




Article

# A Method of Free-Space Point-of-Regard Estimation Based on 3D Eye Model and Stereo Vision

Zijing Wan <sup>1,2</sup> , Xiangjun Wang <sup>1,2,\*</sup> , Lei Yin <sup>1,2</sup>  and Kai Zhou <sup>1,2</sup>

<sup>1</sup> State Key Laboratory of Precision Measurement Technology and Instruments, Tianjin University, No. 92 Weijin Road, Nankai District, Tianjin 300072, China; zjwan@tju.edu.cn (Z.W.); yinlei@tju.edu.cn (L.Y.); zhoukrwx@163.com (K.Z.)

<sup>2</sup> MOEMS Education Ministry Key Laboratory, Tianjin University, No. 92 Weijin Road, Nankai District, Tianjin 300072, China

\* Correspondence: tjuxjw@126.com; Tel.: +86-022-2740-3395

Received: 16 August 2018; Accepted: 26 September 2018; Published: 30 September 2018



**Featured Application:** A 3D Point-of-Regard estimation system is proposed in this paper. The results of this research can be applied to head-mounted eye tracking devices or augmented reality devices.

**Abstract:** This paper proposes a 3D point-of-regard estimation method based on 3D eye model and a corresponding head-mounted gaze tracking device. Firstly, a head-mounted gaze tracking system is given. The gaze tracking device uses two pairs of stereo cameras to capture the left and right eye images, respectively, and then sets a pair of scene cameras to capture the scene images. Secondly, a 3D eye model and the calibration process are established. Common eye features are used to estimate the eye model parameters. Thirdly, a 3D point-of-regard estimation algorithm is proposed. Three main parts of this method are summarized as follows: (1) the spatial coordinates of the eye features are directly calculated by using stereo cameras; (2) the pupil center normal is used to the initial value for the estimation of optical axis; (3) a pair of scene cameras are used to solve the actual position of the objects being watched in the calibration process, and the calibration for the proposed eye model does not need the assistance of the light source. Experimental results show that the proposed method can output the coordinates of 3D point-of-regard more accurately.

**Keywords:** gaze estimation; 3D point-of-regard; binocular eye model; stereo vision

## 1. Introduction

Gaze tracking technology is a potential human-computer interface (HCI) technology. The application of this technology has expanded to the gaze-based remote controlling, human visual attention analysis, psychological analysis, and virtual reality/augmented reality.

After decades of development, recent gaze tracking devices are mainly based on computer vision [1]. The human visual attention information is obtained by analyzing the human eye images, and the gaze direction or the coordinates of point-of-regard (PoR) are further estimated. However, the gaze estimation method cannot accurately provide the visual information what people really see. This is because that current methods make it impossible for people to directly obtain the complete information of human visual attention characteristics. The gaze estimation method can only rely on the human's description or eye images captured by cameras. Due to the individual differences of human visual attention and the limitation of the gaze estimation methods, accurate estimation of PoR is still a challenging work.

The head-mounted gaze tracking device is a potential human-computer interface system in free space. However, the accurate estimation of the 3D PoR on classical methods still needs the assistance

of scene information and inertial sensors [2]. This research also focuses on the 3D PoR estimation method based on a head-mounted gaze tracking system.

In this paper, we propose a novel gaze tracking system based on a head-mounted device and stereo vision. The method uses two pairs of stereo cameras to capture eye region images and extract necessary eye features, and directly solves the spatial coordinates of these features in real-time. In addition, a pair of scene cameras are set to capture scene images, and the stereo vision system formed by the scene cameras can solve the spatial coordinates of matching points in the scene. The whole gaze estimation model is simpler than the classical geometry-based model and composed of three parts.

(1) A simplified geometry-based gaze estimation model is proposed, which takes the pupil center normal as the initial value of the eye's optical axis.

(2) A corresponding personal calibration process for calibrating the parameters of the eye model. The personal calibration uses the spatial coordinates of the target point solved by the scene cameras as the calibration reference, and uses common eye features to estimate the position of the eyeball center and the angle between the optical axis and the visual axis.

(3) The spatial coordinate of the PoR is estimated as the middle point of the nearest point of the two visual axes.

This paper also gives several experiments to evaluate the proposed gaze tracking system. The experimental results show that the method can accurately estimate the spatial coordinates of the PoR.

This paper is organized as follows: We start by summarizing related work in Section 2. In Section 3, the main method is explained. Section 4 presents experiment and discussion, and Section 5, the conclusion.

## 2. Related Works

Earlier gaze estimating framework proposed by [3] has already demonstrated the widely application and the framework of this technology. Due to recent advances in computer vision, video camera-based gaze estimation has been developed by many researchers [1]. Gaze tracking methods based on video camera can be divided into two types: model-based method and regression-based method. These two types of methods have been proved effective in different application environments.

Usually, model-based method builds a 3D eye model to calculate the gaze direction according to different eye and facial features [4–6]. These methods have good accuracy, and also have a longer history. But they usually combine strict personal calibration processes, and some with high accuracy are very complicated. Corneal reflection-based methods are proved to be very successful in the past few years [7–9]. This kind of methods are usually with high accuracy, and the error of gaze direction can be lower than 1 degree [10], but the camera and the light source in these gaze tracking systems should be calibrated before utilizing. Moreover, some methods' structures are more complicated [11,12], it is tedious to deploy these proposed gaze tracking systems.

Removal of the environment light noise from pupil images and the recognition of eyelid occlusion are important for the application of eye tracking system [13], and also the research works of making personal calibration more simply [14–16]. Moreover, these methods can also support high accuracy and real-time processing.

Remote camera-based gaze tracking devices have also attracted people's attention because of their wide range of applications, but the need for personal calibration process is a constraint on the spread of these devices. Regression-based methods have prospect to resolve this problem, so they have developed rapidly in recent years, and their generality has been verified in various HCI applications [17]. The research on the regression-based method mainly concentrates on improving the precision and simplifying the steps of personal calibration. Regression-based methods can be divided into two categories [18]: feature-based methods and appearance-based methods. Feature-based methods establish mappings between eye features and eye directions, such as pupils, reflective spots, and so on [19,20]. Appearance-based methods establish mappings between eye appearances and eye directions [21–23]. Moreover, machine learning algorithms for regression-based gaze estimation have been introduced.

Many studies apply different sensors to gaze tracking devices. RGB-D camera can capture facial images and output depth information of facial features for achieving gaze estimation [24–26]. Because of simplified calibration, commercial RGB-D cameras have good application prospects in remote-camera devices [27]. Funes-Mora et al. proposed a line-of-sight estimation for low resolution images [28]. This work presents a framework for integrating current appearance-based strategies into the framework for comparison. In conclusion, depth camera-based gaze tracking technology or utilizing other sensors has attracted more attention.

Head-mounted gaze tracking system can be deployed freely, and obtain clear eye images, its eye direction estimation is more accurate. These are the largest advantages of such devices. In addition, a glasses-style gaze tracking device is able to capture clear eye images [29]. However, it relies on other sensors or systems to measure head posture [30]. The structure of a new wearable device is based on pupil-glint vector and binocular eyeball model [31], although the application is reasonable, but the error compensation is difficult.

The above listed gaze tracking techniques have good accuracy in gaze detection within specific environments, but there are still many shortcomings in free-space 3D PoR estimation. In recent years, 3D gaze tracking research has been focused on how to make the gaze point estimation approach or exceed the depth perception of the human vision while the device is also simple and convenient.

Takemura et al. [2] proposed a 3D PoR estimation method combining scene images. The pupil vector is used as the initial value of gaze direction, then corner feature extraction and Delaunay triangulation are used to analyze scene images then estimate possible PoR. Pirri et al. [32] added a multi-camera view geometry in a simple eye model. This method can obtain accurate 3D PoR without scene information, but the calibration process still needs the spatial coordinates of target point which are calculated from scene images. Our previous work also proposed a novel method for the 3D PoR estimation in free space [33]. Pupillary reflex as an important human visual mechanism was introduced in a simplified line-of-sight estimation model. It was an innovative attempt, but it was restricted by illumination conditions.

