{G_i} = \mathbf{O} + u_i(\mathbf{g_i} - \mathbf{O})$;
  $f_i(R, t) = \frac{(\mathbf{L_i} - \mathbf{G_i}) \bullet (\mathbf{G_i} - \mathbf{C})}{\|\mathbf{L_i} - \mathbf{G_i}\|} - \frac{(\mathbf{O} - \mathbf{G_i}) \bullet (\mathbf{G_i} - \mathbf{C})}{\|\mathbf{O} - \mathbf{G_i}\|}$;
end for
$F(R, t) = \sum_{i=1}^{2} f_i(R, t)^2$;
$R^*, t^* = \text{Index}(\min(F))$;
$\mathbf{C} = \mathbf{O} + t^* * \frac{\mathbf{n_1^{\Pi}} \times \mathbf{n_2^{\Pi}}}{\|\mathbf{n_1^{\Pi}} \times \mathbf{n_2^{\Pi}}\|}$;
return C

pupil imaging ellipse is also not the real pupil, but a virtual pupil. In this paper, the normal vector of the virtual pupil, represented by Π_{P_V}, is calculated using the 3D reconstruction method of the spatial circular target.

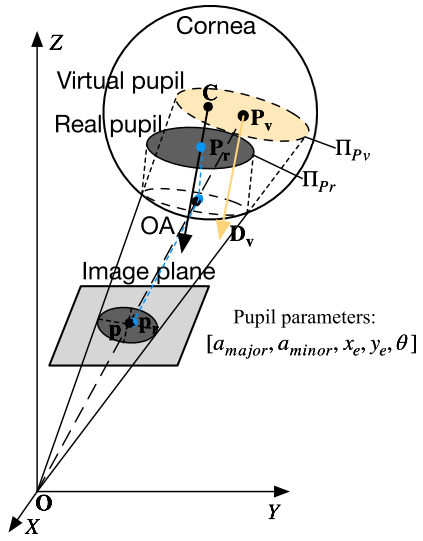


FIGURE 4. Pupil imaging relationship.

According to the parameters of pupil imaging ellipse, the equation of the pupil imaging ellipse in the image coordinate system is expressed as:

$$au^2 + bv^2 + cuv + du + ev + f = 0. \tag{8}$$

Among them,

$$a = \frac{\cos^2\theta}{a_{\text{major}}^2} + \frac{\sin^2\theta}{a_{\text{minor}}^2},$$

$$b = \frac{\sin^2\theta}{a_{\text{major}}^2} + \frac{\cos^2\theta}{a_{\text{minor}}^2},$$

$$c = \sin(2\theta)\left(\frac{1}{a_{\text{major}}^2} - \frac{1}{a_{\text{minor}}^2}\right),$$

$$d = \frac{-2x_e\cos^2\theta - y_e\sin(2\theta)}{a_{\text{major}}^2} - \frac{2x_e\sin^2\theta - y_e\sin(2\theta)}{a_{\text{minor}}^2},$$

$$e = \frac{-x_e\sin(2\theta) - 2y_e\sin^2\theta}{a_{\text{major}}^2} + \frac{2x_e\sin(2\theta) - y_e\cos^2\theta}{a_{\text{minor}}^2},$$

$$f = \frac{(x_e\cos\theta + y_e\sin\theta)^2}{a_{\text{major}}^2} + \frac{(x_e\sin\theta - y_e\cos\theta)^2}{a_{\text{minor}}^2} - 1.$$

Taking the pupil imaging ellipse as the base and the camera optical center as the vertex, an elliptical cone is constructed, then the equation of the elliptical cone is:

$$Ax^2 + By^2 + Cxy + Dxz + Eyz + Fz^2 = 0. \tag{9}$$

The coefficients of the equation can be expressed as $A = af_c^2$, $B = bf_c^2$, $C = cf_c^2$, $D = -df_c$, $E = -ef_c$, $F = f \cdot f_c$ is the focal length of the camera.

Through the coordinate transformation, the equation of the elliptical cone can be converted into the standard form $\begin{bmatrix} x & y & z \end{bmatrix} \mathbf{P} \begin{bmatrix} x & y & z \end{bmatrix}^{\mathrm{T}} = 0$ through a real symmetric matrix $\mathbf{P} = \begin{bmatrix} A & C/2 & D/2 \\ C/2 & B & E/2 \\ D/2 & E/2 & F \end{bmatrix}$. $\mathbf{P}$ has an orthogonal matrix $\mathbf{T}$ such that $\mathbf{T}^{-1}\mathbf{P}\mathbf{T} = diag(\lambda_1, \lambda_2, \lambda_3)$, where λ_1, λ_2, λ_3 are the eigenvalues of the matrix $\mathbf{P}$. According to the three eigenvalues, the three normalized eigenvectors of $\mathbf{P}$ can be obtained correspondingly, which are expressed as: $\mathbf{e_1} = (e_{1x}, e_{1y}, e_{1z})$, $\mathbf{e_2} = (e_{2x}, e_{2y}, e_{2z})$, $\mathbf{e_3} = (e_{3x}, e_{3y}, e_{3z})$. According to the 3D reconstruction method of the spatial circular target [23], the normal vector of the virtual pupil is:

$$\mathbf{D_v} = \begin{bmatrix} e_{1x} & e_{1y} & e_{1z} \\ e_{2x} & e_{2y} & e_{2z} \\ e_{3x} & e_{3y} & e_{3z} \end{bmatrix} \left[\pm\sqrt{\frac{|\lambda_1| - |\lambda_2|}{|\lambda_1| + |\lambda_3|}} \quad 0 \quad -\sqrt{\frac{|\lambda_2| + |\lambda_3|}{|\lambda_1| + |\lambda_3|}} \right]^{\mathrm{T}}. \tag{10}$$

C. PUPIL REFERENCE PLANE DETERMINATION

To eliminate the error caused by the non-coplanarity of the 3D pupil center and the corneal reflection plane, we construct four pupil reference planes using the four corners of the screen, as shown in Fig. 5. The line connecting the screen corner point $\mathbf{V_j}(j = 1, 2, 3, 4)$ and the camera optical center $\mathbf{O}$ intersects the image plane at the point $\mathbf{v_j}$, and the line connecting the 3D corneal center $\mathbf{C}$ and the camera optical center $\mathbf{O}$ intersects the image plane at the point $\mathbf{c}$. With the 3D corneal center $\mathbf{C}$ as the vertex and the quadrilateral connected by four screen corner points as the base, a $\mathbf{C}$-pyramid is formed.

