

## Article

# 2D Gaze Estimation Based on Pupil-Glint Vector Using an Artificial Neural Network

Jianzhong Wang, Guangyue Zhang and Jiadong Shi \*

School of Mechatronical Engineering, Beijing Institute of Technology, 5 South Zhongguancun Street, Haidian District, Beijing 100081, China; cwjzwang@bit.edu.cn (J.W.); zhangguangyue@yeah.net (G.Z.)

\* Correspondence: 3120120135@bit.edu.cn; Tel.: +86-10-6891-5922

Academic Editor: Christian Dawson

Received: 7 April 2016; Accepted: 7 June 2016; Published: 14 June 2016

**Abstract:** Gaze estimation methods play an important role in a gaze tracking system. A novel 2D gaze estimation method based on the pupil-glnt vector is proposed in this paper. First, the circular ring rays location (CRRL) method and Gaussian fitting are utilized for pupil and glnt detection, respectively. Then the pupil-glnt vector is calculated through subtraction of pupil and glnt center fitting. Second, a mapping function is established according to the corresponding relationship between pupil-glnt vectors and actual gaze calibration points. In order to solve the mapping function, an improved artificial neural network (DLSR-ANN) based on direct least squares regression is proposed. When the mapping function is determined, gaze estimation can be actualized through calculating gaze point coordinates. Finally, error compensation is implemented to further enhance accuracy of gaze estimation. The proposed method can achieve a corresponding accuracy of  $1.29^\circ$ ,  $0.89^\circ$ ,  $0.52^\circ$ , and  $0.39^\circ$  when a model with four, six, nine, or 16 calibration markers is utilized for calibration, respectively. Considering error compensation, gaze estimation accuracy can reach  $0.36^\circ$ . The experimental results show that gaze estimation accuracy of the proposed method in this paper is better than that of linear regression (direct least squares regression) and nonlinear regression (generic artificial neural network). The proposed method contributes to enhancing the total accuracy of a gaze tracking system.

**Keywords:** human–computer interaction; gaze tracking; gaze estimation; pupil-glnt vector; direct least squares regression; improved artificial neural network

## 1. Introduction

Human beings acquire 80%–90% of outside information through the eyes. Humans' visual perception of information can be acquired through eye gaze tracking [1–4]. With the increasing development of computer/machine vision technology, gaze tracking technology has been more and more widely applied in the fields of medicine [5], production tests [6], human–machine interaction [7,8], military aviation [9,10], *etc.*

According to differences in dimension of gaze direction estimation, gaze tracking technology can be divided into 2D gaze tracking [11–19] and 3D gaze tracking [20–27]; according to differences in ways of wearing, gaze tracking technology can be classed as intrusive (head-mounted) [12,28–37] or non-intrusive (head-free) [20,23,38–44]. For different gaze tracking systems, gaze tracking methods mainly contain Limbus Tracking [45–47], Pupil Tracking [48–50], Pupil-glnt Vector [51–55], Purkinje Image [24,56,57], *etc.*

For 2D gaze estimation methods, mapping function between gaze points and target plane or regions of interest is firstly established. The mapping function solved is then further utilized to calculate the gaze point on certain targets or regions. For 3D gaze estimation methods, a human eyeball

model is employed to determine the absolute position of eyes in the test space. On this basis, 3D gaze is calculated to acquire the specific staring location or fixation targets of human eyes in the space.

The main purpose of this paper is to estimate the gaze point of the human eye on a monitor screen fixed to the head. The mapping function between gaze points and fixation targets is from plane to plane. Therefore, a novel 2D gaze estimation method based on pupil-glnt vector is proposed to calculate the gaze direction.

In conventional 2D gaze estimation methods, the most widely utilized calculation methods can be divided into two groups: linear regression (direct least squares regression) [58–64] and nonlinear regression (generic artificial neural network) [65–69]. In [58–61], Morimoto *et al.* utilize least squares to calculate the mapping function between calibration markers and corresponding pupil-glnt vectors. Overdetermined linear equations for solving mapping function are composed by a series of 2nd-order polynomials. The number of polynomials depends on the number of calibration markers. Corresponding coordinates of calibration markers, pupil and glnt centers are determined through a calibration process. The pupil-glnt vector is calculated through the subtraction of pupil and glnt center coordinates. Cherif *et al.* [62] propose an adaptive calibration method. A second time calibration is employed for error correction. A polynomial transformation of higher order is utilized to model mapping function by applying a mean square error criterion. The result of single calibration shows that the gaze estimation accuracy will increase with the enhancement of the polynomial order. However, through experimental analyses, Cerrolaza *et al.* [63,64] point out that the gaze estimation accuracy of a gaze tracking system will not increase with the enhancement of polynomial order owing to the factors of head motion, number of calibration markers, and calculating method of pupil-glnt vector, *etc.* When applied for solving mapping function, 2nd-order linear polynomial is the most widely used linear regression solution method with the advantages of fewer calibration markers and better approximation effect.

An artificial neural network is the most widely used nonlinear regression method for solving the mapping function between calibration markers and corresponding pupil-glnt vectors (or pupil centers, eye movements information, *etc.*). Early in 1994, Baluja and Pomerleau [65] proposed the method using simple artificial neural network (ANN) to estimate gaze direction. Multi-group attempts are conducted to find a training network with optimal performance. In the first attempt, images of only the pupil and cornea are utilized as inputs to ANN. The output units are organized with horizontal and vertical coordinates of the gaze point. A single divided layer is used for training in the ANN architecture. In the second attempt, in order to achieve a better accuracy, the total eye socket (including pupil and glnt position) is utilized as an input to ANN. A single continuous hidden layer and a single divided hidden layer are used for training in the ANN architecture. The experimental results show that when the hidden layer units are fewer, the training accuracy of the divided hidden layer is higher than that of the continuous hidden layer. In addition, the training time is short. Furthermore, some of the eye images are employed as training sets and the remainder are employed as testing sets, which provides more accurate experimental results. However, though a higher accuracy can be achieved when the total eye socket (including pupil and glnt position) is utilized as an input to ANN, the training sample data is huge and the training time is long. Piratla *et al.* [66] developed a network-based gaze tracking system. As an auxiliary tool, a strip with black and white bands is mounted on the user's head to facilitate real-time eye detection. Twelve items, consisting of strip edge coordinates at lower ends, eyeball centers coordinates, and eyelid distances, are the input features of the neural network. The X and Y coordinate pair of the point the user is looking at on the screen is the output of the neural network. A 25-neuron hidden layer is utilized between the input and output layer. This method requires a large number of input items and a long detection period. The real-time quality needs to be improved. Demjen *et al.* [67] compare the neural network and linear regression method utilized for estimating gaze direction. The comparison results show that: (1) the calibration procedure of the neural network method is faster as it requires fewer calibration markers, and (2) the neural network method provides higher accuracy. The gaze tracking performance of a neural network is better than

that of linear regression. Multi-layer perceptrons (MLPs) are utilized by Coughlin *et al.* [68] to calculate gaze point coordinates based on electro-oculogram (EOG). The number of input nodes depends on the number of data points chosen to represent the saccadic waveforms. The output nodes of the network provide the horizontal and vertical 2D spatial coordinates of the line-of-sight on a particular training or test trial. In order to determine the number of nodes that can provide the optimal outputs, hidden layers containing different numbers of nodes are selected to train MLP ANN. Initial weights trained on another person are referred to in order to reduce training time. The experimental results show that using MLPs for calibration appears to be able to overcome some of the disadvantages of the EOG and provides an accuracy not significantly different from that obtained with the infrared tracker. In addition, Sesin *et al.* [69] find that MLPs can produce positive effects: jitter reduction of gaze point estimation and enhancing the calculating stability of gaze points. Gneo *et al.* [70] utilize multilayer neural feedforward networks (MFNNs) to calculate gaze point coordinates based on pupil-glint vectors. Two separate MFNNs (each one having the same eye features as inputs, with one single output neuron directly estimating one of the X and Y coordinates of the POG), each containing 10 neurons in the hidden layer, are employed for training to acquire the outputs. The use of MFNNs overcomes the drawbacks of the model-based EGTs and the potential reasons for their failure, which sometimes give ANNs an undeservedly poor reputation. Zhu and Ji [71] utilize generalized regression neural networks (GRNNs) to calculate a mapping function from pupil parameters to screen coordinates in a calibration procedure. The GRNN topology consists of four layers: input layer, hidden layer, summation layer, and output layer. Six factors including pupil-glint vector, pupil ellipse orientation, *etc.* are chosen as the input parameters of GRNNs. The output nodes represent the horizontal and vertical coordinates of the gaze point. Though the use of hierarchical classification schemes simplifies the calibration procedure, the gaze estimation accuracy of this method is not perfect. Kiat and Ranganath [72] utilize two single radial basis function neural networks (RBFNNs) to map the complex and non-linear relationship between the pupil and glint parameters (inputs) to the gaze point on the screen (outputs). Both of the networks have 11 inputs including  $x$  and  $y$  coordinates of left and right pupils, pupil-to-glnt vectors of the left and right eyes, *etc.* The number of network output nodes depends on the number of calibration regions in the horizontal and vertical direction. The weights of the network are stored as calibration data for every subsequent time the user operates the system. As is the case with GRNNs, the gaze estimation accuracy of RBFNNs is not high enough. Wu *et al.* [73] employ the Active Appearance Model (AAM) to represent the eye image features, which combines the shape and texture information in the eye region. The support vector machine (SVM) is utilized to classify 36 2D eye feature points set (including eye contour, iris and pupil parameters, *etc.*) into eye gazing direction. The final results show the independence of the classifications and the accurate estimation of the gazing directions.