Previous methods were limited to application scenarios or hardware structure, and were not suitable as general gaze estimation methods. Model-based method could significantly improve the accuracy but the system calibration steps became more complex; regression-based methods could adapt to a variety of application scenarios and reduce the complexity of utilization, but the accuracy of was insufficient; the combination methods based on the advantages of both usually had the disadvantages of complex structure or high cost, and they were difficult to be the framework for methods within generality.

This paper attempts to propose a general method for 3D PoR estimation. Additionally, the proposed method basically follows the general framework of the geometry-model based method, but some improvements have been made to reduce the operation complexity significantly. This method is suitable for many application scenarios with good accuracy.

### 3. Methods

The proposed gaze estimation method is consisted by four main parts: (1) multi-camera image acquisition; (2) eye feature detection; (3) personal calibration process; and (4) 3D PoR (point-of-regard) estimation. The diagram of the proposed gaze estimation system is shown in Figure 1.

Two pairs of stereo cameras installed in the device capture eye region images and extract necessary eye features, and then directly solve the spatial coordinates of these features in real-time. A 3D eye model and gaze estimation model are proposed in Section 3.3, and also a calibration process. The proposed 3D eye model is a model which is simplified by a classical eye model, and the 3D PoR is considered to be the middle point of the nearest point of the binocular visual axes.

The calibration process is used to estimate the position of the eyeball center and the angle between the visual axis and the optical axis. The spatial coordinates of the pupil, the inner eye corners and the calibration target point are firstly calculated by stereo cameras. We assume that the pupil center

normal is the initial value of the optical axis of the eye, and the spatial coordinates of target point are also involved to the estimate the eyeball features.

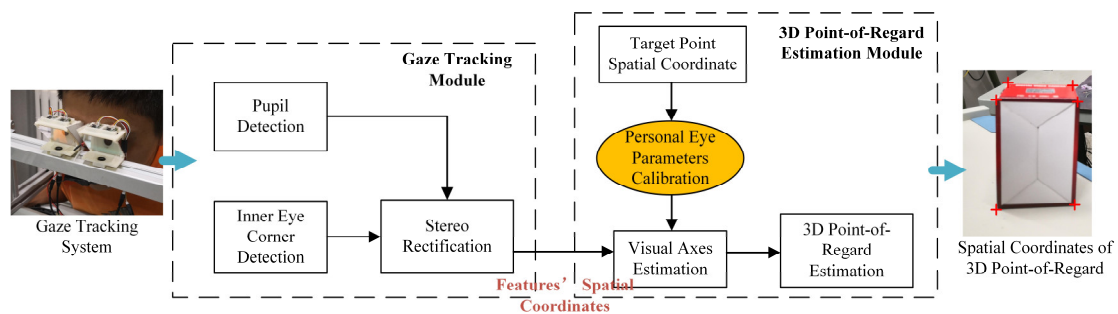


Figure 1. The diagram of the proposed 3D point-of-regard estimation method.

### 3.1. Customized Gaze Tracking System

Gaze tracking devices require different configurations of cameras and light source to satisfy different methods. Three conditions need to be considered when designing a head-mounted gaze tracking system [33]:

1. Ensure a large field of view;
2. Reduce or avoid the influence of environment illumination on images; and
3. The device should be miniaturized, light-weight, and low-cost.

The proposed head-mounted gaze tracking device is based on three pairs of stereo vision cameras and a pair of semi-transparent mirrors. Two pairs of stereo cameras cover left and right eyes, and another pair of stereo cameras are used to capture scene images, all of which are imaged through semi-transparent mirrors. Two near-infrared light emitting panels are, respectively, arranged between the eyes as light sources, which are used for generating dark pupil images and improving the contrast of the eye images. The structure and the photo of the proposed system are shown in Figure 2.

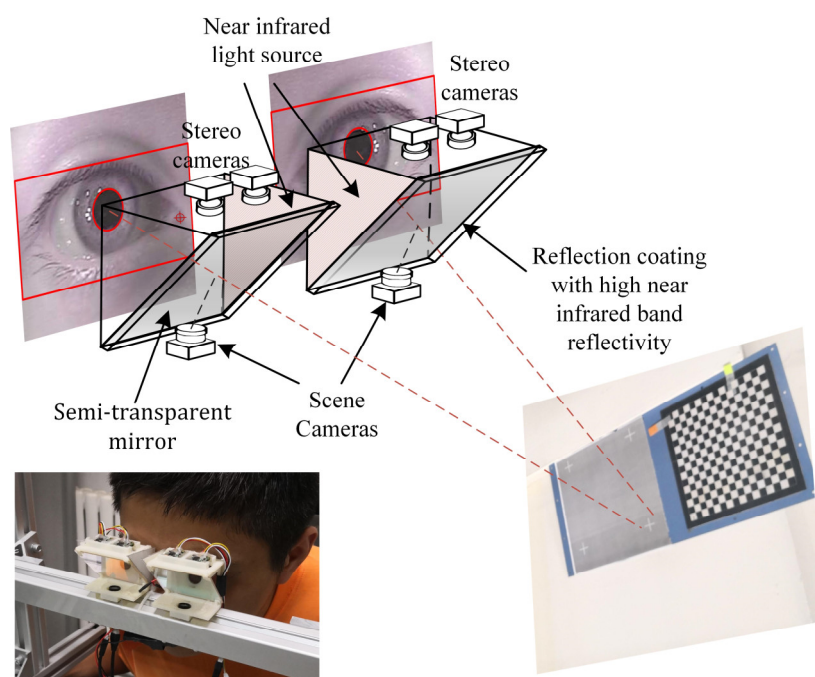


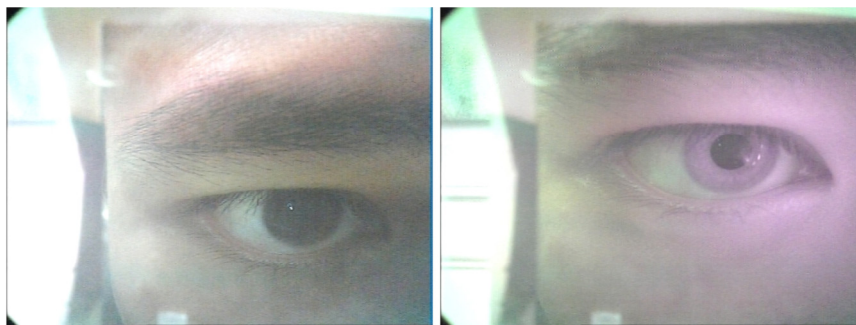
Figure 2. The structure and the photo of the proposed gaze estimation system.

The device is formed by two modules. As shown in Figure 2, each module is equipped with a semi-transparent mirror, which enables the eye cameras to capture eye images taken from the front view, and the scene cameras to obtain the scene images (i.e., scene cameras are fixed under the semi-transparent mirrors and their horizontal field-of-view is 120 degree). The distance between the two modules is not fixed, and can be adjusted according to the pupillary distance. The resolution of a single camera is  $640 \times 580$  pixels, and we choose a smaller focal length in order to get a larger field of view.

The semi-transparent mirror reduces the size of the whole device, and the optical coating of the mirror also has some special features. On the side close to the eye, the coating of the semi-transparent mirror has a reflectivity of over 90% in the near infrared band. The image captured by the eye cameras through the mirror can avoid noise caused by environment light, while the eye area is illuminated by the near infrared light source, making the eye image clearer.

Stereo cameras can acquire the spatial coordinates of eye features accurately. Two pairs of the binocular stereo cameras capture images of both eye regions, respectively. The system calculates the positions of each eye features independently, but these coordinate values are eventually aligned to the device coordinate system. A pair of the scene cameras also compose a stereo vision system. The scene cameras are not directly involved in the gaze estimation in this method, but are used to capture the calibration target images in the calibration process. Additionally, they are also used to evaluate the accuracy of the whole method.

