Accurate Regression-Based 3D Gaze Estimation Using Multiple Mapping Surfaces

ZHONGHUA WAN[1], **CAIHUA XIONG**[1], (Member, IEEE), **QUANLIN LI**[1],
WENBIN CHEN[1], (Member, IEEE), **KELVIN KIAN LOONG WONG**[2],
AND SHIQIAN WU[2], (Senior Member, IEEE)

[1]State Key Laboratory of Digital Manufacturing Equipment and Technology, Institute of Rehabilitation and Medical Robotics, Huazhong University of Science and Technology, Wuhan 430074, China
[2]School of Information Science and Engineering, Institute of Robotics and Intelligent Systems, Wuhan University of Science and Technology, Wuhan 430081, China

Corresponding author: Caihua Xiong (chxiong@hust.edu.cn)

ABSTRACT Accurate 3D gaze estimation using a simple setup remains a challenging issue for head-mounted eye tracking. Current regression-based gaze direction estimation methods implicitly assume that all gaze directions intersect at one point called the eyeball pseudo-center. The effect of this implicit assumption on gaze estimation is unknown. In this paper, we find that this assumption is approximate based on a simulation of all intersections of gaze directions, and it is conditional based on a sensitivity analysis of the assumption in gaze estimation. Hence, we propose a gaze direction estimation method with one mapping surface that satisfies conditions of the assumption by configuring one mapping surface and achieving a high-quality calibration of the eyeball pseudo-center. This method only adds two additional calibration points outside the mapping surface. Furthermore, replacing the eyeball pseudo-center with an additional calibrated surface, we propose a gaze direction estimation method with two mapping surfaces that further improves the accuracy of gaze estimation. This method improves accuracy on the state-of-the-art method by 20 percent (from a mean error of 1.84 degrees to 1.48 degrees) on a public dataset with a usage range of 1 meter and by 17 percent (from a mean error of 2.22 degrees to 1.85 degrees) on a public dataset with a usage range of 2 meters.

INDEX TERMS Head-mounted eye tracking, 3D gaze estimation, gaze direction estimation, eyeball center, mapping surface.

I. INTRODUCTION

Eye gaze is an essential clue for understanding human intentions. Gaze tracking is mainly used for attention analysis [1], [2], human-computer interaction [3], and human-robot interaction (HRI) [4], [5]. Li *et al*. [5] recently reported the first attempt to achieve intuitive HRI using only gaze signals, which is helpful for disabled people with upper limb motor impairment, such as amputees and paralyzed patients, to reconstruct their upper limb motion abilities. In mobile applications, such as gaze-based intuitive HRI, head-mounted 3D gaze trackers are preferred to table-mounted gaze trackers [6].

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

In head-mounted eye tracking, 3D gaze estimation faces three significant challenges, namely the accuracy of gaze estimation, calibration complexity (i.e., the number of calibration points and the burden of calibration procedures), and device complexity (such as the number of cameras and infrared LEDs). Many gaze estimation methods [6] have been proposed to address these limitations, which are generally divided into model-based and regression-based methods. Based on geometric eye models, model-based methods [7]–[9] calculate visual axes to represent gaze directions. Their main advantages are rapid one-point calibration and robustness against the relative drift between heads and eye trackers. However, these methods require sophisticated, fully calibrated devices and are limited in accuracy by geometric eye models, which differ from one individual to another.

In contrast, regression-based 3D gaze estimation methods employ simple devices to estimate gaze points. Most existing paradigms [10]–[13] are based on the implicit assumption that all gaze directions intersect at one point called the *eyeball center*. It is noted that the term in these methods differs from the traditional definition of the *eyeball center*, which is the intersection of all optical axes. To avoid confusion, we adopt the *eyeball pseudo-center* to represent the intersection of all gaze directions. Gaze directions are represented as the eyeball pseudo-center and gaze vectors [13], [14] or 2D mapping points on the imaginary mapping plane [10]. The existing regression-based methods have two limitations that affect the accuracy of 3D gaze estimation. First, the eyeball pseudo-center is challenging to be calibrated accurately, because it is inside the eyeball and outside the field of vision of the scene camera. Second, it is still unknown how the implicit assumption of the eyeball pseudo-center affects the accuracy of gaze direction estimation.

To overcome these two limitations, we present a gaze direction estimation model based on mapping surfaces. By simulating the implicit assumption of the eyeball pseudo-center, we find that all gaze directions intersect within a small region rather than at one point, which implies that this assumption is approximate. The sensitivity analysis of the eyeball pseudo-center on gaze estimation reveals the conditions of the assumption, i.e., configuring the mapping surface and reducing the error of the eyeball pseudo-center. Accordingly, we propose a gaze direction estimation method with one mapping surface that exploits and satisfies the assumption. The eyeball pseudo-center is calibrated by two additional calibration points outside the mapping surface. Furthermore, we propose a gaze direction estimation method with two mapping surfaces that adds another mapping surface to replace the eyeball pseudo-center under the same framework. The gaze direction is thus represented as the line connecting the two points on the two mapping surfaces. Experimental results reveal that the second method is more accurate than the first method, but has more calibration points. The main contributions of this paper are summarized as follows:

1) The simulation and sensitivity analysis reveal that the current implicit assumption of the eyeball pseudo-center is approximate and conditional.
2) A high-quality calibration technique for the eyeball pseudo-center is proposed, which requires only two additional calibration points outside the calibration plane.
3) An accurate gaze direction estimation method with two mapping surfaces is presented to overcome the limitation of the implicit assumption.

The remainder of this paper is organized as follows. Section II introduces the related works. Section III formulates the basic model of gaze direction estimation. Section IV analyzes the implicit assumption of the eyeball pseudo-center. The gaze estimation method with two mapping surfaces is

described in Section V. Experiments are given in Section VI, followed by the conclusion drawn in Section VII.

II. RELATED WORKS

Regression-based 2D gaze estimation methods regress from eye features (i.e., pupil centers or pupil-glint vectors) [15], [16] to gaze positions on a plane, without assuming a geometric eye model [8], [17]. Regression functions are calibrated by simultaneously capturing eye features and calibration points on a calibration plane. Similar to these methods, appearance-based methods use eye appearance [18], [19] as eye features. They leverage large training data with convolutional neural networks to achieve unconstrained gaze estimation. They do not require calibration, but are less accurate than the regression-based methods and are only suitable for table-mounted eye tracking.

When target points are outside the calibration plane, 2D gaze estimation methods result in an error of gaze points, i.e., the parallel error [20], due to the offset between scene cameras and eyes. To this end, some methods use additional input features related to depth coordinates of gaze points, such as pupil distances [5] or Purkinje images [21]. It is highlighted that the depth error of gaze estimation is generally significant, as these additional features have an indirect and weak correlation with depth coordinates of gaze points. Instead of additional eye features, other methods add scene information. Takemura *et al.* [22] used the correspondence between triangles containing 2D gaze points in the scene camera image and triangles containing 3D gaze points in the real world to estimate 3D gaze points.

A. GAZE DIRECTION ESTIMATION USING THE EYEBALL PSEUDO-CENTER

Unlike the above methods of direct mapping from features, most gaze direction estimation methods [10]–[12] implicitly assume that all gaze directions intersect at one point, the eyeball pseudo-center. Then gaze directions are represented as the eyeball pseudo-center and gaze vectors. Based on a simplified eye model of a perfect sphere where optical axes coincide with visual axes, back-projection methods [14], [23] estimate the eyeball center and gaze vectors in the coordinate system of the eye camera by back-projecting pupil image ellipses. Transforming the coordinate system into the scene camera introduces six parameters of homogeneous transformation. To solve these parameters, the 3D-to-3D mapping method in [14] forms a nonlinear optimization problem that minimizes the angular disparity between computed gaze vectors in the eye camera and targets in the scene camera. To reduce the complexity of the search space, Elmadjian *et al.* [24] proposed a two-step iterative optimization algorithm. Each iteration first computes the translation and then the rotation that places gaze vectors into the scene camera space.

Regression-based gaze direction estimation methods regress from eye features to gaze vectors. To solve the mapping parameters and the eyeball pseudo-center, the 2D-to-3D mapping method in [14] formulates a nonlinear

optimization that minimizes the angular error between estimated and ground-truth gaze vectors using at least two calibration planes. To address locally optimal solutions caused by nonlinear optimization, Su *et al.* [13] set sensible initialization and constraints on eye positions to improve accuracy. Our work is inspired by the method proposed by Abbott *et al.* [10], which regresses from eye features to mapping points on a calibration plane instead of gaze vectors. In this way, the eyeball pseudo-center and mapping parameters can be computed separately to obtain analytical solutions, without nonlinear optimization.

