Available online at www.ijpe-online.com

International Journal of Performability Engineering

vol. 15, no. 11, November 2019, pp. 2891-2898
DOI: 10.23940/ijpe.19.11.p7.28912898

Gaze-Tracking Algorithm based on Infrared Gray Image

Nan Xue, Siwen Cai, Yuanyuan Chen, Minglei Shao, and Peng Wang*

School of Electrical and Electronic Engineering, Harbin University of Science and Technology, Harbin, 150080, China

Abstract

Gaze-tracking technology is a new type of HCI (human-computer interaction) technology that has great application potential in the field of intelligent control. According to the characteristics of the reflected spot formed by the human eye under the irradiation of infrared light, a gaze-tracking algorithm was designed based on the pupil corneal reflection principle. Firstly, the human eye is quickly located according to the characteristics of the reflected spot to obtain the eye sub-image, and the eye sub-image is binarized using the image segmentation method based on threshold iteration. Then, the edge detection algorithm is used to obtain the pupil and reflected spot edge, and the geometric center coordinates of the pupil and the reflected spot are obtained by ellipse fitting. Finally, the relationship between the gaze and the gaze point is obtained by view-point mapping based on the polynomial function, so as to realize gaze-tracking. The simulation results show that the gaze point's maximum vertical error is seven pixels and maximum horizontal error is nine pixels. It can track the gaze point in real time and has the characteristics of rapidness and high precision. It has application value in human-computer interaction.

Keywords: gaze-tracking; pupil corneal reflection principle; threshold iteration; view-point mapping

(Submitted on September 16, 2019; Revised on September 29, 2019; Accepted on October 23, 2019)

1. Introduction

Gaze-tracking is achieved by measuring eye movements and states or the movement of the eye relative to the head. Its purpose is to determine the direction of the gaze. It has good characteristics such as directness, objectivity, and rapidity [1]. Due to the development of smart home and virtual reality technology in recent years, its application area extends from early psychology and physiology to human-computer interactions, business, entertainment, and more. [2]. Therefore, gaze-tracking technology has broad application prospects.

Early research on gaze-tracking focused on direct observation and mechanical recording. With the continuous development of science and technology, three methods have been proposed, which include electromagnetic recording, current recording, and optical recording. Electromagnetic recording and current recording need to be in contact with the eyes or around the eyes, so both methods can have an intrusive effect on the eyes. These will also affect the experimental results and experimenter's comfort. Therefore, optical recording methods with non-contact and real-time properties are widely used. Optical recording uses a camera or video camera to record the movement of the eye to determine the direction of the gaze through analysis [3].

With the continuous development of gaze-tracking technology, gaze is input as human-computer interaction, and human-computer interaction can be achieved quickly and efficiently. The existing human-computer interaction methods mainly include mechanical keys, touch panels, speech recognition, and so on. However, all of them have shortcomings in special circumstances. For example, it is difficult to use mechanical keys and touch panels for people with upper limb disabilities. In noisy conditions, the accuracy of speech recognition is lower. The method also does not apply to deaf people. Overall, it can be seen that studying the human-computer interactions of gaze-tracking technology has great research value [4].

This paper proposes a gaze-tracking algorithm based on image processing technology. According to the pupil corneal

* Corresponding author.
E-mail address: wpkunpeng@hrbust.edu.cn

reflection method, the algorithm determines the geometric centre of the pupil and spot of light to achieve gaze-tracking. It has the advantages of high speed and accuracy. The algorithm can be applied to human-machine interactive devices and games. Users can have a more comfortable human-computer interaction experience with this algorithm.

2. The Design of Eye Gaze-Tracking Algorithm

A gaze-tracking algorithm based on pupil corneal reflection principle is presented in this paper, and its flow is shown in Figure 1, including human eye coarse positioning, image binarization, spot and pupil edge extraction, spot and pupil centre positioning, and coordinate mapping.

2.1. Eye Coarse Positioning

When the near-infrared light source with a safe range is used to irradiate the human face, an infrared light spot of great brightness is formed on the cornea of the human eye [5]. Through MATLAB simulation software for the collected infrared image, it was found that due to the existence of the reflection spot in the eye, the brightness of the pixel at the spot is greater than that of other areas, and the distance between the left and right eye reflection spots is more stable. Thus, by this feature, we can perform coarse human eye positioning [6].