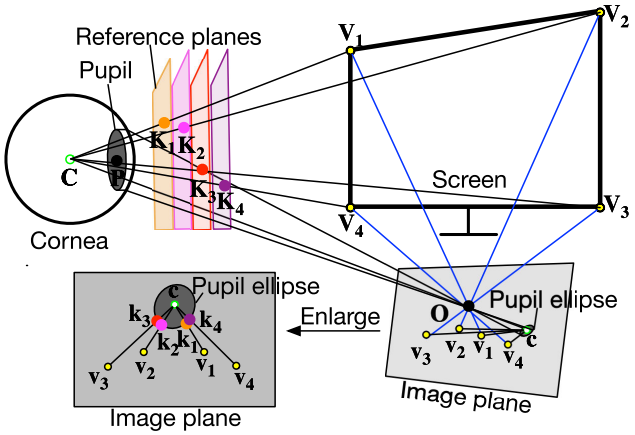


FIGURE 5. Determination of four pupil reference planes.

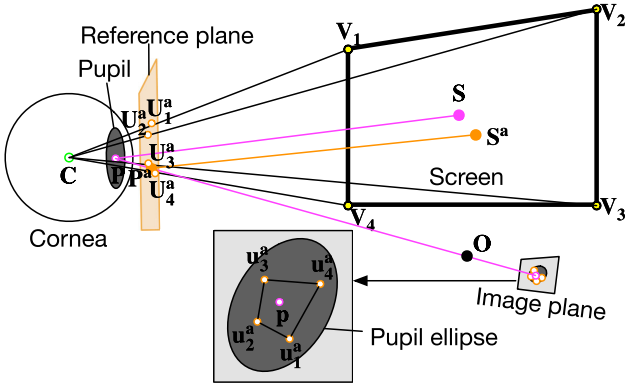


FIGURE 6. Gaze estimation corresponding to the pupil reference plane.

The line connecting $\mathbf{c}$ and $\mathbf{v_j}$ intersects the pupil imaging ellipse at the point $\mathbf{k_j}$, it satisfies:

$$\begin{bmatrix} a & c & d & e & f \end{bmatrix} \begin{bmatrix} (\mathbf{c}[0]+\alpha_j(\mathbf{v_j}[0]-\mathbf{c}[0]))^2 \\ (\mathbf{c}[1]+\alpha_j(\mathbf{v_j}[1]-\mathbf{c}[1]))^2 \\ (\mathbf{c}[0]+\alpha_j(\mathbf{v_j}[0]-\mathbf{c}[0]))(\mathbf{c}[1]+\alpha_j(\mathbf{v_j}[1]-\mathbf{c}[1])) \\ (\mathbf{c}[0]+\alpha_j(\mathbf{v_j}[0]-\mathbf{c}[0])) \\ (\mathbf{c}[1]+\alpha_j(\mathbf{v_j}[1]-\mathbf{c}[1])) \\ 1 \end{bmatrix} = 0. \tag{11}$$

After solving α_j, $\mathbf{k_j}$ is expressed as: $\mathbf{k_j} = \mathbf{c} + \alpha_j(\mathbf{v_j} - \mathbf{c})$.

According to the pinhole imaging model, there is a point $\mathbf{K_j}$ on the corresponding edge $\mathbf{CV_j}$, which satisfies its projection point in the camera as $\mathbf{k_j}$. Thus, a plane passing through the point $\mathbf{K_j}$ and parallel to the virtual pupil plane can be constructed with $\mathbf{K_j}$ and $\mathbf{D_v}$, which is taken as the pupil reference plane. For four corners of the screen, four pupil reference planes can be obtained.

D. CR-BASED GAZE ESTIMATION

After determining the pupil reference planes, the screen points when taking the four pupil reference planes as the medium are estimated based on CR invariance, as shown in Fig. 6. Since the estimation method corresponding to each pupil reference plane is the same, for the convenience of presentation, the pupil reference plane is represented by superscript "a" here. The $\mathbf{C}$-pyramid edge $\mathbf{CV_j}$ intersects the pupil reference plane at the point $\mathbf{U_j^a}$, and $\mathbf{U_j^a}$ is imaged as $\mathbf{u_j^a}$ in the camera image plane. The line connecting the center of pupil imaging ellipse and the camera optical center intersects the pupil reference plane at the point $\mathbf{P^a}$. The CR of the screen is represented using $\mathbf{V_j}$ and $\mathbf{S^a}$, and the CR of the quadrilateral formed by $\mathbf{u_j^a}$ is calculated using $\mathbf{u_j^a}$ and $\mathbf{p}$. Then $\mathbf{S^a}$ can be calculated based on CR invariance represented by Eqs. (1)-(3). In the same way, the screen point corresponding to each pupil reference plane can be calculated, which is represented by $\mathbf{S^{K_j}}$.

E. POA ESTIMATION USING WEIGHTED AVERAGE

The vector from the 3D corneal center $\mathbf{C}$ to the 3D pupil center $\mathbf{P}$ can be represented by the weighted average of the vectors from the 3D corneal center $\mathbf{C}$ to the pupil edge points. In view of this, the POA corresponding to the pupil plane is calculated by taking the weighted average of the screen points corresponding to four pupil reference planes. There is a known point $\mathbf{K_j}$ in the pupil reference plane, and the plane normal vector is $\mathbf{D_v}$. Using the estimated 3D corneal center $\mathbf{C}$, the distance from the 3D corneal center to the pupil reference plane is:

$$d_j = \mathbf{D_v} \bullet (\mathbf{C} - \mathbf{K_j}). \tag{12}$$

To better convert $\mathbf{S^{K_j}}$ to the POA corresponding to the pupil plane, distance weights are adopted. The weight calculation of the pupil reference plane $\mathbf{K_j}$ is

$$w_j = \frac{d_j}{\sum_{t=1}^{4} d_j}. \tag{13}$$

Thus, the POA corresponding to the pupil plane can be calculated as:

$$\mathbf{S_{oa}} = \sum_{t=1}^{4} (w_j * \mathbf{S^{K_j}}). \tag{14}$$

F. POR ESTIMATION USING POLYNOMIAL COMPENSATION

Since the pupil refraction and the kappa angle are not considered, the calculated POA cannot match the POA or the POR corresponding to the real pupil. Therefore, this paper proposes a polynomial compensation method to correct the calculated POA to the real POR.