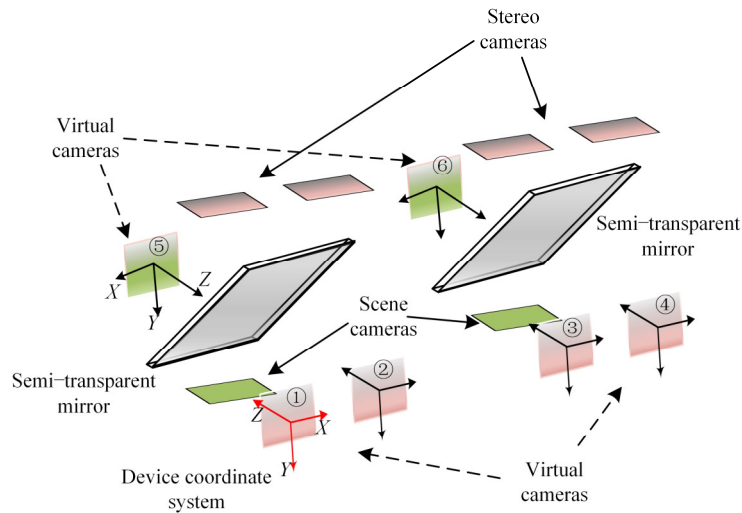
The near infrared light source equipped in this device is a pair of low-power light emitting panels within the near infrared band. Traditional LED bulbs can ensure good reflection spots on the cornea, but the human eyes are usually stimulated, resulting in rapid fatigue of the users. The gaze estimation model in this paper does not use corneal reflection, so there is no need for a point light source. Therefore, the purpose of installing these two low-power light emitting panels is to improve the image contrast. The light emitting panel illuminates the whole eye area through reflection mirrors. The enhancement of the image is shown in Figure 3. Moreover, because of the soft light, the user does not feel uncomfortable when wearing the device for a long time.



**Figure 3.** The left graph is the eye image captured when the near infrared light emitting panel is closed, while the right graph is when the light emitting panel is turned on. When the light emitting panel is turned on, the eye image has obvious dark pupil effect, and the image contrast is also enhanced.

Every camera in the proposed gaze tracking device needs to be calibrated and determine the transformations of all cameras to one device coordinate system. Since all the images of cameras are captured through the semi-transparent mirror, the gaze estimation system in this paper defines the reality of cameras in the mirror as virtual cameras, while all camera calibration refers to the calibration of virtual cameras.

The definition of each virtual camera coordinate system is shown in Figure 4, and the No. 1 camera coordinate system is defined as the device coordinate system.



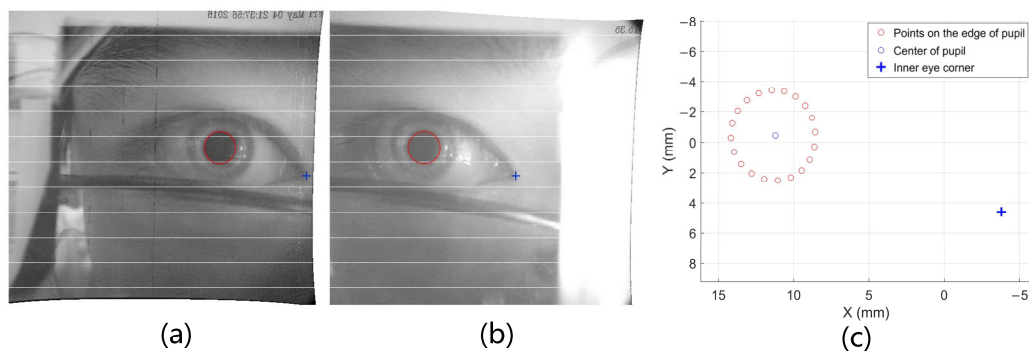
**Figure 4.** The distribution of the virtual cameras with coordinate system definition in the gaze tracking system.

Although the structure of the proposed gaze tracking device is more complicated than the devices based on single or binocular cameras, the proposed device still has a compact structure, and only needs to calibrate stereo cameras. The semi-transparent mirrors can be erected, and it could make the device thinner. In addition, pinhole cameras are miniature devices suitable for wearable devices. The proposed method has the potential applying to the glasses-type device.

### 3.2. Eye Features Extraction

Eye feature extraction is a basic component of most of the gaze tracking systems. Due to the installation of the near infrared light source, the dark pupil effect can be observed clearly from the images captured by the video camera, which makes it easier to extract the pupil contour. The device proposed in this paper is equipped with the stereo cameras. After obtaining the position of the pupil contour in the image, the spatial position of the pupil contour can be solved by using epipolar constraint. The extracting process of the eye features has been detailed explained in our previous work [33].

Due to the installation of the semi-transparent mirror, the eye cameras are able to capture the images of the eye region from the position of the virtual camera, i.e., the front view of the face. Therefore, simple algorithms can be effective for extracting accurate inner eye corner features. In this paper, a multi-scale Harris corner extraction algorithm [34] is used to get the sub-pixel coordinates of the inner eye corner. The result of eye features extraction is shown in Figure 5.



**Figure 5.** (a,b) show the images captured by a pair of eye cameras after epipolar rectification, and the extracted pupil contour and inner eye corner are marked in these two graphs; and (c) represents the spatial coordinates of detected eye features.

### 3.3. 3D Point-of-Regard Estimation

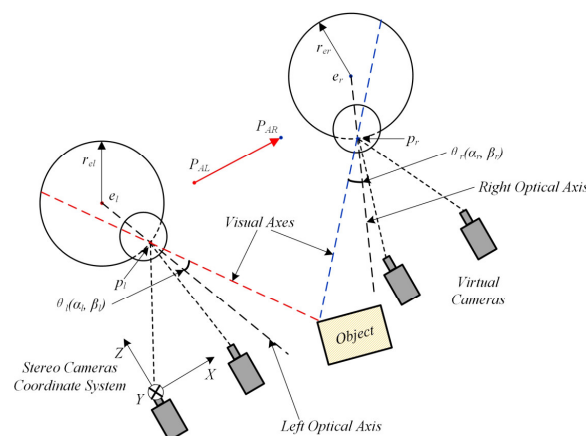
3D model-based gaze estimation method is a kind of precise gaze tracking method. Since the spatial coordinates of the eye features' parameters can be calculated and after combining with the 3D geometric model, the relative position of the head and the viewed object could not be fixed. In addition, the accuracy will not deteriorate even the head is far away from the calibration position. Another advantage is that when this kind of method is applied to remote-camera type gaze tracking device, stereo cameras can also be used to solve head movement.

However, some existing 3D model-based methods, such as the method in [35], because of the need of estimating the corneal center, need at least two light sources to estimate corneal sphere [7]. Moreover, it is also necessary to calibrate the relationship between the light source, the screen and the stereo cameras, which also makes it more difficult to deploy this kind of devices.

Although the 3D model-based gaze estimation has some shortcomings, its line-of-sight model is still fine. The angle between the optical axis and the visual axis is an important factor in human vision mechanism, and this is a major feature of the classic 3D eye model. Therefore, we intend to preserve most of the features of the 3D line-of-sight model while removing some details to enhance our method's practicality. The proposed model is based on a simplified 3D eye model, and then we establish a gaze estimation method suitable for the proposed head-mounted device.

We propose a 3D point-of-regard (PoR) estimation model. The model simplifies corneal spherical center from classic eye model, but estimates the spatial coordinates of the eyeball center by using the pupil center normal through the pupil center. In addition, the pupil center normal is assumed as the initial value for estimating the eyeball center. In order to enable the head-mounted device to be worn repeatedly, the model uses the vector between the left and the right inner eye corners as the aligning reference to calculate the relative poses of the head and the device, and in the following sections, it is named as the inner eye corner vector. Finally, the paper proposes using the middle point of the nearest point of the binocular visual axes as the 3D PoR.

The eyeball center  $e$  and the angle  $\theta$  between optical axis and visual axis of both eyes will be estimated in the personal calibration process. The gaze estimation model is shown in Figure 6. The personal calibration process is realized by the user looking at the calibration target board several times. Theoretically, the eyeball features' parameters can be solved by looking at the calibration target board twice, but the eyeball features' parameters obtained by watching more times are reliable due to the calculation error of stereo vision and the uncertainty of human visual attention. The inner eye corner  $P_A$  is not actually involved in the calculation of binocular visual axes. In fact, only eyeball center  $e$  is needed to calculate its spatial coordinates in this model, and the inner eye corner  $P_A$  is used as a reference to describe the position of the eyeball center relative to the gaze tracking system.