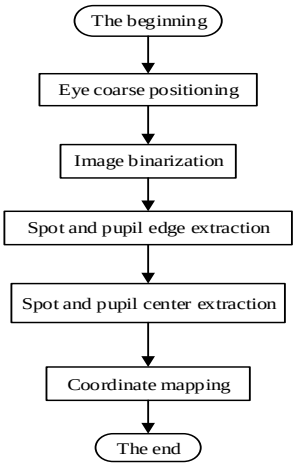
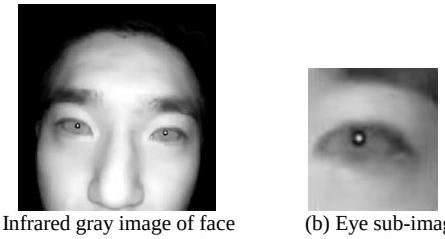


Figure 1. Process of gaze-tracking algorithm

Firstly, a horizontal one-dimensional mask operator [-1,-1,0,1,1] is used to scan each column of the grayscale image to calculate the maximum gradient of the column. Then, the original image around the average is divided into two parts to search the maximum gradient of the brightness, so as to locate the column coordinates of the infrared spot. Secondly, in the column of infrared light spots, the maximum brightness gradient is calculated by using the longitudinal one-dimensional mask operator $[-1,-1,0,1,1]^T$ to locate the raw coordinate of the light spot. The distance between two light spots is calculated, and the fault tolerance processing is based on prior knowledge. Finally, the eye is coarsely positioned.

After the coarse positioning of the eye, in order to improve the processing speed of the system, it is necessary to process the image near the pupil and spot. Therefore, the human eye sub-image is extracted with the centre of the reflected spot. The size of the face image before extraction is 640×360 pixels, as shown in Figure 2(a). After extraction, the size of the eye sub-image is 60×60 pixels. It is depicted in Figure 2(b).

Because the algorithm uses a one-dimensional mask operator, it can be accurately positioned even if the scanning calculation is small. After rough positioning processing, the extracted eye sub-image can be used for the subsequent analysis.



(a) Infrared gray image of face (b) Eye sub-image

Figure 2. The result of coarse localization of eye

2.2. Image Binarization based on Threshold Iteration

After coarse localization of the eye, we need to select the appropriate threshold to obtain the binary image of the eye sub-image to facilitate the extraction of the pupil contour. Among the existing global binarization algorithms, OTSU is the most commonly used [7]. To improve the performance of the algorithm in continuously changing the environment of the scene, this paper proposes a method of image binarization based on threshold iteration for OTSU.

The principle of the OTSU method is to find the best threshold value, divide the image into two parts, and make it reach the maximum variance between classes. In an image, the range of grayscale values is (0, 1, ⋯, *k*-1). The number of pixels with grayscale i is n_i, and the total number of pixels is $N = \sum_{i=0}^{k-1} n_i$. Therefore, the probability of grayscale i pixels is $p_i = n_i /N p_i = n_i/N$.

Selecting threshold t, the gray image pixels are divided into two categories: C_0 and C_1. The possibilities for them are respectively given by Equation (1).

$$p_0(t) = \sum_{i=0}^{t} p_i \quad , \qquad p_1(t) = \sum_{i=t+1}^{k-1} p_i \tag{1}$$

The gray mean value is described by Equation (2).

$$\begin{aligned} u_0(t) &= \sum_{i=0}^{t} ip_i \,/p_0(t) \\ u_1(t) &= \sum_{i=t+1}^{k-1} ip_i \,/p_1(t) \end{aligned} \tag{2}$$

Then, u_T is the total gray level of the image, as shown in Equation (3).

$$u_T = \sum_{i=0}^{k-1} ip_i \tag{3}$$

Thus, the variance between classes C_0 and C_1 is given as Equation (4).

$$\sigma_B^2(t) = p_0(t)[u_0(t) - u_T]^2 + p_1(t)[u_1(t) - u_T]^2 \tag{4}$$

The process of finding the maximum value of the variance between classes is the process of automatically determining the threshold. The threshold t is expressed as Equation (5).

$$t = tan^{-1}\, max\, \sigma_B^2(t) \tag{5}$$

Assuming that both the background and target components in a gray image are normally distributed, the standard deviations of the two components are formulated as Equation (6).

$$\begin{aligned} \sigma_0(t) &= \sqrt{\sum_{i=0}^{t} [i - u_0(t)]^2 p_i / \sum_{i=0}^{t} p_i} \\ \sigma_1(t) &= \sqrt{\sum_{i=t+1}^{k-1} [i - u_1(t)]^2\, p_i / \sum_{i=t+1}^{k-1} p_i} \end{aligned} \tag{6}$$

When the target and background can be separated, the mean and standard deviation can be characterized by Equation (7).

$$u_1 - u_0 > \alpha(\sigma_0 + \sigma_1) \tag{7}$$

The parameter α is based on the gray distribution of the background and the target. The flow chart of the algorithm is shown in Figure 3.

