3-D Gaze-Estimation Method Using a Multi-Camera-Multi-Light-Source System

Jiannan Chi, Jiahui Liu, *Member, IEEE* , Feng Wang, Yingkai Chi, and Zeng-Guang Hou, *Fellow, IEEE*

***Abstract*— The image of the center of a 3-D circular target is not the center of its imaging ellipse due to the imaging distortion of the spatial circular targets; therefore, traditional 3-D gaze-estimation methods with the multi-camera-multi-light-source (MCMLS) systems generally replace the projection of the pupil center with the virtual image of the pupil center for 3-D gaze estimation. However, this introduces a large gaze-estimation error when the oblique angle between the optical axis of the eye and the camera optical axis is large. To eliminate the error caused by using the virtual image of the pupil center, a 3-D gaze-estimation method using an MCMLS system is developed in this study. We first estimate the cornea center and then determine the matching points of the pupil imaging ellipse using the special polar plane. After the cornea radius and the kappa angle are calibrated, the optical axis of the eye is reconstructed by the refraction planes constructed by the edge points of the pupil imaging ellipses. Thus, the 3-D gaze is estimated by the real-time calculated transformation matrix represented by the kappa angle. The feasibility and performance of the method have been analyzed by the simulations and the experiments. Since the estimation of the spatial pupil center or the construction of the refraction plane using the virtual image of the pupil center is not required, the proposed method mitigates the inherent errors caused by optical axis reconstruction in the traditional methods and simplifies the algorithm for gaze estimation, which has a practical value.**

***Index Terms*— 3-D gaze estimation, corneal reflection, multi-camera-multi-light-source (MCMLS) system, optical axis reconstruction, pupil refraction, user calibration.**

I. Introduction

GAZE tracking is a type of visual-based measurement (VBM) based on the visual-based instruments [1]–[3]. A gaze-tracking system may include one or more cameras and one or more light sources. The basic principle of gaze tracking is to capture a facial image through the system camera, extract the facial feature (e.g., eye) parameters from the image, and then measure the user's attention or the object of interest based on these features [4].

Most vision-based gaze-tracking methods fall into one of the two categories [5]: 2-D mapping-based gaze estimation or 3-D model-based gaze estimation. The 2-D gaze-tracking system uses the visual features of the human eye, such as pupil, iris, eye corners, or glints, to establish a 2-D mapping model between the features that change with eye movement and the 2-D point of regard (POR). It normally consists of a single camera and one or more light sources. The common 2-D mapping-based gaze-estimation methods include the pupil-corner technique (PCT)-based methods [6], pupil-cornea-reflection technique (PCRT)-based methods [7]–[15], cross-ratio (CR)-based methods [16]–[19], and homography normalization (HN)-based methods [20], [21].

By comparison, the 3-D gaze-tracking system estimates the real 3-D line-of-sight (LOS) through the 3-D eyeball structure and the eye imaging models. In addition to using a depth sensor such as Kinect to obtain the depth information to achieve 3-D gaze estimation in a one-camera-non-light-source (OCNLS) system [22], [23], most 3-D gaze-estimation methods use a multi-camera-multi-light-source (MCMLS) system. They usually provide the estimates of the 3-D gaze based on pupil refraction and corneal reflection. The gaze-estimation procedure first estimates the 3-D cornea center to reconstruct the optical axis (OA) and then estimates the LOS by the OA and the angle between the OA and the LOS. The intersection of the LOS and the computer screen, which is calculated by system calibration, is the POR [24], [25]. There are two main gaze-estimation methods in the MCMLS systems: 1) the OA of the eye is directly determined by the intersection of two or more refraction planes involving the cornea center, the nodal point of the camera, the pupil center, and the virtual image of the pupil center [26], [27] or 2) the OA of the eye is reconstructed by the estimated cornea center and the pupil center [24], [28]–[32], wherein the cornea center is determined by two or more corneal reflections of different cameras, and the pupil center is estimated by the image of the pupil in two or more camera coordinate systems. The direction of the LOS can, thus, be estimated from the OA once the user-dependent kappa angle has been estimated through a simple calibration procedure. Therefore, the LOS can be determined accordingly with the estimated cornea center.

In the above methods, the refraction plane is inevitably used. The law of refraction states that the incident ray, refracted ray,

Manuscript received January 29, 2020; accepted June 7, 2020. Date of publication July 6, 2020; date of current version November 10, 2020. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFC2001700, in part by the Opening Project of Key Laboratory of Operation Safety Technology on Transport Vehicle, Ministry of Transport, China, in part by the Beijing Municipal Natural Science Foundation under Grant 4172040, and in part by the Science and Technology on Electrooptic Control Laboratory under Grant 301090704. The Associate Editor coordinating the review process was Boby George. *(Corresponding author: Zeng-Guang Hou.)*

Jiannan Chi, Jiahui Liu, and Feng Wang are with the School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China (e-mail: ustbjnc@ustb.edu.cn; ljh@xs.ustb.edu.cn; 1441396590@qq.com).

Yingkai Chi is with the Department of Pathology, General Hospital of Northern Theater Command, Shenyang 110015, China (e-mail: m13898166208@163.com).

Zeng-Guang Hou is with the State Key Laboratory of Complex System Management and Control, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: zengguang.hou@ia.ac.cn).

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

Digital Object Identifier 10.1109/TIM.2020.3006681

and normal at the point of refraction are coplanar [26]; thus, the plane is represented as the coplanarity of the pupil center, the cornea center, the camera optical center, and the virtual image of the pupil center directly by many researchers [25], [26], [29], [32]. However, the image of the center of a 3-D circular target is not the center of its imaging ellipse due to the imaging distortion of the spatial circular targets. There is a deviation between the projection of the 3-D pupil center on the camera image plane and the virtual image of the pupil center.

Shih and Liu [28] demonstrated that the pupil that is observed from a camera is the virtual image of the pupil, not the pupil itself, due to the refraction of the convex lens and listed the expressions of the center of the virtual pupil and the center of the real pupil, respectively. Wei and Zhang [33] established an elliptic center distortion error model in the perspective projection transform and stated that the center calculated by the elliptic curve on the image plane is equal to the real image point of the center of the spatial ellipse only when the plane of the spatial ellipse is parallel to the image plane. It is also suitable for the spatial pupil. The deviation between the center calculated by the pupil imaging ellipse and the real image point of the spatial pupil center decreases as the distance between the pupil and the camera increases, and increases as the angle between the real pupil plane and the camera image plane increases.

Therefore, when the center of the virtual pupil is used to estimate the spatial pupil center, there is the light refraction error inherent to the commonly used OA-reconstruction method in the MCMLS systems. When there is substantial deviation in the pupil orientation from the camera optical axis, there is a large error in the OA reconstruction.

To overcome the inherent errors caused by reconstructing the OA of the eye in the traditional gaze-estimation methods, this article proposes a 3-D gaze-estimation method using an MCMLS system, in which the reconstruction of the OA does not require the estimation of the spatial pupil center or the construction of refraction plane using the virtual image of the pupil center. We first estimate the cornea center according to the geometric relationship of the light sources reflecting on the corneal outer surface and then determine the matching points of the edge of the pupil imaging ellipse. After the cornea radius and the kappa angle of the eye are calibrated during user calibration, the OA of the eye was reconstructed by the refraction planes constructed by the edge points of the pupil imaging ellipses of two cameras. Three OA-reconstruction methods are proposed: the conical axis (CA) method, the elliptical major axis (EMA) method, and the major axis intersection (MAI) method. Once the OA is obtained, the 3-D LOS is expressed by the estimated cornea center and the real-time calculated transformation matrix of the kappa angle.

The primary contributions of this work are as follows.

1) The OA in our method is reconstructed by the edge points of the pupil imaging ellipse, instead of using the virtual image of the pupil center. Thus, the OA-reconstruction error in the traditional 3-D gaze-estimation method [25], [26], [29], [32] can be eliminated.

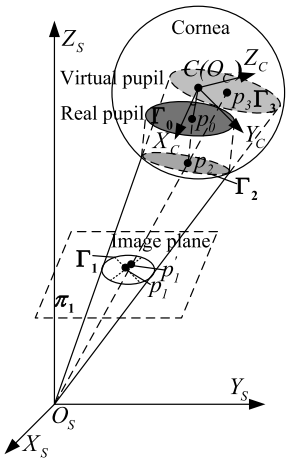


Fig. 1. Pupil imaging model.

2) The cornea radius calibrated in our method is used alone to determine the relative positional relationship between the two glints, so the cornea radius can be loosely calibrated. This reduces the accuracy requirement of the cornea radius calibration.
3) From the perspective of the algorithm, the proposed method is simpler and more operable than the traditional MCMLS method, which gives it a practical application value.

The remainder of this article is organized as follows. Section II provides the problem statement regarding the existing methods based on the MCMLS systems, i.e., the problem to be solved in this study. Section III provides an overview of the proposed method and discusses the proposed gaze-estimation method in detail. Simulations and experiments are discussed in Section IV, and Section V concludes this article.

II. Problem Statement

The image of the center of the spatial pupil is not the center of its imaging ellipse due to imaging distortion. As shown in Fig. 1, O_s-$X_sY_sZ_s$ is the camera coordinate system, in which the origin is the camera optical center O_s and Z_s-axis is the camera optical axis. The camera image plane is π_1. O_c-$X_cY_cZ_c$ is the eyeball coordinate system, in which the origin is the cornea center C. The X_c-axis is directed from the cornea center C to the camera optical center O_s, and the Y_c-axis is determined to keep the OA of the eye in the $X_cO_cY_c$ plane.

