An Eye Gaze Tracking Method of Virtual Reality Headset Using A Single Camera and Multi-light Source

Shengfu Lu 1,2, Richeng Li 1,2, Jinan Jiao 1,2, Jiaming Kang 1,2, Nana Zhao 1,2 and Mi Li 1,2 *

1 Department of Automation, Faculty of Information Technology, Beijing University of Technology, Beijing, China
2 The Beijing International Collaboration Base on Brain Informatics and Wisdom Services, Beijing, China

* Corresponding author: limi@bjut.edu.cn

Abstract: For the problems the hardware configuration of multi-camera and multi-light source is complicated and the eye image of little auxiliary light source is too dark in the eye gaze tracking system, A new eye gaze tracking system based on two-dimensional mapping model is proposed, which achieved the transformation from two-dimensional coordinates to three-dimensional coordinates combined attitude angle of virtual reality headset. In this paper, the mapping relationship between visual features and eye gaze points is established by improving the method of pupil image recognition and eye gaze estimation based pupil–corneal, which the accuracy of eye gaze estimation is further improved. The experimental results show that the error of eye gaze tracking system is less than 1.1 °, which achieved the good effect of eye gaze tracking.

1. Introduction

As a new human-computer interaction mode, eye gaze estimation technology has the characteristics of directness, stability and bidirectional, which can enable users to obtain stable and efficient communication with computers in the most natural state. It has broad application prospects in virtual reality system, assisted driving, disease diagnosis, user experience and other fields.

In order to improve the robustness and accuracy of the eye gaze system, many studies have been carried out by adding multiple cameras and multiple light sources to create special corneal reflection patterns. Ohno et al. [1] used three cameras and one light source to detect the pupil using the method of bright and dark pupil, and used the vector relationship between the center of the pupil and the reflection point on the cornea to estimate the eye gaze. Beymer and Flickner[2] used 4 cameras and two light sources, This system used wide-angle stereoscopic cameras to detect the location of human faces, and then used narrow field stereoscopic cameras to track eye images. The estimation of the eye gaze was realized by building stereoscopic models of eyeballs. Although the above systems have obtained high estimation accuracy, the hardware configuration of the system is very complicated. The most typical eye gaze estimation model under a single camera and a single light source is a second-order polynomial model [3], while Cherif and Motsch use a higher-order polynomial model for the eye gaze estimation[4].

According to the above problem, this paper adopts the hardware configuration of a single camera and a light source is used to estimate the eye gaze. At the same time, aiming at the typical problem of
head movement affecting the accuracy based on polynomial fitting of single pupil-corneal vector, the virtual reality helmet can better keep the relative position of the head and the camera unchanged, which can better meet the system's head motion tolerance error ability.

2. A eye gaze tracking method of virtual reality headset

This paper aims at the problem of eye gaze estimation in virtual reality scene with a single camera and eight infrared light sources.

The experimental device: a virtual reality headset, a CMOS camera, a circle aperture of eight infrared light sources, a computer, a computer monitor. The camera resolution is 1280*720 pixels, and the wavelength of infrared light source is 850nm. Place the camera directly under the lens inside the helmet and fix the round infrared aperture around the lens. This system is mainly composed of eye image capture module, eye image feature detection module, eye gaze feature extraction module and eye gaze point estimation module.

3. Eye image feature detection

Accurate extraction of eye image feature parameters is a prerequisite for accurate eye gaze tracking estimation. In this paper, pupil ellipse parameters and light spots were extracted from eye images of infrared light source.

3.1. The pupil detection and fitting

According to the obvious difference between the gray value, size, shape of the pupillary region and the surrounding region, pupil parameters are extracted through the following image processing methods:

Step 1, as shown in Figure 1. Setting starting point for A (left and top) is 400 pixels(x-coordinate) and 300 pixels(y-coordinate), setting ending point for B (right and bottom) is 900 pixels(x-coordinate) and 700 pixels(y-coordinate). First, the eye image with a resolution of 500 * 400 is obtained by cutting processing; then, the eye image with a resolution of 250 * 200 is obtained by twice down sampling.