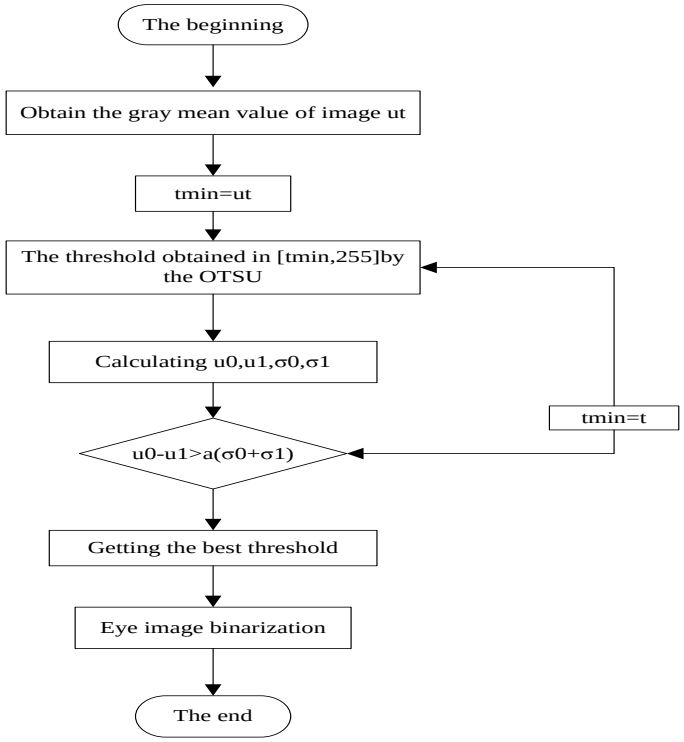


Figure 3. Improved OTSU image segmentation process

When the image is segmented, the mean and standard deviation of gray distribution of both the target and the background must meet Equation (7). Then, the target and the background of image are completely divided. If the condition is not satisfied, the current threshold value t is set to the lower limit of the gradation. The threshold t_{nt} in the new gradation range is obtained, and the segmentation is performed again. Until Equation (7) is satisfied, the correct segmentation results are obtained. Eye segmentation results are presented in Figure 4. Figure 4(a) shows the traditional OTSU algorithm. As a single threshold, it cannot adapt to changes in ambient light. Therefore, the eyelid and other parts of the face cannot be removed. Figure 4(b) shows that the threshold after three iterations satisfies Equation(7), removing the other parts of the pupil and the spot.

(a) Traditional OTSU algorithm (b) Improved OTSU algorithm

Figure 4. Pupil segmentation

2.3. Pupil and Spot Center Edge Extraction

The edge detection algorithm based on the Canny operator is used to extract the pupil and spot edges. The Canny operator is an optimization operator that integrates functions such as filtering, enhancement, and edge detection. It is more suitable for finding weak edges, with high accuracy and strong anti-interference ability. Then, the edge detection effect is optimal. During the filtering process of the Canny operator, the image is first smoothed by the Gauss function, and then the result of the filtering is to solve the maximum of the first derivative, and the maximum value is denoted as the edge point [8]. Its steps are as follows:

Step 1 The purpose of using the Gauss template is to eliminate noise. The process is image convolution. The expression of the template in the continuous space of the Gauss is shown as Equation (8).

$$G(x,y) = exp|-(x^2+y^2)/2\sigma^2| \tag{8}$$

Step 2 The partial derivatives of the image gray level in the *x* and *y* directions are calculated respectively, and then the size of the luminance gradient is obtained by Equation (9).

$$|G| = \sqrt{G_x^2 + G_y^2} \tag{9}$$

Step 3 The vector of gradient direction is shown as Equation (10).

$$\theta = tan^{-1}(G_x/G_y) \tag{10}$$

Finally, we find the horizontal, vertical, 45 °, 135 °, and adjacent pixels of the four gradient directions.

Step 4 The edge of the image is determined by the double threshold algorithm. Firstly, a threshold interval is established, and then the pixel gradient is compared with the threshold interval. Larger than the upper limit of the interval is edge and less than the lower limit of the interval is not edge. In the interval, we need to see if the adjacent pixel gradient of the current pixel is greater than the upper limit of the interval, and if so, then determine the edge. Otherwise, it is not.