Γ_0 is the real pupil and p_0 is the center of the real pupil. The ray that comes from each edge point of the real pupil refracts at a point on the corneal surface, and the refracted ray passes through the camera optical center and intersects with the camera image plane. The refraction points at the cornea–air interface caused by the refraction of the real pupil edge are still coplanar [28], so Γ_2 is the virtual image of the pupil at the cornea–air interface and its center is p_2. Assume that the angle between the normal vector of the real pupil and the X_c-axis of the eyeball coordinate system is φ,

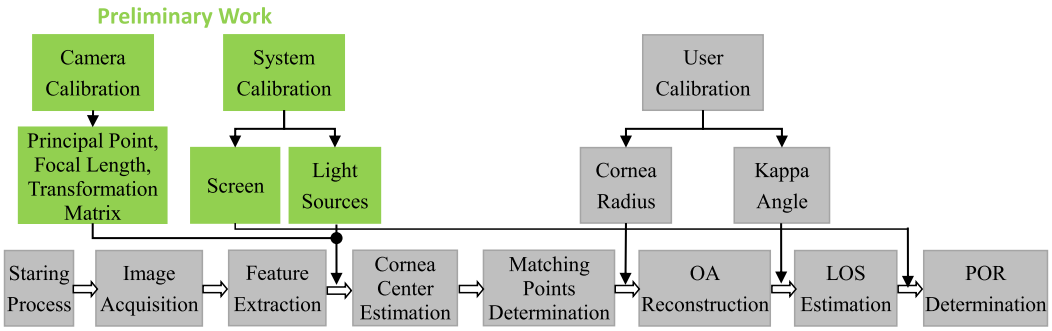


Fig. 2. Procedure of the proposed method.

the pupil radius is r_p, the cornea radius is r_c, and the distance between the cornea center C and the pupil center p_0 is d. Shih and Liu [28] demonstrated that the center of the real pupil can be expressed as $p_0 = (d\cos\varphi, d\sin\varphi, 0)$; the center of the virtual pupil can be determined as $p_2 = (r_c - (n'(r_c - d\cos\varphi))/(n + ((n-n')(r_c - d\cos\varphi))/(r_c)), (dn\sin\varphi))/(n + ((n-n')(r_c - d\cos\varphi))/(r_c)), 0)$, where n is the refraction index of the cornea and n' is the refraction index of the air. The ray that comes from p_0 refracts at p_2 on the corneal surface, and the refracted ray passes through O_s and intersects with π_1 at p_1'. Therefore, p_1' is not only the image point of p_0 but also the image point of p_2.

Γ_1 is the imaging ellipse of Γ_0 and p_1 is the center of the pupil imaging ellipse. In the camera coordinate system O_s-$X_sY_sZ_s$, an oblique elliptic cone is formed by O_s and Γ_1. On the basis of pupil refraction, Γ_2 is a section plane of the oblique elliptic cone. Γ_3 is also a section plane of the oblique elliptic cone. It is the virtual pupil with the same radius as the real pupil and its center is p_3. Therefore, Γ_2 and Γ_3 have the same image in the camera, namely, Γ_1. Since the normal vectors of Γ_2 or Γ_3 can be estimated by the eigenvalues and the corresponding normalized eigenvectors of the real symmetric matrix obtained from the coefficients of the same oblique elliptic cone equation [34], the same normal vectors of Γ_2 and Γ_3 indicate that the planes contained Γ_2 and Γ_3 are parallel. The pupil radius is r_p; thus, the radius of Γ_3 is r_p, and its center can be expressed as

$$p_3 = r_p \begin{pmatrix} \pm e_{1x}\sqrt{\dfrac{|\lambda_3|(|\lambda_1|-|\lambda_2|)}{|\lambda_1|(|\lambda_1|+|\lambda_3|)}} + e_{3x}\sqrt{\dfrac{|\lambda_1|(|\lambda_2|+|\lambda_3|)}{|\lambda_3|(|\lambda_1|+|\lambda_3|)}} \\ \pm e_{1y}\sqrt{\dfrac{|\lambda_3|(|\lambda_1|-|\lambda_2|)}{|\lambda_1|(|\lambda_1|+|\lambda_3|)}} + e_{3y}\sqrt{\dfrac{|\lambda_1|(|\lambda_2|+|\lambda_3|)}{|\lambda_3|(|\lambda_1|+|\lambda_3|)}} \\ \pm e_{1z}\sqrt{\dfrac{|\lambda_3|(|\lambda_1|-|\lambda_2|)}{|\lambda_1|(|\lambda_1|+|\lambda_3|)}} + e_{3z}\sqrt{\dfrac{|\lambda_1|(|\lambda_2|+|\lambda_3|)}{|\lambda_3|(|\lambda_1|+|\lambda_3|)}} \end{pmatrix}. \tag{1}$$

Since the center of Γ_3 is proportional to its radius, it can be expressed as $p_3 = r_p(p_{3x}, p_{3y}, p_{3z})$, and the linear equation connecting O_s and p_3 is

$$\frac{x}{p_{3x}} = \frac{y}{p_{3y}} = \frac{z}{p_{3z}}. \tag{2}$$

Similarly, p_2 can be obtained with the radius of Γ_2. It indicates that p_2 is on the line $O_s p_3$. Therefore, p_2 and p_3 have the same image point in the camera, which is p_1'.

Since the projection of p_3 can be estimated by the intersection of the line $O_s p_3$ and the camera image plane $z = -f$, it can be expressed as $p_1' = ((q_1l_1 - m_1p_1)/(k_1p_1 - n_1l_1), (m_1n_1 - k_1q_1)/(k_1p_1 - n_1l_1), -f)$ [33]. p_1' is also the image point of p_0; thus, the projection of the center of the real pupil is obtained. By contrast, Γ_1 is the image of Γ_3; thus, the center of Γ_1 on the image plane can be expressed according to the elliptic center distortion error model in the perspective projection transform [33], that is, the equation can be derived, as shown at the bottom of next page. p_1 and p_1' are equal only when the plane of Γ_0 is parallel to the camera image plane π_1.

Many researchers directly used the virtual image of the pupil center to construct the refraction plane or estimate the spatial pupil center, so as to reconstruct the OA of the eye [26], [27], [29], [32]. Therefore, there are important errors inherent to the commonly used OA-reconstruction method in the MCMLS systems. When there is substantial deviation in the pupil orientation from the camera optical axis, there is a large error in the OA reconstruction.

III. Proposed Method

In this article, a 3-D gaze-estimation method using an MCMLS system is proposed, wherein the OA is reconstructed by three proposed methods to overcome the deficiencies of OA reconstruction in the traditional 3-D gaze-estimation methods that the virtual image of the pupil center is used. A flowchart of the proposed method is given in Fig. 2.

The preliminary work includes camera calibration and system calibration. Camera calibration serves to determine the principle point and the focal length of each camera and to calculate the relative position relationship of the two system cameras, which is represented by a transformation matrix. System calibration is performed to obtain the positions of the light sources and the screen in the camera coordinate system [35]. User calibration was performed in this study to calibrate the cornea radius and the kappa angle, which have individual differences among users. During 3-D gaze estimation, the users stared at the points on the screen and the eye images were captured by the system cameras. After feature extraction, the cornea center is estimated, and the matching points of the edge of the pupil imaging ellipse in each camera are determined. Then, the OA is reconstructed by the estimated

Fig. 3. Helmet-mounted gaze-tracking system. (a) Helmet-mounted gaze tracker. (b) Eye detection component.

cornea center and the matching points of the pupil edges of the two cameras. The LOS is estimated by the reconstructed OA and the calibrated kappa angle. Therefore, the POR is determined by the LOS and the screen equation. The details of each part are explained in the following sections.

A. System Setup

头戴式

The system used in this study is a helmet-mounted gaze-tracking system designed by the authors, as shown in Fig. 3(a). Its eye-detection component is a two-camera-multi-light-source system that is connected to the computer through a USB port [Fig. 3(b)]. The two system cameras are the CMOS camera sensors (OV7251). The lens focal length of the system camera is 2.8 mm, the resolution of the camera is 640×480, and the pixel size is 3 μm. The relationship between the two system cameras was calibrated by the coordinate transformation matrix in advance. The light sources are composed of 0.8-mm-height flat top infrared LEDs (EVERLIGHT, HIR19-21C/L11/TR8) evenly distributed on the square box. The peak wavelength of the light sources is 850 nm, and the brightness of the light source is adjustable. The proposed gaze-estimation method only needs two light sources, but in order to obtain more experimental data and improve the gaze estimation accuracy, five light sources were used in our experiments, as shown in Fig. 3(b). System calibration was implemented previously to determine the positions of the light sources and the screen plane in the coordinate systems of the two system cameras.

B. Feature Extraction

Eye images are captured by the two system cameras as the user gazes at any on-screen point. The images are processed in order to extract the eye feature parameters for gaze tracking, as shown in Fig. 4.

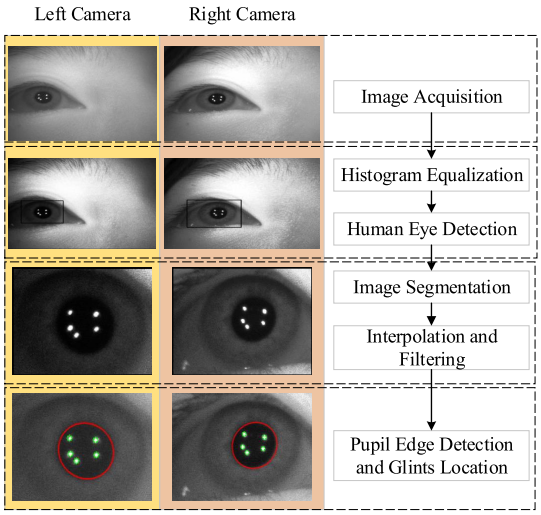


Fig. 4. Procedure of image processing.