III. MODEL FORMULATION

Gaze directions are strongly correlated with rotations of the eyeball, such that they can be inferred from image features of the eye. Fig. 1 illustrates the regression-based gaze direction estimation. The gaze direction is represented as the line connecting the eyeball pseudo-center $\mathbf{O} \in \mathbb{R}^{3\times1}$ to the corresponding point $\mathbf{P} \in \mathbb{R}^{3\times1}$ on the imaginary mapping surface Π_P that has a one-to-one mapping with the eye feature (u, v), as follows:

$$\mathbf{G} = w\mathbf{O} + (1 - w)\,\mathbf{P}, \tag{1}$$

where $\mathbf{G} \in \mathbb{R}^{3\times1}$ denotes an arbitrary point on the gaze direction, and w is a proportional coefficient. The gaze direction of one eye intersects the gaze direction of the other eye or the reconstructed object surface at the gaze point. Therefore, this problem has transformed to calculate the eye pseudo-center $\mathbf{O}$ and the imaginary mapping surface Π_P.
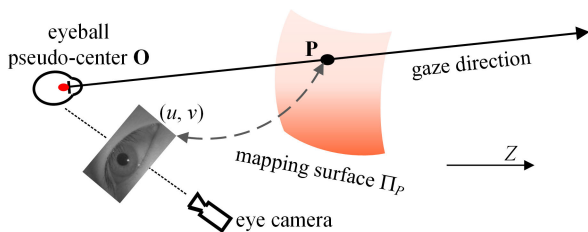

FIGURE 1. Illustration of regression-based gaze direction estimation. Each input eye feature (u, v) can calculate the corresponding point on the mapping surface Π_P. Note that the eyeball pseudo-center is assumed as the intersection of all gaze directions in the eyeball coordinate system, and it is non-coincident with the eyeball center.

A. IMAGINARY MAPPING SURFACE

The choice of the mapping function depends on extrapolation errors [25], which emerge when gaze points are outside the calibration area. Most regression-based methods employ polynomial mapping functions [3], [6], and it is found in [26], [27] that functions higher than two orders do not significantly reduce extrapolation errors. The typical polynomial mapping function is the second-order binary polynomial function that calculates the target point $\mathbf{P}(x_P, y_P, z_P)$ on the mapping plane at a constant distance of z_P from the input eye feature (u, v),

as follows:

$$\begin{cases} x_P = a_1 u^2 + a_2 v^2 + a_3 uv + a_4 u + a_5 v + a_6 \\ y_P = b_1 u^2 + b_2 v^2 + b_3 uv + b_4 u + b_5 v + b_6, \end{cases} \tag{2}$$

where $a_1, a_2, \ldots, a_6$ and $b_1, b_2, \ldots, b_6$ are the unknown mapping parameters. However, this mapping function requires users' heads to remain stationary during calibration and restricts all calibration points on a plane parallel to human faces.

To overcome this limitation, we replace Equation (2) with the second-order ternary polynomial function whose input is the same but output is the target point $\mathbf{P}(x_P, y_P, z_P)$ on the mapping surface, as follows:

$$\begin{cases} x_P = a_1 u^2 + a_2 v^2 + a_3 uv + a_4 u + a_5 v + a_6 \\ y_P = b_1 u^2 + b_2 v^2 + b_3 uv + b_4 u + b_5 v + b_6 \\ z_P = c_1 u^2 + c_2 v^2 + c_3 uv + c_4 u + c_5 v + c_6. \end{cases} \tag{3}$$

This function is calibrated with the least-square method by simultaneously capturing eye features and calibration points on a calibration plane. The advantage of this function is that users' heads are free during calibration since z_P of calibration points can be obtained by scene cameras. Since the basic coordinate system is located in the head-mounted eye tracker and moves with the head, the coordinates of calibration points are on a surface rather than a plane. After calibration, the mapping relationship remains unchanged, even though users look at different distances or move their faces. Therefore, the mapping surface is imaginary and changes consistently as the human face moves.

B. THE EYEBALL PSEUDO-CENTER

The meanings of the eyeball pseudo-center are two-fold. First, it refers to the intersection of all gaze directions in the eye coordinate system, as opposed to the eyeball center, which is the intersection of all optical axes. Second, it refers to an approximate assumption instead of an actual point. Existing papers [10]–[13] make this assumption implicitly and indicate this point as the eyeball center. To clarify the difference, we use the eyeball pseudo-center rather than the eyeball center. Since the eyeball pseudo-center is inside the eye, the position is difficult to measure. More importantly, it is still unknown how this assumption affects the accuracy of gaze direction estimation.

IV. ERROR ANALYSIS AND CALIBRATION OF THE EYEBALL PSEUDO-CENTER

To address the limitations of the assumption of the eyeball pseudo-center, we theoretically analyze the assumption and calibrate the eyeball pseudo-center.

A. INTERSECTIONS OF GAZE DIRECTIONS

According to [28], the assumption of the eyeball pseudo-center does not hold, i.e., all gaze directions do not intersect at one point. Thus, we simulate all intersections of gaze

directions based on the geometric eye model [28], which consists of two spheres, namely the cornea and the retina. The optical axis is the line connecting the eyeball center **E** to the cornea center through the cornea vertex **V** (see Fig. 2). The gaze direction refers to the visual axis, which is the line connecting the fovea to the gaze point by way of the nodal points **N**, **N′**. Model-based gaze estimation methods [7], [17] always assume that **N**, **N′** are coincident with the cornea center. In contrast, we abandon this assumption such that the results are more reliable and meaningful. When two visual axes are non-coplanar, the intersection refers to the midpoint of the common perpendicular of these two visual axes. Thus, we simulate all intersections of coplanar visual axes and all midpoints of common perpendiculars of non-coplanar visual axes.

First, we set the parameters and initial state of the eye model. Due to individual differences, we set typical values for the parameters according to accommodated schematic eyes of [17], [28]: $\mathbf{VN} = 6.6$ mm, $\mathbf{VE} = 13.1$ mm. Then $\mathbf{EN} = \mathbf{VE} - \mathbf{VN} = 6.5$ mm. The angle between the optical axis and the visual axis is always denoted as κ, which is individual-dependent and constant for each individual, as it is determined by the fovea location. Considering the mean value of κ, we set $\kappa = 5°$. In the coordinate system, the origin is the eyeball center **E**, and the Z-axis points to the horizontally forward direction. The X-axis points to the horizontally right eye, and the Y-axis points to the vertically downward direction (see Fig. 2). The initial state is that the optical axis is parallel to the Z-axis direction, and the visual axis is in the X-Z plane (see the red points in Fig. 3 (a)). Thus, the initial position of **N** is $\mathbf{N_0} = [0, 0, 6.5]^\mathrm{T}$, and the initial vector of the visual axis is $\mathbf{v_0} = \mathbf{R}_y(\kappa)\mathbf{z} = [\sin(\kappa), 0, \cos(\kappa)]^\mathrm{T}$.
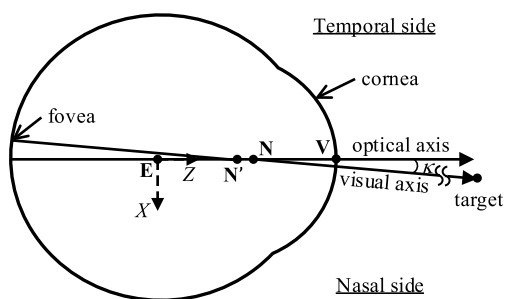
FIGURE 2. Structure of the geometric eye model (top view of the left eye) [28]. The optical axis goes through the eyeball center E and the cornea vertex V. The visual axis goes through the fovea and the gaze point by way of the nodal points N, N′. This axis is not a straight line, but two line segments.