The kappa angle is an eye invariant parameter, it can be compensated by a displacement vector on the screen [8]. In addition, the CR is a linear transformation. Therefore, a linear polynomial is used to compensate for the POR. Assume that the linear polynomial is expressed by $\mathbf{S_k} = \mathcal{H}(\mathbf{S_{oa,k}}) = \mathbf{w}^{\mathrm{T}}\mathbf{S_{oa,k}} + \mathbf{b}$, it satisfies the minimum sum of distance between

the obtained POR and the calibration point $\mathbf{S_c}$ after the POA $\mathbf{S_{oa}}$ is compensated by the linear polynomial $\mathcal{H}$. That is:

$$\mathcal{H}(\mathbf{w}^*, \mathbf{b}^*) = \arg\min_{(\mathbf{w},\mathbf{b})} \sum_{k=1}^{N} (\mathcal{H}(\mathbf{S_{oa,k}}) - \mathbf{S_{c,k}})^2. \quad (15)$$

where $\mathbf{w}$, $\mathbf{b}$ are the polynomial coefficient matrices, and N is the number of calibration points. $\mathbf{S_{oa,k}}$ and $\mathbf{S_{c,k}}$ represent the kth calculated POA and calibration point respectively.

To obtain the polynomial coefficients, we adopt the least squares method. The polynomial coefficient matrix containing $\mathbf{w}$ and $\mathbf{b}$ is represented by Θ, then the cost function is

$$J = \|\mathbf{S_{oa}}\Theta - \mathbf{S_c}\|^2. \quad (16)$$

By taking the first derivative of the cost function, its optimal solution is

$$\hat{\Theta} = (\mathbf{S_{oa}}^{\mathrm{T}}\mathbf{S_{oa}})^{-1}\mathbf{S_{oa}}^{\mathrm{T}}\mathbf{S_c}. \quad (17)$$

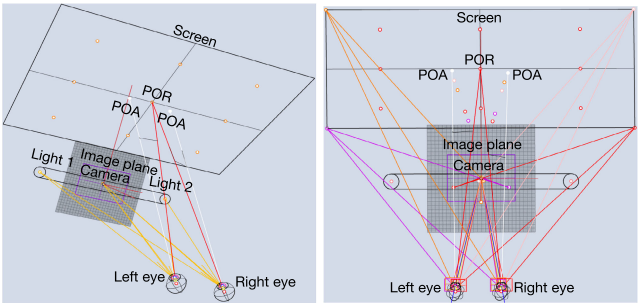
When predicting the POR of a new image, according to the calculated POA $\mathbf{S_{oa,new}}$, the POR can be compensated by Eq. (18) to obtain the accurate POR.

$$\mathbf{S_{new}} = \mathbf{S_{oa,new}}\hat{\Theta}. \quad (18)$$

IV. EXPERIMENTAL VALIDATION

A. COMPUTER SIMULATIONS

To verify the effectiveness of the proposed method, we carried out computer simulations based on simulation data. The simulation model was drawn in Rhino 6 to simulate the scene of the subject sitting in front of the screen and looking at the screen point naturally, as shown in Fig. 7. The gaze tracking system was placed below the screen, and the screen size was 400 mm×300 mm. The camera principal point was (0,0), and the focal length was 3.66 mm. Two light sources were located at (−116.79, −2.3, 8.16) and (116.69, −2.62, 7.68). The eyeball parameters were set according to the Gullstrand-Le Grand eyeball model and the human average [24]. Among them, the eyeball radius was 12 mm, the corneal radius was 7.8 mm, the pupil radius was 2 mm, the distance from the corneal center to the pupil center was 4.2 mm, and the kappa angle was 5° in horizontal and 1° in vertical. The distance from the subject to the screen was about 400 mm. We carried out algorithm verification and error analysis using the corresponding parameters when two eyes look at the nine preset points on the screen.



(a) Corneal reflection (b) Geometric imaging

FIGURE 7. Simulation model.

1) ALGORITHM VERIFICATION

According to the system parameters in the simulation model and the visual features of the eyeball on the camera image plane, including: pupil imaging ellipse and two glints, the screen point corresponding to each pupil reference plane was calculated, as shown in Fig. 8. The screen points corresponding to four pupil reference planes were collinear. Since our simulation model was set according to the real-world scene, the pupil was refracted on the corneal surface. At this time, the parameters of pupil imaging ellipse obtained by the pupil refraction are different from those obtained by pinhole imaging. Therefore, the real POA was not on the line determined by the screen points corresponding to the four pupil reference planes. Pupil refraction effects were considered in subsequent polynomial compensation.

After calculating the weights of the four pupil reference planes, the POA corresponding to the pupil plane was calculated. As shown in Fig. 9(a), they have roughly the same distribution as the real PORs, but they have a certain scaling ratio and position deviation from the real PORs. Fig. 9(b) shows the comparison between the PORs after polynomial compensation and the real PORs. The PORs of the left and right eyes were close to the real PORs. The specific data are shown in Table 1. According to the real PORs in the

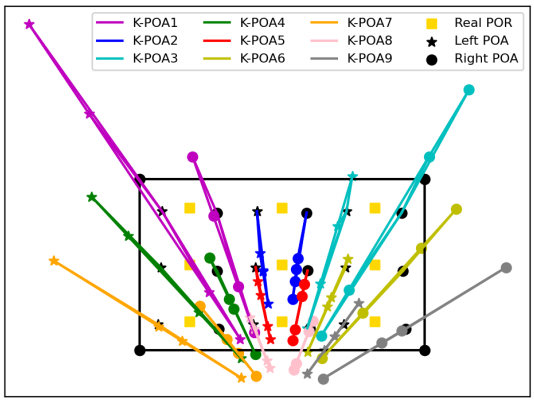
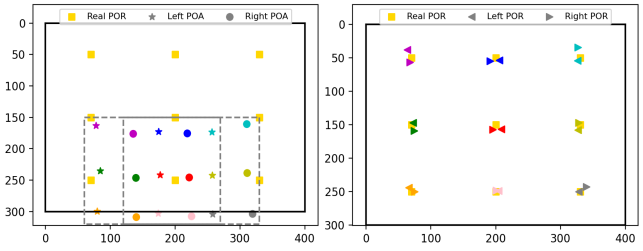


FIGURE 8. Calculated screen points corresponding to the pupil reference planes.

TABLE 1. Calculated PORs of left and right eyes.