Image processing is a four-step process: image acquisition, image preprocessing, image segmentation, and target detection.

1) Based on the gaze-tracking system, eye images are captured by the two system cameras synchronously from different directions. The two images captured simultaneously are synchronously transmitted to the host computer for subsequent image processing.
2) The eye features are highlighted and the local contrast enhanced followed by histogram to improve the overall image quality. The Haar model is used to detect the human eye and identify the ROI for subsequent image segmentation.
3) The pupil image is obtained by image segmentation of the ROI region. The segmented image is interpolated by cubic spline to ensure accurate image detection, and then, the bilateral filtering method is applied to filter the noise.
4) During target detection, the appropriate threshold is first set to identify the glints. The pixel coordinates of the glints in the image coordinate system are determined by the centroid method and sorted according to the position relationship of their pixel coordinates to ensure correspondence to the light sources in the system. Next, edge detection and ellipse fitting are performed on the pupil to obtain the feature parameters of the pupil.

$$p_1 = \left(\begin{array}{c} \dfrac{(l^2+p^2-s^2)(mk+nq-rt)-(kl+np-rs)(ml+pq-st)}{(kl+np-rs)^2-(k^2+n^2-r^2)(l^2+p^2-s^2)}, \\ \dfrac{(k^2+n^2-r^2)(ml+pq-st)-(kl+np-rs)(mk+nq-rt)}{(kl+np-rs)^2-(k^2+n^2-r^2)(l^2+p^2-s^2)}, \\ -f \end{array} \right)^T .$$

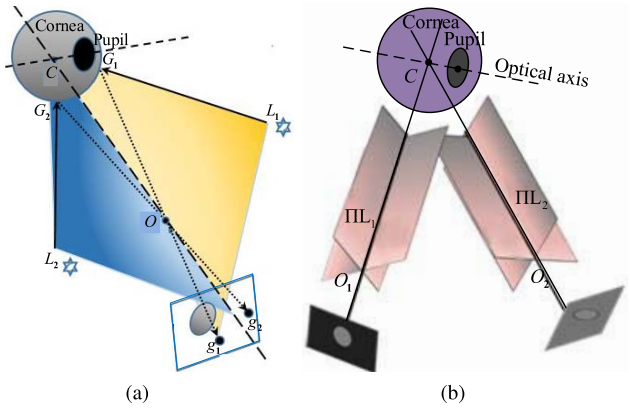


Fig. 5. Corneal reflection for cornea center estimation. (a) One-camera-two-light-source system. (b) Two-camera-two-light-source system.

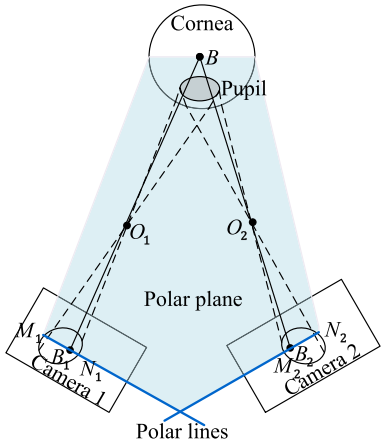


Fig. 6. Matching points on the pupil imaging ellipses of the two cameras.

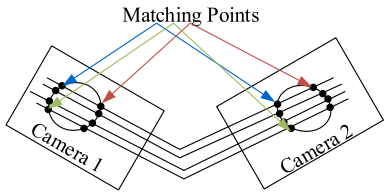


Fig. 7. Multiple pairs of matching points of two cameras.

C. Cornea Center Estimation

The cornea is regarded as a sphere with C as its center and r_c as its radius [26], [36]. As shown in Fig. 5(a), the ray that comes from the light source L_i reflects at a point G_i on the cornea surface so that the reflected ray passes through the camera optical center O and intersects the camera image plane at a point g_i. The law of reflection states that the incident ray, the reflected ray, and the normal at the point of reflection are coplanar [26]. Since the cornea center C is on the normal, the reflection plane satisfies 垂直

$$(L_i - O) \times (g_i - O) \cdot (C - O) = 0. \tag{3}$$

When the number of the light sources is not less than 2, two or more reflection planes can be obtained. The intersection line of these reflection planes is the line connecting the cornea center C and the camera optical center O. When the number of the cameras is 2, the line connecting the cornea center C and the camera optical center $O_j (j = 1, 2)$ can be obtained. Therefore, the cornea center C can be estimated by the intersection of the two lines [Fig. 5(b)].

D. Matching Point Determination

When two cameras are used to capture the images of the same object, image matching is necessary to determine the corresponding relationship between the pixels in their respective images. The pupil is imaged in the cameras after corneal refraction, and then, the pupil imaging ellipses of the two cameras are the images of the virtual pupils that are different captured in different camera positions. Therefore, the edge points of the pupil imaging ellipse cannot be directly used for image matching. In this study, we developed a matching method for determining the matching points of the pupil imaging ellipses of two cameras.

As shown in Fig. 6, suppose that B_1 is a point in the pupil imaging ellipse on the image plane of Camera 1, which is the projection point of a point B. The polar plane of B is then formed by the line B_1O_1 and the camera optical center O_2, and intersects with the two image planes to reveal the polar line. M_1, N_1 are the intersection points of the polar line and the pupil imaging ellipse on the image plane of Camera 1; M_2, N_2 are the intersection points of the polar line and the pupil imaging ellipse on the image plane of Camera 2. Thus, M_1, M_2 and N_1, N_2 are two pairs of matching points on the image planes of Camera 1 and Camera 2, respectively, which are the matching points of the pupil imaging ellipses on the two camera imaging planes.

The corresponding intersection points on the pupil imaging ellipses formed by the polar lines are a pair of matching points if any point on the image plane of Camera 1 is chosen and its polar plane is made. Therefore, multiple pairs of matching points on the pupil imaging ellipses can be obtained, as shown in Fig. 7.

E. OA Reconstruction

In this study, the edge points of the pupil imaging ellipses of the two cameras are used to reconstruct the OA of the eye. After obtaining the matching points on the pupil imaging ellipses of the two cameras, the OA of the eye is reconstructed based on multiple pairs of matching points using three different methods, namely, the CA method, the EMA method, and the MAI method.

1) CA Method: As shown in Fig. 8, the ray that comes from the pupil edge point $E_k (k = 1, 2, \ldots, n)$ refracts at one point on the corneal surface such that the refracted ray passes through the camera optical center O_j and intersects the camera image plane at a point E_{jk}. The law of refraction states that the incident ray, the refracted ray, and the normal at the point of refraction are coplanar [26]. Thus, the coplanarity of E_k, C, O_j, and E_{jk} is

$$\left(O_j - E_{jk}\right) \times \left(C - O_j\right) \cdot \left(E_k - C\right) = 0. \tag{4}$$

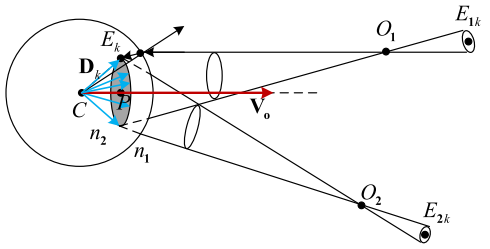


Fig. 8. Model of OA reconstruction.

coe1k coe2k

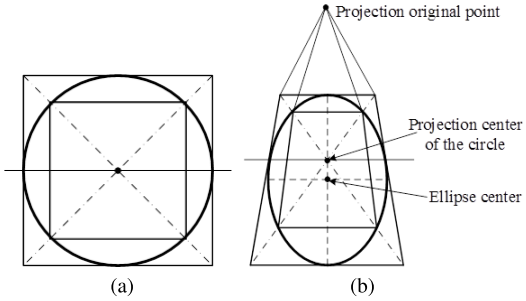


Fig. 9. Perspective construction of a circle. (a) Plane view of the circle. (b) Perspective view of the circle.

When two cameras are used, the line connecting the cornea center C and the pupil edge point E_k can be expressed as the intersection of the two refraction planes. The unit direction of the intersection line can be expressed as

$$\mathbf{D}_k = [(O_1 - E_{1k}) \times (C - O_1)] \times [(O_2 - E_{2k}) \times (C - O_2)]. \tag{5}$$

A series of $E_k(k = 1, 2, \ldots, n)$ determines a series of the intersection lines connecting C and E_k. Since these intersection lines pass through the cornea center, they construct a spatial cone with C as its vertex. The rotation axis of the spatial cone is the OA of the eye. Suppose that the number of the selected pupil edge point E_k is 2T pairs and E_k satisfies $\|E_{2T-1}E_{2T}\| = r_p, T = 1, 2, \ldots, m$, wherein r_p is the pupil radius, the OA direction can be estimated by

$$\mathbf{V}_o = \sum_{k=1}^{2T} \frac{\mathbf{D}_k}{\|\mathbf{D}_k\|}. \tag{6}$$

2) EMA Method: The projection of a 3-D circle is an ellipse, but the center of the image ellipse does not correspond to the center of the 3-D circle [33]. As shown in Fig. 9(a), there are construction lines in the plane view of the circle. The circle center is the intersection of the diagonals of two squares that have affine invariance. The center of the ellipse lies far from the projection of the real center of the circle, as shown in Fig. 9(b). The ellipse passes through the edge points where the ellipse intersects the circumscribed square. The major axis of the ellipse corresponds to the line connecting the points of tangency of the square with the circle. The real center of the projection of the 3-D circle lies on the major axis of the imaging ellipse. The major axis of the pupil imaging ellipse continues to correspond to the diameter of the spatial pupil even after pupil refraction and camera projection. According to this, we studied the major axis of the pupil imaging ellipse in each camera.