After Canny edge extraction to obtain the binary image of eye sub-images is performed, the results shown in Figure 5 are used to obtain clear pupil and spot edges.

Figure 5. The result of pupil and spot edge extraction

2.4. Pupil and Spot Center Coordinates Extraction

After obtaining the edge of the pupil and spot, it is necessary to ellipse the edge to find the geometric center of the pupil and spot [9-10]. The fitting method uses the least squares method. The equation of plane quadratic curve of this method is given as Equation (11).

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0 \tag{11}$$

Consider that the zero solution needs to be avoided to ensure that the synthesized curve is an ellipse. The square sum of algebraic distances from point to quadratic curve needs to be minimized.

$$f(A,B,C,D,E,F) = \sum_{i=1}^{N} (Ax^2 + Bxy + Cy^2 + Dx + Ey + F)^2 \tag{12}$$

Infer from the extreme principle to minimize Equation (12), therefore:

$$\frac{\partial f}{\partial B} = \frac{\partial f}{\partial C} = \frac{\partial f}{\partial D} = \frac{\partial f}{\partial E} = \frac{\partial f}{\partial F} = 0 \tag{13}$$

A multi-parameter linear system is derived from the above Equation (13). The parameters of the ellipse, such as the focus, major axis, minor axis, and rotation angle can be represented by the coefficients of the equations.

The above-mentioned pupil center localization algorithm can also be applied to the solution of the spot coordinates. As a result of the larger difference of brightness gradient between the human eye image pupil and spot, by image binarization, the pupil and spot are presented as black and white. In addition, the radius of the spot is much smaller than that of the pupil, and the pupil and spot can be easily distinguished. The geometric center coordinates of the pupil and spot are extracted as shown in Figure 6. The yellow cross is marked as the pupil geometric center, and the red cross is marked as the center of the spot.

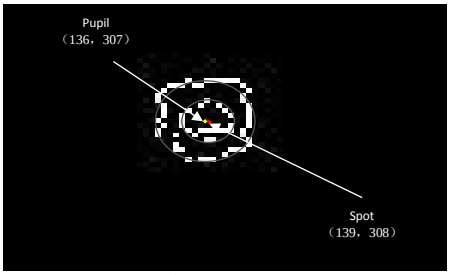


Figure 6. The extraction result of pupil and spot geometric center coordinates

2.5. View-Point Mapping

After obtaining the geometric centre coordinates of the pupil and spot, a view-point mapping function needs to be established to completely map the line of sight features to the display device, thereby realizing the visualization of human eye gaze-tracking [11].

The process of view-pointing mapping is firstly calibrated, giving some definite reference fixation points on the screen. When the user looks at these reference points, the eye tracking system detects the relative displacement vector of the centre of the spot and pupil (x_e, y_e), and the expression is shown as Equation (14).

$$\begin{cases} x_e = x_p - x_r \\ y_e = y_p - y_r \end{cases} \tag{14}$$

Among them, the pupil centre coordinate is (x_p, y_p), and the reflected light spot centre coordinate is (x_r, y_r), Then the least squares method is used to solve the mapping function of the relative displacement vector of the centre of the pupil and spot and gaze point. The relative displacement vector of the centre of the pupil and spot in the eye image is brought into the mapping function that has been obtained, and the coordinates of the view-point of the human eye are calculated and superimposed on the display device. Too many points of the mapping coordinates will inconvenience users. In order to improve the efficiency of the algorithm, a six-parameter fitting function and a nine-point mapping algorithm are selected in the algorithm to obtain the mapping function [12]. The mapping function is shown as Equation (15).

$$\begin{aligned} x_s &= a_0 + a_1 x_e + a_2 y_e + a_3 {x_e}^2 + a_4 x_e y_e + a_5 {y_e}^2 \\ y_s &= b_0 + b_1 x_e + b_2 y_e + b_3 {x_e}^2 + b_4 x_e y_e + b_5 {y_e}^2 \end{aligned} \tag{15}$$

Where a_0, a_1, ⋯, a_4, a_5 and b_0, b_1, ⋯, b_4, b_5 are undetermined coefficients of the variables x_e and y_e. Thus, the coefficients a_0, a_1, ⋯, a_4, a_5 and b_0, b_1, ⋯, b_4, b_5 of the equations need to be solved to build a complete mapping function.