Step 2: image filtering, as shown in Figure 2. The interested region of pupil obtained in step 1 was first morphologically processed and then smoothed. Morphological and smooth processing of pupil interest area obtained in step 1. Eight infrared light sources mounted inside the helmet emit infrared light that bounces off the iris of the eye, creating light spots around the pupil image. In the process of pupil fitting, in order to avoid being affected by the spots, we firstly expanded them and then corrode them out through morphological process, and then carried out secondary denoising by smooth filtering to eliminate the noise, which made preliminary preparation for subsequent extraction of pupil parameters, light spot detection and other eye feature parameters.

Step 3: pupil extraction based on seed filling, as shown in Figure 3. Seed filling algorithm is also known as boundary filling algorithm. Its basic idea is: starting from one or a group of points in the
polygon region, the neighboring pixel points similar to the characteristics of the seed (such as color range and gray level) are added to the seed from the inside out, and the seed grows repeatedly until the specific conditions are met. Seed filling algorithm commonly used four connected domain and eight connected domain.

![Seed filling process (four-connected domain)](image)

**Figure 3.** Seed filling process (four-connected domain).

Step 4: Edge detection was performed on the pupil area obtained in step 3 to detect the edge points. The least square method was used to fit the ellipse of pupil region, and the center position of pupil ellipse \(A(\mathbf{X}_A, \mathbf{Y}_A)\) and the long, short axes were obtained. Ellipse fitting results for pupil region are shown in figure 4.

![Pupil parameter extraction](image)

**Figure 4.** Pupil parameter extraction.

### 3.2. Light spot detecting and fitting

How to detect these light spots is indispensable for line of sight tracking. The detection process is shown in Figure 5 below.

For the light spots are too small, it is difficult to extract light spot by ellipse fitting method and the accuracy is not high, we used the methods based on connected domain to extract the spot. First, after pupil detecting, a certain area is set around the pupil according to the size of its diameter to continue cutting, and the approximate area containing the spot is obtained. Secondly, choosing the appropriate threshold value to binarize the cropped eye image, obtaining the light spot area image and extracting the spots. Thirdly, through the connected component detection of the light spot image, the centroid of the extracted facula is calculated, and the noise points on the edge of the image are eliminated.

![Light spot detection process](image)

**Figure 5.** Light spot detection process.

### 4. Eye gaze estimation model

#### 4.1. Visual feature extraction

In order to improve the accuracy of the eye gaze estimation system, the least square method was used to fit the outer great circle. The coordinates of center \(B(\mathbf{X}_B, \mathbf{Y}_B)\) and the coordinates of pupillary center \(A(\mathbf{X}_A, \mathbf{Y}_A)\) in the image are. If the center \(B\) is taken as the reference point, then the relative offset
of the pupil center A is \((X_{BA}, Y_{BA})\), \(X_{BA} = X_A - X_B\), \(Y_{BA} = Y_A - Y_B\). The final visual feature \(S_{gaze} = (X_{BA}, Y_{BA})\), as shown in figure 6.

![Eye image](image1)

![The visual feature is extracted](image2)

**Figure 6.** The visual feature is extracted.

### 4.2. Eye gaze estimation (Establishment of eye gaze estimation model)

After extracting the visual feature, different mapping relations can be established to estimate the eye gaze[4]. The typical second-order polynomial model is as follows:

\[ S_x = a_0 + a_1 \cdot x_v + a_2 \cdot y_v + a_3 \cdot x_v \cdot y_v + a_4 \cdot x_v^2 + a_5 \cdot y_v^2 \]  

\[ S_y = b_0 + b_1 \cdot x_v + b_2 \cdot y_v + b_3 \cdot x_v \cdot y_v + b_4 \cdot x_v^2 + b_5 \cdot y_v^2 \] 

\((x_v, y_v)\) is visual feature \((X_{BA}, Y_{BA})\), \((S_x, S_y)\) is eye gaze coordinates. \(a=(a_0, a_1, a_2, a_3, a_4, a_5)\), \(b=(b_0, b_1, b_2, b_3, b_4, b_5)\) is the coefficient to be solved.