Points	Left PORs/mm	Right PORs/mm	Real PORs/mm
1	(63.4438,38.0776)	(67.6204,56.5285)	(70,50)
2	(205.4055,53.5409)	(191.2443,55.1671)	(200,50)
3	(325.8432,54.1574)	(326.8874,34.1496)	(330,50)
4	(72.4327,147.2502)	(74.3220,158.9976)	(70,150)
5	(208.3749,156.6857)	(195.4379,157.4315)	(200,150)
6	(326.6954,157.7294)	(328.1937,147.0922)	(330,150)
7	(65.9503,244.0286)	(75.1911,249.6795)	(70,250)
8	(204.2307,248.5608)	(201.0771,248.2103)	(200,250)
9	(327.6235,249.9696)	(340.0262,242.7438)	(330,250)

simulation model, the overall standard deviations (SDs) of the PORs were calculated. For the left eye, the SDs were 1.63 mm (0.23°) in the horizontal and 1.99 mm (0.29°) in the vertical; for the right eye, the SDs were 1.80 mm (0.26°) in the horizontal and 2.54 mm (0.36°) in the vertical. This demonstrates the effectiveness of weighted average and polynomial compensation.



(a) POA (b) POR

FIGURE 9. Comparison of the estimated POAs and the PORs with the ground-truth.

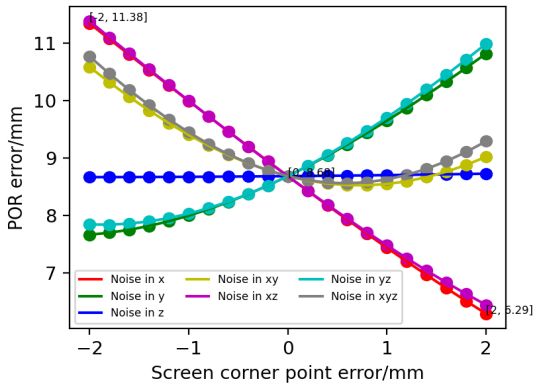
On this basis, we calculated the average of the compensated PORs of left and right eyes. Compared with the real PORs, the overall SDs of the average PORs was 0.99 mm (0.14°) in the horizontal and 1.37 mm (0.20°) in the vertical. It shows that the gaze estimation accuracy can be further improved using the average of the PORs of left and right eyes.

2) ERROR ANALYSIS

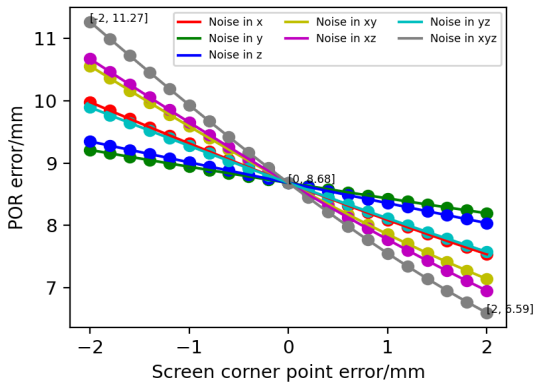
The proposed method uses the four screen corner points and the pupil imaging ellipse to determine the pupil reference planes, and calculates the screen points corresponding to the pupil reference planes to represent the POA and compensate the POR, which reflects the importance of screen calibration and pupil detection. Therefore, we focused on analyzing the influence of screen corner point error and pupil imaging ellipse error on gaze estimation.

(1) Analysis of screen corner point error: The noise with an amplitude of -2 mm-2 mm and a step size of 0.2 mm was added to the screen corner points to simulate the error of screen calibration. When the same noise was added to the four screen corner points at the same time, the variation trend of the POR error is shown in Fig. 10(a). When the x-direction error of the four screen corner points changed from -2 mm to 2 mm, the POR error was negatively correlated with it; when the y-direction error of the four screen corner points changed from -2 mm to 2 mm, the POR error was positively correlated with it; and the z-direction error had little effect on the POR estimation. When the error amplitude of the screen corner points in a certain direction was less than 2 mm, the increment of the POR error did not exceed 2.7 mm (0.39°).

In the real-world scene, the calibration errors of the four screen corner points may be different. Therefore, we also tested the case where only a single screen corner point has error, as shown in Fig. 11(b). The x-direction error had a greater effect on the POR estimation than the y- and z-direction errors. When the errors of the screen corner point



(a) Four screen corner points have same error



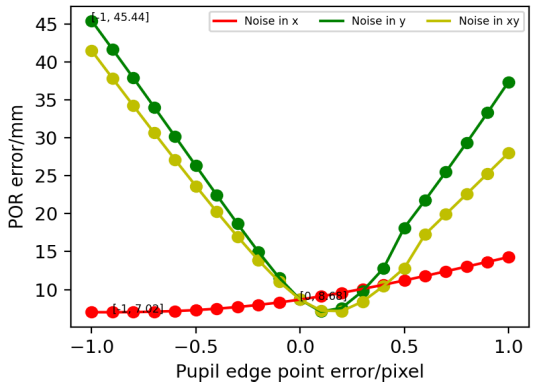
(b) Single screen corner point has error

FIGURE 10. Screen corner point error on POR estimation.

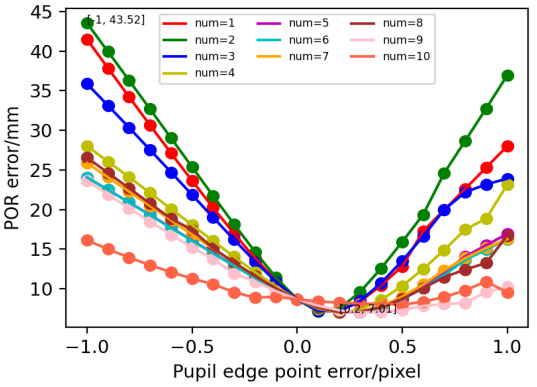
in three directions were all no more than 2 mm, the increment of the POR error did not exceed 2.59 mm (0.37°). Comparing the influence of the error in a single screen corner point and the error in four screen corner points on the POR estimation, it can be seen that part of the POR error was offset.

(2) Analysis of pupil imaging ellipse error: The error of subpixel edge detection is often within one pixel, therefore, we focused on the case where the pupil imaging ellipse error did not exceed one pixel. The pupil imaging ellipse is obtained by extracting the pupil edge points and then fitting the ellipse. To simulate the pupil imaging ellipse error, we added noise ranging from -1 to 1 pixels with a step size of 0.1 pixels to the pupil imaging edge points. When the noise was added to a pupil imaging edge point, the obtained POR error is shown in Fig. 11(a). When there was a 1-pixel error at the pupil imaging edge point, compared with the x-direction error, the y-direction error had a greater influence on the POR estimation.