As shown in Fig. 10, suppose that the major axis of the pupil imaging ellipse in Camera 1 is $p_{11}p_{12}$ and the major axis of the pupil imaging ellipse in Camera 2 is $p_{23}p_{24}$. They are the projections of the pupil diameters P_1P_2 and P_3P_4 in Cameras 1 and 2, respectively. The matching points of p_{11}, p_{12} can be found in Camera 2, which are represented by p_{21}, p_{22} (Section III-D). Similarly, the matching points of p_{23}, p_{24} can be found in Camera 1, which are represented by p_{13}, p_{14}. According to the coplanarity of the incident ray, the refracted ray, and the normal at the refraction point, the unit direction of

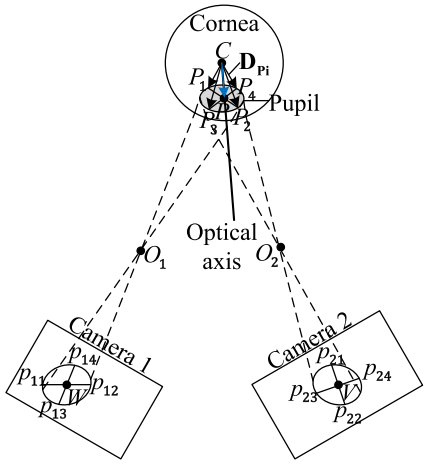


Fig. 10. Imaging model of the elliptical major axis.

the cornea center C and the pupil edge point $P_i(i = 1, 2, 3, 4)$ can be expressed as

$$\mathbf{D}_{Pi} = [(O_1 - p_{1i}) \times (C - O_1)] \times [(O_2 - p_{2i}) \times (C - O_2)]. \tag{7}$$

In the spatial cone with C as its vertex and the pupil as the bottom, CP_i, which is the generatrix of the cone, can be reconstructed. Therefore, the OA direction can be expressed as

$$\mathbf{V}_o = \sum_{k=1}^{4} \frac{\mathbf{D}_{Pi}}{\|\mathbf{D}_{Pi}\|}. \tag{8}$$

3) MAI Method: The matching points of p_{11}, p_{12} can be found as p_{21}, p_{22} in Camera 2, and the matching points of p_{23}, p_{24} can be found as p_{13}, p_{14} in Camera 1. As shown in Fig. 10, W is the intersection of the major axis $p_{11}p_{12}$ and the line $p_{13}p_{14}$, and V is the intersection of the major axis $p_{23}p_{24}$ and the line $p_{21}p_{22}$. W and V are a pair of matching points. W is the real projection point of the pupil center in Camera 1 and V is the real projection point of the pupil center in Camera 2. Therefore, W and V can be used to construct the refraction planes that contain the pupil center, the cornea center, the camera optical center, and the real image point of

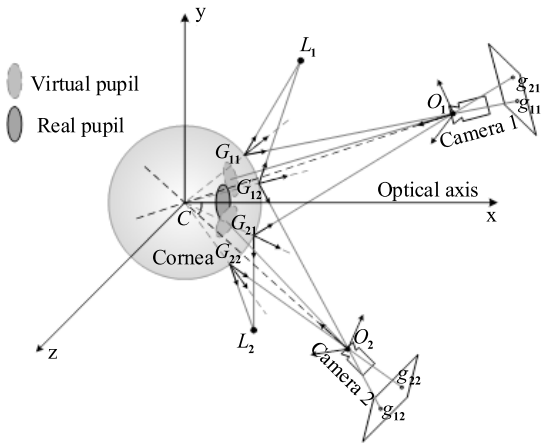


Fig. 11. Pupil imaging and corneal reflection of light sources in a two-camera-two-light-source system.

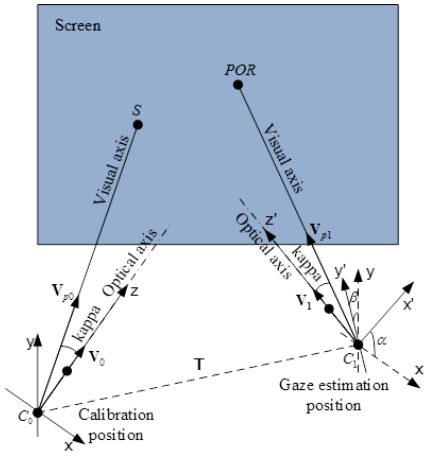


Fig. 12. Kappa angle relationship in two different positions.

the pupil center, that is

$$\begin{cases} (W - O_1) \times (C - O_1) \cdot (P - C) = 0 \\ (V - O_2) \times (C - O_2) \cdot (P - C) = 0. \end{cases} \tag{9}$$

Since the two refraction planes pass through the cornea center C and the pupil center P, the OA can then be reconstructed by the intersection of the refraction planes of the two cameras directly. The OA direction can be expressed as

$$\mathbf{V}_o = \frac{[(W - O_1) \times (C - O_1)] \times [(V - O_2) \times (C - O_2)]}{\|[(W - O_1) \times (C - O_1)] \times [(V - O_2) \times (C - O_2)]\|}. \tag{10}$$

F. User Calibration

To account properly for the differences in the eyes of different users, certain user-dependent eye-invariant parameters are estimated during user calibration. Each user is asked to sit in front of the computer screen and turn their gaze to one predefined calibration point on the screen. Eye images are captured by the system cameras in order to extract the feature parameters of pupil and glints on the images. According to the geometric relationship among the eye, the cameras, and the computer screen, the cornea radius and the kappa angle are calibrated, as discussed below.

1) Cornea Radius Calibration: Fig. 11 shows where the ray that comes from the light source L_i reflects at a point G_{ij} on the cornea surface such that the reflected ray passes through the camera optical center O_j and intersects the camera image plane at a point g_{ij}. Since the law of reflection states that the angles of incidence and reflection are equal [26], the unit vector of the incident light $\mathbf{I}_{ij}$ and the unit vector of the reflection light $\mathbf{R}_{ij}$ satisfies

$$\mathbf{R}_{ij} = \mathbf{I}_{ij} - 2\mathbf{n}_{ij}(\mathbf{n}_{ij} \cdot \mathbf{I}_{ij}) \tag{11}$$

where $\mathbf{n}_{ij}$ is the unit vector of the reflection normal. Assume that the reflection point is $G_{ij} = O_j + k_{ij}(O_j - g_{ij})$, then $\mathbf{I}_{ij}$, $\mathbf{n}_{ij}$, and $\mathbf{R}_{ij}$ are, respectively, represented as

$$\mathbf{I}_{ij} = \frac{(L_i - G_{ij})}{\|L_i - G_{ij}\|}, \quad \mathbf{n}_{ij} = \frac{(G_{ij} - C)}{\|G_{ij} - C\|}, \quad \mathbf{R}_{ij} = \frac{(O_j - G_{ij})}{\|O_j - G_{ij}\|}. \tag{12}$$

The cornea center C is estimated by the intersection of the reflection planes of two cameras, as discussed in Section III-C; thus, k_{ij} can be estimated by (11) and (12). G_{ij} can then be obtained by substituting k_{ij} into the expression of G_{ij}. The distance between the reflection point on the cornea surface G_{ij} and the cornea center C is a constant: the cornea radius r_c. Thus

$$\|G_{ij} - C\| = r_c. \tag{13}$$

To obtain the more accurate cornea radius r_c, a system of quadratic equations with unknown variables r_c is established as follows:

$$f_t = \|O_j - g_{ij}\|^2 k_{ij}^2 + 2(O_j - g_{ij}) \cdot (O_j - C) k_{ij} + \|O_j - C\|^2 - r_c^2 \tag{14}$$

where $i, j = 1, 2$; $t = 1, 2, 3, 4$. The unconstrained optimization method is used in this study to optimize the cornea radius r_c, and it exists as

$$r_c^* = \arg\min_{r_c} \sum_{t=1}^{4} f_t^2. \tag{15}$$

The calibrated cornea radius can be used to determine the relative positional relationship between the reflection points on the corneal surface.

2) Kappa Angle Calibration: The kappa angle is the angle between the OA and VA of the eye. Due to the invisibility of the fovea, it is an eye-invariant parameter of great significance in terms of 3-D gaze estimation when transforming the OA to the LOS. The kappa angle is calibrated as discussed below.

When the eye is in the calibration position, an eyeball coordinate system (x-y-z coordinate system) is set (Fig. 12). The cornea center C_0 is the original point and the OA of the eye is the z-axis. Selecting any line perpendicular to the z-axis as the x-axis allows the y-axis to be determined uniquely.

During user calibration, the user looks to the calibration point S on the screen as the OA direction of the eye is estimated by the method presented in Section III-E. The VA direction of the eye can be obtained by the estimated cornea center C_0 and the known calibration point S. Therefore, the kappa angle can be calibrated as follows:

$$\cos(\kappa) = \frac{\mathbf{V}_o \cdot (S - C_0)}{\|S - C_0\|}. \tag{16}$$

G. 3-D Gaze Estimation

During the 3-D gaze-estimation process, the eye moves from the calibration position to another position while the user stares at any point on the screen. The LOS of the eye is estimated by the OA reconstructed by the methods, as discussed in Section III-E, and the kappa angle. However, the VA direction when the user is in any position cannot be determined uniquely while knowing the OA direction and the value of the kappa angle. Therefore, a transformation matrix of the OA and the VA is deployed in real time in this study.