**Figure 6.** The diagram of the 3D eye model and the definition of the eyeball features. All the 3D points refer to positions in the device coordinate system (as the stereo cameras coordinate system shown in the figure). The 3D point-of-regard is considered as the intersection point of the left and right eyes' visual axes.

### 3.3.1. Eyeball Features and Personal Calibration

In this paper, pupil center normal is introduced to the estimation of the eye model parameters. The vector of pupil center normal is seemed as the vector from the eyeball center to the pupil center (i.e., the direction of the eye optical axis). We extract the pupil contour and calculate the coordinates of the space circle where the pupil contour is located, and then calculate the pupil center normal. In the personal calibration process, the eyeball center is estimated by using the spatial intersection of the lines where the multiple pupil center normal is located.

There is a fixed angle  $\theta$  between the eye's visual axis and the eye's optical axis, which varies from person to person and is usually about five degrees. The optical axis is defined by the spatial line from the eyeball center  $e$  to the pupil center  $p$ , while the visual axis is a straight line passing through the pupil center  $p$ , and the visual axis is defined as the spatial line-of-sight. Since the actual calculated spatial lines-of-sight do not necessarily intersect, this paper takes the midpoint between the nearest points of the binocular visual axes as the point-of-regard (PoR). The single 3D eye model of this method is shown in Figure 7.

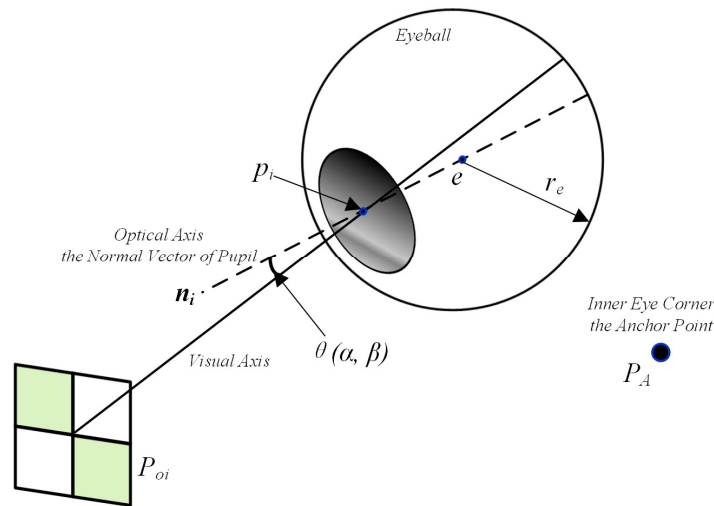


Figure 7. Single 3D eye model in this method and the main eyeball features.

Firstly, we calculate the spatial coordinate of the eyeball center  $e$ . The pupil center  $p$  and the corresponding pupil normal vector  $\vec{n}$  at each fixation need to be obtained first. It is assumed that  $n$  times fixation steps are included in one calibration. We calculate the sum of the nearest distance between each pupil normal and other normal lines, and remove the pupil normal that is far away from other lines. The pupil normal vector is defined as  $\vec{n}$ , and then adding labels to each line  $h_i = 1, i \in [1, \dots, n]$ , at last calculating the sum of the nearest distance between each pupil normal and other normal lines. It can be calculated as:

$$sum_{di} = \sum_{j=1}^n \frac{(\vec{n}_i \times \vec{n}_j) \bullet \vec{p}_i \vec{p}_j}{\|\vec{n}_i \times \vec{n}_j\|}, i \in [1, \dots, n] \tag{1}$$

Then set the label of the line whose distance is greater than the average distance to 0.

$$h_i = \begin{cases} 0, sum_{di} < \frac{1}{n} \sum_{i=1}^n sum_{di} \\ 1, otherwise \end{cases}, i \in [1, \dots, n] \tag{2}$$



The nearest points between each line and other lines are defined as  $pn_{ij} (x_{ij}, y_{ij}, z_{ij})$ . Then the calculated the eyeball center  $e$  is:

$$e(x_e, y_e, z_e) = \operatorname{argmin} \left( \sum_{i=1}^n \sum_{j=1}^n h_i \|e - pn_{ij}\| \right) \tag{3}$$

We solve the  $x_e, y_e$ , and  $z_e$  by using the centroid formula:

$$x_e = \frac{\sum_{i=1}^n \sum_{j=1}^n h_i x_{ij}}{\sum_{i=1}^n \sum_{j=1}^n h_i}, y_e = \frac{\sum_{i=1}^n \sum_{j=1}^n h_i y_{ij}}{\sum_{i=1}^n \sum_{j=1}^n h_i}, z_e = \frac{\sum_{i=1}^n \sum_{j=1}^n h_i z_{ij}}{\sum_{i=1}^n \sum_{j=1}^n h_i} \tag{4}$$

After obtaining the coordinate of the eyeball center  $e$ , the pupil center  $p$  and the spatial coordinate of object  $P_o$  are all used to calculate both eyes. The estimation method of  $\theta$  is to calculate the relative spatial position between the fixation point and the pupil center. After multiple fixations, we can obtain multiple sets of fixation points and pupil center data, and then use the nonlinear optimization method to obtain better value of  $\theta$ .

As shown in Figure 7, the visual axis  $\vec{p_i P_{oi}}$  is formed by the connection of the spatial coordinates of  $p_i$  and  $P_{oi}$ , and the direction of the optical axis is defined as the vector  $e \vec{p_i}$ . Normalizing these two vectors, get normalized vectors  $\vec{v_{si}}$  and  $\vec{v_{oi}}$ , and then defining the rotation of these two vectors by using rotation matrix. In this paper, we use  $R_{\alpha_i}$  and  $R_{\beta_i}$  to represent the rotation matrix of the vector  $\vec{v_{oi}}$  around the X axis and the Y axis in the stereo camera coordinate system, respectively, obtaining:

$$\vec{v_{si}} = R_{\beta_i} R_{\alpha_i} \vec{v_{oi}} \tag{5}$$

Then we can easily get the angle  $\theta_i (\alpha_i, \beta_i)$  calculated in each fixation. The average value  $\theta_0 (\alpha_0, \beta_0)$  is taken as the initial value of the later optimized eyeball feature parameter:

$$\theta_0 = (\alpha_0, \beta_0) = \left( \frac{1}{n} \sum_{i=1}^n \alpha_i, \frac{1}{n} \sum_{i=1}^n \beta_i \right) \tag{6}$$

In the following, we deduce the objective function for optimizing the left and right angles between the optical axis and the visual axis.

Left and right eyes' angles between the visual axis and the optical axis are  $\theta_l (\alpha_l, \beta_l)$  and  $\theta_r (\alpha_r, \beta_r)$ , the left and right eyeball centers are  $e_l = (x_{el}, y_{el}, z_{el})^T$  and  $e_r = (x_{er}, y_{er}, z_{er})^T$ , and the pupil centers are written as  $p_{li} = (X_{li}, Y_{li}, Z_{li})^T$  and  $p_{ri} = (X_{ri}, Y_{ri}, Z_{ri})^T$ .