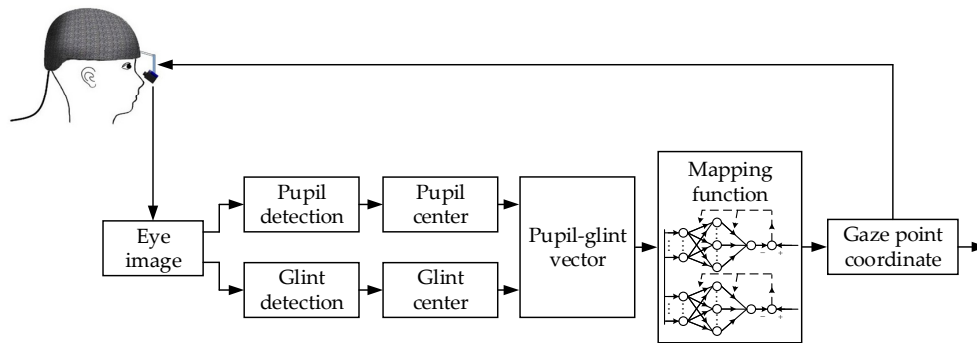
In this paper, considering the high speed of direct least squares regression and the high accuracy of artificial neural network, we propose an improved artificial neural network based on direct least squares regression (DLSR-ANN) to calculate the mapping function between pupil-glnt vectors and actual gaze points. Different from general artificial neural networks, coefficient matrix elements of direct least squares regression are employed as connection coefficients in the input and hidden layers of DLSR-ANN. The error cost function and continuous-time learning rule of DLSR-ANN are defined and calculated according to the constraint condition of solving direct least squares regression. The initial condition of an integrator associated with the learning rule of DLSR-ANN is acquired through linear polynomial calculation of direct least squares regression. The learning rate parameter is limited to a range determined by the maximal eigenvalue of auto-correlation matrix composed by input vector of direct least squares regression. The proposed method contains advantages of both direct least squares regression and artificial neural network.

The remainder of this paper is organized as follows: Section 2 presents the proposed neural network method for gaze estimation in detail. Section 3 describes the experimental system and shows the results. Section 4 concludes the whole work. The experimental results show that the training process of the proposed method is stable. The gaze estimation accuracy of the proposed method in

this paper is better than that of conventional linear regression (direct least squares regression) and nonlinear regression (generic artificial neural network). The proposed method contributes to enhance the total accuracy of a gaze tracking system.

## 2. Proposed Methods for Gaze Estimation

According to the respective characteristics of linear and nonlinear regression, a novel 2D gaze estimation method based on pupil-glnt vector is proposed in this paper. An improved artificial neural network (DLSR-ANN) based on direct least squares regression is developed to solve the mapping function between pupil-glnt vector and gaze point and then calculate gaze direction. The flow-process of gaze direction estimation is shown in Figure 1. First, when gazing at the calibration markers on the screen, corresponding eye images of subjects are acquired through a camera fixed on the head-mounted gaze tracking system. Second, through preprocessing such as Otsu optimal threshold binarization and opening-and-closing operation, pupil and glnt centers are detected by utilizing circular ring rays location (CRRL) method. As inputs of the proposed DLSR-ANN, pupil-glnt vector is calculated through the subtraction of pupil and glnt center coordinates. Third, a three-layer DLSR-ANN (input layer, hidden layer, and output layer) is developed to calculate the mapping function between pupil-glnt vectors and corresponding gaze points. Finally, gaze points on the screen can be estimated according to the mapping function determined.



**Figure 1.** Flow-process of gaze direction estimation.

Pupil-glnt vector is calculated through the subtraction of pupil and glnt center coordinate. 2nd linear gaze mapping function based on pupil-glnt vector is expressed as Equation (1).

$$\begin{cases} x_{ci} = a_1 + a_2x_{ei} + a_3y_{ei} + a_4x_{ei}y_{ei} + a_5x_{ei}^2 + a_6y_{ei}^2 \\ y_{ci} = b_1 + b_2x_{ei} + b_3y_{ei} + b_4x_{ei}y_{ei} + b_5x_{ei}^2 + b_6y_{ei}^2 \end{cases} \quad (1)$$

where  $i = 1, 2, \dots, N$ .  $N$  is the number of calibration markers.  $(x_{ci}, y_{ci})$  is the coordinate of gaze calibration markers on screen coordinate system.  $(x_{ei}, y_{ei})$  is the coordinate of pupil-glnt vector on image coordinate system. Least squares, as conventional linear methods, is utilized to solve the gaze mapping function shown in Equation (1). Residual error is defined as:

$$R^2 = \sum_{i=1}^N \left[ x_{ci} - \left( a_1 + a_2x_{ei} + a_3y_{ei} + a_4x_{ei}y_{ei} + a_5x_{ei}^2 + a_6y_{ei}^2 \right) \right]^2. \quad (2)$$

By calculating a partial derivative of  $a_j$  ( $j = 1, 2, 3, 4, 5, 6$ ) in Equation (2), the constraint condition can be obtained as in Equation (3).

$$\frac{\partial R^2}{\partial a_j} = -2 \sum_{i=1}^N \sigma_j \left[ x_{ci} - \left( a_1 + a_2x_{ei} + a_3y_{ei} + a_4x_{ei}y_{ei} + a_5x_{ei}^2 + a_6y_{ei}^2 \right) \right] = 0, \quad (3)$$

where  $\sigma_1 = 1$ ,  $\sigma_2 = x_{ei}$ ,  $\sigma_3 = y_{ei}$ ,  $\sigma_4 = x_{ei}y_{ei}$ ,  $\sigma_5 = x_{ei}^2$ ,  $\sigma_6 = y_{ei}^2$ . The value of  $a_j$  can be calculated according to Equation (4).

$$\begin{bmatrix} \sum_{i=1}^N x_{ci} \\ \sum_{i=1}^N x_{ci}x_{ei} \\ \sum_{i=1}^N x_{ci}y_{ei} \\ \sum_{i=1}^N x_{ci}x_{ei}y_{ei} \\ \sum_{i=1}^N x_{ci}x_{ei}^2 \\ \sum_{i=1}^N x_{ci}y_{ei}^2 \end{bmatrix} = \begin{bmatrix} N & \sum_{i=1}^N x_{ei} & \sum_{i=1}^N y_{ei} & \sum_{i=1}^N x_{ei}y_{ei} & \sum_{i=1}^N x_{ei}^2 & \sum_{i=1}^N y_{ei}^2 \\ \sum_{i=1}^N x_{ei} & \sum_{i=1}^N x_{ei}^2 & \sum_{i=1}^N x_{ei}y_{ei} & \sum_{i=1}^N x_{ei}^2y_{ei} & \sum_{i=1}^N x_{ei}^3 & \sum_{i=1}^N x_{ei}y_{ei}^2 \\ \sum_{i=1}^N y_{ei} & \sum_{i=1}^N x_{ei}y_{ei} & \sum_{i=1}^N y_{ei}^2 & \sum_{i=1}^N x_{ei}y_{ei}^2 & \sum_{i=1}^N x_{ei}^2y_{ei} & \sum_{i=1}^N y_{ei}^3 \\ \sum_{i=1}^N x_{ei}y_{ei} & \sum_{i=1}^N x_{ei}^2y_{ei} & \sum_{i=1}^N x_{ei}y_{ei}^2 & \sum_{i=1}^N x_{ei}^2y_{ei}^2 & \sum_{i=1}^N x_{ei}^3y_{ei} & \sum_{i=1}^N x_{ei}y_{ei}^3 \\ \sum_{i=1}^N x_{ei}^2 & \sum_{i=1}^N x_{ei}^3 & \sum_{i=1}^N x_{ei}^2y_{ei} & \sum_{i=1}^N x_{ei}^3y_{ei} & \sum_{i=1}^N x_{ei}^4 & \sum_{i=1}^N x_{ei}^2y_{ei}^2 \\ \sum_{i=1}^N y_{ei}^2 & \sum_{i=1}^N x_{ei}y_{ei}^2 & \sum_{i=1}^N y_{ei}^3 & \sum_{i=1}^N x_{ei}y_{ei}^3 & \sum_{i=1}^N x_{ei}^2y_{ei}^2 & \sum_{i=1}^N y_{ei}^4 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \end{bmatrix} \quad (4)$$

As with  $a_j$ , the value of  $b_j$  ( $j = 1, 2, 3, 4, 5, 6$ ) can be calculated. In fact, the relationship between the number of coefficients in mapping function ( $r$ ) and polynomial order ( $s$ ) is as follows:

$$r = 1 + \sum_{t=1}^s (t + 1). \quad (5)$$

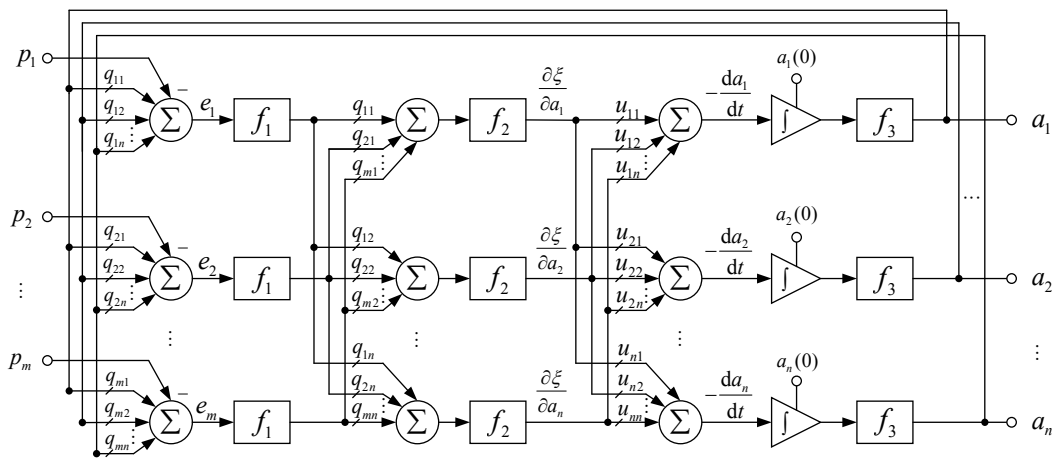
According to Equation (5), when an  $s$  order polynomial is utilized to solve the gaze mapping function, at least  $r$  gaze calibration markers are required. For a head-mounted (intrusive) gaze tracking system, the relative position of the monitor screen and the user's head and eyes remains nearly fixed. In this case, the higher-order terms in the mapping function are mainly utilized to compensate for error between the estimated and actual gaze direction. The higher the polynomial order, the higher the calculation accuracy. However, the number of polynomial coefficients to be solved will increase at the same time (Equation (5)). In addition, the number of calibration markers required also increases. This not only makes the calibration time longer; the cumbersome calibration process also adds to the user's burden. Users are prone to be fatigued, thus affecting the calibration accuracy. In order to further enhance the mapping accuracy and realize precise estimation of gaze direction, a novel artificial neural network (DLSR-ANN) based on direct least squares regression is proposed to solve a mapping function between pupil-glint vectors and calibration markers.