We also analyzed the influence of the number of pupil imaging edge points containing errors on the POR estimation. When the same noise acted on the x and y components of the pupil imaging edge points at the same time, the POR errors corresponding to the errors of one to ten pupil imaging edge points are shown in Fig. 11(b). When multiple pupil imaging edge points had errors at the same time, the POR error decreased significantly. When there was 1-pixel error in



(a) Different components of a pupil imaging edge point have error



(b) Different numbers of pupil imaging edge points have error

FIGURE 11. Pupil imaging ellipse error on POR estimation.

the x and y components of ten pupil imaging edge points, the increment of the POR error introduced by the pupil imaging ellipse error was 7.45 mm (1.07°).

To sum up, compared with the screen corner point error, the pupil imaging ellipse error had a greater effect on the gaze estimation, which puts forward higher requirements for pupil detection. To ensure the accuracy of gaze estimation, it is necessary to accurately extract the pupil parameters. The computer simulations clarified the influence of screen calibration and pupil detection on the proposed gaze estimation method, which has certain guiding significance for the practical experiments.

B. PRACTICAL EXPERIMENTS

1) EXPERIMENTAL APPARATUS AND PROCESS

To verify the performance of the proposed method in real-world scene, we built a practical system, as shown in Fig. 12(a). An eye tracker was placed under the screen. The device is a SCTLS system, where the CMOS camera was located in the middle and two infrared light sources were located on the both sides of the camera. The imaging resolution of the camera was 1920 × 1080 and the pixel size was 2.2 um. The focal length of the lens was 10.17 mm. Two light sources were both composed of three light emitting diodes with a wavelength of 850 nm. The screen resolution of the screen was 1440 × 900. Their positions were obtained through system calibration [25].

(a) Experimental apparatus (b) Experimental process

FIGURE 12. Experiments in the practical system.

The experimental process is shown in Fig. 12 (b). First, the subjects were asked to sit in front of the screen at a distance between 350 mm and 600 mm. They previewed the face image on the visual interface, and determined an initial optimal effect of pupil and glint detection by adjusting the threshold. Then the fixation task can be performed. They controlled the display of preset points on the screen by pressing the key, and only one on-screen point was displayed at a time. The subjects were asked to look at the displayed point until it disappeared. During fixation, they were able to move their heads naturally to create the displacement between the head and the camera. The experiment was completed when the images of the subjects looking at the nine preset points had been collected.

2) EYE FEATURE EXTRACTION

In the experimental process, when the subjects looked at each on-screen point, image acquisition and image processing were carried out simultaneously. We extracted the pupil imaging ellipse of left and right eyes and two glints. First, Yolov5 network was used to train an eye detection model for predicting the eye region from the face image in real time, which is more stable than the eye location algorithm based on haar feature and adaboost [26].

Pupil detection: On the predicted eye region, threshold segmentation was performed using the Otsu method [27], and the pupil region was roughly located. Then the pupil region was filtered to enhance the pupil edge. Finally, the least squares method was applied for pupil edge fitting on the filtered images.

Glint detection: The glint region was segmented from the eye image using a high threshold. Since the glint is small and its edge information is easily lost during the process of pixelation, the centroid method was used to determine its geometric center after canny edge detection [28].

3) PERFORMANCE ANALYSIS OF GAZE ESTIMATION

The performance of the proposed method for gaze tracking was evaluated using the known camera, light source and

screen parameters, as well as the parameters of pupil imaging ellipse and glints obtained in the experimental process. The evaluation included the accuracy analysis of gaze estimation and the performance quantification of binocular strategy on gaze estimation.

(1) Accuracy Analysis of Gaze Estimation: The on-screen points were preset points, so their coordinates are known. They were used as calibration points to learn the linear polynomial from the POA to the POR, in addition, they were used as the ground-truth of the POR to calculate the gaze estimation error and gaze accuracy. The corresponding POR when the subjects looked at each on-screen point was calculated using the proposed method, and the distribution of the subjects' PORs is shown in Fig. 13.

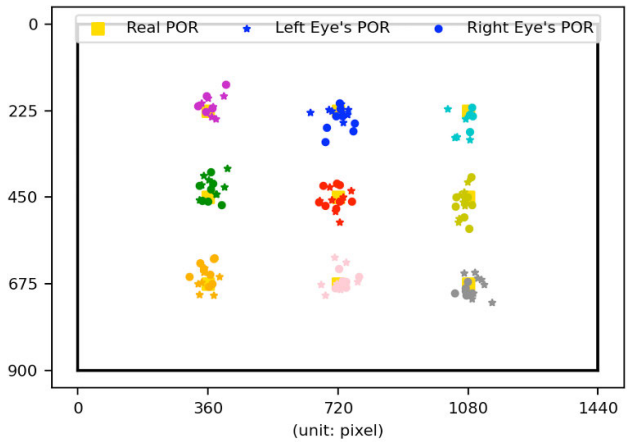


FIGURE 13. Distribution of subjects' PORs.

The POR deviation on the screen was represented by the Euclidean metric between the estimated PORs and the ground-truth of preset points. Assume that the distance between the subject and the screen is approximately represented by the distance from the 3D corneal center to the screen $\mathbf{D}$, then the deviation angle used to indicate the gaze accuracy is

$$\delta = \arctan \frac{\sqrt{\frac{1}{N}\sum_{k=1}^{N}(\mathbf{S_{new,k}}) - \mathbf{S_{c,k}})^2}}{\mathbf{D}}. \tag{19}$$

The gaze accuracy of each subject's left and right eyes is shown in Fig. 14, where the number on the bar is the average distance from the subject to the screen. Among these subjects, the results with the largest and smallest errors are listed in Table 2. The average gaze accuracy of a single eye was within 1.6°, and it can reach 1.33° if the more accurate one was selected from the left and right eyes. The minimum deviation angle was 0.87°, and the maximum is 2.09°. It is within an acceptable margin of error as the accuracy of gaze tracking in X and Y directions is usually 0.5°-2°.

The main reasons for the large deviation are as follows: 1) Inaccurate pupil fitting: the intersection of $\mathbf{cv_j}$ and the pupil imaging ellipse was used to determine $\mathbf{k_j}$ in this paper, which makes for a high demanding pupil fitting. As mentioned in Section IV-A-2), the pupil imaging ellipse error was the larger error source of the proposed method. 2) Rolling movement of the eyeball: the rolling of the eyeball around the OA changes the relative relationship between the OA and the VA, and cannot be characterized only by the eye features. At this time, the learned linear polynomial cannot accurately compensate for the POR deviation.

TABLE 2. Estimated PORs with largest and smallest errors.