The normalized left and right eyes' optical axis vectors  $\vec{v_{oli}}$  and  $\vec{v_{ori}}$  can be written as follows:

$$\begin{aligned} \vec{v_{oli}} &= (X_{li} - x_{el}, Y_{li} - x_{el}, Z_{li} - z_{el})^T = (a_1, a_2, a_3)^T \\ \vec{v_{ori}} &= (X_{ri} - x_{er}, Y_{ri} - x_{er}, Z_{ri} - z_{er})^T = (b_1, b_2, b_3)^T \end{aligned} \tag{7}$$

The normalized left and right eyes' visual axis vectors  $\vec{v_{sli}}$  and  $\vec{v_{sri}}$  can be calculated as follows:

$$\begin{aligned} \vec{v_{sli}} &= \begin{bmatrix} a_1 \cos \beta_l + a_2 \sin \alpha_l \sin \beta_l - a_3 \cos \alpha_l \sin \beta_l \\ a_2 \cos \alpha_l + a_3 \sin \alpha_l \\ a_1 \sin \beta_l - a_2 \sin \alpha_l \cos \beta_l + a_3 \cos \alpha_l \cos \beta_l \end{bmatrix} = \begin{bmatrix} a_1' \\ a_2' \\ a_3' \end{bmatrix} \\ \vec{v_{sri}} &= \begin{bmatrix} b_1 \cos \beta_r + b_2 \sin \alpha_r \sin \beta_r - b_3 \cos \alpha_r \sin \beta_r \\ b_2 \cos \alpha_r + b_3 \sin \alpha_r \\ b_1 \sin \beta_r - b_2 \sin \alpha_r \cos \beta_r + b_3 \cos \alpha_r \cos \beta_r \end{bmatrix} = \begin{bmatrix} b_1' \\ b_2' \\ b_3' \end{bmatrix} \end{aligned} \tag{8}$$

Therefore, the spatial lines of left and right visual axis are represented by using the line's direction vector and the coordinate of the point on the line, so the coordinates of the points on the lines  $P_{li} = (x_{li}, y_{li}, z_{li})^T$  and  $P_{ri} = (x_{ri}, y_{ri}, z_{ri})^T$  can be represented as follows:

$$\begin{aligned} \frac{x_{li} - X_{li}}{a_{1i}'} &= \frac{y_{li} - Y_{li}}{a_{2i}'} = \frac{z_{li} - Z_{li}}{a_{3i}'} \\ \frac{x_{ri} - X_{ri}}{b_{1i}'} &= \frac{y_{ri} - Y_{ri}}{b_{2i}'} = \frac{z_{ri} - Z_{ri}}{b_{3i}'} \end{aligned} \tag{9}$$

The spatial coordinate of the target being watched is defined as  $P_{oi} = (X_{oi}, Y_{oi}, Z_{oi})^T$ . When  $z_l = z_r = Z_{oi}$ , it means that the intersection points of the two lines and the target plane can be written as:

$$P_{li} = \begin{bmatrix} \frac{a_{1i}'}{a_{3i}'}(Z_{oi} - Z_{li}) + X_{li} \\ \frac{a_{2i}'}{a_{3i}'}(Z_{oi} - Z_{li}) + Y_{li} \\ Z_{oi} \end{bmatrix}, P_{ri} = \begin{bmatrix} \frac{b_{1i}'}{b_{3i}'}(Z_{oi} - Z_{ri}) + X_{ri} \\ \frac{b_{2i}'}{b_{3i}'}(Z_{oi} - Z_{ri}) + Y_{ri} \\ Z_{oi} \end{bmatrix} \tag{10}$$

Then, the distances between these points and the corresponding watched target points are added, obtaining two objective functions about  $\theta_l$  and  $\theta_r$  as follows:

$$\begin{aligned} f(\theta_l) &= \sum_{i=1}^n \left( \left( \frac{a_{1i}'}{a_{3i}'}(Z_{oi} - Z_{li}) - (X_{oi} - X_{li}) \right)^2 + \left( \frac{a_{2i}'}{a_{3i}'}(Z_{oi} - Z_{li}) - (Y_{oi} - Y_{li}) \right)^2 \right) \\ f(\theta_r) &= \sum_{i=1}^n \left( \left( \frac{b_{1i}'}{b_{3i}'}(Z_{oi} - Z_{ri}) - (X_{oi} - X_{ri}) \right)^2 + \left( \frac{b_{2i}'}{b_{3i}'}(Z_{oi} - Z_{ri}) - (Y_{oi} - Y_{ri}) \right)^2 \right) \end{aligned} \tag{11}$$

In this paper, Levenberg-Marquardt method is used to optimize  $\theta_l$  and  $\theta_r$ . The initial values of  $\theta_l$  and  $\theta_r$  can be decided by Equations (5) and (6).

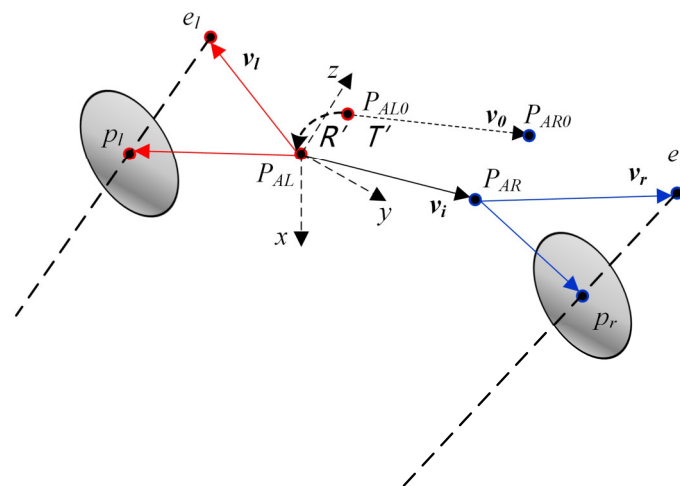
Then we analyze the influence of the user's fixation times on the calibration results through experiment. In this experiment, we assume that the increase of the user's fixation times will reduce the calibration error. Therefore, we compare the calibration results in different fixation times with the real position of the corresponding calibration target to analyze how much fixation times can take stable personal calibration results. Detailed experimental results are described in Section 3.3.3.

### 3.3.2. Eyeball Features Coordinate Alignment

The inner eye corner is a stable facial feature and can be used as a reference for facial posture. We firstly establish an eye coordinate system, the structure and the definition are shown in Figure 8, the origin of the eye coordinate system is established with the  $P_{AL}$ , and its coordinate axis direction and the scale factor are the same as the device coordinate system. Meanwhile, we form the inner eye corner vector from  $P_{AL}$  to  $P_{AR}$ , and it should be used as a reference to describe the position of the eyeball center and the pupil center relative to the gaze tracking system.

Firstly, we get a pair of the coordinates of the eyeball centers and a pair of the coordinates of the inner eye corners in the calibration. Then we describe how to use the calibrated eyeball center when we only know the coordinates of the inner eye corners during utilizing the proposed system. Take the coordinate alignment of the eyeball centers  $P_{AL}$  and  $P_{AR}$  as an example.

The positions of the eyeball center  $e_l$  and  $e_r$  are calculated in the calibration process, but these two coordinates could be changed when the head posture changes. The left and right inner eye corners are obtained at the same time when calibrating, and they can also be calculated when using. Therefore, the coordinates of the eyeball center can be recovered by calculating the relative position between the inner eye corner and the eyeball center. In this paper, we choose the inner corner vector as the reference for alignment.



**Figure 8.** The origin of the eye coordinate system is built as shown in the figure; the axis direction and the scale factor are consistent with the device coordinate system. When the head posture changes, the eyeball center coordinates could be changed when compared with the calibration results. Therefore, the inner eye corner vector needs to be used as the reference to use the calibrated 3D eyeball model parameters when using the gaze tracking system.

The vector  $\vec{v}_0$  obtained during calibration is the initial inner corner vector, and the eye coordinate system is established at the left eye corner  $P_{AL}$ . As shown in Figure 8, the vector  $\vec{v}_{er}$  and the vector  $\vec{v}_{el}$  can be represented by  $\vec{v}_0$  and rotation matrix.

Get:

$$\vec{v}_{el} = R_{el}\vec{v}_0 = P_{AL}\vec{e}_l, \vec{v}_{er} = R_{er}\vec{v}_0 = P_{AL}\vec{e}_r \tag{12}$$

When in use, we first get the  $\vec{v}_i$  in the eye coordinate system, then calculate the rotation matrix  $R'$  from  $\vec{v}_0$  to  $\vec{v}_i$ :

$$\vec{v}_i = R'\vec{v}_0 \tag{13}$$

Then align the left corner coordinate of the  $i$  time with the initial left corner coordinate, and at the same time get the translation vector  $T'$ .

