IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 12, 2018, accepted April 13, 2018, date of publication April 19, 2018, date of current version May 16, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2828400

A Geometry-Appearance-Based Pupil Detection Method for Near-Infrared Head-Mounted Cameras

JIANFENG LI[1,2], (Member, IEEE), SHIGANG LI[2,3], (Member, IEEE), TONG CHEN[1,2], (Member, IEEE), AND YIGUANG LIU[4], (Senior Member, IEEE)

[1]School of Electronic and Information Engineering, Southwest University, Chongqing 400715, China
[2]Chongqing Key Laboratory of Non-linear Circuit and Intelligent Information Processing, Southwest University, Chongqing 400715, China
[3]Graduate School of Information Sciences, Hiroshima City University, Hiroshima 731-3194, Japan
[4]College of Computer Science, Sichuan University, Chengdu 610065, China

Corresponding author: Shigang Li (shigangli@hiroshima-cu.ac.jp)

This work was supported by the Fundamental Research Funds for the Central Universities under Grant SWU20710916.

ABSTRACT This paper presents, for the first time, a novel pupil detection method for near-infrared head-mounted cameras, which relies not only on image appearance to pursue the shape and gradient variation of the pupil contour, but also on structure principle to explore the mechanism of pupil projection. There are three main characteristics in the proposed method. First, in order to complement the pupil projection information, an eyeball center calibration method is proposed to build an eye model. Second, by utilizing the deformation model of pupils under head-mounted cameras and the edge gradients of a circular pattern, we find the best fitting ellipse describing the pupil boundary. Third, an eye-model-based pupil fitting algorithm with only three parameters is proposed to fine-tune the final pupil contour. Consequently, the proposed method extracts the geometry-appearance information, effectively boosting the performance of pupil detection. Experimental results show that this method outperforms the state-of-the-art ones. On a widely used public database (LPW), our method achieves 72.62% in terms of detection rate up to an error of five pixels, which is superior to the previous best one.

INDEX TERMS Pupil detection, head-mounted camera, image-based, eye-model-based, eyeball center calibration.

I. INTRODUCTION

Estimation of gaze direction in everyday lives has become an increasingly active research topic in the field of computer vision, as it plays an important role in many applications such as computer-human interaction. Usually, a gaze estimation device consists of a scene camera and an eye camera. The eye camera is used to capture eye image, estimate the gaze direction of users at the coordinate system of the eye camera, and then, the estimate gaze direction is mapped to the scene camera to determine the gaze point in observed scenes. To cope with significant variability in the eye appearance caused by illumination, reflections, occlusions, and other factors, near-infrared light sources are usually used to illuminate eyes so as to acquire a clearer image in which pupil region become much darker than the surrounding area, but it is still challenging as shown in Fig. 1.

To estimate the gaze direction from such near-infrared eye images, the first step is to find the contour of pupils. Since the error of pupil detection directly influences the accuracy of gaze estimation, how to find the accurate contour of pupils is a basic and key technology. Early researches about pupil detection mainly focus on detecting pupil in laboratory environment [1], [2]. Due to a variety of difficulties occurring when using eye trackers, such as illumination changing, motion blur, recording errors and eyelashes covering the pupil [3], These methods cannot effectively extract the pupil contour in our everyday lives, such as driving [4], [5], shopping [6], or simply walking around [7].

As for our best knowledge, in almost all researches except [29], which uses a 3D model of the eye that constantly updates based on observations of the eye, an eye-model is not exploited in the process for pupil detection of near-infrared eye images. We argue that an eye-model-based approach enable us to find more accurate contour of pupils, considering the position of eyeball center at the coordinate

