

Self-Calibrating Gaze Estimation With Optical Axes Projection for Head-Mounted Eye Tracking

Hanyuan Zhang , Shiqian Wu , Senior Member, IEEE, Wenbin Chen , Member, IEEE,
Zheng Gao , and Zhonghua Wan 

Abstract—Gaze estimation suffers from burdensome personal calibration or complex all-device calibration. Self-calibrating methods can meet this challenge but depend on scenes and sacrifice accuracy. We propose a flexible and accurate gaze estimation approach calibrated implicitly with potential gaze patterns. By constructing an optical axis projection (OAP) plane and a visual axis projection (VAP) plane simultaneously, the optical axis and the visual axis can be represented as 2-D points, i.e., the OAP and VAP, which have a similarity transformation, indicating the linear consistency of OAP patterns with gaze patterns. Hence, a 3-D gaze estimation model using the OAP as an eye feature to predict the VAP is built. The unknown parameters are calculated separately by linearly aligning OAP patterns to natural and easily detectable gaze patterns. Experimental results show that the proposed gaze estimation approach is more accurate than state-of-the-art head-mounted gaze estimation methods, which require explicit calibration or multiscene saliency.

Index Terms—Eye tracking, gaze estimation, implicit calibration, optical axis (OA), pattern alignment, personal calibration.

I. INTRODUCTION

HEAD-MOUNTED eye gaze tracking is progressively compelling due to its ability to access gaze and interpret attention. It is widely applied to real and mobile environments, such as human–robot interaction [1], virtual reality [2], augmented reality [3], and artificial intelligence [4].

Manuscript received 10 October 2022; revised 13 February 2023; accepted 27 April 2023. Date of publication 16 May 2023; date of current version 19 January 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 52027806, Grant U1913601, and Grant U1913205, and in part by New Initiative Project under Grant 2018TDX06, Wuhan University of Science and Technology. Paper no. TII-22-4231. (Corresponding author: Zhonghua Wan.)

Hanyuan Zhang is with the Key Laboratory of Metallurgical Equipment and Control Technology, Ministry of Education, Wuhan University of Science and Technology, Wuhan, Hubei 430081, China (e-mail: hyzhang@wust.edu.cn).

Shiqian Wu, Zheng Gao, and Zhonghua Wan are with the Institute of Robotics and Intelligent Systems, School of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan, Hubei 430081, China (e-mail: shiqian.wu@wust.edu.cn; gzheng@wust.edu.cn; zhwan@wust.edu.cn).

Wenbin Chen is with the Institute of Rehabilitation and Medical Robotics, State Key Laboratory of Intelligent Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China (e-mail: wbchen@hust.edu.cn).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TII.2023.3276322>.

Digital Object Identifier 10.1109/TII.2023.3276322

For head-mounted eye tracking, most gaze estimation methods are divided into regression-based and model-based [5]. Regression-based methods [6], [7] typically regress pupil centers in eye images to gaze points on a plane using mapping functions. Computing the parameters of mapping functions inevitably faces the personal calibration problem that each user must stare at a succession of predefined markers to capture gaze positions and the corresponding eye images as calibration data before each use. Such calibration is time-consuming and burdensome. Besides, due to the slippage between the headgear and user's head [8], the calibrated mapping relationship will change over time and require repeated personal calibration, interrupting the experiment process. Furthermore, regression-based 3-D gaze estimation requires estimating a larger depth range than 2-D gaze estimation, resulting in more calibration points and more severe calibration problems.

Model-based gaze estimation methods [9] establish a geometric model of eyes, representing the visual axis (VA) as the gaze direction and the optical axis (OA) as the eye axis. In general, personal calibration is significantly reduced to at least one point since only the angle between the OA and VA is unknown and constant. These methods can tolerate slippage as the OA calculation is robust to the headgear slippage. However, these methods require not only complex devices, including multiple eye cameras and infrared LEDs for each eye, but also cumbersome all-device calibration. Devices may need to be readjusted and recalibrated due to individual differences. As a subclass of model-based methods, pupil-contour-based methods [10], [11] assume that the OA coincides with the pupil axis, i.e., the normal through the pupil center, and calculate the pupil axis by back-projecting the elliptical contour of the pupil image. These methods inherit the advantages of model-based methods and simplify devices, i.e., only one camera per eye. However, to calculate the transformation from pupil axes in the eye camera to gaze directions in the scene camera, existing methods still require personal calibration by minimizing the angular error between the real and estimated gaze directions.

Instead of explicit calibration with predefined markers, self-calibrating gaze estimation methods capture calibration data from natural and potential gaze behavior while looking at images, movies, or scenes. These methods are based on the phenomenon that the gaze has some particular positions or patterns (see Fig. 1). Saliency-based methods assume more salient regions are more likely to be noticed, thereby using saliency

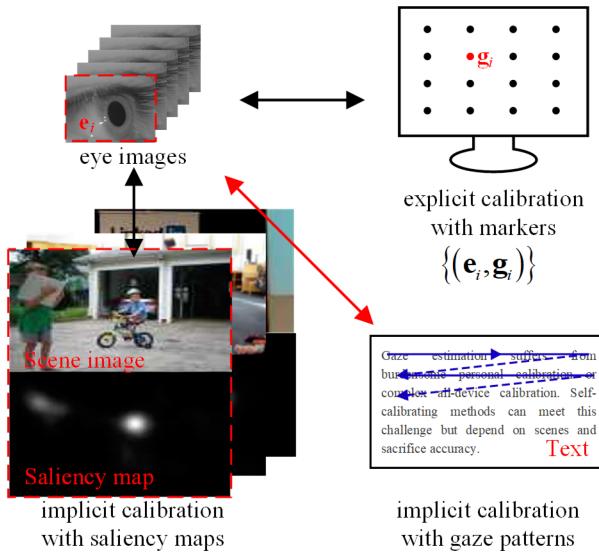


Fig. 1. Comparison of different calibration methods for gaze estimation. Saliency-based methods depend on multiscene saliency maps. This article realizes implicit calibration by pervasive and easily detectable gaze patterns, such as the left-to-right reading pattern.

maps of images or videos to represent gaze probabilities [12], [13]. The calibration problem becomes maximizing the gaze probability. However, one drawback is scene-dependent as some scenes cannot extract valid saliency maps. Another drawback is the sacrifice of accuracy since the underlying assumption that saliency is closely related to the gaze probability is not always satisfied. Moreover, some methods use fixation maps [14] or probable fixation targets [15] instead of saliency maps to represent gaze distribution, but suffer from the same drawbacks. Some methods use gaze patterns to align eye-feature patterns [16] or constrain model parameters [17] but are only suitable for table-mounted rather than head-mounted systems.

To overcome these limitations, we propose a flexible and accurate gaze estimation approach that leverages natural and prominent gaze patterns to calibrate unknown parameters. Inspired by the phenomenon that some gaze patterns are pervasive and easily detectable, such as the left-to-right reading pattern and moving-object pursuit, we adopt gaze patterns and associate eye features as calibration data without explicit calibration. The calibration problem becomes aligning eye-feature patterns with gaze patterns. To simplify and achieve alignment, by constructing the optical axis projection (OAP) plane and the visual axis projection (VAP) plane, we propose a new eye feature, OAP, whose pattern is linearly consistent with the gaze pattern. Specifically, the OAP and VAP have a similarity transformation, including rotation, translation, and scaling. Thus, a pupil-contour-based gaze estimation model using the OAP to predict the VAP is built. Model parameters can be easily calculated by linearly aligning OAP patterns to gaze patterns: OAP and VAP planes are determined by the gaze center; the rotation matrix is calculated by gaze movements; scaling and translation parameters are calibrated by the gaze range. The main contributions of this article are summarized as follows.

- 1) The similarity transformation between the OAP and VAP is revealed by theoretical derivation, implying linear pattern consistency of the OAP and gaze. In contrast, previous methods assume an approximate similarity between some eye features and gaze points.
- 2) An OAP-based 3-D gaze estimation model with few unknown parameters and simple devices is built, combining the advantages of regression-based and model-based methods.
- 3) An implicit calibration method using gaze patterns is proposed, which is natural, accurate, and scene-independent.

The rest of this article is organized as follows. Section II introduces the related works. Section III forms the mathematical problem and overview, followed by the 3-D gaze estimation approach using OAP in Section IV. Experiments are given in Section V. Finally, Section VI concludes this article.

II. RELATED WORKS

A. Gaze Estimation With Implicit Calibration

Reducing calibration complexity, such as decreasing the number of calibration points and simplifying the calibration process, can reduce the burden on users and is desirable [18]. Implicit calibration has received increasing attention in recent years (see Fig. 1). Saliency-based gaze estimation methods were first proposed for table-mounted eye tracking. The gaze probability is represented as the image's saliency map when viewing static images [19] or movies [20]. Saliency-based methods also extend to head-mounted eye tracking, in which saliency maps are extracted from the images of the scene camera. Different gaze estimation methods are proposed, including appearance-based [12], model-based [21], and pupil-contour-based [13], [22]. However, saliency-based methods depend on scenes and undergo low accuracy.