The left and right eyeball vectors where  $\vec{v}_i$  is located are:

$$\vec{v}'_{er} = R_{er}\vec{v}_i = R_{er}R'\vec{v}_0, \vec{v}'_{el} = R_{el}\vec{v}_i = R_{el}R'\vec{v}_0 \tag{14}$$

Therefore, when in use, the  $e_l$  and  $e_r$  coordinates can be written as follows:

$$\begin{aligned} e_r &= R_{er}R'\vec{v}_0 \frac{\|\vec{v}_{er}\|}{\|\vec{v}'_{er}\|} + T' \\ e_l &= R_{el}R'\vec{v}_0 \frac{\|\vec{v}_{el}\|}{\|\vec{v}'_{el}\|} + T' \end{aligned} \tag{15}$$

Similarly, the pupil center coordinates are also calculated using the above derivation process. When using the proposed gaze tracking system, the pupil center coordinates solved in real-time do not need to be aligned. However, in the experiments described below, the pupil centers acquired at different times need to be aligned. If necessary, the  $p_l$  and  $p_r$  coordinates after alignment can be written as follows:

$$\begin{aligned} p_r &= R_{pr}R'\vec{v}_0 \frac{\|\vec{v}_{pr}\|}{\|\vec{v}'_{pr}\|} + T' \\ p_l &= R_{pl}R'\vec{v}_0 \frac{\|\vec{v}_{pl}\|}{\|\vec{v}'_{pl}\|} + T' \end{aligned} \tag{16}$$

wherein,  $\vec{v}_{pl} = R_{pl}\vec{v}_0 = P_{AL}p_l, \vec{v}_{pr} = R_{pr}\vec{v}_0 = P_{AL}p_r, \vec{v}'_{pl} = R_{pr}R'\vec{v}_0, \vec{v}'_{pr} = R_{pl}R'\vec{v}_0$ .

### 3.3.3. 3D Point-of-Regard Estimation and Calibration Experiment

Theoretically, the intersection of the two visual axes is the gaze point, but in practice, it is almost impossible for these two lines to intersect in space. Therefore, the middle point of the nearest points of the binocular visual axes is considered as the estimated point-of-regard (PoR). The binocular pupil center coordinates  $p_l(X_l, Y_l, Z_l), p_r(X_r, Y_r, Z_r)$  calculated by the stereo camera are used in the PoR estimating algorithm, and also the calibrated eyeball feature parameters  $e_l(x_{el}, y_{el}, z_{el}), e_r(x_{er}, y_{er}, z_{er}), \theta_l(\alpha_l, \beta_l)$  and  $\theta_r(\alpha_r, \beta_r)$ .

Firstly, get the spatial line equations of the visual axes according to the Equations (8) and (9):

$$\frac{x_l - X_l}{a_1'} = \frac{y_l - Y_l}{a_2'} = \frac{z_l - Z_l}{a_3'} \tag{17}$$

$$\frac{x_r - X_r}{b_1'} = \frac{y_r - Y_r}{b_2'} = \frac{z_r - Z_r}{b_3'}$$

wherein:

$$\begin{bmatrix} a_1' \\ a_2' \\ a_3' \end{bmatrix} = \begin{bmatrix} (X_l - x_{el}) \cos \beta_l + (Y_l - y_{el}) \sin \alpha_l \sin \beta_l - (Z_l - z_{el}) \cos \alpha_l \sin \beta_l \\ (Y_l - y_{el}) \cos \alpha_l + (Z_l - z_{el}) \sin \alpha_l \\ (X_l - x_{el}) \sin \beta_l - (Y_l - y_{el}) \sin \alpha_l \cos \beta_l + (Z_l - z_{el}) \cos \alpha_l \cos \beta_l \end{bmatrix} \tag{18}$$

$$\begin{bmatrix} b_1' \\ b_2' \\ b_3' \end{bmatrix} = \begin{bmatrix} (X_r - x_{er}) \cos \beta_r + (Y_r - y_{er}) \sin \alpha_r \sin \beta_r - (Z_r - z_{er}) \cos \alpha_r \sin \beta_r \\ (Y_r - y_{er}) \cos \alpha_r + (Z_r - z_{er}) \sin \alpha_r \\ (X_r - x_{er}) \sin \beta_r - (Y_r - y_{er}) \sin \alpha_r \cos \beta_r + (Z_r - z_{er}) \cos \alpha_r \cos \beta_r \end{bmatrix}$$

The nearest points of these two visual axes lines  $p_1(x_1, y_1, z_1)$  and  $p_2(x_2, y_2, z_2)$  can be named as:

$$p_1 = \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = \begin{bmatrix} a_1't_1 + X_l \\ a_2't_1 + Y_l \\ a_3't_1 + Z_l \end{bmatrix}, p_2 = \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} = \begin{bmatrix} b_1't_2 + X_r \\ b_2't_2 + Y_r \\ b_3't_2 + Z_r \end{bmatrix} \tag{19}$$

wherein:

$$t_1 = \frac{(X_r - X_l)(b_2'n_3 - b_3'n_2) + (Y_r - Y_l)(b_1'n_3 + b_3'n_1) - (Z_r - Z_l)(b_1'n_2 + b_2'n_1)}{a_1'(b_2'n_3 - b_3'n_2) + a_2'(b_3'n_1 - b_1'n_3) + a_3'(b_1'n_2 - b_2'n_1)} \tag{20}$$

$$t_2 = \frac{(X_r - X_l)(a_2'n_3 - a_3'n_2) + (Y_r - Y_l)(a_3'n_1 - a_1'n_3) + (Z_r - Z_l)(a_1'n_2 - a_2'n_1)}{a_1'(b_2'n_3 - b_3'n_2) + a_2'(b_3'n_1 - b_1'n_3) + a_3'(b_1'n_2 - b_2'n_1)}$$

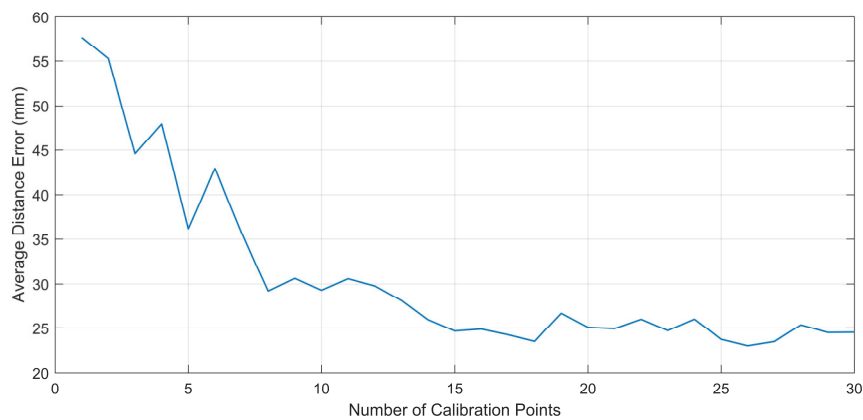
The coordinate of the 3D PoR is described as:

$$P_{POR} = (p_1 + p_2) / 2 \tag{21}$$

We use the calibration data to verify the effectiveness of the proposed algorithms, and also determine how many times fixations are needed in the calibration process, i.e., the calibration points. We use the calibration results estimated within different calibration points to take 3D PoR estimation experiments, and compare the experimental results with the actual target point spatial coordinates to determine the suitable number of the calibration points.

We invited four people to participate in the personal calibration experiment. Each person carries out a personal calibration process, and in one calibration process a person needs to observe 30 different target points. In each observation, the calibration target is randomly placed at any position 0.8 m to 1.5 m away from the person, and it is necessary to ensure that the person can observe the target in comfort. After collecting these experimental data, the calibration results are first calculated, and then these data are used to calculate each person's spatial coordinates of the PoR during calibration, and these calculated spatial coordinates of the PoR will be used to calculated the distances between themselves and the ground truth of the target point. Finally, we plot these experimental results in

a graph according to the number of the calibration points. The experimental results are shown in Figure 9.