Points	Real PORs/pixel	Estimated PORs with largest error/pixel	Estimated PORs with smallest error/pixel
1	(360,225)	(382.08,246.24)	(370.96,218.76)
2	(720,225)	(642.17,230.01)	(713.69,238.58)
3	(1080,225)	(1024.06,220.48)	(1091.52,215.37)
4	(360,450)	(341.05,416.63)	(344.05,457.56)
5	(720,450)	(756.62,433.10)	(758.60,459.76)
6	(1080,450)	(1055.47,504.20)	(1046.41,449.63)
7	(360,675)	(390.81,656.10)	(347.27,637.80)
8	(720,675)	(742.48,619.41)	(732.93,667.83)
9	(1080,675)	(1145.26,723.82)	(1074.55,704.70)

(2) Performance Quantification of Binocular Strategy: The VA of the left and right eyes should converge at one point. Therefore, we analyzed the performance of binocular strategy on the basis of monocular gaze estimation. Two methods using binocular information were evaluated. One was to calculate the average POR of the left and right eyes; the other was to learn a joint linear polynomial in the polynomial compensation, expressed by $\mathcal{H}'(\mathbf{S^L_{oa,k}}, \mathbf{S^R_{oa,k}}) = (\mathbf{w^L})^{\mathrm{T}}\mathbf{S^L_{oa,k}} + (\mathbf{w^R})^{\mathrm{T}}\mathbf{S^R_{oa,k}} + \mathbf{b}$, then the joint POR was calculated using the joint polynomial. The accuracy of the average POR and the joint POR were calculated using Eq. (19), the results are shown in Fig. 15. The average gaze accuracy of the former was 1.17°, while that of the latter was 1.01°. Compared with monocular gaze estimation, the accuracy was improved by 12% by calculating the average POR of the left and right eyes, while the accuracy was improved by 24% by calculating the joint POR. It can be seen that if the PORs of both eyes can be calculated, the binocular strategy can effectively improve the gaze estimation performance.

4) COMPARISON WITH STATE-OF-THE-ART METHODS

In addition to the CR-based method, there are many gaze estimation methods such as appearance-based methods and 3D model-based methods. We made a comparison of several state-of-the-art methods to further assess the proposed method in this section. The respective prerequisites including the system configuration (numbers of the cameras and the lights), gaze tracking technique, allowable head movement, and reported gaze accuracy of these methods are listed in Table 3.

Due to the diversity of samples, the gaze accuracy is usually low when the gaze estimation model trained by the appearance-based method is validated on cross-datasets or cross-subjects. Iris-comer technique-based methods

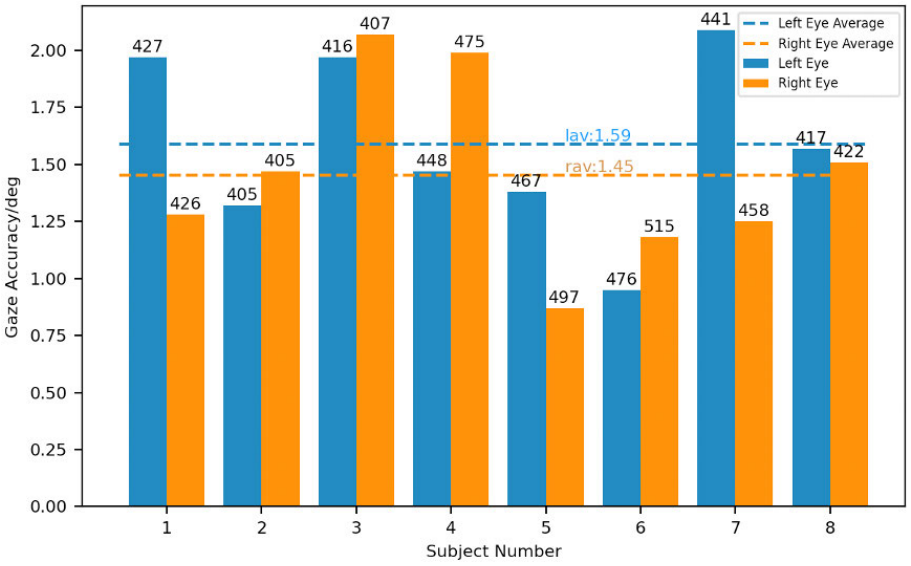


FIGURE 14. Gaze accuracy of different subjects.

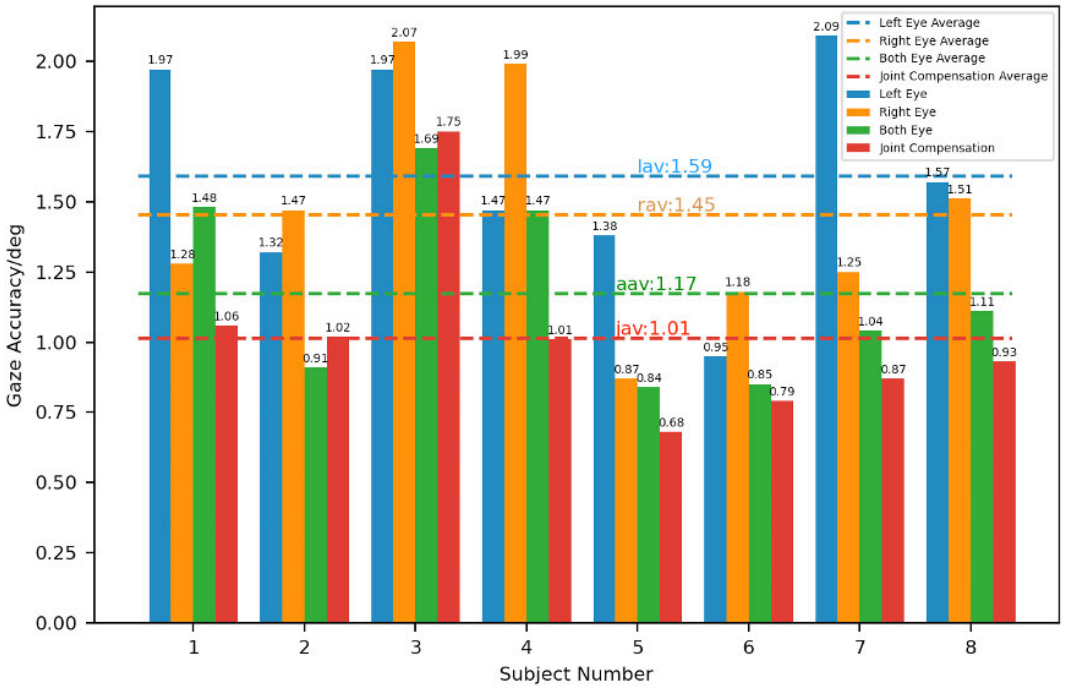


FIGURE 15. Comparison between monocular gaze estimation and binocular gaze estimation.