To address the uncertainty of the saliency map and gaze probability, some papers propose new gaze probability/position presentations, including fixation maps [14], probable fixation targets [15], smooth pursuits [23], and mouse-clicked positions [24]. Since the above methods still rely on specific scenes, some methods employ prominent and potential gaze patterns for calibration. As humans have similar gaze patterns when they look at the same stimulus, Alnajar et al. [16] used the gaze patterns of other viewers to determine the gaze pattern of a new viewer. To reduce the person dependence, Lu et al. [25] proposed a nonlinear dimension reduction that reduces the eye appearance to a 2-D space manifold, which is approximately linear to the gaze pattern. To allow free head movement, Wang and Ji [17] implicitly calibrated the angle between the OA and VA by assuming the following four natural constraints:

- 1) complementary gaze constraint of two gaze estimation methods;
- 2) center prior constraint;
- 3) display boundary constraint of the screen region;
- 4) angular constraint of eye parameters.

However, these methods are used for table-mounted rather than head-mounted systems. Moreover, they also suffer from low accuracy due to the approximate relationship between gaze

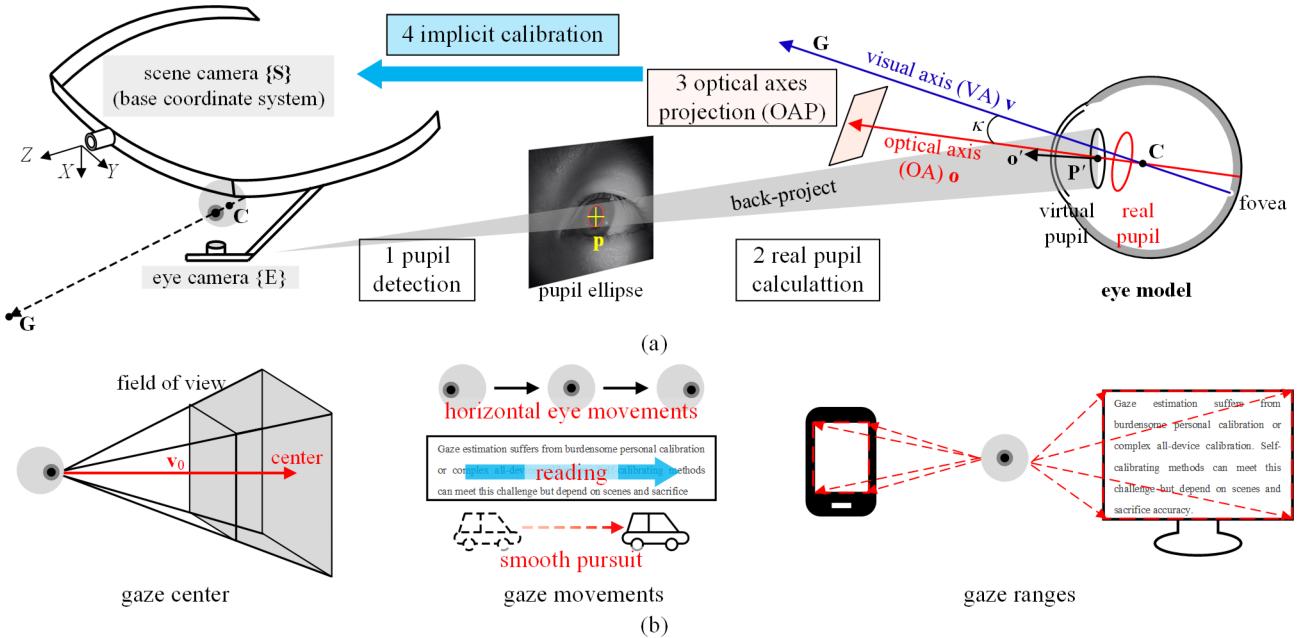


Fig. 2. Proposed self-calibrating gaze estimation approach for head-mounted eye tracking. (a) Schematic diagram of pupil-contour-based gaze estimation. The eye image captured by the eye camera {E} can detect and fit a pupil ellipse, which can be backprojected to the virtual pupil axis vector \mathbf{o}' , and \mathbf{o}' can be rotated to the real pupil axis vector \mathbf{o} using corneal refraction theory. Since the eye OA and VA have a constant angle κ for each eye, the key problem is how to calibrate the transformation from the OA in {E} to the VA in the scene-camera coordinate system {S}. To this end, this article proposes optical axis projection (OAP) and gaze-pattern-based implicit calibration. (b) Easily detectable gaze patterns for implicit calibration. The gaze center means when an eye is looking forward, which can be represented as the mean of all eye directions. Gaze movements refer to the main movement direction of the gaze point, such as the left-to-right saccade of reading and smooth pursuit of moving objects. The gaze range means the observation range on the target object.

patterns and eye movement patterns. In this article, a direct linear relationship between eye-movement patterns and gaze patterns is established for the first time using theoretical derivations, thereby improving the accuracy.

B. Pupil-Contour-Based Gaze Estimation

Pupil-contour-based gaze estimation calculates visual axes from pupil axes, which are derived from pupil contours in eye images using the inverse projection of circles. The achieved pupil is not the real pupil but the virtual pupil due to the cornea refraction. Wan et al. [10] built the optical system of the cornea to calculate real pupil axes from virtual pupil axes. Since the pupil axis is based on the eye-camera coordinate system, gaze estimation requires transforming the coordinate system into the scene camera, introducing six homogeneous transformation parameters of the following two coordinate systems (CS): 1) the translation vector and 2) the rotation matrix. Existing methods calculate this transformation by personal calibration rather than device calibration, considering the pose adjustment of eye cameras. The cost function minimizes angular errors between the estimated and real gaze directions. This nonlinear optimization problem has many algorithms, such as the nonlinear optimization algorithm [26], the two-step nonlinear optimization algorithm [11], the two-step analytical optimization algorithm [10], and the simplified algorithm by ignoring the translation vector [22]. This article calculates the transformation in 2-D projection space for

the first time, thereby converting nonlinear optimization to linear optimization and reducing the number of parameters.

III. PROBLEM FORMATION AND OVERVIEW

Inspired by [10], this article focuses on the following two main problems: 1) gaze estimation solution and 2) implicit calibration.

A. Problem Formation

The schematic diagram of pupil-contour-based gaze estimation for head-mounted eye tracking is shown in Fig. 2(a). The elliptical contour of the pupil image captured by the eye camera {E} is projected backward as an elliptical cone. Combining the pupil image center \mathbf{p} and the eye camera matrix \mathbf{K} , the virtual pupil center \mathbf{P}' is constrained to be on a line, i.e., $\mathbf{EP}' = Z_{\mathbf{P}'} \mathbf{K}^{-1} \hat{\mathbf{p}}$, where $\hat{\mathbf{p}}$ is the homogeneous coordinates of \mathbf{p} and $Z_{\mathbf{P}'}$ is the unknown depth of \mathbf{P}' . Similarly, the virtual pupil normal vector \mathbf{o}' is solved for two feasible solutions, and a unique solution can be filtered out by combining many eye images to fit the eye model [27]. Then, the real pupil normal vector \mathbf{o} in {E} is derived in [10] as $\mathbf{o} = f(\mathbf{o}', \mathbf{EP}')$. Overall, an eye image can derive \mathbf{o} . Since the real pupil axis coincides with the OA, which has a constant angle κ to the VA for each eye, the VA in {S} is represented as $\mathbf{v} = {}_E^S \mathbf{R}(\kappa) \mathbf{o} = \mathbf{R}(\kappa)$ with $\|\mathbf{v}\| = 1$, where $\mathbf{R}(\kappa)$ denotes the rotation from the OA to the VA, and ${}_E^S \mathbf{R}$ is the rotation from the coordinate system {E} to

TABLE I
COMPARISON OF THE TYPICAL GAZE ESTIMATION METHODS WITH THE PROPOSED METHOD

Method	Position	Gaze estimation model	Problem to solve	Calibration
Ours	Head	Pupil-contour-based 3D gaze	2D: OAP \rightarrow VAP	Flexible gaze patterns
Explicit calibration [10]	Head	Pupil-contour-based 3D gaze	3D: OA \rightarrow VA	Explicit predefined markers
Saliency-based [13], [22]	Head	Pupil-contour-based 3D gaze	3D: OA \rightarrow VA	Scene saliency maps
Mouse operations [24]	Table	Appearance-based 2D gaze	Nonlinear gaze mapping	Mouse-clicked positions
Gaze patterns [25]	Table	Appearance-based 2D gaze	2D manifold \rightarrow 2D gaze	Flexible gaze patterns
Gaze patterns [16]	Table	Appearance-based 2D gaze	Gaze sequence matching	Multi-person gaze sequences

$\{\mathbf{S}\}$. The gaze point \mathbf{G} is derived as

$$\mathbf{G} = \mathbf{C} + \lambda \mathbf{R} \mathbf{o} \quad (1)$$

where \mathbf{C} is the eye center and λ is a coefficient. To solve this model, it is necessary to calculate \mathbf{C} and \mathbf{R} , which satisfy

$$\mathbf{v} = \mathbf{R} \mathbf{o} = \frac{\mathbf{C} \mathbf{G}}{\|\mathbf{C} \mathbf{G}\|}. \quad (2)$$

The traditional solution [10] uses explicit calibration to obtain calibration data $\{(\mathbf{o}_i, \mathbf{G}_i)\}_{i=1}^N$ and minimizes the angular deviation of the true and estimated gaze directions by

$$(\mathbf{R}^*, \mathbf{C}^*) = \arg \max_{\mathbf{R}, \mathbf{C}} \sum_{i=1}^N \mathbf{R} \mathbf{o}_i \cdot \frac{\mathbf{C} \mathbf{G}_i}{\|\mathbf{C} \mathbf{G}_i\|}. \quad (3)$$

This solution has the following two limitations: 1) nonlinear optimization is difficult to ensure the optimal value, reducing the accuracy and 2) explicit calibration is time-consuming and burdensome.