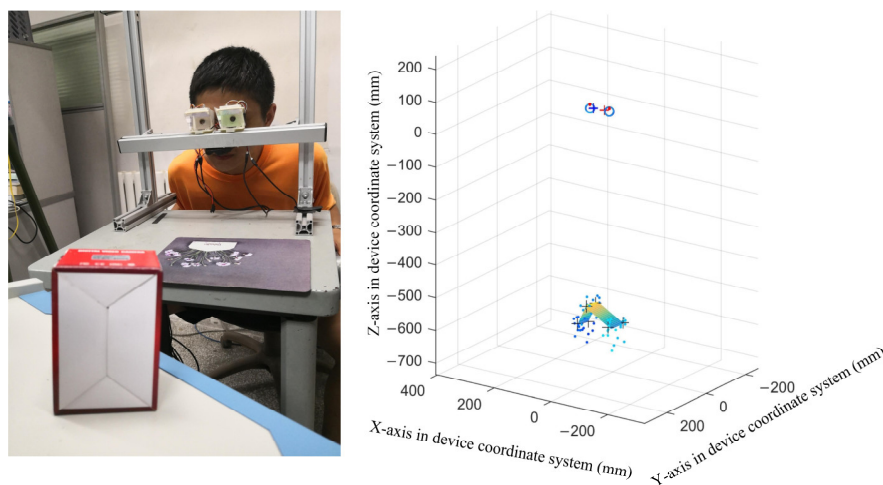


**Figure 9.** The average distance error of calibration results in different numbers of calibration points.

The experimental results show that when the number of the calibration points exceeds 15, the errors drops to a low and stable level. However, too many calibration points will make the calibration process complicated. We have also learned from the calibration processes of some commercial eye tracking devices, which usually use nine calibration points. Therefore, in subsequent experiments and practical use, the proposed gaze tracking system usually set the number of the calibration points to nine.

#### 4. Experiments

According to the previous sections, the proposed method can directly output the estimated coordinate of the point-of-regard (PoR). Therefore, we directly compared the calculated the spatial coordinate of the PoR with the actual position of the object being watched in space. Although the common gaze estimation methods can provide 2D or 3D gaze point information, the error in the direction of the line-of-sight is a more direct evaluation criterion. In this paper, we directly compared the estimated PoR with the spatial coordinates of the object being watched, which is also a direct comparison method. The actual scene during experiment and experimental results are shown in Figure 10.



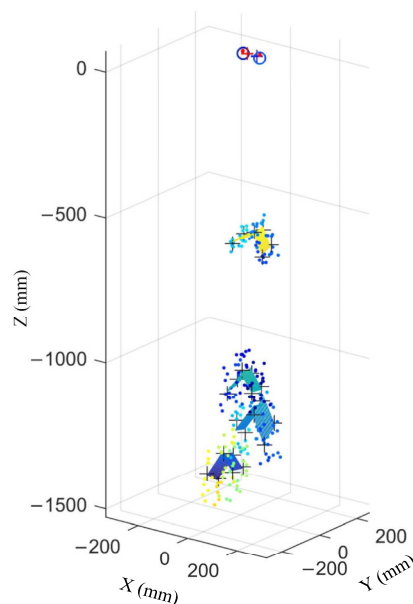
**Figure 10.** The actual scene during experiment. The left picture shows a person is watching the six visible corners of a box. And the right graph shows the 3D point-of-regard estimation result of this observation experiment.

The gaze tracking device proposed in this paper is equipped with a pair of scene cameras. We accurately calculated the spatial coordinates of the object being watched by using these two cameras, and this measurement was independent of the PoR estimation. In the experiments, all coordinate values are in the device coordinate system, and each fixation time is set to 1 s.

#### 4.1. 3D Point-of-Regard Estimation Experiment in Free Space

In this experiment, we used a box to test the effect of our method. The actual scene during experiment is shown as the left image in Figure 10.

We randomly placed a box anywhere from 0.6 m to 1.5 m away from the system, while ensuring that the box could be viewed through the system in comfort. Then a person was invited to look at the six visible corners of the box and get the observation data. We compared the experimental result with the real spatial condition of the corner of the box. This experiment proves that the gaze tracking system proposed in this paper can be applied to common scenes. The result of this experiment is shown in Figure 11.



**Figure 11.** Placing a box in four different positions in front of the tester, and allowed this tester to watch the six visible corners of the box. The crosses in the graph represent the six visible corners of the box, and the colored grid planes show the two visible surfaces of the box. The scattered dots represent the location of the calculated points-of-regard by each fixation. This experiment shows that our method can calculate the 3D PoR coordinates when the system is used in a common scene.

#### 4.2. Error of the Method in Different Distances

We evaluated the accuracy of the 3D PoR estimation at different distances. In the experiment, we carried out three main steps:

(1) We firstly set up a necessary experimental environment. We fixed the gaze tracking system in an empty room, and then fixed six plane boards at six different distances. Each plane board were drawn with a cross.

(2) We invited four people to participate in this experiment, and everyone were allowed to complete a personal calibration. Everyone had to look at each board through the gaze tracking system, and each person was allowed to look at each board 15 times.

(3) Finally, the spatial coordinate of each board was calculated accurately by using scene cameras, and they were used to count the average errors of our method at different distances. At the same time,

we also compared our method with another 3D PoR estimation method described in [32]. The experimental results are shown in Table 1, and all coordinate values are in the device coordinate system.

**Table 1.** Average errors in centimeters at different distances of our method and the comparison with the method in [32].

Unit: cm	0.8 m	1.5 m	2 m	2.5 m	3 m	4 m
Error on X	0.4 ± 0.4	0.6 ± 0.4	0.7 ± 0.6	0.9 ± 0.6	0.9 ± 0.5	1.3 ± 0.9
Error on Y	0.4 ± 0.3	0.6 ± 0.5	0.7 ± 0.4	0.8 ± 0.6	1.0 ± 0.6	1.3 ± 1.1
Error on Z	3.3 ± 2.7	4.2 ± 3.1	5.3 ± 4.0	7.6 ± 6.1	8.8 ± 6.3	10.7 ± 8.5
Overall	3.4 ± 2.7	4.3 ± 3.1	5.4 ± 4.1	7.8 ± 6.1	8.9 ± 6.4	10.8 ± 8.5
Overall in [32]	2.1 ± 0.3	/	3.6 ± 0.3	/	5.7 ± 0.5	/

By analyzing the experimental data of this experiment, our method is proved to be able to output the 3D PoR coordinates of the human eye when observing object at different distances with good accuracy. The error on Z is much bigger than X and Y directions. However, the results are still in line with our expectations for the proposed method.

#### 4.3. Discussion

The evaluation of this method is described in the experiment section. The experimental results are similar to the design expectation, but there are still shortcomings in the accuracy of depth perception extraction of the object being watched. This may be due to the inadequacy of the estimation strategy spatial coordinate of the PoR. We use the middle point of the nearest points of the binocular visual axes as the result of 3D PoR estimation. Although this strategy is easy to use, it is also relatively brute. If we only consider the errors on X or Y direction, it can be found that the accuracy of the visual axes' directions has reached a high level. But in our method, the 3D PoR estimation is formed by the spatial intersection of the binocular visual axes, so the slight angle deviation of the visual axis can increase the error of the depth estimation which is much more than the errors on X and Y directions.

Since the depth perception in human visual attention is often inaccurate, the system is difficult to learn human vision accurately. Therefore, it is the focus of future work to propose a more effective gaze estimation strategy.

## 5. Conclusions

In this paper, a simplified 3D eye model is used to realize 3D point-of-regard (PoR) estimation in free space, and the method is proven by experiments. We also analyze the application potential of this method in the head-mounted gaze tracking device, and propose a corresponding structure of a novel gaze tracking system.

The proposed gaze tracking system is constituted by three pairs of stereo cameras. In Sections 3.1 and 3.2, we describe the features of the semi-transparent mirror and the layout of the cameras. The system structure proposed in this paper has the ability to obtain clearer images. Camera calibration and system coordinate system determination are simpler than the previous methods. Clear images also make the extraction of pupil and inner eye corner more accurate from the image.

In Section 3.3, we focus on describing the 3D PoR estimation model and personal calibration process carried out by the system. The personal calibration process of the system is based on the geometric model adopted by the system. We take full account of reducing the complexity of personal utilization, and reduce the number of personal eyeball parameters to be calibrated to two types. This makes the proposed method not only ensure certain accuracy, but also reduce the user's operation steps as much as possible. Moreover, the validity of calibration process is verified by experiments.

We evaluated the whole method by free observation experiments. Experimental results show that this method can be applied to real scenes and has sufficient accuracy under indoor environment. This method has the potential to be applied to wearable devices. The purpose of further improving its

accuracy is to enable such device to be applied to the spots of industry to assist workers in controlling automated equipment in workshops.