We rewrite the matrix equation in Equation (4) as:

$$Qa = p, \quad (6)$$

where  $a = [a_1 \ a_2 \ a_3 \ a_4 \ a_5 \ a_6]^T$ ,  $p = [\sum_{i=1}^N x_{ci} \ \sum_{i=1}^N x_{ci}x_{ei} \ \sum_{i=1}^N x_{ci}y_{ei} \ \sum_{i=1}^N x_{ci}x_{ei}y_{ei} \ \sum_{i=1}^N x_{ci}x_{ei}^2 \ \sum_{i=1}^N x_{ci}y_{ei}^2]^T$ ,  $Q$  is the coefficient matrix.

Figure 2 shows the scheme framework of an improved artificial neural network based on direct least squares regression. The DLSR-ANN is a three-layer neural network with input layer, hidden layer, and output layer. Elements of matrix  $p$  including pupil-glint vectors gazing at calibration markers are determined as the input of a neural network. Elements of matrix  $a$  are determined as the output of a neural network. The input, output, and hidden layers contain one, one, and three nodes, respectively.



**Figure 2.** Scheme framework of improved artificial neural network based on direct least squares regression.

As shown in Figure 2, coefficient matrix elements of direct least squares regression are employed as connection coefficients in the input and hidden layers of DLSR-ANN. According to the respective characteristics of input, hidden, and output layers and the relationship among them, appropriate weighting functions  $g_1(t)$ ,  $g_2(t)$ ,  $g_3(t)$  are determined. Derivatives of  $g_1(t)$ ,  $g_2(t)$ , and  $g_3(t)$ , respectively, are calculated ( $f_1(t)$ ,  $f_2(t)$ ,  $f_3(t)$  ( $f(t) = dg(t)/dt$ )) as the transfer function of the neuron. The selection of specific parameters is described in Section 3.4. As a three-layer neural network, its output layer carries an integrator. The integrator's initial condition  $a(0) = a_0 = [a_1(0) \ a_2(0) \ a_3(0) \ a_4(0) \ a_5(0) \ a_6(0)]^T$  is calculated through a linear polynomial solution utilizing direct least squares regression.

In the proposed method, to solve the mapping function in Equation (6), the steepest gradient descent method [74] is adopted as the training method of the neural network. To determine the relationship between hidden layer and output layer, the error cost function and continuous-time learning rule of DLSR-ANN are defined according to the constraint condition of solving direct least squares regression. According to the error distribution characteristics of gaze estimation, the Euclid norm ( $L_2$  norm) is selected to acquire the minimal error cost function, which is in the same form as the error solving criterion of direct least squares regression, as defined in Equation (7):

$$\xi(a) = \frac{1}{2} \|e\|^2 = \frac{1}{2} e^T e, \quad (7)$$

where  $e = Qa - p$  is the solution error of Equation (6) in direct least squares regression.

Equation (7) can be further expressed as follows:

$$\begin{aligned} \xi(a) &= \frac{1}{2} (Qa - p)^T (Qa - p) \\ &= \frac{1}{2} (a^T Q^T Q a - a^T Q p - p^T Q a + p^T p). \end{aligned} \quad (8)$$

According to an error cost function based on the constraint condition of direct least squares regression, the learning rule of a continuous-time neural network is set as Equation (9). The function of the learning rule is to modify the weights of DLSR-ANN adaptively to acquire the optimal solution.

$$\frac{da}{dt} = -\mu \nabla_a \xi(a) = -\mu \frac{\partial \xi(a)}{\partial a}, \quad (9)$$

where  $\mu$  is the learning rate parameter. As a positive-definite matrix,  $\mu$  ( $\mu = [\mu_{vw}]$ ,  $v, w = 1, 2, \dots, n$ ) is generally selected as a diagonal matrix. In general,  $\mu$  is determined by experience. If  $\mu$  is set too



small, the weights of the neural network will be modified by the learning rule slowly. More iterations will be needed to reach the error bottom. If  $\mu$  is set too large, the learning rule will show numerical instability. To ensure the stability of the differential equation in Equation (9) and the convergence of its solution, a small enough  $\mu$  is chosen according to Equation (10):

$$0 < \mu < \frac{2}{\lambda_{max}}, \quad (10)$$

where  $\lambda_{max}$  is the maximal eigenvalue of auto-correlation matrix composed by input vector  $p$  in direct least squares regression. When the eigenvalue is unavailable, the auto-correlation matrix can replace it.

By calculating a partial derivative of variable  $a$  in Equation (8), the learning rule of a continuous-time neural network for solving matrix equation  $Qa = p$  can be deduced as:

$$\frac{da}{dt} = -\mu \frac{\partial \xi(a)}{\partial a} = -\mu \cdot \frac{1}{2} (2Q^T Qa - 2Q^T p) = -\mu \cdot Q^T (Qa - 2Q^T p) = -\mu Q^T e. \quad (11)$$

### 3. Experimental System and Results

#### 3.1. Experimental System

In this study, we develop a wearable gaze tracking system composed of a helmet, a monitor, an array of four near-infrared light emitting diodes (NIR LEDs), and a microspur camera, as shown in Figure 3. The screen size of the monitor is 75 mm  $\times$  50 mm. Considering the imaging distance is limited between 3 cm and 7 cm, a microspur camera is adopted to acquire the eye image. The image resolution is 640  $\times$  480 pixels (CCD sensor). As described in [75], when the wavelength of NIR LED is located within the range of 760 nm–1400 nm, the pupil absorbs nearly all the near-infrared light and the iris obviously reflects it. The wavelength of NIR LED employed in this paper is 850 nm and the power is less than 5 mw. The experimental system brings no harm to human eyes [76]. An NVIDIA Jetson TK1 embedded development board (Figure 4) is utilized for image acquiring and processing (NVIDIA: NVIDIA Corporation (Santa Clara, California, CA, USA). TK1: Tegra K1. Jetson TK1 is a code of embedded development board manufactured by NVIDIA Corporation).

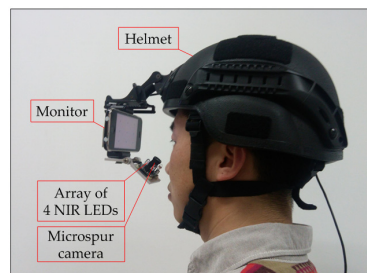


Figure 3. Wearable gaze tracking system.

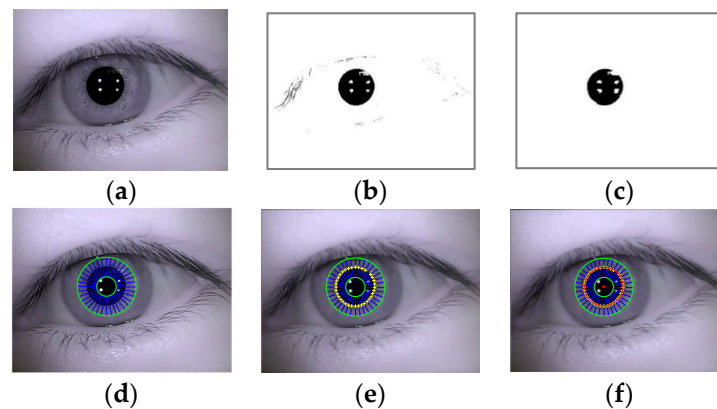


Figure 4. NVIDIA Jetson TK1 embedded development board.

### 3.2. Pupil and Glint Detection

#### 3.2.1. Pupil Detection

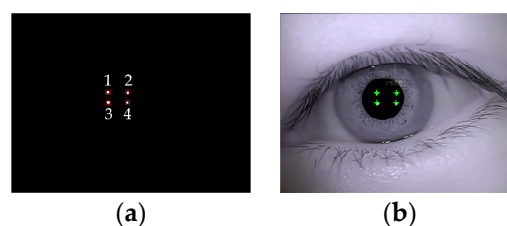
The circular ring rays location (CRRL) method [77] is utilized for pupil center detection, for the reason that it is more robust and accurate than conventional detection methods. As shown in Figure 5, in the CRRL method, improved Otsu optimal threshold binarization is utilized on a gray-scale eye image to eliminate the influence caused by illumination change. Through an opening-and-closing operation, rough location of pupil area, and circular ring rays, and pupil boundary points and center can be detected accurately when interference factors such as eyelashes, glint, and natural light reflection are located on the pupil contour. The CRRL method contributes to enhance the stability, accuracy, and real-time quality of a gaze tracking system.



**Figure 5.** Pupil detection: (a) original eye image; (b) eye binary image utilizing improved Otsu optimal threshold; (c) results of opening-and-closing operation; (d) rough location of pupil region; (e) extraction of pupil boundary points; (f) pupil contour fitting.

#### 3.2.2. Glint Detection

For the reason that the glint's illumination intensity is suitable for Gaussian distribution, Gaussian function deformation solved by improved total least squares [77] is utilized to calculate the glint center. The detection result of glint is shown in Figure 6.



**Figure 6.** Glint detection: (a) rough location of glint; (b) glint detection results.

As a sample, some of the pupil and glint centers detected are shown in Table 1.



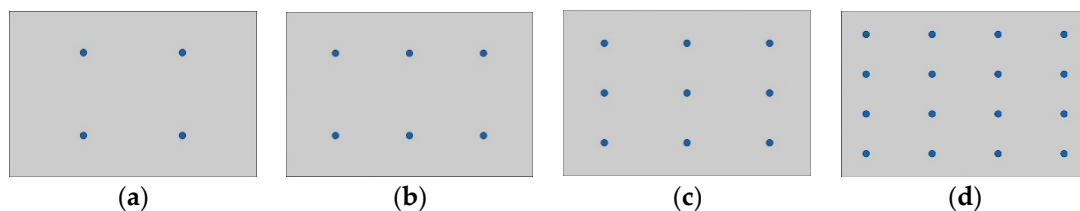
**Table 1.** A sample of the pupil and glint centers.