The z-axis is determined by the OA of the eye, so the real-time estimation of the kappa angle is only a translation of the eyeball coordinate system and the rotation of the x- and y-axes. Suppose that α and β are, respectively, the rotation angle around the x-axis and the y-axis; the spatial kappa angle of the eye can be estimated at any position. Assume that the unit vector of the OA in the calibration position is $\mathbf{V}_0 = (x_0\ y_0\ z_0)^T$ and the unit vector of the VA in the calibration position is $\mathbf{V}_{p0} = (x_{p0}\ y_{p0}\ z_{p0})^T$. The cornea center in the calibration position is represented as $C_0 = (x_{c0}\ y_{c0}\ z_{c0})^T$. When the eye is at any position, the unit vector of the OA in this position is represented as $\mathbf{V}_1 = (x_1\ y_1\ z_1)^T$, the unit vector of the VA in this position is represented as $\mathbf{V}_{p1} = (x_{p1}\ y_{p1}\ z_{p1})^T$, and the cornea center in this position is $C_1 = (x_{c1}\ y_{c1}\ z_{c1})^T$. Thus, the translation of the original point in the eyeball coordinate system is $\mathbf{T} = C_1 - C_0$ and the rotation matrix between the OA of the eye in the calibration position and that in any position is

$$\mathbf{V}_1 = \mathbf{R}\mathbf{V}_0$$
$$\text{where } \mathbf{R} = \begin{pmatrix} \cos\beta & 0 & -\sin\beta \\ -\sin\alpha\sin\beta & \cos\alpha & -\sin\alpha\cos\beta \\ \cos\alpha\sin\beta & \sin\alpha & \cos\alpha\cos\beta \end{pmatrix} \tag{17}$$

where α and β can be estimated in real time. Both α and β have two different solutions. The eyes can only focus on one point, so the only correct solution can be determined by whether the LOSs of our eyes estimated by different α and β intersect at one point. $\mathbf{R}$ can be obtained accordingly. The OA and VA of the eye have a certain positional relationship in the eyeball coordinate system, so the rotation matrix $\mathbf{R}$ estimated by α and β is also applicable to the relationship between the VA directions in these two positions. The VA direction of the calibration position is determined by the cornea center C_0 and the known calibration point S; thus, the VA direction of any position can be expressed as

$$\mathbf{V}_{p1} = \mathbf{R}\mathbf{V}_{p0}. \tag{18}$$

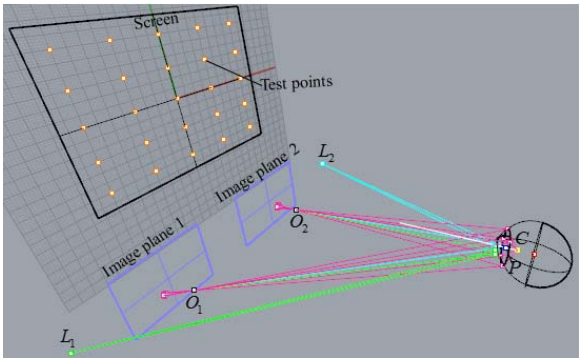


Fig. 13. Simulation model of proposed gaze-tracking system.

The 3-D LOS can be obtained by the cornea center C_1 and the VA direction $\mathbf{V}_{p1}$. The POR can be expressed as

$$\text{POR} = C_1 + \eta\mathbf{V}_{p1} \tag{19}$$

where η is a ratio factor that can be determined by the screen equation.

IV. Experimental Evaluation

The analyses we conducted in this study include simulations on exact data and tests in a real-world scenario. Computer simulations were performed to analyze the possible error sources of the proposed method, and practical experiments were carried out to test the performance of the gaze-tracking system in the real world.

A. Simulations on Exact Data

The simulation model was established using the Rhino5.0 3-D software, as shown in Fig. 13. The eyeball was drawn based on the Gullstrand-Le Grand eye model. The relative positions of the light sources, the screen, and the system cameras were determined according to the scale of the real system. C is the cornea center of the eyeball and P is the pupil center. L_1, L_2 are the two light sources and O_1, O_2 are the optical centers of the two cameras.

The parameters needed in the gaze-tracking system were determined based on the simulation model including the positions of the light sources, the calibration points, the feature parameters of the pupil, the glints, and the focal length of the cameras. They were plugged into MATLAB to analyze certain measurement errors that are unavoidable in the real scenarios.

The possible error sources considered in this study include the glints, matching points, and positions of the light sources. Gaussian noise was added to the possible error sources in order to simulate the real-world conditions effectively. The glint and matching point errors were generated by adding the white Gaussian noise with a standard deviation (SD) from 0.25 pixels (1 pixels = 3 μm) to 2.5 pixels. The position errors of the light sources were generated by adding the random unit vectors multiplied by 0.5 mm. A total of 100 tests were conducted, and the OA was reconstructed by the CA method accordingly. The influence of the error sources on OA reconstruction is shown in Table I.

TABLE I
INFLUENCE OF THE CONSIDERED ERROR SOURCES ON OA RECONSTRUCTION

Glint noises [pixel]	OA errors[°]	Matching point noises[pixel]	OA errors[°]	Light source noises[mm]	OA errors[°]
0.25	0.2060	0.25	0.0904	0.5	1.5240
0.5	0.3056	0.5	0.2340	1	2.1621
0.75	0.3503	0.75	0.1898	1.5	2.7486
1	0.4412	1	0.1991	2	3.3100
1.25	0.4908	1.25	0.6317	2.5	3.6899
1.5	0.5003	1.5	0.2931	3	4.1201
1.75	0.5650	1.75	0.1327	3.5	4.3787
2	0.5700	2	1.3079	4	5.2547
2.25	0.6181	2.25	0.3790	4.5	5.5315
2.5	0.6802	2.5	0.0925	5	5.6870

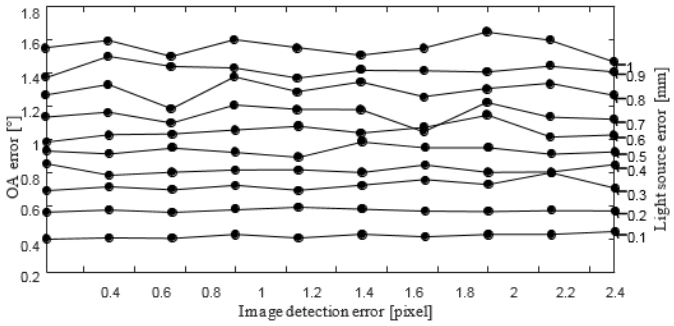


Fig. 14. Comprehensive influence of the error sources on OA reconstruction.

Table I shows where the error of the OA increased as the glint error increased in an approximately linear correlation. When the glint error was 2 pixels, the OA error was less than 0.6°. When the glint error was 1 pixel, the OA error was less than 0.5°. In the practical experiments, glint detection used the subpixel method in which error is usually less than 1 pixel. Therefore, the influence of the glints on the OA reconstruction would not exceed 0.5°. When the Gaussian noise of varying SD was added to any pair of matching points, the OA-reconstruction errors under different matching point errors were estimated. The OA-reconstruction errors were randomly distributed and less than 0.7° when the matching point errors were within 2.5 pixels. The OA-reconstruction errors nearly linearly increased as the light-source error increased. When the light-source error was 1 mm, the OA error was 3.5°. When the light-source error was 5 mm, the OA error exceeded 5.5°. Since the kappa angle is generally 5°, the accuracy of the light sources markedly influences the OA reconstruction. It should be kept below 0.5 mm during the light-source-calibration process.

The glints and the matching points are all in the image plane, so their errors are attributable to the image-detection error. When the image-detection and light-source-calibration errors exist simultaneously, the error of the reconstructed OA is as shown in Table II.

Fig. 14 shows the comprehensive influence of the error sources on OA reconstruction. Each curve represents different light-source errors. Glint and matching point errors appear to have a little effect on OA reconstruction, but the light-source error markedly impacts the OA.

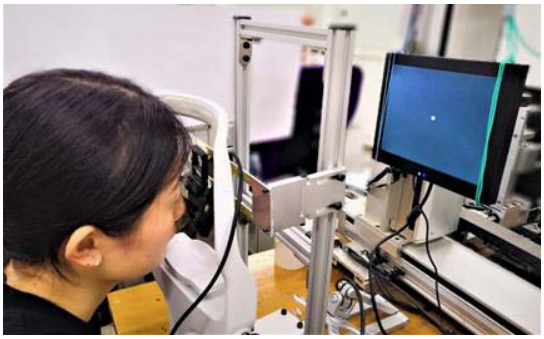


Fig. 15. User-calibration experiments.

B. Experiments in Real Scene

We tested the performance of the proposed gaze-tracking method on the system shown in Fig. 3. We performed both user-calibration experiments and 3-D gaze-estimation experiments. During the user-calibration experiments, the eye-invariant parameters involving the cornea radius and the kappa angle were calibrated as per the method described in Section III-F. In the 3-D gaze-estimation experiments, the OA of the eye was reconstructed, as discussed in Section III-E, on the basis of the calibrated cornea radius. The LOS of the eye was estimated by the reconstructed OA and the estimated transformation matrix that represents the kappa angle. Finally, the gaze-estimation results of the proposed method were compared with the traditional methods [26], [28] to confirm its effectiveness.