Different visual axes are generated by eye rotations. Through rotating both a pan angle α (left-right rotation) and a tilt angle β (up-down rotation) from the initial eyeball state, the nodal point is $\mathbf{N}(\alpha, \beta) = \mathbf{R}_y(\alpha)\mathbf{R}_x(\beta)\mathbf{N_0}$, and the vector of the visual axis is $\mathbf{v}(\alpha, \beta) = \mathbf{R}_y(\alpha)\mathbf{R}_x(\beta)\mathbf{v_0}$, where $\mathbf{R}_y(\alpha)$ and $\mathbf{R}_x(\beta)$ are rotation matrices around y-axis and x-axis respectively. According to [29], the horizontal field of view is 100° for the temporal side, but 60° for the nasal side due to the
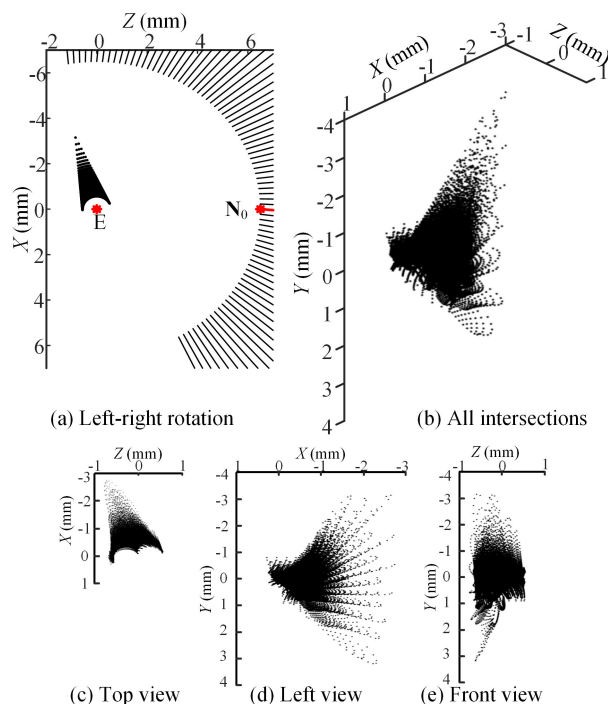
FIGURE 3. The simulation of the intersections of visual axes. (a) The intersections of visual axes in the left-right rotation. Each line is a visual axis, and the red point N_0 and the red line refer to the initial state of the visual axis. A large number of points in the middle are the intersections. (b) All intersections of coplanar visual axes and all midpoints of common perpendiculars of non-coplanar visual axes. Note that only a random part of all sample points are displayed for better observation. Three projection views of these points are shown in (c), (d), (f). The top view (c) is similar to (a).

cut-off of noses. Hence, $\alpha \in -100° \sim 60°$. As the vertical visual field is 60° to the brow side and 75° to the cheek side, $\beta \in -75° \sim 60°$. The sampling interval for the two angles of α and β is 2°. Note that since it is uncommon for the eye to rotate around the optical axis, torsion is ignored.

The simulation of the intersections of visual axes is shown in Fig. 3. For the left-right rotation (about the Y-axis), the intersections of visual axes lie within an approximately triangular area, whose range is 3.19 mm × 1.35 mm ($X × Z$) (see Fig. 3 (a)). For the up-down rotation (about the X-axis), visual axes form a part of a conical surface, such that they intersect at the apex of the cone, i.e., $[-0.57, 0, 0]^\mathrm{T}$. Besides the above two particular rotations, most eye rotations combine both pan and tilt. The area range of all intersections is 3.03 mm × 6.32 mm × 1.34 mm ($X × Y × Z$) (see Fig. 3 (b)). The distribution of all intersections is $[-0.52 \pm 0.22, 0.01 \pm 0.25, -0.15 \pm 0.31]$ mm (*mean ± standard deviation*). Therefore, all visual axes intersect in a small region. The assumption of the eyeball pseudo-center is approximate. Although the above simulation is based on the left eye model with typical parameters and the practical rotation range of the visual axis, the achieved region of intersections is similar by considering individual differences and the right eye.

B. SENSITIVITY ANALYSIS OF THE EYEBALL PSEUDO-CENTER IN GAZE ESTIMATION

Although all visual axes intersect in a small region, it is still insufficient to accept the assumption of the eyeball pseudo-center, unless this assumption has little effect on gaze estimation. To this end, we analyze the impact of the eyeball pseudo-center on gaze estimation. Note that the following analysis ignores the extrapolation error of the mapping surface, and we focus on the parallel error.
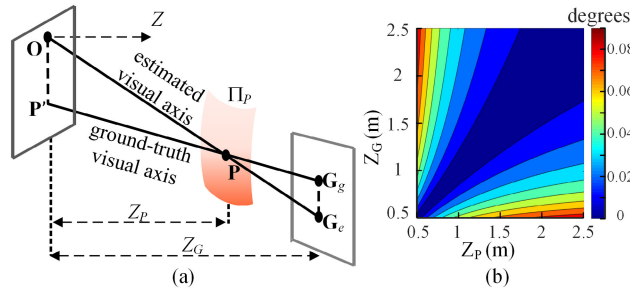

FIGURE 4. Sensitivity analysis of the eyeball pseudo-center in gaze estimation. (a) G_g is a ground-truth gaze point, and P is the corresponding point on the mapping surface Π_P. O is the eyeball pseudo-center. The estimated visual axis PO intersects with the vertical plane through G_g at G_e. The ground-truth visual axis PG_g intersects the vertical plane through O at P'. The Z-axis is the horizontal forward direction. (b) The relationship between the angular error E and Z_P, Z_G when O is the ideal eyeball pseudo-center, and $\|OP'\| = 1$ mm.

Fig. 4 shows the sensitivity analysis of the eyeball pseudo-center in gaze estimation. When a user is looking at a ground-truth point G_g, there are two kinds of visual axes through the mapping point P, namely the ground-truth visual axis PG_g and the estimated visual axis PG_e (see Fig. 4 (a)). According to [15], [16], the angular error of gaze points is calculated approximately by

$$E \approx \arctan\left(\frac{\|G_g G_e\|}{Z_G}\right), \qquad (4)$$

where $\|G_g G_e\|$ is the planer error of gaze estimation, and Z_G denotes the depth distance of G_g. Based on triangular similarity, $\|G_g G_e\|$ can be computed by

$$\frac{\|G_g G_e\|}{\|OP'\|} = \frac{\|PG_e\|}{\|OP\|} = \frac{|Z_G - Z_P|}{Z_P}. \qquad (5)$$

By substituting Equation (5) into Equation (4),

$$E \approx \arctan\left(\left|\frac{1}{Z_P} - \frac{1}{Z_G}\right| \|OP'\|\right). \qquad (6)$$

We should first investigate the effect of the ideal eyeball pseudo-center on gaze estimation. The ideal eyeball pseudo-center O is the average position of the intersecting region in Section IV-A. Then, $\|OP'\|$ is in the interval $[-3\sigma, 3\sigma]$ with a high probability, about 99.7% (assuming a Gaussian distribution), where σ is the standard deviation on the vertical plane in the intersecting region. Hence, $\|OP'\| \leq 3 \times \sqrt{0.22^2 + 0.25^2} \approx 1$ mm. Since E is quite small, less than

0.1° (see Fig. 4 (b)), it is reasonable to use the ideal eyeball pseudo-center to approximately represent the intersection of all real visual axes. Therefore, the ideal estimated visual axis can represent the ground-truth visual axis, as shown in Fig. 5 (a).
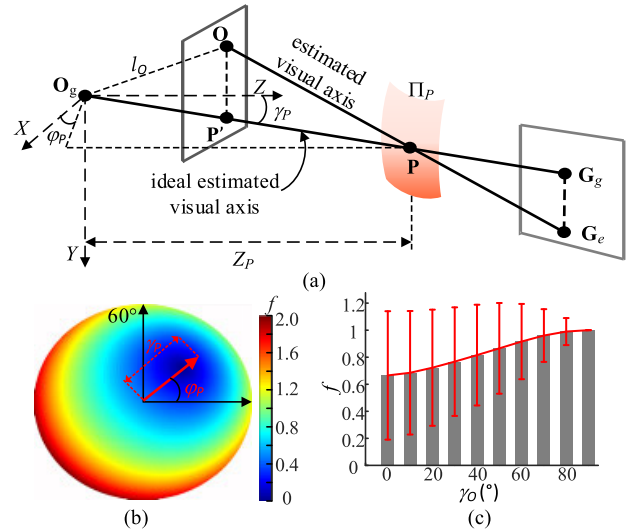

FIGURE 5. Sensitivity analysis of the estimated eyeball pseudo-center in gaze estimation. (a) The ideal estimated visual axis can represent the ground-truth visual axis. The ideal eyeball pseudo-center O_g is the origin, and the Y-axis points to the vertically downward direction. The spherical coordinates of P is $(l_P, \varphi_P, \gamma_P)$. (b) The distribution of f with different φ_P and γ_P when $\varphi_O = 45°$, $\gamma_O = 30°$. γ_P refers to the length. (c) Mean values and standard deviations of f with different γ_O.