B. Overview

To address the above two limitations, we analyze and build the relationship between the OA and VA in 2-D projection space instead of 3-D space to transform the nonlinear optimization (3) into a linear optimization (12). Besides, the parameters of this new model can be calibrated by three flexible and detected gaze patterns without interrupting tasks or looking at multiple salient scenes [see Fig. 2(b)]. Table I compares typical gaze estimation methods with this proposed method. This method belongs to the pupil-contour-based gaze estimation as the existing studies [10], [13], [22], but these methods differ in the following three ways: 1) this method reduces the 3-D homogeneous transformation in [10], [13], [22] between the OA and VA to 2-D projection space so that the nonlinear optimization is transformed into a linear optimization; 2) this method is calibrated by flexible gaze patterns rather than explicit calibration [10] or scene saliency maps [13], [22] and 3) this method uses real pupils as in [10] rather than virtual pupils in [13] and [22]. Moreover, this method is significantly different from the appearance-based methods [16], [24], [25], which are suitable for desktop scenes and calculate 2-D gaze points on a monitor.

IV. GAZE ESTIMATION WITH OAs PROJECTION

A. Similarity Between the OAP and VAP

The relationship between the OAP and VAP is shown in Fig. 3. Let \mathbf{v} and \mathbf{o} denote the VA and OA when the eye is looking

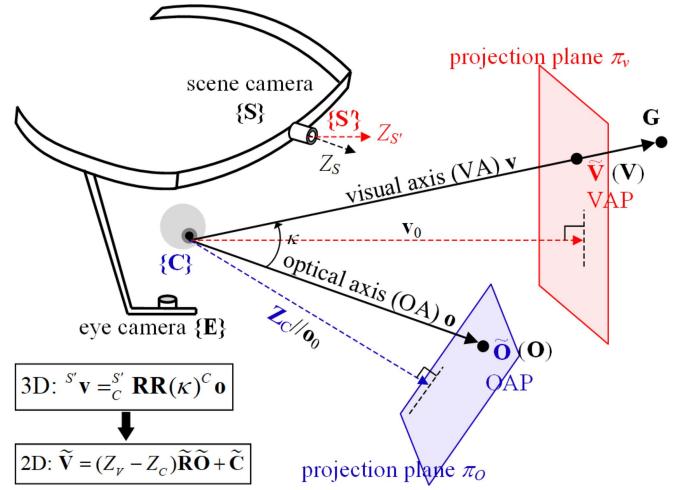


Fig. 3. Reducing the 3-D homogeneous relationship between the OA \mathbf{o} and VA \mathbf{v} to the 2-D relationship between the OAP \mathbf{O} and VAP \mathbf{V} . Given \mathbf{o}_0 , \mathbf{v}_0 , π_o , and π_v , \mathbf{v} intersects with π_v at \mathbf{V} , and \mathbf{o} intersects with π_o at \mathbf{O} . Besides the two basic CS, $\{\mathbf{E}\}$ and $\{\mathbf{S}\}$, two auxiliary CS, $\{\mathbf{C}\}$ and $\{\mathbf{S}'\}$ are also defined. $\tilde{\mathbf{O}}$ and $\tilde{\mathbf{V}}$ are the first 2-D coordinates of ${}^C\mathbf{O}$ and ${}^{S'}\mathbf{V}$, respectively. Note that the marker in the upper left corner of a point or vector indicates the coordinate system.

forward. The VAP plane and OAP plane are π_v and π_o , whose normal vectors are \mathbf{v}_0 and \mathbf{o}_0 , respectively. A VA \mathbf{v} intersects with π_v at \mathbf{V} , and an OA \mathbf{o} intersects with π_o at \mathbf{O} .

To analyze the relationship between \mathbf{O} and \mathbf{V} , we define two base CS and two auxiliary CS. Note that the marker in the upper left corner indicates the CS on which the point or vector is based.

- 1) Eye camera CS $\{\mathbf{E}\}$: Its origin is the optical center of the eye camera. The captured eye images and pupil axes are based on this CS.
- 2) Scene camera CS $\{\mathbf{S}\}$: Its origin is the optical center of the scene camera. Gaze points are based on this CS.
- 3) Eye auxiliary CS $\{\mathbf{C}\}$: Its origin is the eye center \mathbf{C} . Its z -axis satisfies $\mathbf{Z}_C // \mathbf{o}_0$. The OA is

$${}^C\mathbf{o} = w_1 {}^C\mathbf{O} = w_1 \begin{bmatrix} \tilde{\mathbf{O}} \\ 1 \end{bmatrix} \quad (4)$$

where w_1 is the normalized coefficient, and $\tilde{\mathbf{O}} \in \mathbb{R}^2$ denotes the OAP, i.e., the first 2-D coordinates of ${}^C\mathbf{O}$. The distance between \mathbf{C} and π_o is 1.

- 4) Scene camera rectified CS $\{\mathbf{S}'\}$: Its origin is the optical center of the scene camera. Its z -axis satisfies $\mathbf{Z}_{S'} // \mathbf{v}_0$.

The VA can be represented using the VAP $\tilde{\mathbf{V}}$ as

$${}^S \mathbf{v} = \frac{\mathbf{CV}}{\|\mathbf{CV}\|} = \frac{{}^S' \mathbf{V} - {}^S' \mathbf{C}}{\|{}^S' \mathbf{V} - {}^S' \mathbf{C}\|} = w_2 \begin{bmatrix} \tilde{\mathbf{V}} - \tilde{\mathbf{C}} \\ Z_V - Z_C \end{bmatrix} \quad (5)$$

where w_2 is the normalized coefficient, and $\tilde{\mathbf{V}}, \tilde{\mathbf{C}} \in \mathbb{R}^2$ is the first 2-D coordinates of ${}^S' \mathbf{V}, {}^S' \mathbf{C}$.

Given ${}^E \mathbf{o}, \tilde{\mathbf{O}}$ can be calculated by the following OAP equation

$${}^C \mathbf{o} = w_1 \begin{bmatrix} \tilde{\mathbf{O}} \\ 1 \end{bmatrix} = {}^E \mathbf{R} {}^E \mathbf{o} \quad (6)$$

where ${}^E \mathbf{R} = {}^C \mathbf{R}^\top = [{}^E \mathbf{X}_C, {}^E \mathbf{Y}_C, {}^E \mathbf{Z}_C]^\top$, and ${}^E \mathbf{Z}_C = {}^E \mathbf{o}_0 = [Z(1), Z(2), Z(3)]^\top$. Since ${}^E \mathbf{X}_C$ and ${}^E \mathbf{Y}_C$ have no unique solution, we set ${}^E \mathbf{X}_C = \frac{1}{\|Z(1), Z(2)\|_2} [Z(2), -Z(1), 0]^\top$ and ${}^E \mathbf{Y}_C = {}^E \mathbf{Z}_C \times {}^E \mathbf{X}_C$. Then, $\tilde{\mathbf{O}} = \frac{1}{\alpha(3)} [\mathbf{o}(1), \mathbf{o}(2)]^\top$.

Then, \mathbf{o} can be rotated to \mathbf{v} by

$${}^S' \mathbf{v} = {}^S' \mathbf{R} {}^C \mathbf{v} = {}^S' \mathbf{R} \mathbf{R}(\kappa) {}^C \mathbf{o} \quad (7)$$

where $\mathbf{R}(\kappa)$ represents the constant rotation from the OA to the VA for one eye. Let $\mathbf{R} = {}^S' \mathbf{R} \mathbf{R}(\kappa)$. If ${}^C \mathbf{o} = \mathbf{Z}_C$, then ${}^S' \mathbf{v} = \mathbf{Z}_{S'}$. Thus, $\mathbf{R} = \begin{bmatrix} \tilde{\mathbf{R}} & \mathbf{0} \\ \mathbf{0}^\top & 1 \end{bmatrix}$. Substituting (4) and (5) into (7) obtains

$$w_2 \begin{bmatrix} \tilde{\mathbf{V}} - \tilde{\mathbf{C}} \\ Z_V - Z_C \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{R}} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix} w_1 \begin{bmatrix} \tilde{\mathbf{O}} \\ 1 \end{bmatrix} = w_1 \begin{bmatrix} \tilde{\mathbf{R}} \tilde{\mathbf{O}} \\ 1 \end{bmatrix}. \quad (8)$$

This equation can derive the similarity transformation between the OAP and VAP as

$$\tilde{\mathbf{V}} = (Z_V - Z_C) \tilde{\mathbf{R}} \tilde{\mathbf{O}} + \tilde{\mathbf{C}} \quad (9)$$

where $\tilde{\mathbf{R}}, \tilde{\mathbf{C}}$, and $(Z_V - Z_C)$ represent rotation, translation, and scaling, respectively. Thus, this similarity indicates the linear consistency between the OAP pattern and VAP pattern, i.e., the gaze pattern.