Eye Image	Pupil Center (x, y)	Glint Center (x, y)			
		1	2	3	4
1	(290.15, 265.34)	(265.31, 298.65)	(294.56, 300.87)	(266.41, 310.28)	(296.25, 312.49)
2	(251.42, 255.93)	(245.58, 292.36)	(276.54, 295.13)	(246.19, 305.67)	(277.51, 307.26)
3	(203.34, 260.81)	(221.95, 297.32)	(252.49, 298.61)	(221.34, 309.17)	(253.65, 310.28)
4	(297.74, 275.62)	(271.25, 300.56)	(301.58, 300.67)	(270.91, 315.66)	(300.85, 315.46)
5	(247.31, 277.58)	(243.25, 302.62)	(273.55, 303.46)	(242.81, 317.54)	(274.26, 318.19)

### 3.3. Calibration Model

As expressed in Equation (5), at least three, six, and 10 polynomial coefficients are required to be calculated, respectively, when a 1st, 2nd, and 3rd order linear polynomial is utilized for calibration, which means that at least three, six, and 10 calibration markers are required. When the number of calibration markers needed is too large, unessential input items can be removed according to principal component analysis to reduce the number of polynomial coefficients to be solved. Generally, based on an overall consideration of the real-time quality and accuracy of a gaze tracking system, four- and five-marker calibration models are most widely employed for 1st order calculation, while six- and nine-marker calibration models are most widely employed for 2nd order calculation [78,79].

Considering that there is some motion between the wearable gaze tracking system and the user's head, error of gaze point data will occur along with a drifting motion. In this paper, position coordinates of quadrangular NIR LEDs are considered as inputs of gaze estimation model to compensate for error caused by drifting motion. As shown in Figure 7, for the purpose of comparison, four-, six-, nine-, and 16-marker calibration models are employed in the process of calculating mapping function. Gaze direction is estimated with and without error compensation. The gaze tracking accuracy of the two cases is compared.



**Figure 7.** Calibration model: (a) four-marker calibration model; (b) six-marker calibration model; (c) nine-marker calibration model; (d) 16-marker calibration model.

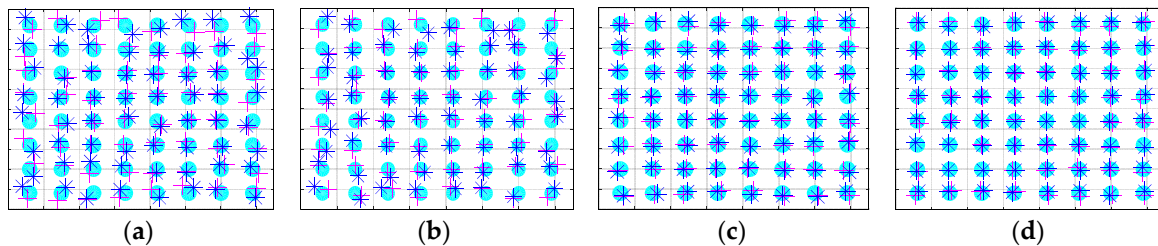
### 3.4. Gaze Point Estimation

An improved artificial neural network (DLSR-ANN) based on direct least squares regression is developed to calculate the mapping function between pupil-glint vectors and calibration markers. For four-, six-, nine-, and 16-marker calibration models, the number of training samples is selected as 180. The number of hidden nodes is equal to the number of training samples. The  $x$  (or  $y$ ) coordinate is set as the output of the neural network. Two separate DLSR-ANNs are utilized to estimate the  $x$  and  $y$  coordinates of the gaze point on the screen. Each separate neural network has the same inputs. Weighting function  $g_1(t)$ ,  $g_2(t)$ ,  $g_3(t)$  is respectively determined as  $g_1(t) = \frac{1-e^{-\beta_1 t^2}}{1+e^{-\beta_1 t^2}}$ ,  $g_2(t) = \begin{cases} \frac{t^2}{2}, t \leq \beta_2 \\ \beta_2 |t| - \frac{\beta_2^2}{2}, t > \beta_2 \end{cases}$ ,  $g_3(t) = \beta_3^2 \ln \left( \cosh \frac{t}{\beta_3} \right)$ . The transfer function for input, hidden, and output layers is selected as the derivative of  $g_1(t)$ ,  $g_2(t)$ ,  $g_3(t)$ , which is respectively calculated as

$$f_1(t), f_2(t), f_3(t): f_1(t) = \frac{dg_1(t)}{dt} = \frac{4\beta_1 t e^{-\beta_1 t^2}}{(1+e^{-\beta_1 t^2})^2}, f_2(t) = \frac{dg_2(t)}{dt} = \begin{cases} -\beta_2, t \leq -\beta_2 \\ t, |t| \leq \beta_2 \\ \beta_2, t > \beta_2 \end{cases}, f_3(t) = \frac{dg_3(t)}{dt} =$$

$\beta_3 \tanh \frac{t}{\beta_3}$ . Learning rate parameter  $\mu$  is determined by  $\mu = \mu_j = 0.0025$  (when a four-marker calibration model is employed,  $j = 1, 2, 3, 4$ ; when a six-, nine-, or 16-marker calibration model is employed,  $j = 1, 2, 3, 4, 5, 6$ ). In order to acquire optimal learning and training results,  $\beta_1, \beta_2, \beta_3$  is respectively determined as  $\beta_1 = 0.8, \beta_2 = 0.7, \beta_3 = 0.7$  through a process of trial and error. The initial condition  $a(0)$  of an integrator associated with learning rules is acquired through linear polynomial calculation in direct least squares regression.

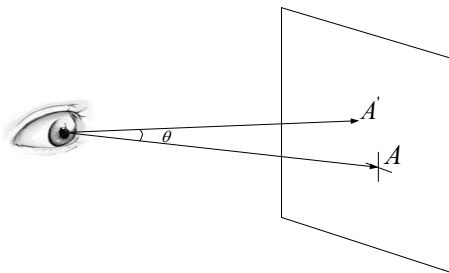
In the developed wearable gaze tracking system, an array of four near-infrared light emitting diodes (NIR LEDs) is employed instead of the conventional single one. The NIR LEDs array can form well-distributed illumination around the human eye, which contributes to extract pupil and glint characteristics more stably and precisely. In addition the center position coordinates of quadrangular NIR LEDs, considered as inputs of the neural network, can further compensate for error caused during the process of gaze point calculation. When a calibration process is accomplished, a model with  $8 \times 8$  test markers is employed to validate the calculation accuracy of the gaze point. Figure 8a–d shows the gaze point estimated through the proposed method with/without considering error compensation, utilizing a four-, six-, nine-, or 16-marker calibration model, respectively. The cyan “●” symbols represent actual reference gaze points on the monitor screen. The magenta “+” symbols represent gaze points estimated through the proposed method without considering error compensation. The blue “\*” symbols represent gaze points estimated through the proposed method considering error compensation.



**Figure 8.** Gaze point estimation with/without considering error compensation: (a) gaze point estimation utilizing a four-marker calibration model; (b) gaze point estimation utilizing a six-marker calibration model; (c) gaze point estimation utilizing a nine-marker calibration model; (d) gaze point estimation utilizing a 16-marker calibration model.

### 3.5. Gaze Estimation Accuracy Comparison of Different Methods

As shown in Figure 9, gaze estimation accuracy is expressed as intersection angle  $\theta$  between actual gaze direction ( $A$  as gaze point) and estimated gaze direction ( $A'$  as gaze point).



**Figure 9.** Definition of gaze estimation accuracy.

Angle  $\theta$  can be calculated through Equation (12), where  $L$  is the distance between the human eye and the monitor screen:

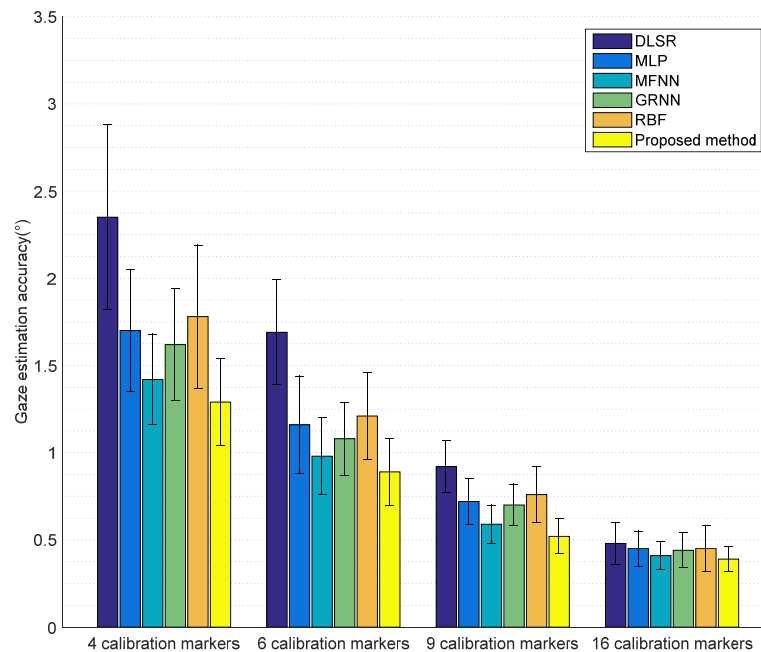
$$\theta = \arctan\left(\frac{\sqrt{(x_A - x_{A'})^2 + (y_A - y_{A'})^2}}{L}\right). \quad (12)$$

The standard deviation of gaze estimation accuracy  $\theta$  is defined as Equation (13), where  $\bar{\theta}$  represents the mean value of  $\theta_j$  ( $j = 1, 2, \dots, K$ ) and  $K$  is the total number of gaze points estimated:

$$\Delta_{std} = \sqrt{\frac{1}{K} \sum_{j=1}^K (\theta_j - \bar{\theta})^2}. \quad (13)$$

### 3.5.1. Gaze Estimation Accuracy without Considering Error Compensation

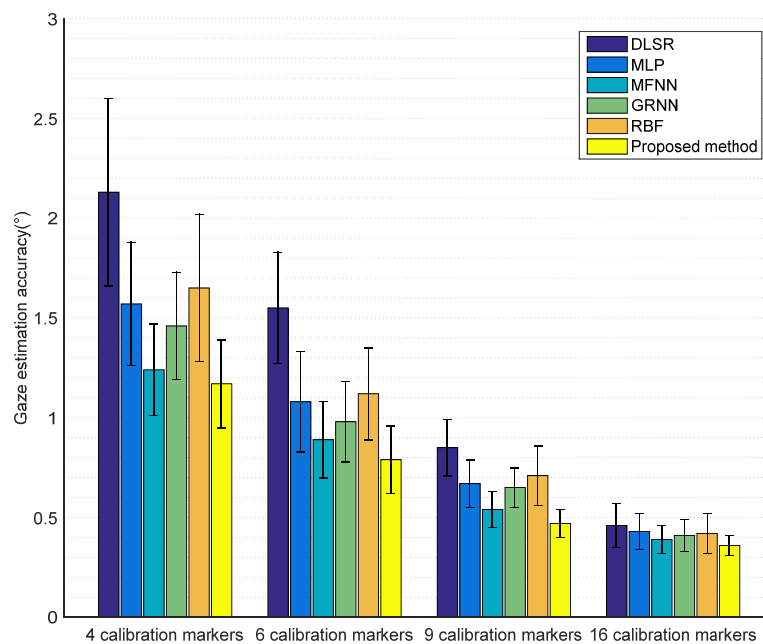
Figure 10 shows a comparison of gaze estimation accuracy and standard deviation calculated through the proposed method and other neural network methods, respectively, without considering error compensation. The proposed method can provide an accuracy of  $1.29^\circ$ ,  $0.89^\circ$ ,  $0.52^\circ$ , and  $0.39^\circ$  when a four-, six-, nine-, or 16-marker calibration model is utilized for calibration, respectively. The maximum gaze estimation error through the proposed method for a four-, six-, nine-, or 16-marker calibration model is, respectively,  $2.45^\circ$ ,  $1.98^\circ$ ,  $1.21^\circ$ , and  $0.82^\circ$ . The specific results are shown in Table A1.