We should then study the influence of the estimated eyeball pseudo-center on gaze estimation. In the spherical coordinate system, $P = (l_P, \varphi_P, \gamma_P)$, and $O = (l_O, \varphi_O, \gamma_O)$. As $P' = \frac{Z_O}{Z_P}P$, $\|OP'\| = l_O f(\gamma_P, \varphi_P; \gamma_O, \varphi_O)$, where

$$f(\gamma_P, \varphi_P; \gamma_O, \varphi_O)$$
$$= |\cos \gamma_O|$$
$$\times \sqrt{\tan^2 \gamma_P + \tan^2 \gamma_O - 2 \tan \gamma_P \tan \gamma_O \cos(\varphi_P - \varphi_O)}. \qquad (7)$$

φ_P and φ_O determine the phase of f, and γ_P and γ_O determine the amplitude of f. The minimum f is equal to 0 when O, P, and O_g are collinear, i.e., $\varphi_P = \varphi_O$, and $\gamma_P = \gamma_O$. The maximum is $f = |\cos \gamma_O|(\tan \gamma_P + \tan \gamma_O)$ when $\varphi_P = 180° + \varphi_O$, and γ_P is set to its maximum. Note that $f = 1$ when $\gamma_O = 90°$. Fig. 5 (b) shows the distribution of f with different γ_P and φ_P. Besides, Fig. 5 (c) shows that mean values of f change slightly when γ_O increases. Therefore, $f(\gamma_P, \varphi_P; \gamma_O, \varphi_O)$ can be regarded as the phase item of E. According to Equation (5), $\|G_g G_e\| = \frac{|Z_G - Z_P|}{Z_P - Z_O}\|OP'\| \approx \frac{|Z_G - Z_P|}{Z_P}\|OP'\|$, since $Z_O \ll Z_P$. By substituting Equation (7) into Equation (4),

$$E \approx \arctan\left(m(Z_P, Z_G, \gamma_O)f(\gamma_P, \varphi_P, \gamma_O, \varphi_O)\right), \qquad (8)$$

where

$$m(Z_P, Z_G, \gamma_O) = \left|\frac{1}{Z_P} - \frac{1}{Z_G}\right| l_O. \qquad (9)$$

$m(Z_P, Z_G, \gamma_O)$ and $f(\gamma_P, \varphi_P; \gamma_O, \varphi_O)$ approximately represent the amplitude term and phase term of E respectively. $m(Z_P, Z_G, \gamma_O)$ is similar to Equation (6), except that the amplitude is magnified by a factor of l_O. If l_O is larger than 10 mm, E may exceed 1°. Therefore, the estimated error of the eyeball pseudo-center has a significant effect on gaze estimation.

In summary, the assumption of the eyeball pseudo-center is conditional. To be specific, it is acceptable only if

- the distance between Z_P and Z_G decreases, and $Z_P Z_G$ increases.
- the error of the estimated eyeball pseudo-center, l_O, is small.

C. CALIBRATION OF THE EYEBALL PSEUDO-CENTER

To reduce the error of the estimated eyeball pseudo-center, we should build the calibration model of the eyeball pseudo-center. After calibrating the mapping surface, at least two additional calibration points outside the mapping surface are required to calibrate the eyeball pseudo-center $\mathbf{O}$. With n additional calibration points $\{\mathbf{C}_i\}_{i=1}^n$, $n \geq 2$, $\mathbf{C}_i \notin \Pi_P$, we can use the corresponding eye features to calculate the corresponding mapping points $\{\mathbf{P}_i\}_{i=1}^n$, which have extrapolation errors [25], as mapping points outside the calibration area are extrapolated. According to [15], extrapolation errors are not uniform along the mapping surface but increase towards the edges of the surface. The main reason is that detecting the limbus is more difficult at the edges than in the middle. Hence, it is assumed that the maximum radius of the extrapolation error is $\Delta P \approx \varepsilon \|\mathbf{OP}\|$, where ε is a small constant, such as $\varepsilon = \tan 0.5° \approx 0.0087$. Thus, each line $\mathbf{C}_i \mathbf{P}_i$ forms an error cone. Increasing the accuracy of the eyeball pseudo-center is equivalent to decreasing the intersection volume of all error cones.

Assuming that there are only two additional calibration points that have the same depths and lengths, $\|\mathbf{OC}_1\| = \|\mathbf{OC}_2\|$, error analysis of eyeball pseudo-center calibration with two additional calibration points is illustrated in Fig. 6. There are two auxiliary axes (see Fig. 6 (a)): the Z-axis refers to the horizontal forward direction; the d-axis refers to the line connecting $\mathbf{O}$ to the midpoint of $\mathbf{P}_1 \mathbf{P}_2$. The angle between the d-axis and the Z-axis is γ_d. The angle between d-axis and $\mathbf{OP}_1$ is γ_P. Then $\|\mathbf{OP}_1\| = Z_P/(\cos \gamma_P \cos \gamma_d)$. The two error cones formed by $\mathbf{C}_1 \mathbf{P}_1$ and $\mathbf{C}_2 \mathbf{P}_2$ intersect to form a diamond-like region (see Fig. 6 (b)). Based on triangular similarity, the region width h_O satisfies

$$\frac{h_O}{\Delta P} = \frac{\|\mathbf{C}_1 \mathbf{O}\|}{\|\mathbf{C}_1 \mathbf{P}_1\|} = \frac{d_C}{d_C - d_P} = \frac{Z_C}{Z_C - Z_P}, \quad (10)$$

then

$$h_O = \frac{Z_C Z_P}{(Z_C - Z_P) \cos \gamma_P} \frac{\varepsilon}{\cos \gamma_d}$$
$$= \frac{1}{\left(\frac{1}{Z_P} - \frac{1}{Z_C}\right) \cos \gamma_P} \frac{\varepsilon}{\cos \gamma_d}. \quad (11)$$
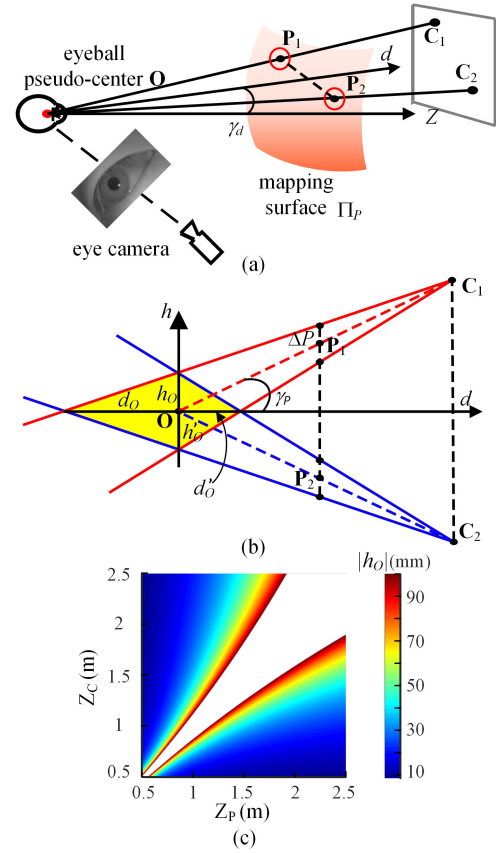
FIGURE 6. Error analysis of eyeball pseudo-center calibration with two additional calibration points. (a) Two additional calibration points are $\mathbf{C}_1$ and $\mathbf{C}_2$. The Z-axis refers to the horizontal forward direction; the d-axis refers to the line connecting $\mathbf{O}$ to the midpoint of $\mathbf{P}_1 \mathbf{P}_2$. The angle between the d-axis and the Z-axis is γ_d. (b) The yellow diamond-like region is the intersection of the two error cones formed by $\mathbf{C}_1 \mathbf{P}_1$ and $\mathbf{C}_2 \mathbf{P}_2$. The angle between d-axis and $\mathbf{OP}_1$ is γ_P. (c) Error analysis of $|h_O|$ with different Z_P, Z_C when $\gamma_d = 0$, and $\gamma_P = 45°$. Note that the white region in heat maps means the error is larger than 100 mm, and we do not show the detail for better observation.