B. Gaze Estimation With OAP

Since the OA of the scene camera is not parallel to the forward direction of the eye, the coordinate system needs to be rotated from $\{\mathbf{S}'\}$ to $\{\mathbf{S}\}$ as follows:

$${}^S \mathbf{V} = {}_{S'}^S \mathbf{R} {}^S' \mathbf{V} = {}_{S'}^S \mathbf{R} \begin{bmatrix} \tilde{\mathbf{V}} \\ Z_V \end{bmatrix} \quad (10)$$

where ${}_{S'}^S \mathbf{R} = [{}^S \mathbf{X}_{S'}, {}^S \mathbf{Y}_{S'}, {}^S \mathbf{Z}_{S'}]$. ${}^S \mathbf{Z}_{S'}$ is the normal vector of π_v . ${}^S \mathbf{Y}_{S'} = {}^S \mathbf{Z}_{S'} \times {}^S \mathbf{X}_S$. ${}^S \mathbf{X}_{S'} = {}^S \mathbf{Y}_{S'} \times {}^S \mathbf{Z}_{S'}$.

Using the OAP as a new eye feature, the similarity transformation (9) can be generalized to any depth to obtain the VA as

$$\mathbf{G} = {}_{S'}^S \mathbf{R} \begin{bmatrix} \tilde{\mathbf{G}} \\ Z_G \end{bmatrix}, \tilde{\mathbf{G}} = (Z_G - Z_C) \tilde{\mathbf{R}} \tilde{\mathbf{O}} + \tilde{\mathbf{C}}. \quad (11)$$

This function can intersect the target surface or another VA to obtain a 3-D gaze point. The unknown parameters, namely $\tilde{\mathbf{R}}, \tilde{\mathbf{C}}$, and Z_C , have four degrees of freedom, which are less than six degrees of freedom in (1). They can be calibrated by at least two calibration points $\{(\mathbf{o}_i, \mathbf{G}_i)\}_{i=1}^N, N \geq 2$ with the following

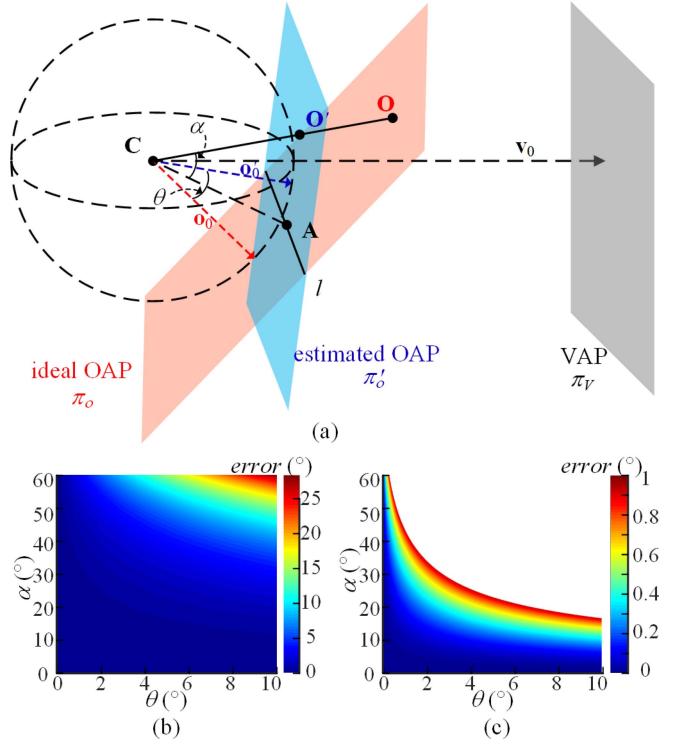


Fig. 4. Sensitivity analysis of an OAP. (a) Relationship between the ideal OAP plane π_o with \mathbf{o}_0 and the estimated OAP plane π'_o with \mathbf{o}'_0 . θ is the angle between \mathbf{o}_0 and \mathbf{o}'_0 , and α is the angle between \mathbf{CA} and \mathbf{CO} . (b) Relationship between error and θ, α . (c) Only show the range of error to $[0, 1^\circ]$.

optimization function:

$$(Z_C^*, \tilde{\mathbf{R}}^*, \tilde{\mathbf{C}}^*) = \arg \min_{Z_C, \tilde{\mathbf{R}}, \tilde{\mathbf{C}}} \sum_{i=1}^N \left\| \tilde{\mathbf{G}}_i - (Z_G - Z_C) \tilde{\mathbf{R}} \tilde{\mathbf{O}}_i - \tilde{\mathbf{C}} \right\|^2. \quad (12)$$

This function can be solved to the analytical solution with the Umeyama algorithm [28] to avoid nonlinear optimization (3). Next, we propose a self-calibrating method to calculate the parameters separately.

Furthermore, when the test depth is the same as the calibration depth, (11) is degenerated to a 2-D gaze estimation function with two scaling parameters to improve gaze estimation accuracy

$$\mathbf{G} = {}_{S'}^S \mathbf{R} \begin{bmatrix} \tilde{\mathbf{G}} \\ Z_G \end{bmatrix}, \tilde{\mathbf{G}} = \begin{bmatrix} d_x & d_y \end{bmatrix} \tilde{\mathbf{R}} \tilde{\mathbf{O}} + \tilde{\mathbf{C}}. \quad (13)$$

Overall, the VA equation (11) is used for 3-D gaze estimation, and (13) is used for 2-D gaze estimation.

C. Sensitivity Analysis of OAP

Since the OAP plane is determined by \mathbf{o}_0 , it is necessary to discuss the sensitivity analysis of \mathbf{o}_0 on gaze estimation. The estimated \mathbf{o}'_0 deviates from the ideal \mathbf{o}_0 , and the ideal OAP plane π_o and the estimated OAP plane π'_o are tangent planes of the ball whose origin is \mathbf{C} and radius is 1 [see Fig. 4(a)]. The ideal OA \mathbf{CO} intersects with π'_o at \mathbf{O}' . Thus, calibration will align \mathbf{O}' on π'_o and \mathbf{V} on π_v , which is equivalent to aligning \mathbf{O}' and \mathbf{O} on

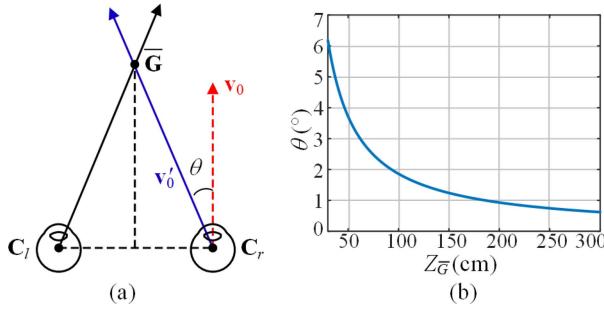


Fig. 5. Calculate the base OA \mathbf{o}_0 with the gaze center when looking forward. (a) Relationship between the ideal VA \mathbf{v}_0 and the estimated VA \mathbf{v}'_0 . (b) Relationship between θ and $Z_{\bar{G}}$.

π_o . As the relationship between CO' and CO is antisymmetric about the intersection line l of the two planes, the points on these two planes should align around l . Here we consider that CO is in the plane constructed by \mathbf{o}_0 and \mathbf{o}'_0 , which intersects l at point A. When CO is not this case, the analysis is similar. Let θ denote the angle between \mathbf{o}_0 and \mathbf{o}'_0 , and let α denote the angle between CA and CO. Thus

$$\begin{aligned} \|\Delta \tilde{\mathbf{O}}\| &= \|\mathbf{AO}\| - \|\mathbf{AO}'\| \\ \|\mathbf{AO}\| &= \tan(\alpha + \frac{\theta}{2}) - \tan(\frac{\theta}{2}) = \frac{\sin \alpha}{\cos \frac{\theta}{2} \cos(\alpha + \frac{\theta}{2})} . \\ \|\mathbf{AO}'\| &= \tan(\alpha - \frac{\theta}{2}) + \tan(\frac{\theta}{2}) = \frac{\sin \alpha}{\cos \frac{\theta}{2} \cos(\alpha - \frac{\theta}{2})} \end{aligned} \quad (14)$$

This equation can be simplified to

$$\|\Delta \tilde{\mathbf{O}}\| = \frac{2 \sin^2 \alpha \sin \frac{\theta}{2}}{\cos \frac{\theta}{2} (\cos^2 \alpha - \sin^2 \frac{\theta}{2})} . \quad (15)$$

Assuming that the unknown parameters, namely $\tilde{\mathbf{R}}$, $\tilde{\mathbf{C}}$, and Z_C , are accurate, gaze deviation is $\Delta \tilde{\mathbf{G}} = (Z_G - Z_C) \tilde{\mathbf{R}} \Delta \tilde{\mathbf{O}}$. Since $Z_G \gg Z_C$, the angular error of gaze estimation is

$$\text{error} = \arctan \frac{\|\Delta \tilde{\mathbf{G}}\|}{Z_G} \approx \arctan \|\Delta \tilde{\mathbf{O}}\| . \quad (16)$$

The relationship between error and θ and α shown in Fig. 4(b) and (c) indicates that error increases with θ and α and is more sensitive to θ than α . Besides, error may be larger than 1° , such as when $\alpha = 30^\circ$, $\theta = 4^\circ$. In conclusion, error is sensitive to θ , and decreasing θ can increase gaze estimation accuracy.