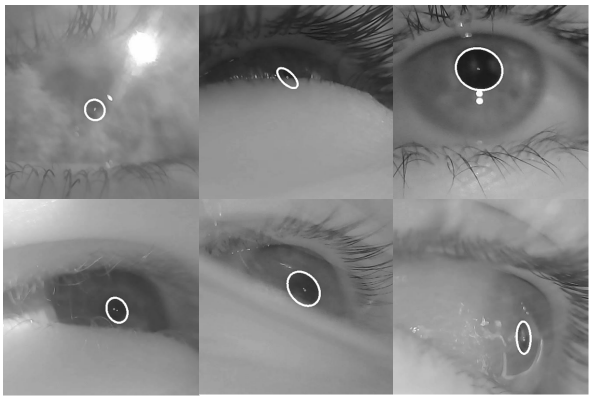


FIGURE 1. Detection samples by our proposed pupil detection method on LPW.

system of head-mounted eye camera does not change for a head-mounted head camera.

In this paper, we introduce an eye-model to pupil detection of near-infrared eye images. The contribution of this paper is as follows.

- We propose an eyeball center calibration method for head-mounted near-infrared cameras by multi-frames strategy and geometry principle.
  In this paper, considering that in consecutive frames, different pupil contour projections correspond to a same eye model, we propose an eyeball center calibration method to obtain the eyeball center position. The pupil contour appears as an ellipse on images, and from the five parameters of an ellipse, we can infer the unit vector of gaze direction [8]. Meanwhile, a ray from camera center passes through the 2D pupil center also passes through the 3D pupil center. However, as the true scale of pupil is variable for an image, we only achieve a set of possible 3D pupil centers on the ray even we know the scale range. Therefore, from one frame, we obtain a set of possible 3D eyeball centers. By multi-frames strategy and geometry principle, we derive an objective function to solve the eyeball center position.
- We propose a pupil selection method by using the pupil deformation model and the edge gradients of a circular pattern.
  For head-mounted camera, once the eyeball is determined, the movement of pupil is also determined. Based on the biological structure, pupil moves on the surface of the eyeball in a fixed model, so the pupil contour moves on the image also in a fixed deformation model. In this paper, we propose a pupil detection method using both the knowledge of 3D spatial projection and 2D image information. Instead of only using the appearance on images, our method combines the deformation model with the edge gradients of a circular pattern. Unlike the existing methods, the dependence of the threshold is greatly reduced. Moreover, principle of our method takes into account biological conditions, which makes selecting the best fitting ellipse describing the pupil boundary more logical and reliable.
- We propose an eye-model based pupil fitting algorithm with only three parameters to fine-tune the final pupil contour.
  After selecting the pupil boundary, the only way to obtain the final pupil contour is ellipse fitting since the pupil shows as an ellipse on images [9]. However, researches show that the selected boundary sometimes could not perform as well as what pupil contour should be, due to various kinds of interference. In these situations, the accuracy of ellipse fitting will vary depending on the situation of the interference points. With the help of eye model, a pupil fitting algorithm with only three parameters is proposed to fine-tune the final pupil contour.

The remainder of the paper is organized as follows. Section 2 describes related works. Section 3 introduces the principle of our method, Section 4 presents the experimental results, and Section 5 concludes the work.

II. RELATED WORK

In this section, we first introduce the related works focused on eyeball calibration, and then introduce pupil detection in head-mounted trackers.

A. RESEARCH ON EYEBALL CALIBRATION

Eyeball calibration is a preliminary and necessary step for eye-model based approach. For the moment, eyeball calibration is more common in remote trackers. When a remote camera is used, users' head moves freely at the camera coordinate system. To estimate the gaze, it needs to estimate the center position of eyeball continuously. Conventional methods are that calibrate the eyeball center in head coordinate system, and then estimate the head pose continuously [10].

Xiong *et al.* [11] use a simple onetime calibration procedure to obtain the eyeball center. 9 points are predefined on the monitor screen to be looked at; the calibration is achieved by minimizing the sum of angles between predicted gaze directions and the ground truth. Although 9 points pattern method is a very common way for eyeball calibration, there are various principles in different researches. Chen *et al.* [12] compute the 3D position of the eyeball center based on the middle point of two eye corners. While Jianfeng and Shigang [13] use the principle that the optical axis and visual axis have a fixed angle.