1) User-Calibration Experiments: We performed one-point user calibration in the first set of experiments. Eight subjects (gender: 1 male and 7 females; age: 24–27 years old; physical condition: healthy) participated in the experiments. The subjects were asked to keep their chin on the system's fixed bracket and look at an on-screen point that was preset as the calibration point, as shown in Fig. 15. The fixation time was about 1 s. The eye images were captured by the two system cameras throughout this time; these images are used for eye recognition, pupil segmentation, and glint segmentation so as to extract the eye feature parameters needed for user calibration, namely, the coordinates of the pupil edge points and glints.

According to the method for determining the matching points proposed in Section III-D, the matching points on the pupil edges in the two cameras were obtained to reconstruct the OA of the eye by the CA method described in Section III-E. The cornea center was estimated by the intersection of the reflection planes of the two cameras, as shown in Section III-C, so the cornea radius can be calibrated by the method in Section III-F. The gaze point during user calibration was preset and known; thus, the VA of the eye could be estimated by the estimated cornea center and the calibration point. The calibrated cornea radius and the kappa angle values are listed in Table III.

Since the cornea radius ranges from 7.0 to 8.5 mm and the kappa angle is about 5° [37], the calibrated cornea radius and the kappa angle are all within the normal range. Their accuracy is verified in the following gaze-estimation experiments.

TABLE II
COMPREHENSIVE EFFECT OF THE CONSIDERED ERROR SOURCES ON OA RECONSTRUCTION

OA errors[°]		Light source errors[mm]									
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Image detection errors [pixel]	0.25	0.3999	0.5618	0.6897	0.8492	0.9319	0.9811	1.1328	1.2656	1.3738	1.5520
	0.5	0.4075	0.5764	0.7128	0.7868	0.9146	1.0270	1.1618	1.3287	1.5002	1.5959
	0.75	0.4043	0.5600	0.6946	0.8036	0.9499	1.0315	1.0980	1.1831	1.4411	1.4987
	1	0.4265	0.5775	0.7212	0.8153	0.9235	1.0569	1.2045	1.3743	1.4324	1.6011
	1.25	0.4066	0.5944	0.6918	0.8170	0.8931	1.0791	1.1789	1.2854	1.3729	1.5529
	1.5	0.4271	0.5813	0.7217	0.8015	0.9874	1.0377	1.1773	1.3437	1.4179	1.5064
	1.75	0.4135	0.5687	0.7545	0.8431	0.9518	1.0729	1.0464	1.2558	1.4152	1.5492
	2	0.4261	0.5655	0.7255	0.8023	0.9529	1.1485	1.2200	1.3044	1.4088	1.6464
	2.25	0.4263	0.5721	0.7956	0.8044	0.9125	1.0121	1.1331	1.3356	1.4421	1.5972
	2.5	0.4420	0.5702	0.7055	0.8447	0.9268	1.0253	1.1205	1.2652	1.4094	1.4681

TABLE III
USER-CALIBRATION RESULTS OF EACH SUBJECT

Subjects	Cornea radius[mm]	Kappa angle[°]
1	7.46	6.14
2	7.22	6.41
3	8.03	5.58
4	7.59	5.77
5	8.11	5.06
6	7.15	4.89
7	7.38	5.78
8	7.74	5.53

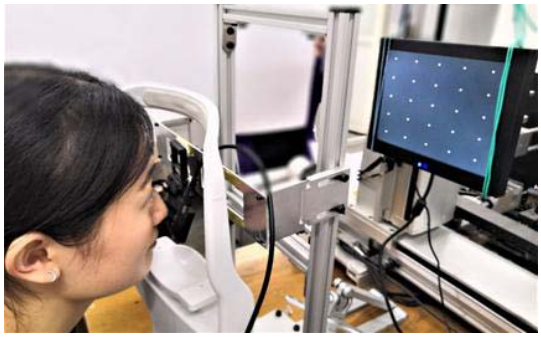


Fig. 16. 3-D gaze-estimation experiments.

2) 3-D Gaze-Estimation Experiments: In the 3-D gaze-estimation experiments, 25 test points with uniform distribution were set on the screen, as shown in Fig. 16. The screen in the camera coordinate system was calibrated beforehand, so the coordinates of these test points in the camera coordinate system were known. The points were used to calculate the deviation between the estimated gaze point and the real gaze point so as to evaluate the accuracy of the system.

The subjects whose cornea radius and kappa angle had been calibrated were also asked to perform the 3-D gaze-estimation experiments. They again kept their chin on the system's fixed bracket and looked at each on-screen test point. They were able to move slightly their heads in translation, pitch, and yaw to create slight displacement between the head and the camera. Eye images of the subjects were captured by the two system cameras throughout this process at a sampling frequency of ten frames/s.

The matching points of two cameras were determined after extracting the pupil edge points and glints. The OA was then reconstructed by the three proposed methods, namely, CA method, EMA method, and MAI method (Section III-E). Among them, 20 matching points were used to reconstruct the OA by The CA method. The LOS and the POR on the screen were estimated in combination with the calibrated kappa angle of each subject. The experimental results of one of the subjects are provided in Table IV.

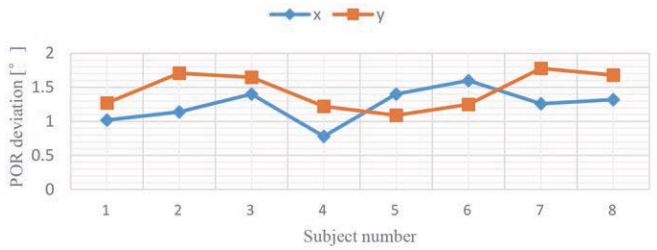
TABLE IV
GAZE-ESTIMATION RESULTS OF SUBJECT 4

Points	Real PORs[mm]	Proposed CA method[mm]	Proposed EMA method[mm]	Proposed MAI method[mm]
1	(21.20,109.55)	(20.17,104.33)	(25.77,113.35)	(10.62,119.49)
2	(63.60,109.55)	(62.98,116.88)	(61.17,115.17)	(63.62,109.59)
3	(106.00,109.55)	(106.39,113.97)	(101.60,110.11)	(104.72,113.77)
4	(148.40,109.55)	(148.11,109.91)	(146.74,115.77)	(135.84,121.26)
5	(190.80,109.55)	(191.19,116.94)	(186.13,127.09)	(209.39,144.70)
6	(21.20,15.65)	(39.41,10.72)	(41.39,17.61)	(38.38,11.79)
7	(63.60,15.65)	(60.88,22.17)	(59.36,25.50)	(53.42,13.07)
8	(106.00,15.65)	(101.93,16.29)	(101.86,19.65)	(100.17,15.55)
9	(148.40,15.65)	(146.38,14.57)	(144.01,23.31)	(138.18,16.57)
10	(190.80,15.65)	(182.68,15.61)	(183.34,18.56)	(185.09,19.94)

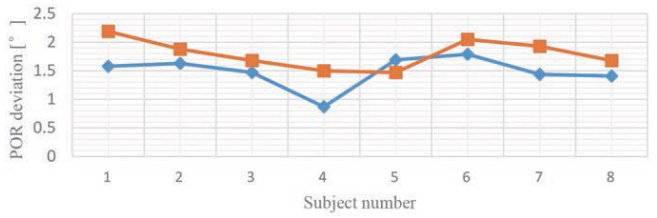
To analyze the 3-D gaze-estimation results intuitively, we calculated the deviation between the estimated POR and the real POR as a root-mean-square error (RMSE). Fig. 17 shows the gaze accuracy of eight subjects in the x- and y-directions, respectively, using the three methods we tested.

The average deviations of the PORs in the x- and y-directions using the CA method were 1.24° and 1.46°, respectively. The average deviations of the PORs in the x- and y-directions using the EMA method were 1.49° and 1.80°, respectively. The average deviations of the PORs in the x- and y-directions using the MAI method were 1.95° and 1.84°, respectively. The CA method showed higher OA-reconstruction accuracy and more stable results than the other two methods. The deviation between the estimated PORs and the real PORs based on the CA method is shown in Fig. 18.

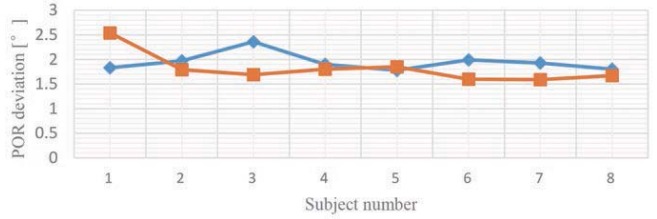
3) Comparisons With Other 3-D Gaze-Estimation Methods: The method proposed in this article was established to resolve the shortcomings of the traditional methods using multiple cameras and multiple light sources. Based on the extracted features of the pupil and the glints, we also used the traditional method using the virtual image of the pupil center to estimate the 3-D LOS. The SD of the kappa angle, mean value (MV)



(a) CA method



(b) EMA method



(c) MAI method

Fig. 17. Deviation of the estimated PORs in the x- and y-directions of each subject. (a) CA method, (b) EMA method, (c) MAI method.