D. Implicit Calibration With Gaze Patterns

Based on the linear pattern consistency between the OAP and VAP, the unknown parameters can be calculated separately by aligning the OAP to flexible and natural gaze patterns [see Fig. 2(b)], which are detected from the eye data captured by the eye camera and the scene images captured by the scene camera.

- 1) Calculate the base OA \mathbf{o}_0 with the gaze center, i.e., when an eye is looking forward along the face direction. Due to the convergence of two eyes, the two eyes' gaze directions intersect at a gaze point \bar{G} in the middle of both eyes and are not parallel to the face direction [see Fig. 5(a)]. Thus, the estimated VA satisfies $\mathbf{v}'_0 // \bar{G}$. The angle between

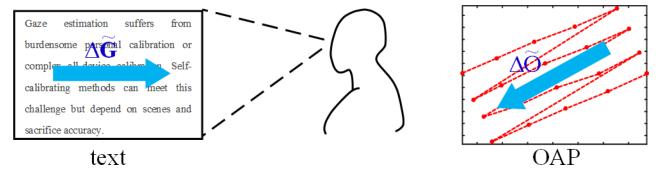


Fig. 6. Gaze movements and OAP movements when reading.

\mathbf{v}'_0 and the ideal VA \mathbf{v}_0 is equal to the error angle θ between \mathbf{o}'_0 and \mathbf{o}_0 , which satisfies

$$\theta = \arctan \frac{\|\mathbf{C}_l \mathbf{C}_r\|}{2Z_{\bar{G}}} . \quad (17)$$

Section IV-C shows that θ should be decreased to improve gaze estimation accuracy. Fig. 5(b) shows the relationship between θ and $Z_{\bar{G}}$ when the two-eye distance $\|\mathbf{C}_l \mathbf{C}_r\|$ is set to a typical value 6.5 cm. The farther $Z_{\bar{G}}$, the smaller θ . If $Z_{\bar{G}} > 200$ cm, then $\theta < 1^\circ$. Therefore, although error is sensitive to θ , calculating \mathbf{o}_0 by looking to the far center can decrease θ and increase gaze estimation accuracy. Moreover, since free-viewing gaze points obey a central normal distribution [19], \mathbf{o}_0 can also be represented as the mean value of all OAs. This operation only requires accumulating eye data without gaze points.

- 2) Calculate $\tilde{\mathbf{R}}$ with common and easily detectable gaze movements, i.e., the main movement direction of the gaze point, such as the left-to-right saccade of reading and smooth pursuit of moving objects. The actual gaze position \mathbf{G} is difficult to obtain without calibration, but the gaze movement $\Delta \tilde{\mathbf{G}}$ is easy to estimate from scenes. For instance, when a person looks at a text for a while, he is likely to be reading. For reading a normal text, $\Delta \tilde{\mathbf{G}}$ is the main direction, i.e., from left to right, and the OAP movement $\Delta \tilde{\mathbf{O}}$ can be detected by fitting the difference of \mathbf{O} (see Fig. 6). The VA equation (11) can derive the relationship between gaze movements and OAP movements, i.e., $\Delta \tilde{\mathbf{G}} // \tilde{\mathbf{R}} \Delta \tilde{\mathbf{O}}$. Hence, by extracting $\Delta \tilde{\mathbf{G}}$ and $\Delta \tilde{\mathbf{O}}$ during the scene with prominent gaze movements, $\tilde{\mathbf{R}}$ can be solved. Moreover, eyes can easily perform horizontal or vertical movements to calculate $\tilde{\mathbf{R}}$ even without a scene.
- 3) Calculate $d_x, d_y, \tilde{\mathbf{C}}$ with the gaze range, which refers to the observation range of the eye on the target object when performing a task. The gaze range is typically the valid boundary of the target object. For instance, the gaze range when reading is the text range. With the gaze range, 2-D gaze estimation function (13) is valid and reformulated with $\mathbf{B} = \tilde{\mathbf{R}} \tilde{\mathbf{O}} \in \mathbb{R}^2$ as

$$\tilde{\mathbf{G}} = [\text{diag}(\mathbf{B}) \quad \mathbf{I}] \begin{bmatrix} d_x \\ d_y \\ \tilde{\mathbf{C}} \end{bmatrix} \quad (18)$$

where $\text{diag}(\mathbf{B})$ is the diagonal matrix of \mathbf{B} . Suppose the diagonal points of the gaze range are $\tilde{\mathbf{G}}_1$ and $\tilde{\mathbf{G}}_2$, and the corresponding points for \mathbf{B} are \mathbf{B}_1 and \mathbf{B}_2 . Then, the

Algorithm 1: OAP-Based Gaze Estimation.

Input: The base OA \mathbf{o}_0 when an eye is looking forward.
An OA set $\mathcal{O}_h = \{\mathbf{o}_i\}$ with detected gaze movement $\Delta \mathbf{G}$. An OA set $\mathcal{O}_{range} = \{\mathbf{o}_i\}$ when the diagonal points of the gaze range are $\tilde{\mathbf{G}}_1, \tilde{\mathbf{G}}_2$. An instance \mathbf{o} .

Output: The rotation matrix $\tilde{\mathbf{R}}$, translation vector $\tilde{\mathbf{C}}$, and scaling coefficients d_x, d_y, Z_C . The gaze point \mathbf{G} .

```

begin
    // 1 OAP transformation
     $\overset{C}{E}\mathbf{R} = transformation(\mathbf{o}_0)$  // Eq. (6)
     $\mathbf{o}_i = \overset{C}{E}\mathbf{R}\mathbf{o}_i, \tilde{\mathbf{O}}_i = \frac{1}{\mathbf{o}_i(3)}[\mathbf{o}_i(1), \mathbf{o}_i(2)]^\top$ 
    // 2 Implicit calibration
     $\Delta \tilde{\mathbf{O}} = fitDirection(\mathcal{O}_h)$ 
     $\tilde{\mathbf{R}} = rotateDirection(\Delta \tilde{\mathbf{O}}, \Delta \tilde{\mathbf{G}})$ 
     $[\mathbf{B}_1, \mathbf{B}_2] = calculateRange(\tilde{\mathbf{R}} \cdot \mathcal{O}_{range})$ 
     $(d_x, d_y, Z_C, \tilde{\mathbf{C}}) = align(\mathbf{B}_1, \mathbf{B}_2, \tilde{\mathbf{G}}_1, \tilde{\mathbf{G}}_2)$  // Eq. (19)
    // 3 Gaze estimation
     $\mathbf{o} = \overset{C}{E}\mathbf{R}\mathbf{o}, \tilde{\mathbf{O}} = \frac{1}{\mathbf{o}(3)}[\mathbf{o}(1), \mathbf{o}(2)]^\top$ 
    if  $Z_G = Z_{range}$  then
         $\tilde{\mathbf{G}} = \begin{bmatrix} d_x \\ d_y \end{bmatrix} \tilde{\mathbf{R}} \tilde{\mathbf{O}} + \tilde{\mathbf{C}}$  // Eq. (13)
    else
         $\tilde{\mathbf{G}} = (Z_G - Z_C) \tilde{\mathbf{R}} \tilde{\mathbf{O}} + \tilde{\mathbf{C}}$  // Eq. (11)
     $\mathbf{G} = \overset{S}{S'}\mathbf{R} \begin{bmatrix} \tilde{\mathbf{G}} \\ Z_G \end{bmatrix}$ 

```

function can be solved to the analytical solution as

$$\begin{bmatrix} d_x \\ d_y \\ \mathbf{C} \end{bmatrix} = \begin{bmatrix} \text{diag}(\mathbf{B}_1) & \mathbf{I} \\ \text{diag}(\mathbf{B}_2) & \mathbf{I} \end{bmatrix}^\dagger \begin{bmatrix} \tilde{\mathbf{G}}_1 \\ \tilde{\mathbf{G}}_2 \end{bmatrix}. \quad (19)$$

The parameters Z_C and $\tilde{\mathbf{C}}$ in the VA equation (11) can be calculated using similar calculations as above.

The above gaze patterns are performed naturally and are detected easily without interrupting tasks. The proposed gaze estimation algorithm is detailed in Algorithm 1.