The region width on the opposite side is $h_O' = -h_O$. Based on a similar deduction, the region depths d_O, d_O' satisfy

$$d_O = \frac{1}{\frac{\varepsilon}{Z_C} - \left(\frac{1}{Z_P} - \frac{1}{Z_C}\right) \sin \gamma_P} \frac{\varepsilon}{\cos \gamma_d}$$

$$d_O' = \frac{1}{\frac{\varepsilon}{Z_C} + \left(\frac{1}{Z_P} - \frac{1}{Z_C}\right) \sin \gamma_P} \frac{\varepsilon}{\cos \gamma_d}. \quad (12)$$

Reducing $|d_O|$ and $|d_O'|$ is equivalent to increasing the denominator. As ε is small, $\frac{\varepsilon}{Z_C} \ll \left|\frac{1}{Z_P} - \frac{1}{Z_C}\right| \sin \gamma_P$. Hence, $|d_O| \approx |d_O'|$, and $|h_O|/|d_O| = \tan \gamma_P$. The error region of $\mathbf{O}$ can be decreased by

- increasing the distance between Z_P and Z_C and decreasing $Z_P Z_C$ (see Fig. 6 (c)).
- decreasing the angle between d-axis and z-axis, i.e., γ_d, whose optimal value is 0, which implies that two calibration points are central symmetry about the Z-axis;

- setting the angle between two additional calibration points, $\gamma_P = 45°$, considering the contradictory relation between $|h_O|$, $|d_O|$ with γ_P.

When the two additional calibration points are not in the above situation, i.e., when their depths or lengths are different, the intersection region of the two error cones increases. In addition, Section VI-A demonstrates that gaze estimation is not significantly different using different numbers of calibration points.

Based on these conclusions and the conclusions in Section IV-B, the optimal configuration of gaze estimation with one mapping surface can be derived, as shown in Fig. 7.
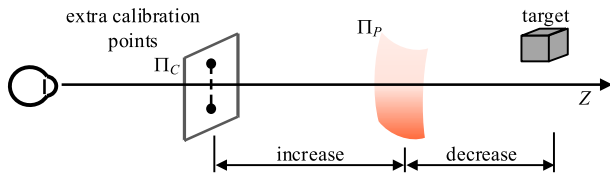

FIGURE 7. The optimal configuration of the method of gaze direction estimation with one mapping surface. The two additional calibration points should be central symmetrical about the Z-axis, and the angle between them and the eyeball should be $40° - 50°$.

V. GAZE ESTIMATION WITH TWO MAPPING SURFACES

Since the implicit assumption of the eyeball pseudo-center is conditional, an alternative method that replaces the assumption of the eyeball pseudo-center with another mapping surface is proposed to further improve accuracy under the same framework (same devices and similar calibration procedures). Thus, two points on the two mapping surfaces are calculated from an input eye feature, and both the two points are on the same visual axis. The gaze direction equation is changed from Equation (1) as follows:

$$\mathbf{G} = w\mathbf{P}_1 + (1 - w)\mathbf{P}_2, \qquad (13)$$

where $\mathbf{P}_1$ and $\mathbf{P}_2$ are corresponding points on the two mapping surfaces, respectively.

To employ this method, we analyze the errors outside mapping surfaces and consider configuring two calibration surfaces, as shown in Fig. 8 (a). According to the extrapolation error in Section IV-C, the maximum error radius on the mapping surface is $r = \varepsilon Z / \cos \gamma$, where γ is the angle between the gaze direction and the Z-axis. Then, the error range in two mapping surfaces forms an error cone whose vertex is $\mathbf{O}$. When the target is outside the two mapping surfaces, E has a maximum value when the error directions of $\mathbf{P}_1$, $\mathbf{P}_2$ are opposite (see red lines in Fig. 8 (a)). According to Equation (4), E is calculated by

$$E \approx \begin{cases} \arctan\left(\dfrac{\varepsilon}{\cos \gamma}\right), & Z_1 \leq Z_G \leq Z_2 \\[2ex] \arctan\dfrac{\varepsilon \left|Z_2 + Z_1 - \dfrac{2\,Z_1\,Z_2}{Z_G}\right|}{\cos \gamma\,(Z_2 - Z_1)}, & \text{otherwise.} \end{cases} \qquad (14)$$
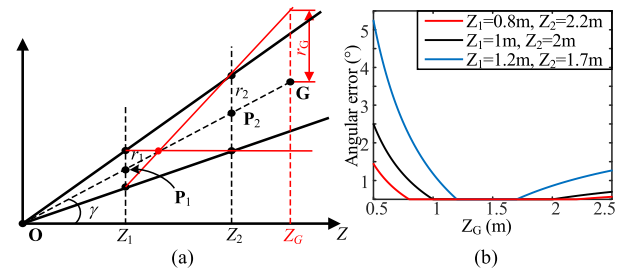

FIGURE 8. Error analysis for gaze estimation with two mapping surfaces. (a) Illustration of the errors outside two mapping surfaces whose depths are Z_1, Z_2. G is the gaze point. P_1, P_2 are ideal mapping points. γ is the angle between the gaze direction and the Z-axis. r_1 and r_2 are the maximum error radius on mapping surfaces. The red lines correspond to the maximum of E. (b) The relationship between E and Z_G with different Z_1 and Z_2 when $\gamma = 0$.
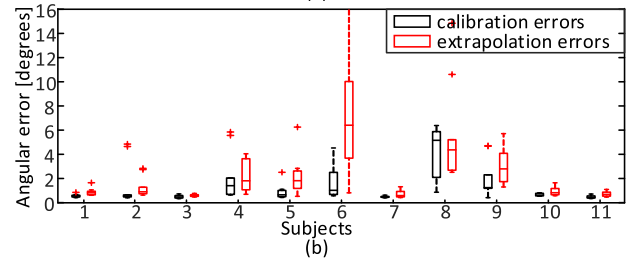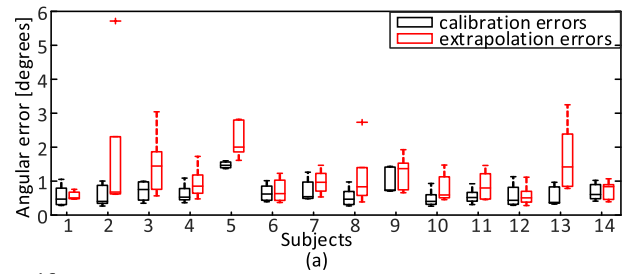

FIGURE 9. Calibration errors and extrapolation errors of 2D gaze estimation on calibration planes. (a) The Mansouryar dataset [14]. (b) The Elmadjian dataset [24].

Since the parallel error outside the two surfaces is larger than that between the two surfaces, it is concluded that the fixation target should be located between the two mapping surfaces. Besides, when the target object is outside the two surfaces, the angular error on the near side is larger than that on the far side (see Fig. 8 (b)).

VI. EXPERIMENTAL RESULTS

Since the proposed model with mapping surfaces requires calibration points on at least two planes, the datasets for table-mounted gaze estimation, such as MPIIGaze [30], are unsuitable. In this work, experiments are conducted on two publicly available datasets for head-mounted gaze estimation, namely the Mansouryar dataset [14] and the Elmadjian dataset [24]. The two datasets have small and large usage ranges, respectively. We first validate and evaluate the proposed methods on the Mansouryar dataset [14]. We then compared the proposed methods with state-of-the-art methods

TABLE 1. Average errors and standard deviations for the method using one mapping surface on the Mansouryar dataset [14] using different calibration combinations.

| Z_P [m]	1.5	1.75	2	1	1	1
Z_C [m]	1	1	1	1.5	1.75	2
Avg. err. [°]	1.90	1.91	1.85	3.41	1.7	**1.64**
Std [°]	1.09	1.18	1.05	1.86	0.84	**0.75**

using multiple calibration planes on the two datasets. Note that methods are evaluated by the angular error calculated by Equation (4), which is a better indicator for gaze direction estimation than the depth error.

The Mansouryar dataset [14] contains monocular pupil features and corresponding ground-truth 3D gaze coordinates recorded from 14 subjects and five planes of small depth range, including 1.0 m, 1.25 m, 1.5 m, 1.75 m, and 2.0 m. Each plane has 25 calibration points and 16 test points. After calibrating each mapping surface with Equation (3), the corresponding point on the mapping surface can be calculated for each pupil feature. The calibration error of the calibration plane is calculated as the average of angular errors between the ground-truth calibration points and the mapping points. The extrapolation error on the calibration plane is calculated as the average of angular errors between the ground-truth test points and the mapping points. Fig. 9 (a) shows that this dataset has small calibration and extrapolation errors.