**Figure 10.** Comparison of gaze estimation accuracy results between proposed method and other methods without considering error compensation.

### 3.5.2. Gaze Estimation Accuracy Considering Error Compensation

Figure 11 shows the comparison of gaze estimation accuracy and standard deviation calculated respectively through the proposed method and other NN (Neural Network) methods considering error compensation. The proposed method can provide an accuracy of  $1.17^\circ$ ,  $0.79^\circ$ ,  $0.47^\circ$ , and  $0.36^\circ$  respectively, when a four-, six-, nine-, or 16-marker calibration model is utilized for calibration. When considering error compensation, the improvement percentage of gaze estimation accuracy for four-, six-, nine-, and 16-marker calibration models is 9.3%, 11.2%, 9.6%, and 7.6%, respectively. The specific results are shown in Table A2 of the Appendix.



**Figure 11.** Comparison of gaze estimation accuracy results between proposed method and other methods considering error compensation.

#### 4. Conclusions

In this paper, a novel 2D gaze estimation method based on pupil-glint vector is proposed on the basis of conventional gaze tracking methods. In order to realize the accurate estimation of gaze direction, an improved artificial neural network (DLSR-ANN) based on direct least squares regression is developed. Learning rate parameter, weighting function, and corresponding coefficients are determined according to trial and experience. Detected coordinates of pupil-glint vectors are applied as inputs to train an improved neural network. The mapping function model is solved and then utilized to calculate gaze point coordinates. An array of four NIR LEDs is employed to form quadrangular glints. The NIR LEDs array can generate well-distributed illumination around the human eye, which contributes to extracting pupil and glint characteristics more stably and precisely. In addition, the center coordinates of quadrangular NIR LEDs, considered as additional inputs of neural network, can further compensate for the error caused during the process of calculating the gaze point, which can enhance the accuracy of gaze point coordinates. When the gaze tracking system is established, calibration models with different numbers of markers are utilized to validate the proposed method. When a four-, six-, nine-, or 16-marker calibration model is employed for the calibration process, the proposed method can achieve an accuracy of  $1.29^\circ$ ,  $0.89^\circ$ ,  $0.52^\circ$ , and  $0.39^\circ$ , respectively. Taking into account error compensation, the proposed method can achieve an accuracy of  $1.17^\circ$ ,  $0.79^\circ$ ,  $0.47^\circ$ , and  $0.36^\circ$ , respectively, when a four-, six-, nine-, or 16-marker calibration model is employed. When considering error compensation, the improvement percentage of gaze estimation accuracy for a four-, six-, nine-, or 16-marker calibration model is 9.3%, 11.2%, 9.6%, and 7.6%, respectively. The experimental results show that the training process of the proposed method is stable. The gaze estimation accuracy of the proposed method in this paper is better than that of conventional linear regression (direct least squares regression) and nonlinear regression (generic artificial neural network). The proposed method contributes to enhance the total accuracy of a gaze tracking system.

**Acknowledgments:** This work is supported by Program for Changjiang Scholars and Innovation Research Team in University under Grant No. IRT1208, Basic Research Fund of Beijing Institute of Technology under Grant (No. 20130242015) and Autonomous Program of State Key Laboratory of Explosion Science and Technology under Grant No. YBKT15-09. The authors would like to thank the editor and all anonymous reviewers for their constructive suggestions.

**Author Contributions:** All authors made significant contributions to this article. Jianzhong Wang was mainly responsible for deployment of the system and revision of the paper; Guangyue Zhang was responsible for developing gaze estimation method, performing the experiments, and writing the paper; Jiadong Shi, the corresponding author, was responsible for conceiving and designing the experiments.

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

## Appendix

**Table A1.** Comparison of gaze estimation accuracy between proposed method and other NN methods without considering error compensation.

Calibration Markers	Method	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Average Err.
4	DLSR [55]	$2.32^\circ \pm 0.54^\circ$	$2.41^\circ \pm 0.58^\circ$	$2.48^\circ \pm 0.62^\circ$	$2.24^\circ \pm 0.45^\circ$	$2.29^\circ \pm 0.48^\circ$	$2.37^\circ \pm 0.51^\circ$	$2.35^\circ \pm 0.53^\circ$
	MLP [66]	$1.71^\circ \pm 0.35^\circ$	$1.74^\circ \pm 0.39^\circ$	$1.64^\circ \pm 0.32^\circ$	$1.72^\circ \pm 0.33^\circ$	$1.76^\circ \pm 0.40^\circ$	$1.62^\circ \pm 0.29^\circ$	$1.70^\circ \pm 0.35^\circ$
	MFNN [67]	$1.38^\circ \pm 0.27^\circ$	$1.45^\circ \pm 0.30^\circ$	$1.43^\circ \pm 0.24^\circ$	$1.49^\circ \pm 0.29^\circ$	$1.34^\circ \pm 0.22^\circ$	$1.40^\circ \pm 0.25^\circ$	$1.42^\circ \pm 0.26^\circ$
	GRNN [68]	$1.63^\circ \pm 0.32^\circ$	$1.52^\circ \pm 0.28^\circ$	$1.69^\circ \pm 0.35^\circ$	$1.72^\circ \pm 0.36^\circ$	$1.55^\circ \pm 0.30^\circ$	$1.61^\circ \pm 0.33^\circ$	$1.62^\circ \pm 0.32^\circ$
	RBF [69]	$1.81^\circ \pm 0.43^\circ$	$1.92^\circ \pm 0.48^\circ$	$1.85^\circ \pm 0.44^\circ$	$1.74^\circ \pm 0.37^\circ$	$1.67^\circ \pm 0.33^\circ$	$1.72^\circ \pm 0.41^\circ$	$1.78^\circ \pm 0.41^\circ$
	Proposed	$1.36^\circ \pm 0.24^\circ$	$1.28^\circ \pm 0.19^\circ$	$1.26^\circ \pm 0.31^\circ$	$1.31^\circ \pm 0.25^\circ$	$1.21^\circ \pm 0.30^\circ$	$1.32^\circ \pm 0.20^\circ$	$1.29^\circ \pm 0.25^\circ$
6	DLSR [55]	$1.68^\circ \pm 0.29^\circ$	$1.62^\circ \pm 0.25^\circ$	$1.64^\circ \pm 0.28^\circ$	$1.72^\circ \pm 0.31^\circ$	$1.71^\circ \pm 0.33^\circ$	$1.74^\circ \pm 0.31^\circ$	$1.69^\circ \pm 0.30^\circ$
	MLP [66]	$1.15^\circ \pm 0.26^\circ$	$1.23^\circ \pm 0.33^\circ$	$1.17^\circ \pm 0.25^\circ$	$1.06^\circ \pm 0.24^\circ$	$1.10^\circ \pm 0.27^\circ$	$1.26^\circ \pm 0.35^\circ$	$1.16^\circ \pm 0.28^\circ$
	MFNN [67]	$0.98^\circ \pm 0.22^\circ$	$0.96^\circ \pm 0.20^\circ$	$1.05^\circ \pm 0.27^\circ$	$1.03^\circ \pm 0.25^\circ$	$0.95^\circ \pm 0.18^\circ$	$0.93^\circ \pm 0.19^\circ$	$0.98^\circ \pm 0.22^\circ$
	GRNN [68]	$1.07^\circ \pm 0.19^\circ$	$1.16^\circ \pm 0.27^\circ$	$1.02^\circ \pm 0.15^\circ$	$1.05^\circ \pm 0.19^\circ$	$1.12^\circ \pm 0.26^\circ$	$1.08^\circ \pm 0.18^\circ$	$1.08^\circ \pm 0.21^\circ$
	RBF [69]	$1.20^\circ \pm 0.26^\circ$	$1.17^\circ \pm 0.24^\circ$	$1.23^\circ \pm 0.27^\circ$	$1.24^\circ \pm 0.29^\circ$	$1.15^\circ \pm 0.19^\circ$	$1.18^\circ \pm 0.25^\circ$	$1.21^\circ \pm 0.25^\circ$
	Proposed	$0.88^\circ \pm 0.16^\circ$	$0.94^\circ \pm 0.19^\circ$	$0.78^\circ \pm 0.25^\circ$	$0.86^\circ \pm 0.14^\circ$	$0.92^\circ \pm 0.21^\circ$	$0.95^\circ \pm 0.18^\circ$	$0.89^\circ \pm 0.19^\circ$
9	DLSR [55]	$0.91^\circ \pm 0.15^\circ$	$0.89^\circ \pm 0.16^\circ$	$0.97^\circ \pm 0.18^\circ$	$0.96^\circ \pm 0.15^\circ$	$0.86^\circ \pm 0.13^\circ$	$0.94^\circ \pm 0.14^\circ$	$0.92^\circ \pm 0.15^\circ$
	MLP [66]	$0.73^\circ \pm 0.13^\circ$	$0.78^\circ \pm 0.16^\circ$	$0.74^\circ \pm 0.16^\circ$	$0.67^\circ \pm 0.11^\circ$	$0.64^\circ \pm 0.10^\circ$	$0.75^\circ \pm 0.14^\circ$	$0.72^\circ \pm 0.13^\circ$
	MFNN [67]	$0.58^\circ \pm 0.09^\circ$	$0.57^\circ \pm 0.12^\circ$	$0.64^\circ \pm 0.11^\circ$	$0.56^\circ \pm 0.14^\circ$	$0.59^\circ \pm 0.09^\circ$	$0.62^\circ \pm 0.13^\circ$	$0.59^\circ \pm 0.11^\circ$
	GRNN [68]	$0.71^\circ \pm 0.11^\circ$	$0.74^\circ \pm 0.12^\circ$	$0.77^\circ \pm 0.16^\circ$	$0.65^\circ \pm 0.09^\circ$	$0.64^\circ \pm 0.10^\circ$	$0.67^\circ \pm 0.12^\circ$	$0.70^\circ \pm 0.12^\circ$
	RBF [69]	$0.77^\circ \pm 0.17^\circ$	$0.72^\circ \pm 0.14^\circ$	$0.84^\circ \pm 0.21^\circ$	$0.80^\circ \pm 0.20^\circ$	$0.76^\circ \pm 0.15^\circ$	$0.70^\circ \pm 0.12^\circ$	$0.76^\circ \pm 0.16^\circ$
	Proposed	$0.51^\circ \pm 0.08^\circ$	$0.49^\circ \pm 0.09^\circ$	$0.48^\circ \pm 0.12^\circ$	$0.56^\circ \pm 0.10^\circ$	$0.51^\circ \pm 0.11^\circ$	$0.47^\circ \pm 0.07^\circ$	$0.52^\circ \pm 0.10^\circ$
16	DLSR [55]	$0.50^\circ \pm 0.12^\circ$	$0.47^\circ \pm 0.10^\circ$	$0.49^\circ \pm 0.13^\circ$	$0.48^\circ \pm 0.15^\circ$	$0.49^\circ \pm 0.09^\circ$	$0.51^\circ \pm 0.14^\circ$	$0.48^\circ \pm 0.12^\circ$
	MLP [66]	$0.44^\circ \pm 0.11^\circ$	$0.48^\circ \pm 0.13^\circ$	$0.49^\circ \pm 0.11^\circ$	$0.46^\circ \pm 0.09^\circ$	$0.44^\circ \pm 0.10^\circ$	$0.46^\circ \pm 0.08^\circ$	$0.45^\circ \pm 0.10^\circ$
	MFNN [67]	$0.39^\circ \pm 0.09^\circ$	$0.42^\circ \pm 0.08^\circ$	$0.44^\circ \pm 0.12^\circ$	$0.39^\circ \pm 0.07^\circ$	$0.40^\circ \pm 0.07^\circ$	$0.42^\circ \pm 0.08^\circ$	$0.41^\circ \pm 0.08^\circ$
	GRNN [68]	$0.46^\circ \pm 0.12^\circ$	$0.41^\circ \pm 0.09^\circ$	$0.45^\circ \pm 0.10^\circ$	$0.47^\circ \pm 0.13^\circ$	$0.40^\circ \pm 0.08^\circ$	$0.43^\circ \pm 0.11^\circ$	$0.44^\circ \pm 0.10^\circ$
	RBF [69]	$0.48^\circ \pm 0.15^\circ$	$0.46^\circ \pm 0.13^\circ$	$0.41^\circ \pm 0.11^\circ$	$0.42^\circ \pm 0.12^\circ$	$0.46^\circ \pm 0.14^\circ$	$0.44^\circ \pm 0.15^\circ$	$0.45^\circ \pm 0.13^\circ$
	Proposed	$0.36^\circ \pm 0.06^\circ$	$0.42^\circ \pm 0.09^\circ$	$0.38^\circ \pm 0.08^\circ$	$0.40^\circ \pm 0.07^\circ$	$0.43^\circ \pm 0.08^\circ$	$0.37^\circ \pm 0.06^\circ$	$0.39^\circ \pm 0.07^\circ$