**Author Contributions:** All authors were involved in the research of this article. Z.W. proposed the algorithms, design the experiments, analyzed the experimental data, and wrote this paper; X.W. designed the gaze tracking device and guided the research group; and L.Y. and K.Z. carried out the calibration experiments. All authors reviewed and approved the manuscript.

**Funding:** This research was funded by National Natural Science Foundation of China, grant number 51575388.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Hansen, W.; Ji, Q. In the Eye of the Beholder: A Survey of Models for Eyes and Gaze. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 478–500. [[CrossRef](#)] [[PubMed](#)]
2. Takemura, K.; Takahashi, K.; Takamatsu, J.; Ogasawara, T. Estimating 3-D Point-of-Regard in a Real Environment Using a Head-Mounted Eye-Tracking System. *IEEE Trans. Hum. Mach. Syst.* **2014**, *44*, 531–536. [[CrossRef](#)]
3. Talmi, K.; Liu, J. Eye and gaze tracking for visually controlled interactive stereoscopic displays. *Signal Process. Image Commun.* **1999**, *14*, 799–810. [[CrossRef](#)]
4. Kang, R.P.; Lee, J.J.; Kim, J. Gaze position detection by computing the three dimensional facial positions and motions. *Pattern Recognit.* **2002**, *35*, 2559–2569.
5. Guestrin, E.D.; Eizenman, M. Remote point-of-gaze estimation requiring a single-point calibration for applications with infants. In Proceedings of the Symposium on Eye Tracking Research & Applications, Savannah, GA, USA, 26–28 March 2008; pp. 267–274.
6. Wang, K.; Ji, Q. Real Time Eye Gaze Tracking with 3D Deformable Eye-Face Model. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 1003–1011.
7. Guestrin, E.D.; Eizenman, M. General theory of remote gaze estimation using the pupil center and corneal reflections. *IEEE Trans. Biomed. Eng.* **2006**, *53*, 1124–1133. [[CrossRef](#)] [[PubMed](#)]
8. Zhu, Z.; Ji, Q. Novel eye gaze tracking techniques under natural head movement. *IEEE Trans. Biomed. Eng.* **2007**, *54*, 2246–2260. [[PubMed](#)]
9. Villanueva, A.; Cabeza, R. A Novel Gaze Estimation System with One Calibration Point. *IEEE Trans. Syst. Man Cybern. B Cybern.* **2008**, *38*, 1123–1138. [[CrossRef](#)] [[PubMed](#)]
10. Cho, D.C.; Yap, W.S.; Lee, H.; Kim, W. Long range eye gaze tracking system for a large screen. *IEEE Trans. Consum. Electron.* **2013**, *58*, 1119–1128. [[CrossRef](#)]
11. Dong, H.Y.; Chung, M.J. A novel non-intrusive eye gaze estimation using cross-ratio under large head motion. *Comput. Vis. Image Underst.* **2005**, *98*, 25–51.
12. Beymer, D.; Flickner, M. Eye gaze tracking using an active stereo head. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Madison, WI, USA, 16–22 June 2003; pp. 451–458.
13. Cho, C.W.; Ji, W.L.; Shin, K.Y.; Lee, E.C.; Park, K.R.; Lee, H.; Cha, J. Gaze Detection by Wearable Eye-Tracking and NIR LED-Based Head-Tracking Device Based on SVR. *ETRI J.* **2012**, *34*, 542–552. [[CrossRef](#)]
14. Model, D.; Eizenman, M. An Automatic Personal Calibration Procedure for Advanced Gaze Estimation Systems. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 1031–1039. [[CrossRef](#)] [[PubMed](#)]
15. Chen, J.; Ji, Q. Probabilistic gaze estimation without active personal calibration. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, CO, USA, 20–25 June 2011; pp. 609–616.
16. Chi, J.N.; Xing, Y.Y.; Liu, L.N.; Gou, W. Calibration method for 3D gaze tracking systems. *Appl. Opt.* **2017**, *56*, 1536. [[CrossRef](#)]
17. Zhang, X.; Sugano, Y.; Fritz, M.; Bulling, A. Appearance-based gaze estimation in the wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 4511–4520.
18. Wang, K.; Ji, Q. 3D Gaze Estimation without Explicit Personal Calibration. *Pattern Recognit.* **2018**, *79*, 216–227. [[CrossRef](#)]
19. Morimoto, C.H.; Mimica, M.R.M. Eye gaze tracking techniques for interactive applications. *Comput. Vis. Image Underst.* **2005**, *98*, 4–24. [[CrossRef](#)]



20. Cheng, H.; Liu, Y.; Fu, W.; Ji, Y.; Yang, L.; Zhao, Y.; Yang, J. Gazing point dependent eye gaze estimation. *Pattern Recognit.* **2017**, *71*, 36–44. [[CrossRef](#)]
21. Tan, K.H.; Kriegman, D.J.; Ahuja, N. Appearance-based Eye Gaze Estimation. In Proceedings of the IEEE Workshop on Applications of Computer Vision, Orlando, FL, USA, 3–4 December 2002; p. 191.
22. Lu, F.; Chen, X.; Sato, Y. Appearance-Based Gaze Estimation via Uncalibrated Gaze Pattern Recovery. *IEEE Trans. Image Process.* **2017**, *26*, 1543–1553. [[CrossRef](#)] [[PubMed](#)]
23. Durna, Y.; Ari, F. Design of a Binocular Pupil and Gaze Point Detection System Utilizing High Definition Images. *Appl. Sci.* **2017**, *7*, 498. [[CrossRef](#)]
24. Sun, L.; Liu, Z.; Sun, M.T. Real time gaze estimation with a consumer depth camera. *Inf. Sci.* **2015**, *320*, 346–360. [[CrossRef](#)]
25. Li, J.; Li, S. Gaze Estimation from Color Image Based on the Eye Model with Known Head Pose. *IEEE Trans. Hum-Mach. Syst.* **2016**, *46*, 414–423. [[CrossRef](#)]
26. Cazzato, D.; Leo, M.; Distanto, C. An investigation on the feasibility of uncalibrated and unconstrained gaze tracking for human assistive applications by using head pose estimation. *Sensors* **2014**, *14*, 8363–8379. [[CrossRef](#)] [[PubMed](#)]
27. Sun, L.; Song, M.; Liu, Z.; Sun, M. Real-time gaze estimation with online calibration. *IEEE Multimedia* **2014**, *21*, 28–37. [[CrossRef](#)]
28. Funes-Mora, K.A.; Odobez, J.M. Gaze Estimation in the 3D Space Using RGB-D Sensors. *Int. J. Comput. Vis.* **2016**, *118*, 194–216. [[CrossRef](#)]
29. Li, B.; Fu, H.; Wen, D.; LO, W. Etracker: A Mobile Gaze-Tracking System with Near-Eye Display Based on a Combined Gaze-Tracking Algorithm. *Sensors* **2018**, *18*, 1626. [[CrossRef](#)] [[PubMed](#)]
30. Huang, C.W.; Tan, W.C. An approach of head movement compensation when using a head mounted eye tracker. In Proceedings of the IEEE International Conference on Consumer Electronics-Taiwan 2016, Nantou, Taiwan, 27–29 May 2016.
31. Su, D.; Li, Y.F. Toward flexible calibration of head-mounted gaze trackers with parallax error compensation. In Proceedings of the IEEE International Conference on Robotics and Biomimetics 2016, Qingdao, China, 3–7 December 2016; pp. 491–496.
32. Pirri, F.; Pizzoli, M.; Rudi, A. A general method for the point of regard estimation in 3D space. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, CO, USA, 20–25 June 2011; pp. 921–928.
33. Wan, Z.; Wang, X.; Zhou, K.; Chen, X.; Wang, X. A Novel Method for Estimating Free Space 3D Point-of-Regard Using Pupillary Reflex and Line-of-Sight Convergence Points. *Sensors* **2018**, *18*, 2292. [[CrossRef](#)] [[PubMed](#)]
34. Zhang, X.; Li, B.; Yang, D. A novel harris multi-scale corner detection algorithm. *J. Electron. Inf. Technol.* **2007**, *29*, 1735–1738.
35. Haro, A.; Flickner, M.; Essa, I. Detecting and tracking eyes by using their physiological properties, dynamics, and appearance. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head, SC, USA, 15 June 2000; pp. 163–168.