The Elmadjian dataset [24] contains binocular pupil features and 3D gaze coordinates recorded from 11 subjects and five planes of large depth range, including 0.75 m, 1.25 m, 1.75 m, 2.25 m, and 2.75 m. Each plane has 20 calibration points and 12 test points. The size of the plane increases with the depth to ensure that different planes have the same field of view in the scene camera. As each calibration or test point contains 30 continuous frames, we filtered pupil features to reduce noise. Fig. 9 (b) shows that five subjects (number 4,5,6,8,9) have large calibration and extrapolation errors ($> 2°$). Thus, we discarded these data in the following experiments.

A. EVALUATION OF THE METHOD WITH ONE MAPPING SURFACE

After calibrating the mapping surface using a calibration plane, the eyeball pseudo-center is calibrated by two extra calibration points that have the largest distance on another plane. Fig. 10 shows gaze estimation errors for the method using one mapping surface with different calibration combinations, and Table 1 shows average errors and standard deviations. With the same third plane ($Z_P = 1.5$ m) as the calibration plane, Fig. 10 (a) shows that increasing the distance of calibration $|Z_P - Z_C|$ can reduce angular errors, and Z_C seems to be better on the near side ($Z_C < Z_P$) than on the far side ($Z_C > Z_P$). Besides, decreasing $|Z_G - Z_P|$ yields smaller angular errors, and Z_G is better on the far side ($Z_G > Z_P$) than on the near side ($Z_G < Z_P$). These results are consistent with the
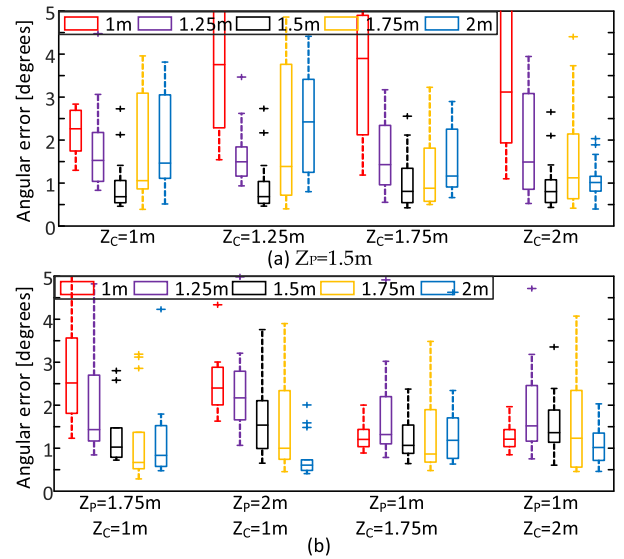
FIGURE 10. Gaze estimation errors for the method using one mapping surface on the Mansouryar dataset [14] with different calibration combinations. Note that it is not displayed when the angular error is larger than 5°. (a) Combinations using different additional calibration points and the same third plane ($Z_P = 1.5$ m) as the calibration plane. (b) Combinations when $|Z_P - Z_C|$ is large.
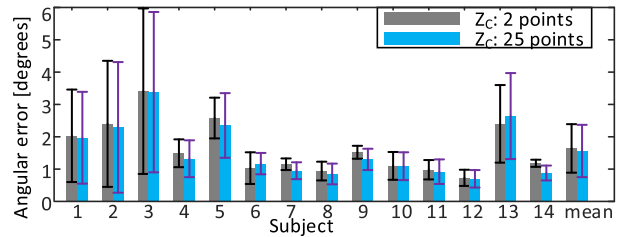
FIGURE 11. The impact of the number of additional calibration points on gaze estimation when $Z_P = 1$ m, and $Z_C = 2$ m. Bars show mean errors across all test planes; error bars indicate standard deviations.

theoretical analysis in Section IV-B. When $|Z_P - Z_C|$ is large, Fig. 10 (b) shows that the average errors of four combinations are approximate, but the standard deviation when ($Z_P > Z_C$) is larger than that when ($Z_P < Z_C$), because the angular error before the mapping surface is larger than that after the mapping surface. Overall, the optimal combination is $Z_P = 1$ m and $Z_C = 2$ m, whose error is $1.64° \pm 0.75°$.

To investigate the effect of the number of additional calibration points on gaze estimation, we used all 25 calibration points on a calibration plane instead of only two points. Fig. 11 shows that there is no significant difference between these two situations (paired-t test: *tstat* $= -2.075$, $p = 0.058$).

B. EVALUATION OF THE METHOD WITH TWO MAPPING SURFACES

Fig. 12 shows that the angular error is smallest between the two mapping surfaces ($Z_1 < Z_G < Z_2$), followed by after
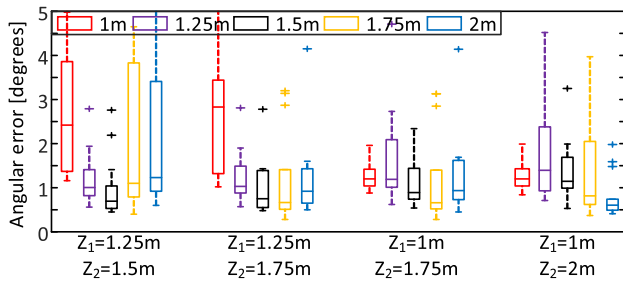
FIGURE 12. Gaze estimation errors for the method using two mapping surfaces on the Mansouryar dataset [14] with different combinations of two calibration planes. Note that it is not displayed when the angular error is larger than 5°.

TABLE 2. Average errors and standard deviations for the method using two mapping surfaces on the Mansouryar dataset [14] using different combinations of two calibration planes.

Z_1 [m]	1	1.5	1.25	1.25	1	1
Z_2 [m]	1.5	1.75	1.5	1.75	1.75	2
Avg. err. [°]	1.67	3.09	1.98	1.66	1.54	**1.48**
Std [°]	1.02	2.79	1.27	1.02	0.84	**0.83**

the second mapping surface ($Z_G > Z_2$), and then before the first mapping surface ($Z_G < Z_1$). These results are consistent with the theoretical analysis in Section V. Overall, the optimal combination is $Z_1 = 1$ m and $Z_2 = 2$ m, whose error is $1.48° \pm 0.83°$ (see Table 2). Compared to the method with one mapping surface, this method has better accuracy and similar precision. We can conclude that the assumption of the eyeball pseudo-center is conditional and affects gaze estimation.

C. COMPARISON WITH OTHER METHODS
We implemented and evaluated the following baseline methods of 3D gaze estimation using multiple calibration planes:

- *2D-to-2D Mapping*: As a basic baseline method, the method of 2D-to-2D mapping [14] employs multiple calibration planes to calibrate the regression relationship between pupil features and 2D mapping points. To achieve a comparison of the same conditions, we calibrated Equation (3) using the first and fifth calibration planes.
- *Fixed Eyeball Center (FEC)*: Abbott *et al.* [10] used a fixed eyeball pseudo-center and a mapping surface to estimate the gaze direction. We used the fifth plane for surface calibration, and the fixed eyeball center was set to $(0, 0, -0.05m)$ empirically.
- *Nonlinear Optimization*: The method recently proposed by Su *et al.* [13] used the constrained nonlinear optimization to calculate multiple parameters, including the eyeball pseudo-center and the mapping parameters between pupil features and gaze vectors. We used the first and the fifth planes for calibration. The initial 3D eyeball center was computed by the 2D eyeball center and the intrinsic matrix of the scene camera.

The constrained search range of the eyeball center was $\pm[0.05m, 0.05m, 0.02m]$. Due to the limitation of calibration points, cross-validated input-space partition was not used.