**Table A2.** Comparison of gaze estimation accuracy between proposed method and other NN methods considering error compensation.

Calibration Markers	Method	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Average Err.
4	DLSR [55]	$2.11^\circ \pm 0.48^\circ$	$2.20^\circ \pm 0.52^\circ$	$2.24^\circ \pm 0.55^\circ$	$2.03^\circ \pm 0.40^\circ$	$2.06^\circ \pm 0.42^\circ$	$2.15^\circ \pm 0.44^\circ$	$2.13^\circ \pm 0.47^\circ$
	MLP [66]	$1.56^\circ \pm 0.31^\circ$	$1.64^\circ \pm 0.35^\circ$	$1.52^\circ \pm 0.28^\circ$	$1.58^\circ \pm 0.30^\circ$	$1.63^\circ \pm 0.36^\circ$	$1.51^\circ \pm 0.26^\circ$	$1.57^\circ \pm 0.31^\circ$
	MFNN [67]	$1.23^\circ \pm 0.24^\circ$	$1.21^\circ \pm 0.26^\circ$	$1.28^\circ \pm 0.21^\circ$	$1.26^\circ \pm 0.25^\circ$	$1.18^\circ \pm 0.20^\circ$	$1.25^\circ \pm 0.22^\circ$	$1.24^\circ \pm 0.23^\circ$
	GRNN [68]	$1.48^\circ \pm 0.29^\circ$	$1.37^\circ \pm 0.22^\circ$	$1.45^\circ \pm 0.31^\circ$	$1.57^\circ \pm 0.28^\circ$	$1.41^\circ \pm 0.26^\circ$	$1.49^\circ \pm 0.28^\circ$	$1.46^\circ \pm 0.27^\circ$
	RBF [69]	$1.65^\circ \pm 0.39^\circ$	$1.77^\circ \pm 0.43^\circ$	$1.79^\circ \pm 0.40^\circ$	$1.54^\circ \pm 0.34^\circ$	$1.52^\circ \pm 0.29^\circ$	$1.61^\circ \pm 0.36^\circ$	$1.65^\circ \pm 0.37^\circ$
	Proposed	$1.23^\circ \pm 0.21^\circ$	$1.17^\circ \pm 0.18^\circ$	$1.14^\circ \pm 0.26^\circ$	$1.18^\circ \pm 0.22^\circ$	$1.09^\circ \pm 0.28^\circ$	$1.21^\circ \pm 0.19^\circ$	$1.17^\circ \pm 0.22^\circ$
6	DLSR [55]	$1.54^\circ \pm 0.28^\circ$	$1.49^\circ \pm 0.23^\circ$	$1.51^\circ \pm 0.26^\circ$	$1.57^\circ \pm 0.30^\circ$	$1.56^\circ \pm 0.31^\circ$	$1.61^\circ \pm 0.29^\circ$	$1.55^\circ \pm 0.28^\circ$
	MLP [66]	$1.06^\circ \pm 0.24^\circ$	$1.15^\circ \pm 0.30^\circ$	$1.08^\circ \pm 0.21^\circ$	$1.01^\circ \pm 0.22^\circ$	$1.03^\circ \pm 0.23^\circ$	$1.14^\circ \pm 0.29^\circ$	$1.08^\circ \pm 0.25^\circ$
	MFNN [67]	$0.87^\circ \pm 0.21^\circ$	$0.88^\circ \pm 0.18^\circ$	$0.96^\circ \pm 0.23^\circ$	$0.94^\circ \pm 0.21^\circ$	$0.86^\circ \pm 0.16^\circ$	$0.84^\circ \pm 0.17^\circ$	$0.89^\circ \pm 0.19^\circ$
	GRNN [68]	$0.95^\circ \pm 0.19^\circ$	$1.05^\circ \pm 0.24^\circ$	$0.91^\circ \pm 0.15^\circ$	$0.94^\circ \pm 0.19^\circ$	$1.01^\circ \pm 0.23^\circ$	$0.99^\circ \pm 0.18^\circ$	$0.98^\circ \pm 0.20^\circ$
	RBF [69]	$1.11^\circ \pm 0.23^\circ$	$1.09^\circ \pm 0.21^\circ$	$1.15^\circ \pm 0.25^\circ$	$1.14^\circ \pm 0.27^\circ$	$1.07^\circ \pm 0.18^\circ$	$1.18^\circ \pm 0.22^\circ$	$1.12^\circ \pm 0.23^\circ$
	Proposed	$0.78^\circ \pm 0.13^\circ$	$0.82^\circ \pm 0.17^\circ$	$0.71^\circ \pm 0.23^\circ$	$0.73^\circ \pm 0.12^\circ$	$0.81^\circ \pm 0.20^\circ$	$0.87^\circ \pm 0.17^\circ$	$0.79^\circ \pm 0.17^\circ$
9	DLSR [55]	$0.84^\circ \pm 0.14^\circ$	$0.81^\circ \pm 0.15^\circ$	$0.89^\circ \pm 0.17^\circ$	$0.88^\circ \pm 0.13^\circ$	$0.80^\circ \pm 0.12^\circ$	$0.86^\circ \pm 0.13^\circ$	$0.85^\circ \pm 0.14^\circ$
	MLP [66]	$0.68^\circ \pm 0.12^\circ$	$0.74^\circ \pm 0.14^\circ$	$0.70^\circ \pm 0.15^\circ$	$0.62^\circ \pm 0.10^\circ$	$0.61^\circ \pm 0.09^\circ$	$0.69^\circ \pm 0.12^\circ$	$0.67^\circ \pm 0.12^\circ$
	MFNN [67]	$0.53^\circ \pm 0.08^\circ$	$0.52^\circ \pm 0.10^\circ$	$0.60^\circ \pm 0.09^\circ$	$0.51^\circ \pm 0.11^\circ$	$0.54^\circ \pm 0.08^\circ$	$0.56^\circ \pm 0.09^\circ$	$0.54^\circ \pm 0.09^\circ$
	GRNN [68]	$0.66^\circ \pm 0.09^\circ$	$0.69^\circ \pm 0.11^\circ$	$0.71^\circ \pm 0.15^\circ$	$0.61^\circ \pm 0.08^\circ$	$0.60^\circ \pm 0.09^\circ$	$0.62^\circ \pm 0.10^\circ$	$0.65^\circ \pm 0.10^\circ$
	RBF [69]	$0.71^\circ \pm 0.16^\circ$	$0.66^\circ \pm 0.13^\circ$	$0.78^\circ \pm 0.18^\circ$	$0.73^\circ \pm 0.19^\circ$	$0.70^\circ \pm 0.14^\circ$	$0.65^\circ \pm 0.11^\circ$	$0.71^\circ \pm 0.15^\circ$
	Proposed	$0.46^\circ \pm 0.07^\circ$	$0.45^\circ \pm 0.06^\circ$	$0.47^\circ \pm 0.10^\circ$	$0.51^\circ \pm 0.08^\circ$	$0.48^\circ \pm 0.09^\circ$	$0.43^\circ \pm 0.05^\circ$	$0.47^\circ \pm 0.07^\circ$
16	DLSR [55]	$0.43^\circ \pm 0.09^\circ$	$0.48^\circ \pm 0.12^\circ$	$0.45^\circ \pm 0.10^\circ$	$0.43^\circ \pm 0.09^\circ$	$0.49^\circ \pm 0.13^\circ$	$0.46^\circ \pm 0.10^\circ$	$0.46^\circ \pm 0.11^\circ$
	MLP [66]	$0.47^\circ \pm 0.11^\circ$	$0.42^\circ \pm 0.09^\circ$	$0.40^\circ \pm 0.08^\circ$	$0.44^\circ \pm 0.07^\circ$	$0.45^\circ \pm 0.10^\circ$	$0.41^\circ \pm 0.11^\circ$	$0.43^\circ \pm 0.09^\circ$
	MFNN [67]	$0.36^\circ \pm 0.08^\circ$	$0.41^\circ \pm 0.07^\circ$	$0.38^\circ \pm 0.05^\circ$	$0.40^\circ \pm 0.09^\circ$	$0.41^\circ \pm 0.06^\circ$	$0.38^\circ \pm 0.08^\circ$	$0.39^\circ \pm 0.07^\circ$
	GRNN [68]	$0.42^\circ \pm 0.11^\circ$	$0.39^\circ \pm 0.07^\circ$	$0.42^\circ \pm 0.10^\circ$	$0.38^\circ \pm 0.06^\circ$	$0.41^\circ \pm 0.06^\circ$	$0.44^\circ \pm 0.10^\circ$	$0.41^\circ \pm 0.08^\circ$
	RBF [69]	$0.38^\circ \pm 0.09^\circ$	$0.44^\circ \pm 0.12^\circ$	$0.45^\circ \pm 0.13^\circ$	$0.43^\circ \pm 0.09^\circ$	$0.39^\circ \pm 0.08^\circ$	$0.41^\circ \pm 0.11^\circ$	$0.42^\circ \pm 0.10^\circ$
	Proposed	$0.33^\circ \pm 0.05^\circ$	$0.35^\circ \pm 0.04^\circ$	$0.39^\circ \pm 0.06^\circ$	$0.38^\circ \pm 0.08^\circ$	$0.35^\circ \pm 0.04^\circ$	$0.37^\circ \pm 0.05^\circ$	$0.36^\circ \pm 0.05^\circ$