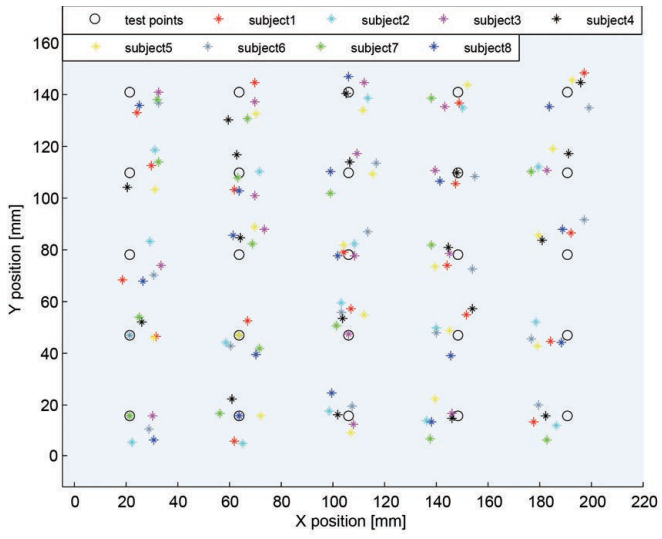


Fig. 18. Deviation between the estimated PORs and the real PORs.

of VA, SD of VA, and the accuracy of 3-D gaze-estimation are listed in Table V.

The SD of the kappa angle of the proposed CA method is smaller than that of the traditional method, reflecting that the OA reconstructed by the proposed CA method has higher stability. Considering the accuracy of the VA direction, the proposed CA method, EMA method, and MAI method are all more accurate than the traditional method, but their stability is slightly poor. The main reason for the worse stability of the proposed CA method is the error introduced by more matching points. The proposed CA method showed higher accuracy than the traditional method, while the accuracy of the proposed EMA method and the MAI method was less accurate than the traditional method. The EMA method has relatively poor accuracy, because the orientation and angle of the major axis of the pupil imaging ellipse are easily disturbed by the random errors. The MAI method has relatively poor accuracy for many reasons such as the influence of pupil ellipse fitting and the unreasonable structure of the two-camera system. For example, if the angle between the OAs of the two cameras is too small, the pupil imaging ellipses in the two cameras are similar; thus, the error of the intersection of the major axes of the pupil imaging ellipse is magnified. The measurement error caused by the inadequate baseline distance of the dual cameras also affects the gaze-estimation accuracy.

TABLE V
COMPARISON WITH THE TRADITIONAL 3-D GAZE-ESTIMATION METHOD

Methods	SD of kappa angle[°]	MV of VA[°]	SD of VA[°]	Accuracy[°]	
				X	Y
Traditional method [26], [28]	0.66	2.3	1.04	1.47	1.78
Proposed CA method	0.52	1.9	1.31	1.24	1.46
Proposed EMA method	0.69	2.01	1.19	1.49	1.8
Proposed MAI method	0.86	1.93	1.2	1.95	1.84

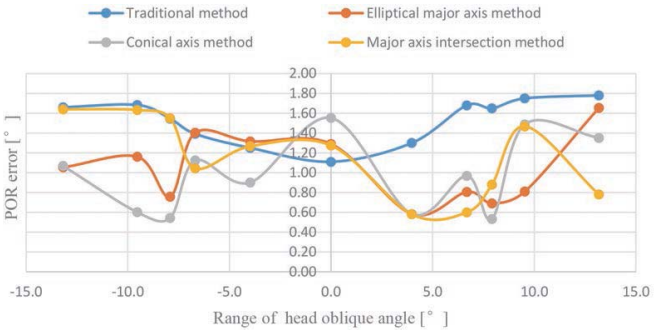


Fig. 19. Comparison of POR errors with different oblique angles.

We carried out the 3-D gaze-estimation experiments under different oblique angles. Assume that the left oblique angle is negative and the right oblique angle is positive, and the POR errors under the oblique angles in the range of $-13.2°$ to $+13.2°$ were estimated according to the method proposed in this article and the traditional method in [26], [27], [29], and [32], as shown in Fig. 19. The POR errors of the traditional method were relatively large when the oblique angle of the LOS was large, which indicates that the proposed method is valid and does improve upon the traditional method.

C. Discussion

Our simulations indicate that light-source deviation has a substantial impact on the reconstruction of the OA of the eyeball, which is consistent with the traditional method. The matching error of the pupil edge points in two camera images is also an important error source in the OA-reconstruction process. Accurate determination of the matching points of the

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

Methods	Number of cameras	Number of lights	Head movement	Accuracy[°]	Categories
[38]	1	0	Slight	2.37±1.42	Appearance-based
[39]	1	0	Fixed	6.91	Appearance-based
[40]	1	0	15.5°×15.5°×5°	2.27	PCT-based
[41]	1	2	Free	1.4	RCNN-based
[42]	1	5	Free	0.7	CR-based
[43]	1	4	Free	1.23±0.54	HN-based
[22]	1	0	Free	1.28-2.71	Kinect-based
[23]	1	0	11.63°×13.44°×9.43°	1.99	Kinect-based
[44]	1	0	Free	X:3.0;Y:4.5	Kinect-based
[45]	2	2	50mm×25mm×50mm	1.18	Remote gaze tracker
[46]	4	4	Free	1.5	Head-mounted gaze tracker
Proposed CA method	2	5	-13.2°-13.2°	X:1.24;Y:1.46	Helmet-mounted gaze tracker
Proposed EMA method	2	5	-13.2°-13.2°	X:1.49;Y:1.8	Helmet-mounted gaze tracker
Proposed MAI method	2	5	-13.2°-13.2°	X:1.95;Y:1.84	Helmet-mounted gaze tracker

pupil imaging ellipses of two cameras is crucial in practical applications. It is necessary to reduce the pair number of the accurate matching points necessary for high-precision OA reconstruction.

We tested three different OA-reconstruction methods in this study: the CA method, EMA method, and MAI method. These methods are all feasible; however, our results show that only the CA method has a higher accuracy than the traditional method. The CA method mitigates the inherent shortcomings of the traditional method, but lacks utility value due to the large amount of computation necessary to operate it. By comparison, the EMA method is simple in operation and has a practical value. However, the gaze-estimation accuracy depends heavily on the accuracy of pupil ellipse fitting. The MAI method is the most practical method, but unfortunately, it has lower accuracy in our practical experiments than the traditional method for many reasons. In addition to the influence of pupil ellipse fitting, the unreasonable structure of the two-camera system can drive down the accuracy of the MAI method. The measurement error caused by the inadequate baseline distance of the two cameras also affects the gaze-estimation accuracy. In the case of a large eyeball squint angle, the accuracy of the proposed method is much higher than that of the traditional method.

Taken together, the results of our experiments well validate the rationality of our work and the effectiveness of the proposed method. However, further research and system optimizations are yet needed.

D. Comparison With State of the Art

There are many different gaze-tracking methods depending on different system configurations. In this study, several state-of-the-art methods in recent five years have been studied. These state-of-the-art methods are compared in terms of system configuration, influence of head movement, and the gaze estimation accuracy, as shown in Table VI. It can be seen from Table VI that the three OA-reconstruction methods proposed in this article meet the accuracy requirements in the application of the gaze-tracking system.

V. CONCLUSION

A new 3-D gaze-estimation method using an MCMLS system is proposed in this article. From the perspective of algorithm, relatively few pairs of matching points of the pupil imaging ellipses are needed to reconstruct the OA of the eye accurately due to the high precision of pupil detection in an infrared-light-source-based gaze-tracking system. Unlike the traditional MCMLS gaze-tracking method using the virtual image of the pupil center to reconstruct the OA, the proposed method obtains the OA of the eye by the refraction planes constructed by the edge points of the pupil imaging ellipses of the two cameras. The estimation of the spatial pupil center or the construction of the refraction plane using the virtual image of the pupil center does not required. This mitigates the inherent errors caused by OA reconstruction in the traditional method and prevents the light refraction error introduced by calculating the spatial pupil center directly. In addition, the cornea radius calibrated during user calibration is used alone to determine the relative positional relationship between the two glints in the proposed method, so the cornea radius can be loosely calibrated. From a systematic perspective, both the traditional multi-camera gaze-tracking method and the proposed method require calibration of the relative positional relationship between the cameras; thus, the proposed method does not add additional requirements to the traditional method. Therefore, the method proposed in this article is simple and has a practical application value.

Our experiments showed that the accuracy of the proposed CA method was 1.24° and 1.46° in the x- and y-directions, the accuracy of the proposed EMA method was 1.49° and 1.80° in the x- and y-directions, and the accuracy of the proposed MAI method was 1.95° and 1.84° in the x- and y-directions, respectively. The CA method is more accurate than the traditional method among the proposed OA-reconstruction methods. This demonstrates the deficiency of the traditional method and the effectiveness of the proposed method. However, the CA method cannot satisfy the real-time performance of the algorithm due to the large amount of computation necessary to operate it. We also attempted to improve the CA method for practical application in this study by establishing the EMA method and the MAI method. When the oblique angle of the LOS was large, the advantage of the proposed OA-reconstruction methods is more obvious.

Although the essence of these three methods is consistent, the EMA method and the MAI method allow for fast estimation of the OA and have a noteworthy practical application

value despite their lower OA-reconstruction accuracy than the CA method in our experiments. These two methods are more susceptible to various error factors (e.g., unreasonable structure of the two-camera system and influence of pupil ellipse fitting). In the future, we will work to control further the error in each step of the gaze-estimation process. We seek especially to optimize the design of the light-source distribution and the dual-camera system to reduce the system-calibration error and make the EMA and MAI methods more practical.

REFERENCES

[1] H. Duan and Q. Zhang, "Visual measurement in simulation environment for vision-based UAV autonomous aerial refueling," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 9, pp. 2468–2480, Sep. 2015.

[2] S. He and X. Dong, "High-accuracy localization platform using asynchronous time difference of arrival technology," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1728–1742, Jul. 2017.