Different from a remote camera, a head-mounted eye camera has a direct view of eye and no head pose is needed. Moreover, the gaze point can be easily mapped to the viewed scene based on the center of the pupil and a user-specific calibration routine. Thus, eyeball calibration isn't being paid much attention in head-mounted trackers. Indeed, when determining the gaze point, eyeball center may not be so important in head-mounted trackers as well as remote trackers, but there are still some researches on this

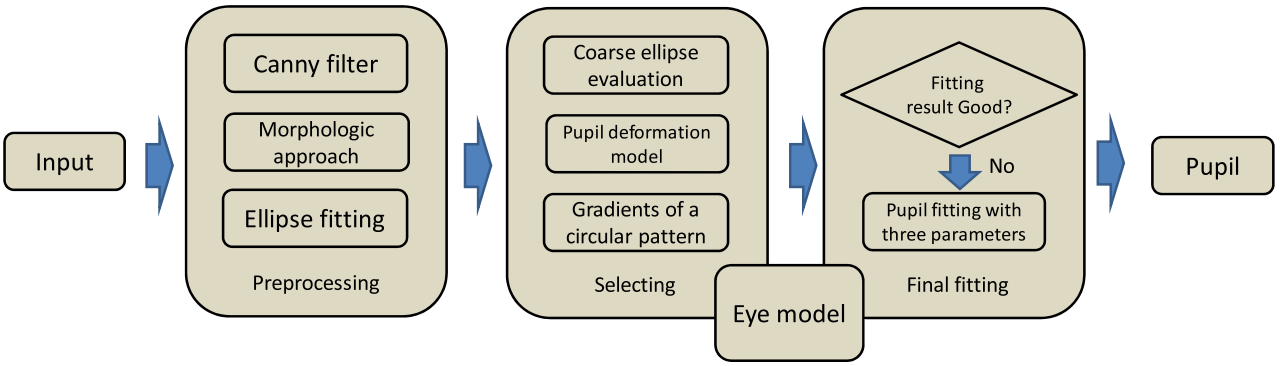


FIGURE 2. Flowchart of our proposed pupil detection method.

part, Swirsky and Dodgson [27] presented a 3D eye model fitting algorithm for gaze estimation, that operated on pupil ellipse geometry alone. Villanueva and Cabeza [28] proposed a model based on pupil contour and multiple lighting. Relying on the back project of 2D pupil contour, they solved the 3D pupil center and estimated the gaze direction. Recently, Li and Li [14] propose an algorithm of automatic online eyeball center calibration in visible light. Their works above inspire us that an accurate eye model can be helpful in pupil detection.

B. RESEARCH ON PUPIL DETECTION

Over the last years, there has been a large body of work on pupil detection for near-infrared head-mounted cameras [9]. Probably the most popular algorithm in this realm is Starburst [15]. They estimate the pupil contour by detecting edges along a limited number of rays that extend from a central best guess of the pupil center. Pupil fitting is finally performed following a Random Sample Consensus (RANSAC) paradigm to find the best fitting ellipse describing the pupil boundary. Then, Swirski *et al.* [16] proposed an algorithm, which starts with a coarse positioning by using Haar-like features, and refines the pupil center position by an image-aware RANSAC ellipse fitting. Pupil Labs also comes with a pupil detection algorithm by looking for darker areas and a final ellipse fit found through an augmented combinatorial search [17]. Besides, the SET approach [18] applies a region size threshold to extract pupil candidates, and the fitted ellipse that is closest to a circle is considered as the pupil. ExCuSe [19] and ElSe [20] are recently proposed; they are all based on edge filtering, ellipse evaluation, and pupil contour validation. No matter how different the process is, the final step of methods above is ellipse fitting, the least squares fitting of an ellipse is a common choice [21], however, it would be easily influenced by outliers.