## References

1. Jacob, R.J.K. What you look at is what you get: Eye movement-based interaction techniques. In Proceedings of the Sigchi Conference on Human Factors in Computing Systems, Seattle, WA, USA, 1–5 April 1990; pp. 11–18.
2. Jacob, R.J.K. The use of eye movements in human-computer interaction techniques: What you look at is what you get. *ACM Trans. Inf. Syst.* **1991**, *9*, 152–169. [[CrossRef](#)]
3. Schütz, A.C.; Braun, D.I.; Gegenfurtner, K.R. Eye movements and perception: A selective review. *J. Vis.* **2011**, *11*, 89–91. [[CrossRef](#)] [[PubMed](#)]
4. Miriam, S.; Anna, M. Do we track what we see? Common *versus* independent processing for motion perception and smooth pursuit eye movements: A review. *Vis. Res.* **2011**, *51*, 836–852.
5. Blondon, K.; Wipfli, R.; Lovis, C. Use of eye-tracking technology in clinical reasoning: A systematic review. *Stud. Health Technol. Inform.* **2015**, *210*, 90–94. [[PubMed](#)]
6. Higgins, E.; Leinenger, M.; Rayner, K. Eye movements when viewing advertisements. *Front. Psychol.* **2014**, *5*, 1–15. [[CrossRef](#)] [[PubMed](#)]
7. Spakov, O.; Majaranta, P.; Spakov, O. Scrollable keyboards for casual eye typing. *Psychol. J.* **2009**, *7*, 159–173.
8. Nouredin, B.; Lawrence, P.D.; Man, C.F. A non-contact device for tracking gaze in a human computer interface. *Comput. Vis. Image Underst.* **2005**, *98*, 52–82. [[CrossRef](#)]
9. Biswas, P.; Langdon, P. Multimodal intelligent eye-gaze tracking system. *Int. J. Hum. Comput. Interact.* **2015**, *31*, 277–294. [[CrossRef](#)]
10. Lim, C.J.; Kim, D. Development of gaze tracking interface for controlling 3d contents. *Sens. Actuator A-Phys.* **2012**, *185*, 151–159. [[CrossRef](#)]
11. Ince, I.F.; Jin, W.K. A 2D eye gaze estimation system with low-resolution webcam images. *EURASIP J. Adv. Signal Process.* **2011**, *1*, 589–597. [[CrossRef](#)]
12. Kumar, N.; Kohlbecher, S.; Schneider, E. A novel approach to video-based pupil tracking. In Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Hyatt Regency Riverwalk, San Antonio, TX, USA, 11–14 October 2009; pp. 1255–1262.
13. Lee, E.C.; Min, W.P. A new eye tracking method as a smartphone interface. *KSII Trans. Internet Inf. Syst.* **2013**, *7*, 834–848.
14. Kim, J. Webcam-based 2D eye gaze estimation system by means of binary deformable eyeball templates. *J. Inform. Commun. Converg. Eng.* **2010**, *8*, 575–580. [[CrossRef](#)]
15. Dong, L. Investigation of Calibration Techniques in video based eye tracking system. In Proceedings of the 11th international conference on Computers Helping People with Special Needs, Linz, Austria, 9–11 July 2008; pp. 1208–1215.
16. Fard, P.J.M.; Moradi, M.H.; Parvaneh, S. Eye tracking using a novel approach. In Proceedings of the World Congress on Medical Physics and Biomedical Engineering 2006, Seoul, Korea, 27 August–1 September 2006; pp. 2407–2410.
17. Yamazoe, H.; Utsumi, A.; Yonezawa, T.; Abe, S. Remote gaze estimation with a single camera based on facial-feature tracking without special calibration actions. In Proceedings of the 2008 symposium on Eye tracking research and applications, Savannah, GA, USA, 26–28 March 2008; pp. 245–250.
18. Lee, E.C.; Kang, R.P.; Min, C.W.; Park, J. Robust gaze tracking method for stereoscopic virtual reality systems. *Hum.-Comput. Interact.* **2007**, *4552*, 700–709.
19. Lee, H.C.; Lee, W.O.; Cho, C.W.; Gwon, S.Y.; Park, K.R.; Lee, H.; Cha, J. Remote gaze tracking system on a large display. *Sensors* **2013**, *13*, 13439–13463. [[CrossRef](#)] [[PubMed](#)]
20. Ohno, T.; Mukawa, N.; Yoshikawa, A. Free gaze: A gaze tracking system for everyday gaze interaction. In Proceedings of the Symposium on Eye Tracking Research and Applications Symposium, New Orleans, LA, USA, 25–27 March 2002; pp. 125–132.
21. Chen, J.; Ji, Q. 3D gaze estimation with a single camera without IR illumination. In Proceedings of the International Conference on Pattern Recognition, Tampa, Florida, FL, USA, 8–11 December 2008; pp. 1–4.
22. Sheng-Wen, S.; Jin, L. A novel approach to 3D gaze tracking using stereo cameras. *IEEE Trans. Syst. Man Cybern. Part B-Cybern.* **2004**, *34*, 234–245.



23. Ki, J.; Kwon, Y.M.; Sohn, K. 3D gaze tracking and analysis for attentive Human Computer Interaction. In Proceedings of the Frontiers in the Convergence of Bioscience and Information Technologies, Jeju, Korea, 11–13 October 2007; pp. 617–621.
24. Ji, W.L.; Cho, C.W.; Shin, K.Y.; Lee, E.C.; Park, K.R. 3D gaze tracking method using purkinje images on eye optical model and pupil. *Opt. Lasers Eng.* **2012**, *50*, 736–751.
25. Lee, E.C.; Kang, R.P. A robust eye gaze tracking method based on a virtual eyeball model. *Mach. Vis. Appl.* **2009**, *20*, 319–337. [[CrossRef](#)]
26. Wang, J.G.; Sung, E.; Venkateswarlu, R. Estimating the eye gaze from one eye. *Comput. Vis. Image Underst.* **2005**, *98*, 83–103. [[CrossRef](#)]
27. Ryoung, P.K. A real-time gaze position estimation method based on a 3-d eye model. *IEEE Trans. Syst. Man Cybern. Part B-Cybern.* **2007**, *37*, 199–212.
28. Topal, C.; Dogan, A.; Gerek, O.N. A wearable head-mounted sensor-based apparatus for eye tracking applications. In Proceedings of the IEEE Conference on Virtual Environments, Human-computer Interfaces and Measurement Systems, Istanbul, Turkey, 14–16 July 2008; pp. 136–139.
29. Ville, R.; Toni, V.; Outi, T.; Niemenlehto, P.-H.; Verho, J.; Surakka, V.; Juhola, M.; Lekkahla, J. A wearable, wireless gaze tracker with integrated selection command source for human-computer interaction. *IEEE Trans. Inf. Technol. Biomed.* **2011**, *15*, 795–801.
30. Noris, B.; Keller, J.B.; Billard, A. A wearable gaze tracking system for children in unconstrained environments. *Comput. Vis. Image Underst.* **2011**, *115*, 476–486. [[CrossRef](#)]
31. Stengel, M.; Grogorick, S.; Eisemann, M.; Eisemann, E.; Magnor, M. An affordable solution for binocular eye tracking and calibration in head-mounted displays. In Proceedings of the ACM Multimedia conference for 2015, Brisbane, Queensland, Australia, 26–30 October 2015; pp. 15–24.
32. 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](#)]
33. Schneider, E.; Dera, T.; Bard, K.; Bardins, S.; Boening, G.; Brand, T. Eye movement driven head-mounted camera: It looks where the eyes look. In Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Waikoloa, HI, USA, 12 October 2005; pp. 2437–2442.
34. Min, Y.K.; Yang, S.; Kim, D. Head-mounted binocular gaze detection for selective visual recognition systems. *Sens. Actuator A-Phys.* **2012**, *187*, 29–36.
35. Fuhl, W.; Kübler, T.; Sippel, K.; Rosenstiel, W.; Kasneci, E. Excuse: Robust pupil detection in real-world scenarios. In Proceedings of the 16th International Conference on Computer Analysis of Images and Patterns (CAIP), Valletta, Malta, 2–4 September 2015; pp. 39–51.
36. Fuhl, W.; Santini, T.; Kasneci, G.; Kasneci, E. PupilNet: Convolutional Neural Networks for Robust Pupil Detection, arXiv preprint arXiv: 1601.04902. Available online: <http://arxiv.org/abs/1601.04902> (accessed on 19 January 2016).
37. Fuhl, W.; Santini, T.; Kübler, T.; Kasneci, E. ElSe: Ellipse selection for robust pupil detection in real-world environments. In Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications (ETRA' 16), Charleston, SC, USA, 14–17 March 2016; pp. 123–130.
38. 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.
39. Mohammadi, M.R.; Raie, A. Selection of unique gaze direction based on pupil position. *IET Comput. Vis.* **2013**, *7*, 238–245. [[CrossRef](#)]
40. Arantxa, V.; Rafael, C. A novel gaze estimation system with one calibration point. *IEEE Trans. Syst. Man Cybern. Part B-Cybern.* **2008**, *38*, 1123–1138.
41. Coutinho, F.L.; Morimoto, C.H. Free head motion eye gaze tracking using a single camera and multiple light sources. In Proceedings of the SIBGRAPI Conference on Graphics, Patterns and Images, Manaus, Amazon, Brazil, 8–11 October 2006; pp. 171–178.
42. Beymer, D.; Flickner, M. Eye gaze tracking using an active stereo head. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Madison, WI, USA, 18–20 June 2003; pp. 451–458.