V. EXPERIMENTAL RESULTS

Experiments are undertaken on a publicly accessible dataset for head-mounted gaze estimation, the Elmadjian dataset [11], which was captured by the modified Pupil Labs eye tracker that includes two eye cameras with 480p resolution at 30 Hz and an Intel Realsense R200 RGB-D camera as the scene camera. This dataset was recorded by reading-like experiments from 11 subjects (five females) aged between 22 to 35 years, using five observation planes including 0.75 m, 1.25 m, 1.75 m, 2.25 m, and 2.75 m. These planes are denoted as D1, D2, D3, D4, and D5, respectively. The data of Subjects 6 and 8 are discarded due to invalidation. Each plane has 4×5 calibration points and 3×4 test points, shown orderly along horizontal lines (see the black dashed lines in Fig. 7 BOTTOM). The size of the five planes varies with depth to ensure that their field of view is essentially the same. While staring at a point, pupil image centers, the vectors of virtual pupil axes, and 3-D gaze point coordinates were

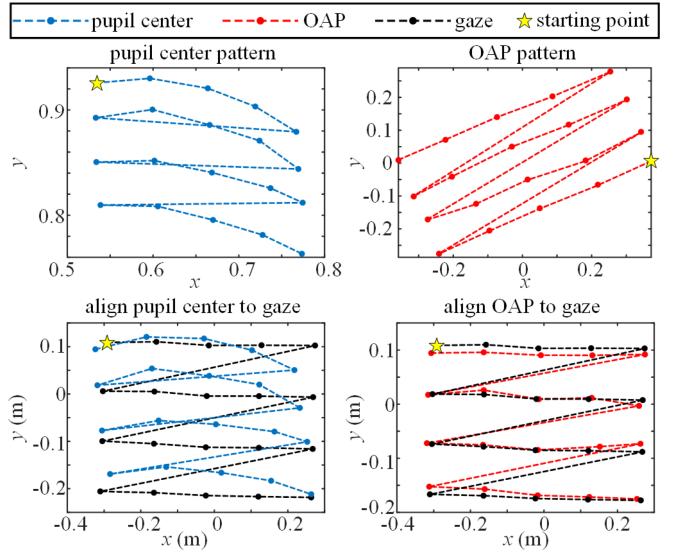


Fig. 7. Comparison of the pupil center with the OAP as input features to align gaze patterns. The data are from the left eye of Subject 1 looking at the D1 plane. The dashed lines indicate the transfer between points. TOP: Raw pupil center and OAP patterns. BOTTOM: aligned pupil center and OAP patterns.

recorded. By the method in [10], real pupil axes are calculated, and its noise and outliers are filtered. We first evaluated the proposed OAP patterns and gaze estimation method. Then, we compared the proposed method with state-of-the-art methods.

A. Evaluation of OAP Patterns

The benefit of the OAP as an eye feature is the similarity relationship with gaze points, implying that OAP patterns and gaze patterns are identical. To clarify this, we compare the pupil center with the OAP as input features to align gaze patterns using the same function (9). Fig. 7 shows that the pupil center pattern is nonlinear with the gaze pattern, but the OAP pattern is linear with the gaze pattern. By testing all subjects' data after calibrating each plane explicitly, the angular errors of gaze estimation using the pupil center and OAP are $1.95 \pm 0.25^\circ$ (mean \pm standard deviation) and $1.33 \pm 0.37^\circ$, respectively. Thus, the OAP is significantly better than the pupil center in terms of aligning gaze patterns.

B. Evaluation of 3-D Gaze Estimation

The proposed gaze estimation approach needs to be validated by ablation experiments. Based on this dataset, the unknown parameters are calibrated with three detected gaze patterns separately rather than the exact gaze points: \mathbf{o}_0 is the mean of OAs when viewing the farthest plane; $\tilde{\mathbf{R}}$ is calculated by horizontal gaze movements when viewing one or more horizontal lines; d_x, d_y , and $\tilde{\mathbf{C}}$ are calculated by the gaze range, i.e., the boundary of the viewing plane. Through the control variables technique, the components of the proposed approach are validated by comparing the angular errors on all planes.

First, the effect of horizontal gaze movements on gaze estimation is analyzed. Two solutions are tested, i.e., $\tilde{\mathbf{R}}$ is calculated by

TABLE II
INFLUENCE OF HORIZONTAL GAZE MOVEMENTS ON GAZE ANGULAR ERRORS

Subject	One horizontal line				Four horizontal lines			
	Case 1		Case 2		Case 1		Case 2	
	Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.
S1L	1.03	0.29	1.08	0.24	1.02	0.29	1.03	0.29
S1R	0.81	0.25	1.35	0.18	0.92	0.38	0.81	0.23
S2L	1.31	0.20	1.29	0.19	1.23	0.25	1.29	0.19
S2R	1.17	0.12	1.17	0.13	1.46	0.20	1.37	0.19
S3L	1.41	0.68	1.42	0.73	1.61	0.53	1.41	0.66
S3R	0.86	0.31	0.80	0.30	1.22	0.33	0.89	0.31
S4L	1.91	0.76	2.03	0.70	1.96	0.75	1.95	0.76
S4R	1.43	0.34	1.33	0.35	1.35	0.34	1.54	0.43
S5L	1.75	0.66	2.37	1.11	2.25	0.92	2.38	1.12
S5R	1.75	0.38	1.52	0.68	1.52	0.70	1.53	0.69
S7L	1.67	0.77	1.31	0.95	2.18	0.70	1.31	0.96
S7R	1.74	0.70	1.79	0.65	2.05	0.58	1.98	0.62
S9L	1.56	0.33	1.60	0.33	1.21	0.11	1.69	0.33
S9R	1.07	0.24	0.99	0.18	1.14	0.15	1.20	0.27
S10L	1.49	0.36	1.60	0.32	1.94	0.38	1.47	0.36
S10R	1.77	0.83	1.67	0.86	1.67	0.83	1.72	0.83
S11L	1.18	0.42	1.77	0.38	1.62	0.38	1.42	0.38
S11R	2.49	0.77	2.00	0.52	2.19	0.66	1.99	0.47
Average	1.47	0.47	1.51	0.49	1.58	0.47	1.50	0.50

Note: The unit of angular errors is degrees.

TABLE III
INFLUENCE OF GAZE RANGES ON GAZE ANGULAR ERRORS

Subject	ran1-d1		ran1-d2		ran5-d1		ran5-d2	
	Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.
S1L	1.31	0.66	7.23	5.25	1.04	0.15	1.03	0.29
S1R	1.96	0.43	6.76	3.85	0.92	0.18	0.81	0.25
S2L	2.25	1.53	7.03	5.69	1.47	0.21	1.31	0.20
S2R	2.18	1.10	6.35	3.44	1.51	0.17	1.17	0.12
S3L	3.32	1.59	9.01	5.51	1.75	0.63	1.41	0.68
S3R	2.78	2.03	7.78	5.93	1.18	0.34	0.86	0.31
S4L	5.97	3.87	8.18	4.15	1.95	0.62	1.91	0.76
S4R	3.15	1.11	6.81	4.22	1.74	0.34	1.43	0.34
S5L	2.52	0.96	7.65	4.10	1.50	0.71	1.75	0.66
S5R	2.96	0.99	6.92	3.25	1.64	0.31	1.75	0.38
S7L	1.97	1.35	7.08	2.69	1.97	1.01	1.67	0.77
S7R	2.53	2.09	6.89	3.89	1.89	0.41	1.74	0.70
S9L	4.54	1.83	8.23	5.15	1.50	0.22	1.56	0.33
S9R	3.77	1.80	6.44	4.91	1.29	0.11	1.07	0.24
S10L	3.23	3.46	8.42	6.67	1.56	0.29	1.49	0.36
S10R	3.54	1.09	8.73	3.78	1.93	0.80	1.77	0.83
S11L	3.12	2.60	7.76	7.91	1.47	0.31	1.18	0.42
S11R	5.08	2.77	7.35	4.42	2.64	0.87	2.49	0.77
Average	3.12	1.74	7.48	4.71	1.61	0.43	1.47	0.47

Note: The unit of angular errors is degrees.

fitting the direction of one or four horizontal lines. Each solution tests two cases, i.e., the horizontal lines in D2 (case 1) or D4 (case 2) calibration plane. Table II shows that the four cases do not have a significant difference. Thus, $\tilde{\mathbf{R}}$ is not sensitive to the data of horizontal gaze movements and can be calibrated flexibly, such as reading multiline texts or active horizontal eye movements.

Second, the influence of gaze ranges on gaze estimation is evaluated (see Table III). Let “ran1” denote using the range of one calibration plane with the VA equation (11) and “ran5” denote using five calibration planes with 2-D gaze estimation (13). Let “d1” and “d2” denote using one or two scale parameters in (11) and (13). Thus, four cases can be combined. Using a gaze range for each plane is significantly better than using a gaze range for only one plane. Besides, one scale parameter is enough for “ran1,” but two scale parameters are more suitable for “ran5.” Thus, (13) is used when the gaze is at the gaze range depth;

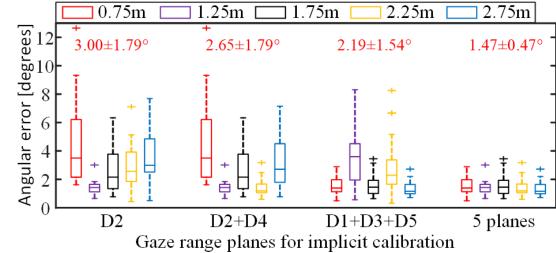


Fig. 8. Influence of gaze range planes on gaze angular errors at different depths. The red numbers at the top are angular errors in degrees (mean \pm standard deviation).