Gaze estimation results and comparison on two datasets are shown in Table 3 and Table 4. The proposed method with two mapping surfaces achieves the best accuracy, followed by the proposed method with one mapping surface. The method of 2D-to-2D mapping [14] has the largest error since it cannot deal with the parallel error. The method with nonlinear optimization [13] has the smallest standard deviation of the error. Compared to the FEC method [10], the proposed method with one mapping surface adds two additional calibration points, and the effect on reducing errors is significant. Compared to the method with nonlinear optimization [13], the proposed method with two mapping surfaces has the same calibration procedure but better accuracy. Besides, as targets may lie within different depth ranges, we evaluated all methods in different usage ranges, including 0.5 m, 1 m, and 2 m. Fig. 13 shows that all methods have better accuracy in a small range than in a large range, and the method with two mapping surfaces has better accuracy cross usage ranges than other methods. We can conclude that the proposed model has improved accuracy in both small and large usage range.
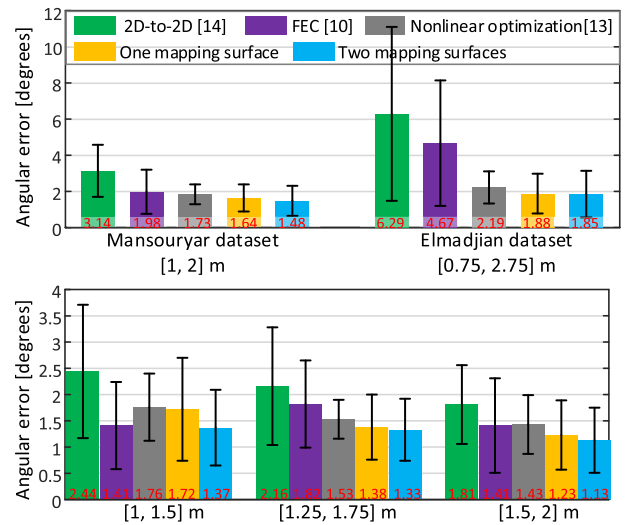
FIGURE 13. Gaze estimation error in different usage ranges. The horizontal coordinate represents the depth interval of the target object. Bars show the mean angular error; error bars indicate standard deviations; numbers on the bottom are the mean angular error in degrees. Top: The usage ranges of the Elmadjian dataset [24] and the Mansouryar dataset [14] are 2 m and 1m, respectively. Bottom: The usage range of 0.5 m on the Mansouryar dataset [14], which means that only three planes are used for calibration and testing.

Fig. 14 shows correlation coefficients between extrapolation errors and gaze estimation methods on the two datasets. Gaze direction estimation methods with multiple calibration planes are relevant, implying their performances are limited by the same underlying factors. For a small usage range, the extrapolation error is the primary error source due to

TABLE 3. Comparison with other methods on the Mansouryar dataset [14].

Subject	2D-to-2D mapping [14]		FEC [10]		Nonlinear optimization [13]		One mapping surface		Two mapping surfaces	
	Avg. err. [°]	Std [°]	Avg. err. [°]	Std [°]	Avg. err. [°]	Std [°]	Avg. err. [°]	Std [°]	Avg. err. [°]	Std [°]
S1	3.40	1.03	3.18	1.57	2.14	0.97	2.03	1.43	1.92	1.46
S2	2.53	1.63	1.06	0.57	2.35	0.91	2.40	1.95	2.28	2.02
S3	4.79	2.35	6.96	5.78	4.43	1.57	3.41	2.56	3.19	2.69
S4	2.57	1.45	0.90	0.48	1.24	0.24	1.49	0.43	1.17	0.63
S5	3.25	0.94	2.76	0.91	2.41	1.16	2.58	0.63	2.42	0.96
S6	2.96	2.04	1.11	0.63	1.25	0.24	1.03	0.49	1.04	0.38
S7	2.55	0.97	1.03	0.42	1.00	0.17	1.15	0.18	0.94	0.26
S8	3.02	1.25	1.90	1.64	1.49	0.26	0.94	0.29	0.84	0.30
S9	2.63	1.34	1.32	1.01	1.43	0.26	1.52	0.20	1.13	0.43
S10	3.30	1.95	1.92	1.75	1.23	0.43	1.10	0.43	1.04	0.41
S11	2.95	1.43	1.19	0.59	1.13	0.37	0.98	0.30	0.88	0.36
S12	2.98	1.47	0.64	0.18	0.83	0.05	0.73	0.25	0.69	0.28
S13	4.26	0.89	2.92	1.21	2.23	1.18	2.40	1.20	2.40	1.16
S14	2.78	1.44	0.83	0.36	1.02	0.12	1.18	0.11	0.77	0.26
Average	3.14	1.44	1.98	1.22	1.73	**0.57**	1.64	0.75	**1.48**	0.83
Calibration	2 planes		1 plane		2 planes		1 plane + 2 points		2 planes	

TABLE 4. Comparison with other methods on the Elmadjian dataset [24]. For the valid six subjects, two eyes are evaluated respectively. "L" and "R" represent the left and right eye.

| Subject	2D-to-2D mapping [14]		FEC [10]		Nonlinear optimization [13]		One mapping surface		Two mapping surfaces	
	Avg. err. [°]	Std [°]	Avg. err. [°]	Std [°]	Avg. err. [°]	Std [°]	Avg. err. [°]	Std [°]	Avg. err. [°]	Std [°]
S1L	6.24	4.63	4.17	2.93	2.16	0.92	1.71	1.05	1.68	1.11
S1R	5.71	5.79	1.66	1.33	0.82	0.22	0.88	0.17	0.65	0.11
S2L	7.30	5.88	4.42	2.77	2.73	1.26	1.47	0.51	1.71	0.85
S2R	7.41	6.06	3.80	2.32	1.15	0.49	1.77	0.81	1.58	0.85
S3L	6.08	5.22	2.48	1.35	1.90	0.79	2.03	1.21	1.64	1.12
S3R	6.55	6.16	3.19	2.13	1.07	0.45	1.48	0.80	1.22	0.66
S7L	6.49	4.90	2.44	1.06	2.11	0.80	1.54	0.70	1.58	0.86
S7R	5.17	3.32	7.32	5.40	1.91	1.06	2.48	1.67	2.71	1.85
S10L	5.39	4.27	4.92	4.64	3.87	1.11	1.16	0.21	1.20	0.23
S10R	5.74	5.05	1.82	0.76	1.17	0.62	1.38	0.68	1.25	0.58
S11L	6.71	3.44	10.41	8.78	3.74	1.26	3.66	3.42	3.76	4.30
S11R	6.74	3.03	9.38	8.13	3.64	1.34	2.94	2.01	3.26	3.00
Average	6.29	4.81	4.67	3.47	2.19	**0.86**	1.88	1.10	**1.85**	1.29
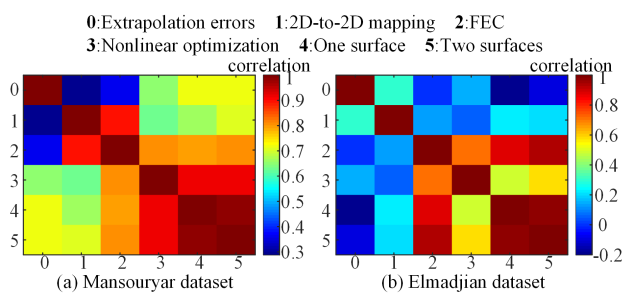
FIGURE 14. Correlation coefficients between extrapolation errors and gaze estimation methods on the two datasets.

the strong correlation between extrapolation errors and the proposed methods (see Fig. 14 (a)). In contrast, for a large usage range, the parallel error seems to be the primary error source due to the weak correlation between extrapolation errors and the proposed methods (see Fig. 14 (b)).

VII. CONCLUSION

In this paper, we propose a regression-based 3D gaze estimation model with high accuracy using a simple, easily calibrated setup in the head-mounted eye tracker. The key idea is to analyze and leverage the implicit assumption of the eyeball pseudo-center. The simulation of visual axes shows that visual axes intersect in a small region rather than at a point, implying the assumption is approximate. The sensitivity analysis of the eyeball pseudo-center reveals that the assumption is acceptable only if the error of the estimated eyeball pseudo-center is small, and the distance between the fixation target and the mapping surface is small. Hence, we propose a gaze direction estimation method with one mapping surface that satisfies the assumption by a high-quality calibration of the eyeball pseudo-center, which only requires two additional calibration points outside the mapping surface. Furthermore, we propose a gaze direction estimation method with two mapping surfaces that replaces the assumption with another mapping surface. Experimental results indicate that although the assumption of the eyeball pseudo-center is met, the accuracy of gaze estimation is still affected, compared to the gaze estimation method that replaces the assumption. Besides, the proposed method with two mapping surfaces has higher accuracy than state-of-the-art methods. Since

extrapolation errors are the basis of error analysis of the proposed model, we should reduce extrapolation errors in further studies. Besides, the robustness of the proposed model should be discussed and improved.

REFERENCES

[1] K. Panetta, Q. Wan, A. Kaszowska, H. A. Taylor, and S. Agaian, "Software architecture for automating cognitive science eye-tracking data analysis and object annotation," *IEEE Trans. Human-Mach. Syst.*, vol. 49, no. 3, pp. 268–277, Jun. 2019.

[2] S. Li and X. Zhang, "Implicit intention communication in human–robot interaction through visual behavior studies," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 4, pp. 437–448, Aug. 2017.

[3] A. T. Duchowski, *Eye Tracking Methodology: Theory and Practice*, 2nd ed. London, U.K.: Springer, 2007.