43. Zhu, Z.; Ji, Q. Eye gaze tracking under natural head movements. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 20–25 June 2005; pp. 918–923.
44. Magee, J.J.; Betke, M.; Gips, J.; Scott, M.R. A human–computer interface using symmetry between eyes to detect gaze direction. *IEEE Trans. Syst. Man Cybern. Part A-Syst. Hum.* **2008**, *38*, 1248–1261. [[CrossRef](#)]
45. Scott, D.; Findlay, J.M. *Visual Search, Eye Movements and Display Units*; IBM UK Hursley Human Factors Laboratory: Winchester, UK, 1991.
46. Wen, Z.; Zhang, T.N.; Chang, S.J. Eye gaze estimation from the elliptical features of one iris. *Opt. Eng.* **2011**, *50*. [[CrossRef](#)]
47. Wang, J.G.; Sung, E. Gaze determination via images of irises. *Image Vis. Comput.* **2001**, *19*, 891–911. [[CrossRef](#)]
48. Ebisawa, Y. Unconstrained pupil detection technique using two light sources and the image difference method. *WIT Trans. Inf. Commun. Technol.* **1995**, *15*, 79–89.
49. Morimoto, C.; Koons, D.; Amir, A.; Flicker, M. Framerate pupil detector and gaze tracker. In Proceedings of the International Conference on Computer Vision, Kerkyra, Greece, 20–27 September 1999; pp. 1–6.
50. Wang, C.W.; Gao, H.L. Differences in the infrared bright pupil response of human eyes. In Proceedings of Etra Eye Tracking Research and Applications Symposium, New Orleans, LA, USA, 25–27 March 2002; pp. 133–138.
51. Dodge, R.; Cline, T.S. The angle velocity of eye movements. *Psychol. Rev.* **1901**, *8*, 145–157. [[CrossRef](#)]
52. Tomono, A.; Iida, M.; Kobayashi, Y. A TV camera system which extracts feature points for non-contact eye movement detection. In Proceedings of the SPIE Optics, Illumination, and Image Sensing for Machine Vision, Philadelphia, PA, USA, 1 November 1989; pp. 2–12.
53. Ebisawa, Y. Improved video-based eye-gaze detection method. *IEEE Trans. Instrum. Meas.* **1998**, *47*, 948–955. [[CrossRef](#)]
54. Hu, B.; Qiu, M.H. A new method for human-computer interaction by using eye gaze. In Proceedings of the IEEE International Conference on Systems, Man, & Cybernetics, Humans, Information & Technology, San Antonio, TX, USA, 2–5 October 1994; pp. 2723–2728.
55. Hutchinson, T.E.; White, K.P.; Martin, W.N.; Reichert, K.C.; Frey, L.A. Human-computer interaction using eye-gaze input. *IEEE Trans. Syst. Man Cybern.* **1989**, *19*, 1527–1534. [[CrossRef](#)]
56. Cornsweet, T.N.; Crane, H.D. Accurate two-dimensional eye tracker using first and fourth purkinje images. *J. Opt. Soc. Am.* **1973**, *63*, 921–928. [[CrossRef](#)] [[PubMed](#)]
57. Glenstrup, A.J.; Nielsen, T.E. Eye Controlled Media: Present and Future State. Master’s Thesis, University of Copenhagen, Copenhagen, Denmark, 1 June 1995.
58. Mimica, M.R.M.; Morimoto, C.H. A computer vision framework for eye gaze tracking. In Proceedings of XVI Brazilian Symposium on Computer Graphics and Image Processing, Sao Carlos, Brazil, 12–15 October 2003; pp. 406–412.
59. Morimoto, C.H.; Mimica, M.R.M. Eye gaze tracking techniques for interactive applications. *Comput. Vis. Image Underst.* **2005**, *98*, 4–24. [[CrossRef](#)]
60. Jian-Nan, C.; Peng-Yi, Z.; Si-Yi, Z.; Chuang, Z.; Ying, H. Key Techniques of Eye Gaze Tracking Based on Pupil Corneal Reflection. In Proceedings of the WRI Global Congress on Intelligent Systems, Xiamen, China, 19–21 May 2009; pp. 133–138.
61. Feng, L.; Sugano, Y.; Okabe, T.; Sato, Y. Adaptive linear regression for appearance-based gaze estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 2033–2046.
62. Cherif, Z.R.; Nait-Ali, A.; Motsch, J.F.; Krebs, M.O. An adaptive calibration of an infrared light device used for gaze tracking. In Proceedings of the IEEE Instrumentation and Measurement Technology Conference, Anchorage, AK, USA, 21–23 May 2002; pp. 1029–1033.
63. Cerrolaza, J.J.; Villanueva, A.; Cabeza, R. Taxonomic study of polynomial regressions applied to the calibration of video-oculographic systems. In Proceedings of the 2008 symposium on Eye Tracking Research and Applications, Savannah, GA, USA, 26–28 March 2008; pp. 259–266.
64. Cerrolaza, J.J.; Villanueva, A.; Cabeza, R. Study of polynomial mapping functions in video-oculography eye trackers. *ACM Trans. Comput.-Hum. Interact.* **2012**, *19*, 602–615. [[CrossRef](#)]
65. Baluja, S.; Pomerleau, D. Non-intrusive gaze tracking using artificial neural networks. *Neural Inf. Process. Syst.* **1994**, *6*, 753–760.

66. Piratla, N.M.; Jayasumana, A.P. A neural network based real-time gaze tracker. *J. Netw. Comput. Appl.* **2002**, *25*, 179–196. [[CrossRef](#)]
67. Demjen, E.; Abosi, V.; Tomori, Z. Eye tracking using artificial neural networks for human computer interaction. *Physiol. Res.* **2011**, *60*, 841–844. [[PubMed](#)]
68. Coughlin, M.J.; Cutmore, T.R.H.; Hine, T.J. Automated eye tracking system calibration using artificial neural networks. *Comput. Meth. Programs Biomed.* **2004**, *76*, 207–220. [[CrossRef](#)] [[PubMed](#)]
69. Sesin, A.; Adjouadi, M.; Cabrerizo, M.; Ayala, M.; Barreto, A. Adaptive eye-gaze tracking using neural-network-based user profiles to assist people with motor disability. *J. Rehabil. Res. Dev.* **2008**, *45*, 801–817. [[CrossRef](#)] [[PubMed](#)]
70. Gneo, M.; Schmid, M.; Conforto, S.; D'Alessio, T. A free geometry model-independent neural eye-gaze tracking system. *J. NeuroEng. Rehabil.* **2012**, *9*, 17025–17036. [[CrossRef](#)] [[PubMed](#)]
71. Zhu, Z.; Ji, Q. Eye and gaze tracking for interactive graphic display. *Mach. Vis. Appl.* **2002**, *15*, 139–148. [[CrossRef](#)]
72. Kiat, L.C.; Ranganath, S. One-time calibration eye gaze detection system. In Proceedings of International Conference on Image Processing, Singapore, 24–27 October 2004; pp. 873–876.
73. Wu, Y.L.; Yeh, C.T.; Hung, W.C.; Tang, C.Y. Gaze direction estimation using support vector machine with active appearance model. *Multimed. Tools Appl.* **2014**, *70*, 2037–2062. [[CrossRef](#)]
74. Fredric, M.H.; Ivica, K. *Principles of Neurocomputing for Science and Engineering*; McGraw Hill: New York, NY, USA, 2001.
75. Oyster, C.W. *The Human Eye: Structure and Function*; Sinauer Associates: Sunderland, MA, USA, 1999.
76. Sliney, D.; Aron-Rosa, D.; DeLori, F.; Fankhauser, F.; Landry, R.; Mainster, M.; Marshall, J.; Rassow, B.; Stuck, B.; Trokel, S.; *et al.* Adjustment of guidelines for exposure of the eye to optical radiation from ocular instruments: Statement from a task group of the International Commission on Non-Ionizing Radiation Protection. *Appl. Opt.* **2005**, *44*, 2162–2176. [[CrossRef](#)] [[PubMed](#)]
77. Wang, J.Z.; Zhang, G.Y.; Shi, J.D. Pupil and glint detection using wearable camera sensor and near-infrared led array. *Sensors* **2015**, *15*, 30126–30141. [[CrossRef](#)] [[PubMed](#)]
78. Lee, J.W.; Heo, H.; Park, K.R. A novel gaze tracking method based on the generation of virtual calibration points. *Sensors* **2013**, *13*, 10802–10822. [[CrossRef](#)] [[PubMed](#)]
79. Gwon, S.Y.; Jung, D.; Pan, W.; Park, K.R. Estimation of Gaze Detection Accuracy Using the Calibration Information-Based Fuzzy System. *Sensors* **2016**. [[CrossRef](#)] [[PubMed](#)]



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).