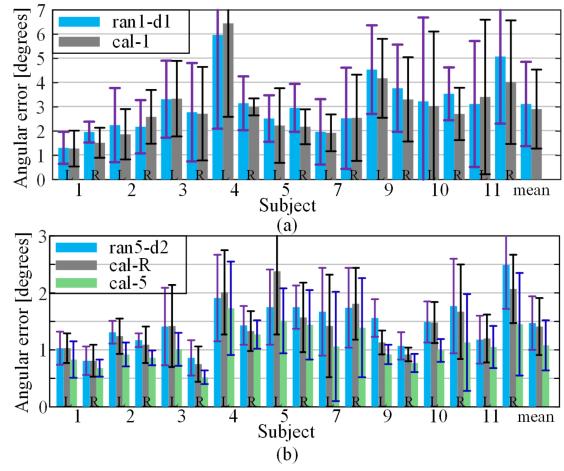


Fig. 9. Comparison of explicit and implicit calibration on gaze estimation. Bars show mean errors across all test planes; error bars indicate standard deviations. The labels at the bottom, “L” and “R,” represent the left and right eyes, respectively. (a) Compare “ran1-d1” and “cal-1.” (b) Compare “ran5-d2,” “cal-R,” and “cal-5.”

otherwise, (11) is used. Fig. 8 shows the effect of using different numbers of gaze range planes on gaze estimation accuracy at different depths. The more gaze range planes, the better the gaze estimation.

Finally, explicit calibration and implicit calibration on gaze estimation are compared. Let “cal-1” denote using one calibration plane to calculate (11) and “cal-5” denote using five calibration planes to calculate (13). As a comparison, let “cal-R” denote the case that $\tilde{\mathbf{R}}$ is calibrated explicitly but other parameters in (13) are calibrated implicitly. Fig. 9(a) shows that “cal-1” and “ran1-d1” are not significantly different. Comparing “ran5-d2” and “cal-R” [see Fig. 9(b)] indicates that horizontal eye movements can produce the same calibration effect as explicit calibration points. Comparing “ran5-d2” and “cal-5” [see Fig. 9(b)] shows that explicit calibration is better than implicit calibration, but explicit calibration points are as large as $20 \times 5 = 100$. Therefore, implicit calibration with gaze patterns can approach the accuracy of explicit calibration and is a suitable alternative to explicit calibration.

C. Comparison With Other Methods

Using the same dataset, several state-of-the-art 3-D gaze estimation methods for head-mounted eye tracking are calibrated explicitly and tested as follows.

TABLE IV
COMPARISON WITH OTHER METHODS USING EXPLICIT CALIBRATION

Subject	2D-to-2D Mapping [26]		Virtual Pupil [11]		Gaze Vector Mapping [29]		Real Pupil [10]		OAP	
	Avg. err.	Std.	Avg. err.	Std.	Avg. err.	Std.	Avg. err.	Std.	Avg. err.	Std.
S1L	6.24	4.64	10.34	0.71	2.16	0.92	1.53	0.76	1.03	0.29
S1R	5.71	5.78	9.96	0.45	0.82	0.22	2.55	0.27	0.81	0.25
S2L	7.30	5.88	10.44	0.98	2.73	1.26	2.13	1.08	1.31	0.20
S2R	7.41	6.06	10.27	0.59	1.15	0.49	2.09	1.29	1.17	0.12
S3L	6.08	5.22	9.28	1.06	1.90	0.79	2.73	0.78	1.41	0.68
S3R	6.55	6.15	9.16	0.47	1.07	0.46	1.72	0.70	0.86	0.31
S4L	6.73	3.82	9.91	1.75	2.79	1.72	4.49	2.51	1.91	0.76
S4R	7.46	3.85	11.65	2.29	2.79	2.02	3.66	2.03	1.43	0.34
S5L	7.28	2.26	8.96	1.21	4.82	1.50	2.58	1.67	1.75	0.66
S5R	7.38	3.43	9.04	1.12	2.36	0.95	2.97	0.54	1.75	0.38
S7L	6.49	4.90	9.52	1.30	2.11	0.80	1.70	0.58	1.67	0.77
S7R	5.17	3.32	9.71	1.10	1.91	1.06	2.28	2.03	1.74	0.70
S9L	6.97	3.65	9.53	0.87	4.17	1.61	2.62	1.00	1.56	0.33
S9R	6.55	5.23	10.18	1.12	1.54	0.63	2.46	1.15	1.07	0.24
S10L	5.39	4.27	10.33	0.63	3.87	1.11	1.81	1.55	1.49	0.36
S10R	5.74	5.05	9.61	0.68	1.17	0.62	1.96	0.93	1.77	0.83
S11L	6.71	3.44	9.40	0.65	3.74	1.26	2.62	1.09	1.18	0.42
S11R	6.74	3.03	10.56	2.23	3.64	1.34	4.16	1.11	2.49	0.77
Average	6.55	4.44	9.88	1.07	2.49	1.04	2.56	1.17	1.47	0.47
Calibration	Two planes		One planes		Two planes		Two planes		Gaze patterns	

Note: The unit of angular errors is degrees.

- 1) *2D-to-2D Mapping*: This pupil-center-based gaze estimation [26] calibrates the regression function from pupil centers to 2-D mapping points by multiple calibration planes.
- 2) *Gaze Vector Mapping*: This pupil-center-based gaze estimation [29] calculates the eyeball center and the regression function from pupil centers to gaze vectors by constrained nonlinear optimization.
- 3) *Virtual Pupil*: This pupil-contour-based gaze estimation [11] treats virtual pupil axes as gaze directions. The six parameters of homogeneous transformation between two CS are calibrated by a two-step iterative nonlinear optimization.
- 4) *Real Pupil*: This recent pupil-contour-based gaze estimation [10] employs a two-step linear optimization to calibrate the 3-D eyeball center and the rotation matrix from the OA in the eye camera to the VA in the scene camera.

The comparison results in Table IV show that the proposed method not only has better accuracy than other methods, but also can be implicitly calibrated. Moreover, all methods are evaluated and compared in different usage ranges, including [0.75, 2.75] m, [0.75, 1.75] m, [1.25, 2.25] m, and [1.75, 2.75] m (see Fig. 10). For the last three cases, the testing planes are reduced to three. The middle calibration plane is used by methods with one calibration plane, whereas the first and last calibration planes are used by methods with two calibration planes. When the usage range is reduced, the angular errors of explicit calibration methods decrease, but are still larger than the proposed method. Therefore, the proposed method with implicit calibration has high accuracy in different usage ranges.

Table V compares the proposed method with other head-mounted gaze estimation methods with implicit calibration.

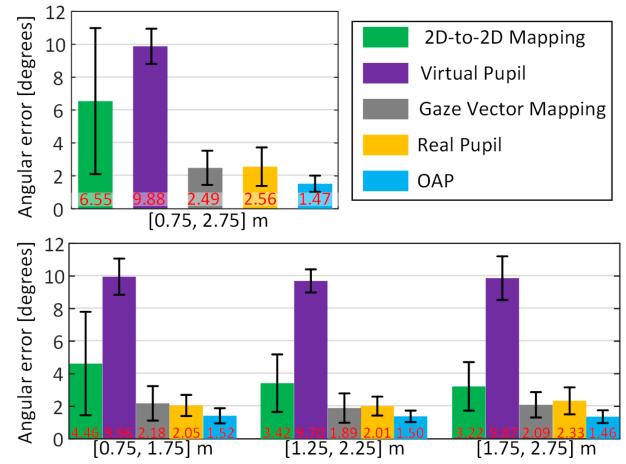


Fig. 10. Comparison of gaze estimation methods in different usage ranges, including 2 m (TOP) and 1 m (BOTTOM). Bars show the mean angular error; error bars indicate standard deviations; numbers on the bottom are the mean angular error in degrees.

TABLE V
COMPARISON WITH OTHER HEAD-MOUNTED GAZE ESTIMATION METHODS
USING IMPLICIT CALIBRATION

Method	Angular error	Implicit calibration
Ours	$1.47 \pm 0.47^\circ$	Align to gaze patterns
Liu et al. [13]	$2.2 \pm 1.0^\circ$	
Sugano et al. [12]	$8.0 \pm 3.5^\circ$	Watch multiple scenes with known saliency
Perra et al. [21]	$1.49 \pm 1.42^\circ$	

Since other methods differ from this article in testing conditions, their results are from the original papers. The other three methods use saliency maps to represent gaze probabilities and thus depend on multiple scenes whose saliency maps are closely

related to gaze probabilities. In contrast, the proposed method is scene-independent and accurate.

The limitation of the proposed method is the same as the pupil-contour-based gaze estimation methods, i.e., requiring high-quality elliptical contours of pupil images. If pupil contours are poorly fitted, pupil axes suffer from large errors, which directly affect gaze estimation.