[4] A. Frisoli, C. Loconsole, D. Leonardis, F. Banno, M. Barsotti, C. Chisari, and M. Bergamasco, "A new gaze-BCI-driven control of an upper limb exoskeleton for rehabilitation in real-world tasks," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 42, no. 6, pp. 1169–1179, Nov. 2012.

[5] S. Li, X. Zhang, and J. D. Webb, "3-D-gaze-based robotic grasping through mimicking human visuomotor function for people with motion impairments," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 12, pp. 2824–2835, Dec. 2017.

[6] D. W. Hansen and Q. Ji, "In the eye of the beholder: A survey of models for eyes and gaze," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 478–500, Mar. 2010.

[7] 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, Dec. 2007.

[8] 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.

[9] 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.

[10] W. W. Abbott and A. A. Faisal, "Ultra-low-cost 3D gaze estimation: An intuitive high information throughput compliment to direct brain–machine interfaces," *J. Neural Eng.*, vol. 9, no. 4, Aug. 2012, Art. no. 046016.

[11] K. Sumi, A. Sugimoto, T. Matsuyama, M. Toda, and S. Tsukizawa, "Active wearable vision sensor: Recognition of human activities and environments," in *Proc. Int. Conf. Informat. Res. Develop. Knowl. Soc. Infrastruct. ICKS*, Mar. 2004, pp. 15–22.

[12] I. Mitsugami, N. Ukita, and M. Kidode, "Estimation of 3D gazed position using view lines," in *Proc. 12th Int. Conf. Image Anal. Process.*, Sep. 2003, pp. 466–471.

[13] D. Su, Y.-F. Li, and H. Chen, "Cross-validated locally polynomial modeling for 2-D/3-D gaze tracking with head-worn devices," *IEEE Trans. Ind. Informat.*, vol. 16, no. 1, pp. 510–521, Jan. 2020.

[14] M. Mansourya, J. Steil, Y. Sugano, and A. Bulling, "3D gaze estimation from 2D pupil positions on monocular head-mounted eye trackers," in *Proc. 9th Biennial ACM Symp. Eye Tracking Res. Appl. ETRA*, 2016, pp. 197–200.

[15] J. Sigut and S.-A. Sidha, "Iris center corneal reflection method for gaze tracking using visible light," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 2, pp. 411–419, Feb. 2011.

[16] 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.

[17] 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.

[18] Y. Cheng, X. Zhang, F. Lu, and Y. Sato, "Gaze estimation by exploring two-eye asymmetry," *IEEE Trans. Image Process.*, vol. 29, pp. 5259–5272, 2020.

[19] P. Kellnhofer, A. Recasens, S. Stent, W. Matusik, and A. Torralba, "gaze360: Physically unconstrained gaze estimation in the wild," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6912–6921.

[20] D. Mardanbegi and D. W. Hansen, "Parallax error in the monocular head-mounted eye trackers," in *Proc. ACM Conf. Ubiquitous Comput. UbiComp*, 2012, pp. 689–694.

[21] J. W. Lee, C. W. Cho, K. Y. Shin, E. C. Lee, and K. R. Park, "3D gaze tracking method using purkinje images on eye optical model and pupil," *Opt. Lasers Eng.*, vol. 50, no. 5, pp. 736–751, May 2012.

[22] K. Takemura, K. Takahashi, J. Takamatsu, and T. Ogasawara, "Estimating 3-D Point-of-Regard in a real environment using a head-mounted eye-tracking system," *IEEE Trans. Human-Machine Syst.*, vol. 44, no. 4, pp. 531–536, Aug. 2014.

[23] L. Świrski and N. A. Dodgson, "A fully-automatic, temporal approach to single camera, glint-free 3D eye model fitting," in *Proc. Pervas. Eye Tracking Mobile Eye-Based Interact. (PETMEI)*, 2013, pp. 1–11.

[24] C. Elmadjian, P. Shukla, A. D. Tula, and C. H. Morimoto, "3D gaze estimation in the scene volume with a head-mounted eye tracker," in *Proc. Workshop Commun. Gaze Interact. COGAIN*, 2018, pp. 1–9.

[25] M. Barz, A. Bulling, and F. Daiber, "Computational modelling and prediction of gaze estimation error for head-mounted eye trackers," *German Res. Center Artif. Intell. (DFKI) Res. Rep.*, vol. 1, p. 10, Jan. 2015.

[26] J. J. Cerrolaza, A. Villanueva, and R. Cabeza, "Study of polynomial mapping functions in video-oculography eye trackers," *ACM Trans. Comput.-Hum. Interact.*, vol. 19, no. 2, p. 10, Jul. 2012.

[27] 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.

[28] D. Atchison and G. Smith, *Optics of the Human Eye*. Oxford, U.K.: Butterworth-Heinemann, 2000.

[29] T. C. Ruch and J. F. Fulton, "Medical physiology and biophysics," *Academic Med.*, vol. 35, no. 11, p. 1067, 1960.

[30] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, "Appearance-based gaze estimation in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 4511–4520.

ZHONGHUA WAN received the B.S. degree in mechanical engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2013. He is currently pursuing the Ph.D. degree in mechanical engineering with the Huazhong University of Science and Technology (HUST), Wuhan, China. His research interests include eye tracking and computer vision.

CAIHUA XIONG (Member, IEEE) received the Ph.D. degree in mechanical engineering from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 1998. From 1999 to 2003, he was with the City University of Hong Kong and The Chinese University of Hong Kong, as a Postdoctoral Fellow, and with Worcester Polytechnic Institute, Worcester, MA, USA, as a Research Scientist. He is a Chang Jiang Professor with HUST. His research interests include biomechatronic prostheses, rehabilitation robotics, and robot motion planning and control. He received the National Science Fund for Distinguished Young Scholars of China.

QUANLIN LI received the B.Eng. degree in mechanical engineering from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2012, where he is currently pursuing the Ph.D. degree in mechanical engineering. His research interests include human–robot interaction and robot motion planning and control.

WENBIN CHEN (Member, IEEE) received the B.S. degree in mechanical engineering from the Wuhan University of Science and Technology, Wuhan, China, in 2004, and the Ph.D. degree in mechatronic engineering from the Huazhong University of Science and Technology (HUST), Wuhan, in 2012. From 2013 to 2014, he was a Research Fellow with the Lincoln School of Computer Science, University of Lincoln, Lincoln, U.K. Since 2015, he has been a Lecturer with the State Key Laboratory of Digital Manufacturing and Equipment Technology, HUST. His research interests include anthropomorphic prosthetic robotics, rehabilitation robotics, and robot motion planning and control.

SHIQIAN WU (Senior Member, IEEE) received the B.Eng. and M.Eng. degrees from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 1985 and 1988, respectively, and the Ph.D. degree from Nanyang Technological University, Singapore, in 2001. He is currently a Professor with the College of Machinery and Automation, Wuhan University of Science and Technology, Wuhan. He was an Assistant Professor, a Lecturer, and an Associate Professor with HUST, from 1988 to 1997. From 2000 to 2014, he was a Research Fellow or Research Scientist with the Agency for Science, Technology and Research, Singapore. He has coauthored the book titled "*Dynamic Fuzzy Neural Networks*" (McGraw-Hill, 2003). He has authored or coauthored over 100 scientific publications (book chapters and journal/conference papers). His current research interests include image processing, pattern recognition, machine vision, fuzzy systems, and neural networks.

• • •

KELVIN KIAN LOONG WONG was born in Singapore, in 1976. He received the B.Eng. degree (Hons.) in mechanical and production engineering from Nanyang Technological University, in 2001, the M.A.I.T. degree in applied information technology from The University of Sydney, in 2003, and the Ph.D. degree in electrical and electronic engineering from The University of Adelaide, in 2009.

He has more than ten years of medical imaging, computational hemodynamics, and simulation modeling experience. From 2006 to 2009, he was doing research work on computerized medical imaging and cardiac flow analysis. Since 2009, he has been involved in computational fluid and solid dynamics. In 2013, he began working on discrete element method with applications in biological systems at The University of Western Australia. He is also an Adjunct Professor with The University of Adelaide. He was the Originator of the spatial game moment concept, in which multi-objective multi-constraint combinatorial optimization problems can be treated as decision-making problems in the game theoretical sense, and solved with high efficiency. In addition, he is the author of two books "*Methods in Research and Development of Biomedical Devices*," and "*Computational Hemodynamics—Theory, Modelling and Applications*." His publications span a diverse range of topics in the medical science and engineering field. He has served as an associate editor and a guest editor for journals in the area of biomedical engineering.