3. Experiment and Results Analysis

The following host configuration is used for the experiment: Inter Core i5 CPU 2.5GHz, and 4GB RAM. In the MATLAB R2012a platform, a number of images of the user's gaze at the reference point of the screen is simulated. The image size is 640×360.

Firstly, the user looks at nine reference points each time, as shown in Figure 7. Six groups are collected to obtain 54 eye images [13]. The coefficients of the coordinate mapping equation are solved by the method proposed in Section 2.5, obtaining the mapping shown in Equation (16).

$$\begin{aligned} x_s &= 54.240311 + 6.45539 x_e + 0.513897 y_e + 0.011126 {x_e}^2 - 0.035126 x_e y_e - 0.009393 {y_e}^2 \\ y_s &= -64.20675 - 0.453292 x_e + 8.888434 y_e + 0.019451 {x_e}^2 - 0.00197 x_e y_e - 0.05826 {y_e}^2 \end{aligned} \tag{16}$$

Secondly, the users pay attention to four reference points. Each reference point collects eye images of six groups. The pupil centre coordinates and the spot centre coordinates are input into the coordinate mapping equation to estimate the true gaze error, as shown in Figure 7.

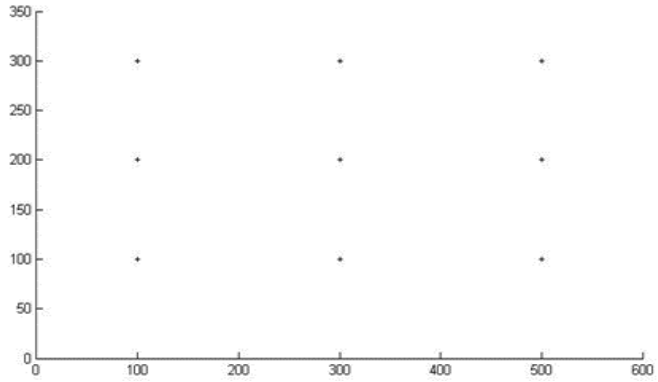


Figure 7. Reference point distribution

The experimental results are shown in Figure 8. This line-of-sight tracking algorithm can achieve the tracking of the line of sight. When the subjects looked at the (100,100), (300,100), and (500,300) points, there was a large deviation. The data is shown in Table 1. The maximum longitudinal error is seven pixels, and the maximum row error is about nine pixels. The main reason is that during the experiment, when the experimenter looked at the reference point at the edge of the screen, the head would inadvertently move slightly, which partially impacted the experimental results.

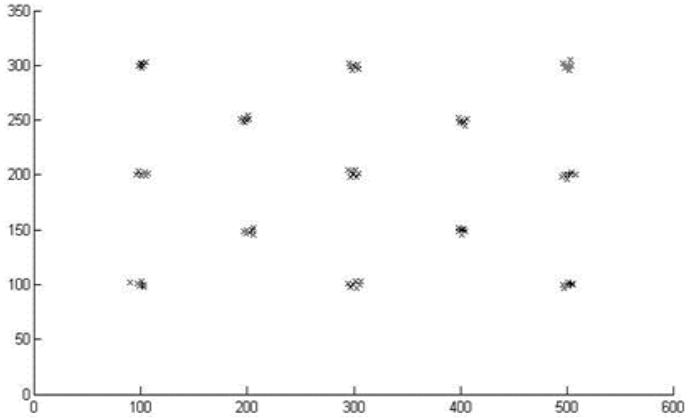


Figure 8. Reference point and gaze point distribution

Table 1. Partial view coordinate record

Marked point	**(100, 100)**	**(100, 200)**	**(500, 300)**
1 fixation	(91,105)	(101, 204)	(504, 393)
2 times watching	(95, 93)	(105, 201)	(496, 305)
3 times watching	(103,96)	(109, 204)	(499, 304)
4 times watching	(103,105)	(106, 203)	(504, 293)
5 times watching	(103,97)	(105, 204)	(506, 295)
6 times watching	(98, 107)	(94, 196)	(501, 304)

4. Conclusions

In this paper, we used the characteristics of human eyes spots under infrared light to design a gaze-tracking algorithm based on threshold iteration. A one-dimensional mask operator was used to scan the human face infrared gray image to quickly and accurately locate the human eye area. An improved OTUS algorithm based on threshold iteration was used to segment the eye sub-images. The Canny edge detection algorithm was used to calibrate the edges of the pupil and spot, and the geometric centre coordinates of the pupil and spot were calculated using the ellipse fitting of the least square method.