The solution to \(a=(a_0, a_1, a_2, a_3, a_4, a_5)\) is transformed into solving the following equations:

\[
\begin{bmatrix}
1 & x_{v1} & y_{v1} & x_{v1}^2 & y_{v1}^2 \\
1 & x_{v2} & y_{v2} & x_{v2}^2 & y_{v2}^2 \\
1 & x_{v3} & y_{v3} & x_{v3}^2 & y_{v3}^2 \\
1 & x_{v4} & y_{v4} & x_{v4}^2 & y_{v4}^2 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
1 & x_{vn} & y_{vn} & x_{vn}^2 & y_{vn}^2 \\
\end{bmatrix} \begin{bmatrix}
a_0 \\
a_1 \\
a_2 \\
a_3 \\
a_4 \\
a_5 \\
\end{bmatrix} = \begin{bmatrix}
x_0 \\
x_1 \\
x_2 \\
x_3 \\
x_4 \\
x_n \\
\end{bmatrix} 
\]  

(3)

For this overdetermined system, the least squares algorithm is used to obtain the optimal solution of \(a\), and similarly, the optimal solution of \(b\) can be obtained.

### 4.3. Nine points calibration positioning method

The eye gaze tracking adopts nine-point positioning method, as shown in figure 7. Eye gaze data was collected successively, and each point lasted for 5s. The second-order polynomial model is calculated after all data collection is completed.

![Nine points calibration process](image3)

**Figure 7.** Nine points calibration process.

According to the parameters obtained after calibration and the visual feature vector, the real-time 2D eye gaze point data is calculated, and the 3D coordinate position is calculated through the coordinate system transformation.
5. Coordinate transformation
The conversion process from 2D coordinates to 3d coordinates is shown in figure 8.

5.1. Coordinate system and transformation relation
The world coordinates of the VR are related to the base position of the helmet. We take the world coordinate system as the coordinate system of the helmet at the base position, as shown in figure 9.

The positive of x-axis is the left to right side of helmet; The positive of y-axis is the bottom to top side of helmet; The positive of z-axis is the back to front side of helmet.

5.2. The transformation of eye gaze coordinates to camera coordinate system
As shown in figure 10, point A($U_A, V_A$) is the point in the UOV. The UOV is perpendicular to the z axis in the camera coordinate system, and the corresponding coordinate point in the camera coordinate system is ($X_A, Y_A, Z_A), X_A=U_A, Y_A=V_A$.

5.3. Camera coordinate system to inertial coordinate system
The helmet attitude angle indicates the rotation from the inertial coordinate system to the camera coordinate system. The rotation of coordinate system is equivalent to the points in space with opposite rotation amount. From the inertial coordinate system to the camera coordinate system, rotate in the order of y-axis, x-axis and z-axis. The reverse rotation is the reverse rotation along z-axis, x-axis and y-axis. The rotation matrix is $R(-\theta_z', -\theta_x', -\theta_y') = R_y(-\theta_y')R_x(-\theta_x')R_z(-\theta_z')$.

$$R_x(-\theta_x') = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_x' & -\sin \theta_x' \\ 0 & \sin \theta_x' & \cos \theta_x' \end{bmatrix}$$

$$R_y(-\theta_y') = \begin{bmatrix} \cos \theta_y' & 0 & \sin \theta_y' \\ 0 & 1 & 0 \\ -\sin \theta_y' & 0 & \cos \theta_y' \end{bmatrix}$$

$$R_z(-\theta_z') = \begin{bmatrix} \cos \theta_z' & -\sin \theta_z' & 0 \\ \sin \theta_z' & \cos \theta_z' & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Finally, the eye gaze points (x, y, z) in the camera coordinate system are transformed into the
inertial coordinate system by rotation:

\[
\begin{bmatrix}
    x_o \\
y_o \\
z_o
\end{bmatrix} = R(\theta_z, \theta_x, \theta_y) \begin{bmatrix}
x \\
y \\
z
\end{bmatrix}
\]  \hspace{1cm} (7)

6. Results

6.1. Eye gaze estimation results of different users

For eye gaze accuracy evaluation test, we use the method of single calibration and single test. In order to better verify the effectiveness of the eye gaze tracking algorithm, four more points are added to the nine points in the calibration during the system accuracy test, marked with black box. The distribution of test points is shown in Figure 11, with red point as the evaluation point and blue point as the user's eye gaze point. During the test, each test point presents 5s in turn. Since the frame rate of the camera is 25Hz and 25 pictures are collected per second, each collection point can collect about 125 pictures, and each tested sample is about 1625 (125 × 13).