and pupil-corneal reflection-technique-based methods can achieve high accuracy when the head is fixed, but the accuracy would decrease significantly when the head deviates from the calibration position. Although facial-feature-based methods and depth-sensor-based methods can obtain the depth information using Kinect, there are many parameters that need to be calibrated, and the accuracy is low in a system with simple configuration. Both CR-based methods and HN-based methods utilize at least four light sources and require all of them to be imaged in the camera, which limits the range of head movement to a certain extent. The proposed method uses the 3D model-based method to estimate the 3D corneal center under the configuration of two light sources, and establishes the pupil reference plane that can ensure the coplanarity, for subsequent gaze estimation based on CR invariance. It can provide competitive gaze accuracy compared to the state-of-the-art method using a more complicated system.

TABLE 3. Comparison with state-of-the-art methods.

Methods	System Configuration		Gaze tracking techniques	Head movement	Reported accuracy/deg
	Camera(s)	Light(s)			
Lu et al. [29]	1	0	AP-based	Slight	2.37±1.42
Sugano et al. [30]	1	0	AP-based	✓	2.9
Yu et al. [31]	1	0	ICT-based	×	X:0.99,Y:1.33
Blignaut [32]	1	1	PCRT-based	×	0.87
Cheng et al. [11]	1	5	CR-based	✓	0.7
Arar et al. [15]	1	5	CR-based	✓	1
Ma et al. [18]	1	4	HN-based	✓	1.23±0.54
Li and Li [33]	1(Kinect)	0	FF-based	✓	H:4.4,V: 5.9
Zhou et al. [34]	1(Kinect)	0	DS-based	✓	1.99
Lai et al. [35]	2	2	CRPR-based	✓	1.18
Wang and Ji [36]	1	4	PCRT-based+3D model-based	✓	1.3
Our	1	2	3D model-based+CR-based	✓	1.33

AP = Appearance, ICT = Iris-corner technique, PCRT = Pupil-corneal reflection technique, FF = Facial-feature, DS = Depth-sensor, CRPR = Corneal-reflection-pupil-refraction.

V. CONCLUSION

This paper proposes an improved CR-based gaze estimation method using a SCTLS system. The 3D corneal center and the normal vector of the virtual pupil are first estimated using the 3D model-based method, then four pupil reference planes are determined according to the relationship between the pupil features and the projections of screen corner points on the image plane. Based on CR invariance, the screen point corresponding to the intersection of each reference plane and the line connecting the camera optical center and the pupil center is calculated respectively. The POA is determined using weighted average, and then the POR is compensated from the POA using the learned linear polynomial. On the basis of simplifying the system configuration of CR-based method, the proposed method avoids the bias introduced by the non-coplanarity of 3D pupil center and corneal reflection plane, compensates the kappa angle, and the gaze estimation performance is improved. We validated our method using computer simulations and practical experiments, and found that the proposed method is more susceptible to the pupil imaging ellipse error compared with the screen corner point error, which puts forward higher requirements for the pupil detection. In addition, the effect of polynomial compensation will become worse if the subject rolls the eyeball during fixation. Nevertheless, the average gaze accuracy of a single eye reached 1.33°, which can provide competitive accuracy compared with some state-of-the-art methods using more complex systems while allowing head movement. Binocular strategy has also been proved to be superior to the monocular method. Compared with monocular gaze estimation, the accuracy was improved by 12% by calculating the average POR of the left and right eyes, while the accuracy was improved by 24% by calculating the joint POR.

In a remote gaze tracking system, the pupil detection error will be amplified. Therefore, expanding upon our research, we are now focusing on a more accurate detection method for pupil to improve the stability of the proposed method. In addition, although the proposed method estimates the 3D corneal center based on the 3D model, which is better than the 2D mapping-based methods (including CR-based method) to deal with the impact of head movement, the accurate representation of the eyeball rolling around the OA is still the key problem to achieve full free head movement. It is urgent to make further in-depth research combined with facial features.

REFERENCES

[1] P. Drakopoulos, G.-A. Koulieris, and K. Mania, "Eye tracking interaction on unmodified mobile VR headsets using the selfie camera," *ACM Trans. Appl. Perception*, vol. 18, no. 3, pp. 1–20, Jul. 2021.

[2] K. Qian, T. Arichi, A. Price, S. Dall'Orso, J. Eden, Y. Noh, K. Rhode, E. Burdet, M. Neil, A. D. Edwards, and J. V. Hajnal, "An eye tracking based virtual reality system for use inside magnetic resonance imaging systems," *Sci. Rep.*, vol. 11, no. 1, p. 16301, Aug. 2021.

[3] Z. Hu, C. Lv, P. Hang, C. Huang, and Y. Xing, "Data-driven estimation of driver attention using calibration-free eye gaze and scene features," *IEEE Trans. Ind. Electron.*, vol. 69, no. 2, pp. 1800–1808, Feb. 2022.

[4] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, "MPIIGaze: Real-world dataset and deep appearance-based gaze estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 1, pp. 162–175, Nov. 2017.

[5] C. Mestre, J. Gautier, and J. Pujol, "Robust eye tracking based on multiple corneal reflections for clinical applications," *J. Biomed. Opt.*, vol. 23, no. 3, 2018, Art. no. 035001.

[6] J. J. Kang, M. Eizenman, E. D. Guestrin, and E. Eizenman, "Investigation of the cross-ratios method for point-of-gaze estimation," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 9, pp. 2293–2302, Sep. 2008.

[7] D. Yoo and M. Chung, "A novel non-intrusive eye gaze estimation using cross-ratio under large head motion," *Comput. Vis. Image Understand.*, vol. 98, no. 1, pp. 25–51, Apr. 2005.

[8] F. L. Coutinho and C. H. Morimoto, "Free head motion eye gaze tracking using a single camera and multiple light sources," in *Proc. 19th Brazilian Symp. Comput. Graph. Image Process.*, Manaus, Brazil, Oct. 2006, pp. 171–178.

[9] F. L. Coutinho and C. H. Morimoto, "Improving head movement tolerance of cross-ratio based eye trackers," *Int. J. Comput. Vis.*, vol. 101, no. 3, pp. 459–481, Feb. 2013.