Finally, by building the mapping function of the relative displacement vector of the pupil spot centre and the reference point of the display screen, the human eye could be tracked. This algorithm meets the requirement of high precision in eye-tracking. At the same time, it can be used in conjunction with eye-opening detection algorithms to allow users to control targets though eye-tracking and active blinking action to achieve a complete human-computer interaction function, thereby broadening the application prospect of eye-tracking technology.

References

1. R. Liu, H. Jin, and X. R. Wu, "Real Time Auto-Focus Algorithm for Eye Gaze Tracking System," *2007 International Symposium on Intelligent Signal Processing and Communications Systems*, pp. 742-745, Xiamen, China, November 2007
2. C. L. Kleinke, "Gaze and Eye Contact: A Research Review," *Psychological Bulletin*, Vol. 100, pp. 78-100, 1986
3. C. Hennessey, "Noncontact Binocular Eye-Gaze Tracking for Point-of-Gaze Estimation in Three Dimensions," *IEEE Transactions on Biomedical Engineering*, Vol. 56, No. 3, pp. 790-799, March 2009
4. A. Wojciechowski and K. Fornalczyk, "Single Web Camera Robust Interactive Eye-Gaze Tracking Method," *Bulletin of the Polish Academy of Sciences*: *Technical Sciences*, Vol. 63, No. 4, pp. 879-886, 2015
5. C. H. Morimoto, D. Koons, A. Amir, and M. Flickner, "Pupil Detection and Tracking using Multiple Light Sources," *Image and Vision Computing*, Vol. 18, No. 4, pp. 331-335, March 2000
6. H. P. Shen, H. J. Feng, and Z. H. Xu, "Gaze Tracking System based on Pupil Detection," *Guangdianzi Jiguang/Journal of Optoelectronics Laser*, Vol. 16, No. 8, pp. 961-964, August 2005
7. R. Liu, X. Zhou, N. Wang, and M. M. Zhang, "Adaptive Regulation of CCD Camera for Real Time Eye Tracking," *Multimedia Tools and Applications*, Vol. 52, No. 1, pp. 33-43, March 2011
8. P. Jinwoo, K. Yong-Moo, and S. Kwanghoon, "Gaze Tracking System using Single Camera and Purkinje Image," in *Proceedings of the 2005 International Conference on Augmented Tele-Existence*, pp. 280, Christchurch, New Zealand, December 2005
9. H. Y. Dong and X. Wang, "Wave Gate Tracking Method based on Locus Fitting for UAV Following the Moving Target," *Applied Mechanics and Materials*, Vol. 513-517, pp. 4366-4371, January 2014
10. K. R. Park, "A Real-Time Gaze Position Estimation Method based on a 3-D Eye Model," *IEEE Transactions on Systems Man and Cybernetics Part B* (*Cybernetics*), Vol. 37, No. 1, pp. 199-212, February 2007
11. K. Ishac and D. Rozado, "Low Cost Human-Robot Gaze Estimation System," in *Proceedings of the 26th Australian Computer-Human Interaction Conference*, pp. 404-413, Sydney, Australia, December 2014
12. M. R. M. Mimica, "A Computer Vision Framework for Eye Gaze Tracking," in *Proceedings of 16th Brazilian Symposium on Computer Graphics and Image Processing*, pp. 406-412, Sao Carlos, Brazil, January 2003
13. C. Hennessey, B. Noureddin, and P. Lawrence, "Fixation Precision in High-Speed Noncontact Eye-Gaze Tracking," *IEEE Transactions on Systems Man and Cybernetics*, Vol. 38, No. 2, pp. 289-298, April 2008

Nan Xue received her M.Sc. degree in 2002 from Harbin University of Science and Technology. She is currently an associate professor at Harbin University of Science and Technology. Her main research interests include signal and information processing and image processing.

Siwen Cai received his B.Sc. degree in 2015 from Harbin University of Science and Technology. He is currently a graduate student at Harbin University of Science and Technology. His main research interests are embedded systems, eye-tracking technology, and smart home technology.

Yuanyuan Chen received her M.Sc. degree in 2019 from Harbin University of Science and Technology. Her main research interests include signal and information processing, eye-tracking technology, and smart home technology.

Minglei Shao received his M.Sc. degree in 2017 from Harbin University of Science and Technology. His main research interests are eye-tracking technology and smart home technology.

Peng Wang received his Ph.D. degree in 2013 from Harbin University of Science and Technology. He is currently an associate professor at Harbin University of Science and Technology. His main research interests include signal and information processing, embedded systems, biomedicine signal detection, and smart home technology.