![Figure 11. Eye gaze accuracy assessment process.](image)

The test results of eye gaze accuracy are shown in Figure 12. In the figure, the red circle represents the standard test point, and the blue dot represents the prediction point calculated by the eye gaze model. It can be seen that most of the prediction points are around the red circle, and there are individual points with large deviation from the actual points. The average accuracy of eye gaze estimation is 1.08 degrees.

![Figure 12. Distribution of eye gaze points for six participants.](image)
6.2. Comparison with other systems

Table 1. Comparison with other eye gaze estimation systems.

| Method | Number of cameras | Number of light sources | Average precision |
|--------|-------------------|-------------------------|-------------------|
| [2]    | 2                 | 5                       | 0.90°             |
| [3]    | 4                 | 2                       | 0.60°             |
| [5]    | 1                 | 1                       | 1.80°             |
| [6]    | 1                 | 2                       | 1.67°             |
| [7]    | 1                 | 2                       | 1.30°             |
| [8]    | 1                 | 2                       | 1.11°             |
| Ours   | 1                 | 8                       | 1.08°             |

We compare this method with other eye gaze estimation methods, as shown in Table 1. The calculation formula of angle error $\theta$ is as follows: $\theta = (\theta_x + \theta_y)/2$, in which the angle error of horizontal direction is $\theta_x = \tan^{-1}(d_x/d_z)$, the angle error of horizontal direction is $\theta_y = \tan^{-1}(d_y/d_z)$. Although the method in reference [2] [3] gives better results than the method in this paper, the method in this paper has more advantages in the number of cameras, while avoiding the head motion error [9], and the accuracy is also satisfactory.

7. Conclusion

Aiming at the problem of complex configuration based on multi cameras and low accuracy of eye gaze estimation in a single camera and a single light source system, a cost-effective solution of eye gaze tracking based on a single camera and multi light sources in virtual reality environment is proposed. Based on the improved eye feature detection method and the improved pupil cornea reflection polynomial model, the mapping relationship between the eye feature and the eye gaze point is established, which further improves the estimation accuracy. At the same time, the virtual reality helmet can also avoid the small-scale movement of the user's head. The experimental results show that the average estimation error of the eye gaze tracking system is 1.08°, which can effectively estimate the eye gaze.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (61602017), the National Basic Research Programme of China (2014CB744600), 'Rixin Scientist' Foundation of Beijing University of Technology (2017-RX(1)-03), the Beijing Natural Science Foundation (4164080), the Beijing Outstanding Talent Training Foundation (2014000020124G039), the National Natural Science Foundation of China (61420106005), the International Science & Technology Cooperation Program of China (2013DFA32180).

Reference

[1] Ohno T, Mukawa N and Kawato S 2003 Just blink your eyes: A head-free gaze tracking system (CHI'03 extended abstracts on Human factors in computing systems) pp 950-957
[2] Beymer D and Flickner M 2003 Eye gaze tracking using an active stereo head vol 2 (Computer Society Conference on Computer Vision and Pattern Recognition) p II-451
[3] Ramanauskas N 2006 Calibration of video-oculographical eye-tracking system (Electronics & Electrical Engineering) 72(8) pp 65-68
[4] Cherif Z R, Motsch J F and Krebs M O 2002 An adaptive calibration of an infrared light device used for gaze tracking vol 2 (IEEE Instrumentation and Measurement Technology Conference) pp 1029-1033
[5] Zhang C, Chi J N and Zhang Z H 2010 A novel eye gaze tracking technique based on pupil center cornea reflection technique (Chinese Journal of Computers) 33(7) pp 1272-1285
[6] Zhu Z and Ji Q 2007 Novel eye gaze tracking techniques under natural head movement (Biomedical Engineering) 54(12) pp 2246-2260

[7] Morimoto C H, Amir A and Flickner M 2002 Detecting eye position and gaze from a single camera and 2 light sources vol 4 (Pattern Recognition 16th International Conference) pp 314-317

[8] Shin Y G, Choi K A and Kim S T 2015 A novel 2-D mapping-based remote eye gaze tracking method using two IR light sources (IEEE International Conference on Consumer Electronics) pp 190-191

[9] Wang K and Ji Q 2017 Real time eye gaze tracking with 3d deformable eye-face model. (IEEE International Conference on Computer Vision) pp 1003-1011