[10] D. W. Hansen, J. S. Agustin, and A. Villanueva, "Homography normalization for robust gaze estimation in uncalibrated setups," in *Proc. Symp. Eye-Tracking Res. Appl.* New York, NY, USA: ACM, 2010, pp. 13–20.

[11] H. Cheng, Y. Liu, W. Fu, Y. Ji, L. Yang, Y. Zhao, and J. Yang, "Gazing point dependent eye gaze estimation," *Pattern Recognit.*, vol. 71, pp. 36–44, Nov. 2017.

[12] M. Sasaki, T. Nagamatsu, and K. Takemura, "Cross-ratio based gaze estimation using polarization camera system," in *Proc. ACM Int. Conf. Interact. Surf. Spaces*. New York, NY, USA: ACM, Nov. 2018, pp. 333–338.

[13] D. H. Yoo, J. H. Kim, B. R. Lee, and M. J. Chung, "Non-contact eye gaze tracking system by mapping of corneal reflections," in *Proc. 5th IEEE Int. Conf. Autom. Face Gesture Recognit.* Washington, DC, USA, May 2002, pp. 101–106.

[14] N. M. Arar, H. Gao, and J.-P. Thiran, "Towards convenient calibration for cross-ratio based gaze estimation," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.* Waikoloa, HI, USA, Jan. 2015, pp. 642–648.
[15] N. M. Arar, H. Gao, and J.-P. Thiran, "A regression-based user calibration framework for real-time gaze estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 12, pp. 2623–2638, Dec. 2017.
[16] J. J. Kang, E. D. Guestrin, W. J. MacLean, and M. Eizenman, "Simplifying the cross-ratios method of point-of-gaze estimation," in *Proc. 30th Can. Med. Biol. Eng. Conf. (CMBEC)*. Toronto, ON, Canada: Oxford Univ. Press, 2007, pp. 14–17.
[17] K. A. Choi, C. Ma, and S. J. Ko, "Improving the usability of remote eye gaze tracking for human-device interaction," *IEEE Trans. Consum. Electron.*, vol. 60, no. 3, pp. 493–498, Aug. 2014.
[18] 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, May 2014, Art. no. 053112.
[19] J.-B. Huang, Q. Cai, Z. Liu, N. Ahuja, and Z. Zhang, "Towards accurate and robust cross-ratio based gaze trackers through learning from simulation," in *Proc. Symp. Eye Tracking Res. Appl.* Safety Harbor, FL, USA: ACM, 2014, pp. 75–82.
[20] Z. Zhang and Q. Cai, "Improving cross-ratio-based eye tracking techniques by leveraging the binocular fixation constraint," in *Proc. Symp. Eye Tracking Res. Appl.* Safety Harbor, FL, USA: ACM, Mar. 2014, pp. 267–270.
[21] C. H. Morimoto, F. L. Coutinho, and D. W. Hansen, "Screen-light decomposition framework for point-of-gaze estimation using a single uncalibrated camera and multiple light sources," *J. Math. Imag. Vis.*, vol. 62, no. 4, pp. 585–605, May 2020.
[22] A. Villanueva and R. Cabeza, "A novel gaze estimation system with one calibration point," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 38, no. 4, pp. 1123–1138, Aug. 2008.
[23] J. Liu, J. Chi, N. Lu, Z. Yang, and Z. Wang, "Iris feature-based 3-D gaze estimation method using a one-camera-one-light-source system," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 7, pp. 4940–4954, Jul. 2020.
[24] A. Villanueva and R. Cabeza, "Evaluation of corneal refraction in a model of a gaze tracking system," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 12, pp. 2812–2822, Dec. 2008.
[25] 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.
[26] M. Rezaei and R. Klette, "Novel adaptive eye detection and tracking for challenging lighting conditions," in *Proc. Asian Conf. Comput. Vis.* Berlin, Germany: Springer, 2012, pp. 427–440.
[27] H. Du, X. Chen, and J. Xi, "An improved background segmentation algorithm for fringe projection profilometry based on Otsu method," *Opt. Commun.*, vol. 453, Dec. 2019, Art. no. 124206.
[28] X. Xu, Z. Fei, Z. Tan, B. Zhao, and J. He, "Improved calibration method based on the RANSAC approach and an improved gray centroid method for a laser-line-based structured light system," *Appl. Opt.*, vol. 58, no. 35, pp. 9603–9613, 2019.
[29] 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.
[30] Y. Sugano, Y. Matsushita, Y. Sato, and H. Koike, "Appearance-based gaze estimation with online calibration from mouse operations," *IEEE Trans. Human-Mach. Syst.*, vol. 45, no. 6, pp. 750–760, Dec. 2015.
[31] M. Yu, Y. Lin, X. Tang, J. Xu, 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, Oct. 2015.
[32] 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, 2014.
[33] 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.
[34] 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. (ICRA)*, Singapore, May 2017, pp. 1646–1653.
[35] 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.
[36] K. Wang and Q. Ji, "3D gaze estimation without explicit personal calibration," *Pattern Recognit.*, vol. 79, pp. 216–227, Jul. 2018.

JIAHUI LIU (Member, IEEE) received the Ph.D. degree in control science and engineering from the University of Science and Technology Beijing, Beijing, China, in 2021, where she is currently pursuing the Ph.D. degree with the School of Automation and Electrical Engineering. Her research interest includes theoretical algorithms and the practical applications of the gaze tracking systems.

JIE WU received the B.S. degree in measurement and control technology and instruments from the University of Science and Technology Beijing, China, in 2020, where he is currently pursuing the M.S. degree in instrument science and technology. His main research interests include system calibration of the gaze tracking systems and experimental data collection.

HUIJIE YANG received the B.S. degree in measurement and control technology and instruments from the University of Science and Technology Beijing, China, in 2020, where he is currently pursuing the M.S. degree in instrumentation engineering. His research interests include appearance-based gaze estimation and 2D mapping-based gaze estimation.

JIANNAN CHI received the B.S. degree in optic instrument engineering from Tianjin University, Tianjin, China, in 1990, and the M.S. degree in detection technology and automatic equipment and the Ph.D. degree in pattern recognition and intelligent system from Northeastern University, in 2002 and 2005, respectively. He was a Postdoctoral Researcher at the University of Science and Technology Beijing, China, from 2005 to 2006. He was also a Visiting Scholar at Nanyang Technological University, Singapore, in 2011. He is currently a Professor with the Department of Instrument Science and Technology, University of Science and Technology Beijing. His current research interests include computer vision, human–computer interaction, and optical measurement.