[3] M. D. Cordea, E. M. Petriu, and D. C. Petriu, "3D head tracking and facial expression recovery using an anthropometric muscle-based active appearance model," *IEEE Trans. Instrum. Meas.*, vol. 57, no. 8, pp. 1578–1588, Aug. 2008.

[4] Q. Hu, Y. Guo, Z. Lin, X. Deng, and A. Wei, "Object tracking using multiple features and adaptive model updating," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 11, pp. 2882–2897, Nov. 2017.

[5] C. Hennessey, B. Noureddin, and P. Lawrence, "Fixation precision in high-speed noncontact eye-gaze tracking," *IEEE Trans. Syst., Man, Cybern. B. Cybern.*, vol. 38, no. 2, pp. 289–298, Apr. 2008.

[6] L. Xia, B. Sheng, W. Wu, L. Ma, and P. Li, "Accurate gaze tracking from single camera using Gabor corner detector," *Multimedia Tools Appl.*, vol. 75, no. 1, pp. 221–239, Jan. 2016.

[7] C. H. Morimoto and M. R. M. Mimica, "Eye gaze tracking techniques for interactive applications," *Comput. Vis. Image Understand.*, vol. 98, no. 1, pp. 4–24, Apr. 2005.

[8] 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.

[9] P. Blignaut, "A new mapping function to improve the accuracy of a video-based eye tracker," in *Proc. South Afr. Inst. Comput. Scientists Inf. Technologists Conf. (SAICSIT)*, 2013, pp. 56–59.

[10] L. Sesma-Sanchez, A. Villanueva, and R. Cabeza, "Gaze estimation interpolation methods based on binocular data," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 8, pp. 2235–2243, Aug. 2012.

[11] Y. Ebisawa, "Improved video-based eye-gaze detection method," *IEEE Trans. Instrum. Meas.*, vol. 47, no. 4, pp. 948–955, Aug. 1998.

[12] C. Zhang, J. Chi, Z. Zhang, X. Gao, T. Hu, and Z. Wang, "Gaze estimation in a gaze tracking system," *Sci. China Inf. Sci.*, vol. 54, no. 11, pp. 2295–2306, Nov. 2011.

[13] Z. Zhu and Q. Ji, "Eye gaze tracking under natural head movements," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 918–923.

[14] S. M. Kim, M. Sked, and Q. Ji, "Non-intrusive eye gaze tracking under natural head movements," in *Proc. 26th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Sep. 2004, p. 2271.

[15] K. P. White, T. E. Hutchinson, and J. M. Carley, "Spatially dynamic calibration of an eye-tracking system," *IEEE Trans. Syst., Man, Cybern.*, vol. 23, no. 4, pp. 1162–1168, Jul./Aug. 1993.

[16] 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.

[17] 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.*, Oct. 2006, pp. 171–178.

[18] 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.*, May 2002, pp. 101–106.

[19] D. H. Yoo and M. J. 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.

[20] 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. (ETRA)*, 2010, pp. 13–20.

[21] 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.

[22] L. Sun, Z. Liu, and M.-T. Sun, "Real time gaze estimation with a consumer depth camera," *Inf. Sci.*, vol. 320, pp. 346–360, Nov. 2015.

[23] 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)*, May 2017, pp. 1646–1653.

[24] C. Hennessey, B. Noureddin, and P. Lawrence, "A single camera eye-gaze tracking system with free head motion," in *Proc. Symp. Eye Tracking Res. Appl. (ETRA)*, San Diego, CA, USA, Mar. 2006, pp. 87–94.

[25] H. Craig and L. Peter, "Noncontact binocular eye-gaze tracking for point-of-gaze estimation in three dimensions," *IEEE Trans. Bio-Med. Eng.*, vol. 56, no. 3, pp. 790–799, Mar. 2009.

[26] 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.

[27] T. Nagamatsu, J. Kamahara, and N. Tanaka, "Calibration-free gaze tracking using a binocular 3D eye model," in *Proc. 27th Int. Conf. Extended Abstr. Hum. Factors Comput. Syst. (CHI EA)*, 2009, pp. 3613–3618.

[28] S.-W. Shih and J. Liu, "A novel approach to 3-D gaze tracking using stereo cameras," *IEEE Trans. Syst., Man, Cybern. B. Cybern.*, vol. 34, no. 1, pp. 234–245, Feb. 2004.

[29] E. D. Guestrin and M. Eizenman, "Remote point-of-gaze estimation with free head movements requiring a single-point calibration," in *Proc. 29th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2007, p. 4556.

[30] D. Beymer and M. Flickner, "Eye gaze tracking using an active stereo head," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, Jun. 2003, pp. 1–8.

[31] N. M. Bakker *et al.*, "Accurate gaze direction measurements with free head movement for strabismus angle estimation," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 11, pp. 3028–3035, Nov. 2013.

[32] K. Zhang, X. Zhao, Z. Ma, and Y. Man, "A simplified 3D gaze tracking technology with stereo vision," in *Proc. Int. Conf. Optoelectron. Image Process.*, Nov. 2010, pp. 131–134.

[33] Z. Wei, "A distortion error model of the perspective projection of ellipse center and its simulation," *Chin. J. Sci. Instrument.*, vol. 24, no. 2, pp. 160–164, 2003.

[34] Y. C. Shiu and S. Ahmad, "3D location of circular and spherical features by monocular model-based vision," in *Proc. IEEE Conf. Int. Conf. Syst., Man Cybern.*, vol. 2, Nov. 1989, pp. 576–581.

[35] J.-N. Chi, Y.-Y. Xing, L.-N. Liu, W.-W. Gou, and G.-S. Zhang, "Calibration method for 3D gaze tracking systems," *Appl. Opt.*, vol. 56, no. 5, p. 1536, 2017.

[36] 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.

[37] Z. Zhu and Q. Ji, "Novel eye gaze tracking techniques under natural head movement," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 12, pp. 2246–2260, 2008.

[38] 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.

[39] 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.

[40] 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.

[41] K. Wang, S. Wang, and Q. Ji, "Deep eye fixation map learning for calibration-free eye gaze tracking," in *Proc. 9th Biennial ACM Symp. Eye Tracking Res. Appl. (ETRA)*, 2016, pp. 47–55.

[42] H. Cheng *et al.*, "Gazing point dependent eye gaze estimation," *Pattern Recognit.*, vol. 71, pp. 36–44, Nov. 2017.

[43] 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.

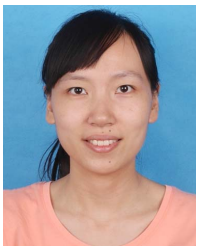
[44] J. Li and S. Li, "Gaze estimation from color image based on the eye model with known head pose," *IEEE Trans. Human-Mach. Syst.*, vol. 46, no. 3, pp. 414–423, Jun. 2016.

[45] 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.

[46] M. Lidegaard, D. W. Hansen, and N. Kruger, "Head mounted device for point-of-gaze estimation in three dimensions," in *Proc. Symp. Eye Tracking Res. Appl.* Safety Harbor, FL, USA: ACM, 2014, pp. 83–86.

Jiannan Chi received the B.S. degree in optic instrument engineering from Tianjin University, Tianjin, China, in 1990, the M.S. degree from Detection Technology and Automatic Equipment, Northeastern University, Shenyang, China, in 2002, and the Ph.D. degree in pattern recognition and intelligent system from Northeastern University, Shenyang, in 2005.

From 2005 to 2006, he was a Post-Doctoral Researcher with the University of Science and Technology Beijing, Beijing, China, where he is currently an Associate Professor with the Department of Instrument Science and Technology. He was a Visiting Scholar with Nanyang Technological University, Singapore, in 2011. His current research interests include computer vision, human–computer interaction, and optical measurement.



Jiahui Liu (Member, IEEE) received the B.S. degree from the Measurement and Control Technology and Instruments, Beijing Information Science and Technology University, Beijing, China, in 2014. She is currently pursuing the Ph.D. degree with the Control science and Engineering, University of Science and Technology Beijing, Beijing.

She is devoted to the research on the theoretical algorithms and the practical applications of the gaze-tracking systems.

Feng Wang received the B.S. degree from the University of Science and Technology Beijing, Beijing, China, in 2017, and the M.S. degree from the Instrument Science and Technology, University of Science and Technology Beijing, in 2020.

Her main work includes computer vision and image processing.

Yingkai Chi received the B.S. degree from China Medical University, Shenyang, China, in 1995, the M.S. degree from Dalian Medical University, in 2002, Dalian, China, and the Ph.D. degree from the Peking University Health Science Center, Beijing, China, in 2008.

He has been engaged in medical health, pathological study, and clinical diagnosis for a long time. He is currently the Deputy Director of the Department of Pathology, General Hospital of Northern Theater Command, Shenyang.

Zeng-Guang Hou (Fellow, IEEE) received the B.S. and M.S. degrees from the Department of Automation, Yanshan University, Qinghuangdao, China, in 1991 and 1994, respectively, and the Ph.D. degree from the Department of Automatic Control, Beijing Institute of Technology, Beijing, China, 1997.

From 1997 to 1999, he was a Post-Doctoral Researcher with the Key Laboratory of System Control, System Science Institute of China Science Academy, Beijing. He is currently a Doctoral Supervisor of the Institute of Automation, Chinese Academy of Sciences, Beijing, and also the Deputy Director of the State Key Laboratory for Complex System Management and Control. His currently research interests include robot and automation, intelligent control, and rehabilitation robot and minimally invasive intervention robot in the field of medical and health automation.