VI. CONCLUSION

This article proposes an accurate and flexible gaze estimation approach that can be implicitly calibrated by natural gaze patterns. The key idea is to leverage easily detectable gaze patterns and corresponding eye features as calibration data. To linearly align eye-feature patterns to gaze patterns, we theoretically derived a new eye feature, OAP, whose pattern is linearly consistent with the gaze pattern. A pupil-contour-based gaze estimation model with few unknown parameters and simple devices is developed by using the OAP to predict the VAP. Model parameters can be easily calculated by linearly aligning OAP patterns to gaze patterns. Based on the sensitivity analysis of OAP, the OAP plane is determined by looking at the far center or the mean of all OAs due to the central normal distribution of gaze points. The rotation matrix is determined by detectable gaze movements. Scaling and translation parameters are determined by the gaze range on the target object. Hence, the proposed implicit calibration is natural and scene-independent. Experimental results show that the proposed approach is more accurate than state-of-the-art head-mounted gaze estimation methods, which require explicit calibration or known scene saliency. Future work should enhance the algorithm of pupil-contour detection since the OA is calculated from the elliptical contour of the pupil image.

REFERENCES

- [1] A. Shafti, P. Orlov, and A. A. Faisal, "Gaze-based, context-aware robotic system for assisted reaching and grasping," in *Proc. IEEE Int. Conf. Rob. Autom.*, 2019, pp. 863–869.
- [2] B. Duinkharjav, P. Chakravarthula, R. Brown, A. Patney, and Q. Sun, "Image features influence reaction time: A learned probabilistic perceptual model for saccade latency," *ACM Trans. Graph.*, vol. 41, no. 4, pp. 1–15, Jul. 2022.
- [3] Z. Wang, Y. Zhao, and F. Lu, "Gaze-vergence-Controlled see-through vision in augmented reality," *IEEE Trans. Visual. Comput. Graph.*, vol. 28, no. 11, pp. 3843–3853, Nov. 2022.
- [4] R. Zhang et al., "Human gaze assisted artificial intelligence: A review," in *Proc. Int. Joint Conf. Artif. Intell.*, 2020, pp. 4951–4958.
- [5] 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.
- [6] J. Sun, Z. Wu, H. Wang, P. Jing, and Y. Liu, "A novel integrated eye-tracking system with stereo stimuli for 3D gaze estimation," *IEEE Trans. Instrum. Meas.*, vol. 72, 2023, Art. no. 5000615.
- [7] D. Su, Y.-F. Li, and H. Chen, "Toward precise gaze estimation for mobile head-mounted gaze tracking systems," *IEEE Trans. Ind. Informat.*, vol. 15, no. 5, pp. 2660–2672, May 2019.
- [8] D. C. Niehorster, T. Santini, R. S. Hessels, I. T. C. Hooge, E. Kasneci, and M. Nyström, "The impact of slippage on the data quality of head-worn eye trackers," *Behav. Res. Methods*, vol. 52, no. 3, pp. 1140–1160, Jan. 2020.
- [9] C. Hennessey and P. Lawrence, "Noncontact binocular eye-gaze tracking for point-of-gaze estimation in three dimensions," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 3, pp. 790–799, Mar. 2009.
- [10] Z. Wan, C. Xiong, W. Chen, H. Zhang, and S. Wu, "Pupil-contour-Based gaze estimation with real pupil axes for head-mounted eye tracking," *IEEE Trans. Ind. Informat.*, vol. 18, no. 6, pp. 3640–3650, Jun. 2022.
- [11] 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.*, 2018, pp. 1–9.
- [12] Y. Sugano and A. Bulling, "Self-calibrating head-mounted eye trackers using egocentric visual saliency," in *Proc. 28th Annu. ACM Symp. User Interface Softw. Technol.*, 2015, pp. 363–372.
- [13] M. Liu, Y. Li, and H. Liu, "Robust 3-D gaze estimation via data optimization and saliency aggregation for mobile eye-tracking systems," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, Art. no. 5008010.
- [14] 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.*, 2016, pp. 47–55.
- [15] P. Kasprowski and K. Harezlak, "Comparison of mapping algorithms for implicit calibration using probable fixation targets," in *Proc. 10th Biennial ACM Symp. Eye Tracking Res. Appl.*, 2018, pp. 1–8.
- [16] F. Alnajar, T. Gevers, R. Valentini, and S. Ghebreab, "Auto-calibrated gaze estimation using human gaze patterns," *Int. J. Comput. Vis.*, vol. 124, no. 2, pp. 223–236, May 2017.
- [17] K. Wang and Q. Ji, "3D gaze estimation without explicit personal calibration," *Pattern Recogn.*, vol. 79, pp. 216–227, Jul. 2018.
- [18] Z. Chen and B. Shi, "Towards high performance low complexity calibration in appearance based gaze estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 1174–1188, Jan. 2023.
- [19] J. Chen and Q. Ji, "A probabilistic approach to online eye gaze tracking without explicit personal calibration," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 1076–1086, Mar. 2015.
- [20] Y. Sugano, Y. Matsushita, and Y. Sato, "Appearance-based gaze estimation using visual saliency," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 2, pp. 329–341, Feb. 2013.
- [21] D. Perra, R. K. Gupta, and J.-M. Frahm, "Adaptive eye-camera calibration for head-worn devices," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 4146–4155.
- [22] M. Liu, Y. Li, and H. Liu, "3D gaze estimation for head-mounted eye tracking system with auto-calibration method," *IEEE Access*, pp. 104 207–104 215, 2020.
- [23] K. Pfeuffer, M. Vidal, J. Turner, A. Bulling, and H. Gellersen, "Pursuit calibration: Making gaze calibration less tedious and more flexible," in *Proc. 26th Annu. ACM Symp. User Interface Softw. Technol.*, 2013, pp. 261–269.
- [24] Y. Sugano, Y. Matsushita, Y. Sato, and H. Koike, "Appearance-based gaze estimation with online calibration from mouse operations," *IEEE Trans. Hum.-Mach. Syst.*, vol. 45, no. 6, pp. 750–760, Dec. 2015.
- [25] 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.
- [26] M. Mansouryar, 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.*, 2016, pp. 197–200.
- [27] L. Świński and N. A. Dodgson, "A fully-automatic, temporal approach to single camera, glint-free 3D eye model fitting," in *Proc. Pervasive Eye Tracking Mobile Eye-Based Interaction*, 2013, pp. 1–11.
- [28] S. Umeyama, "Least-squares estimation of transformation parameters between two point patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 4, pp. 376–380, Apr. 1991.
- [29] 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.



Hanyuan Zhang received the M.S. degree in electronics and communications engineering from the North University of China (NUC), Taiyuan, China, in 2018. She is currently working toward the Ph.D. degree in mechanical engineering with Wuhan University of Science and Technology (WUST), Wuhan, China.

Her research interests include human-machine interaction and gaze tracking.



Shiqian Wu (Senior Member, IEEE) received the B.Eng. and M.Eng. degrees in mechanical engineering from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 1985 and 1988, respectively, and the Ph.D. degree in electrical and electronic engineering from Nanyang Technological University, Singapore, in 2001.

From 1988 to 1997, he was an Assistant Professor, Lecturer, and Associate Professor with HUST. From 2000 to 2014, he was a Research Fellow or Research Scientist with the Agency for Science, Technology, and Research (A-STAR), Singapore. He is currently a Professor with the School of Information Science and Engineering, Deputy Director, Institute of Robotics & Intelligent Systems (IRIS), Wuhan University of Science and Technology (WUST), Wuhan, China, and Director of Hubei Province Key Laboratory of Intelligent Information Processing and Real-Time Industrial Systems, Wuhan, China. He has authored or coauthored two books and more than 280 scientific publications (book chapters and journal/conference papers). His research interests include computer vision, pattern recognition, robotics, and artificial intelligence.

Dr. Wu was recognized as a Most Cited Chinese Researcher (2015–2022, Elsevier). He was the recipient of the BEST 10% PAPER award in ICIP 2015 and the Best Paper Finalist award in ICIEA 2020.



Zheng Gao received the B.E. degree in automation, in 2020, from the School of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan, China, where he is currently working toward the M.S. degree in control science and engineering with the School of Information Science and Engineering.

His research interests include intention inference and human-machine interaction.



Zhonghua Wan received the B.S. degree in mechanical engineering from University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2013, and the Ph.D. degree in mechanical engineering from Huazhong University of Science and Technology (HUST), Wuhan, China, in 2021.

He is currently engaged in Postdoctoral Research with the Wuhan University of Science and Technology, Wuhan, China. His research interests include eye tracking, computer vision, and human-machine interaction.



Wenbin Chen (Member, IEEE) received the Ph.D. degree in mechatronic engineering from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2012.

From 2013 to 2014, he was a Research Fellow with the Lincoln School of Computer Science, University of Lincoln, Lincoln, U.K. He is currently a Professor with the State Key Laboratory of Intelligent Manufacturing Equipment and Technology, HUST. His research interest includes wearable robotics, anthropomorphic prosthetic robotics, soft robotics, and robot motion planning and control.