According to [9], currently ElSe performs best among other existing methods. However, same as other methods, ElSe only explores image appearance and strictly depends on threshold. Sometimes, it could not find a good pupil edge after their proposed selecting strategy. To solve the problem, they apply an additional method that tries to find the pupil by determining a likely location, which is a coarse positioning.

In this paper, different from existing methods, we first propose an eyeball center calibration method for head-mounted near-infrared camera by multi-frames strategy and geometry principle. Then, based on the eyeball center, we detect the pupil by using the pupil deformation model and the edge gradients of a circular pattern. Last, we propose an eye-model based pupil fitting algorithm with only three parameters to fine-tune the final pupil contour.

III. PROPOSED METHOD

The flowchart of the proposed method is shown in Fig. 2. The preprocessing for proposed pupil detection method is similar to traditional image-based methods. We first detect the edges by a Canny filter [22], then using the morphologic approach [20] to filter edges, and do ellipse fitting for each edge to obtain the possible pupil contour set.

With the possible pupil contour set, we carry out a coarse ellipse evaluation, judging by size and aspect ratio. We set the threshold very loose since we just want to eliminate extremes. Then we select the best pupil contour by pupil deformation model (Section 3B) and gradients of a circular pattern (Section 3C).

With the best pupil contour, we compare the original edge with its fitted ellipse contour. If the reappearance rate is above 90%, we think it fits very well; no further processing is needed. Otherwise, there must be some outliers inside, and then we use our proposed pupil fitting algorithm with three parameters to fine-tune the final contour (Section 3D).

A known eye model is the basis of the proposed deformation model and pupil fitting algorithm; thus, we calibrate the eyeball center at the beginning (Section 3A).

A. EYEBALL CENTER CALIBRATION

To build an eye model, the 3D eyeball center position relative to the head-mounted camera is needed. Thus, a calibration is conducted. Here, we first present the constraints of eyeball

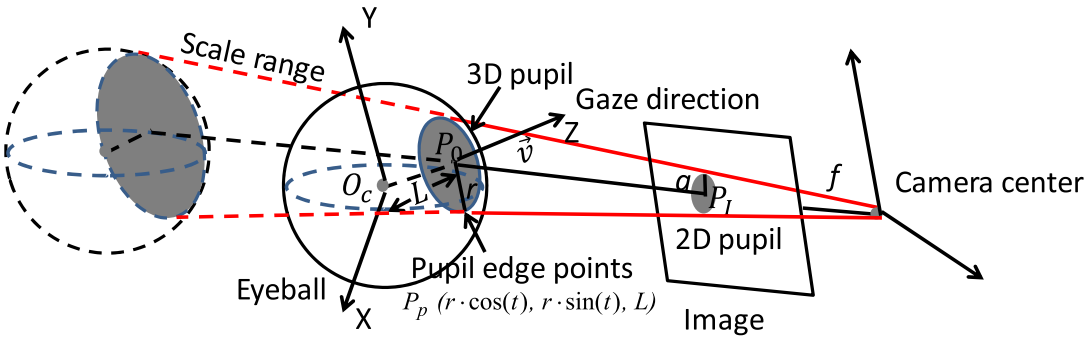


FIGURE 3. Diagram of our eye model and pupil projection principle from a head-mounted camera.

center, and then, explain how to determine the initial values to solve the set of non-linear equations.

1) CONSTRAINTS OF EYEBALL CENTER

As Fig. 3 shows, given an ellipse contour describing the pupil boundary on a frame, we can obtain its ellipse center $P_I(u_p, v_p)$ and compute its corresponding unit vector of gaze direction $\vec{v}(v_x, v_y, v_z)$ [8]. Assuming the eyeball center is $O_C(x_C, y_C, z_C)$, the distance between eyeball center and pupil center is L; here, we use the average biological value as $L = 10.5$ mm [23]. If only considering the affine projection principle, without considering the very small error from cornea refraction [26], a ray from the camera center passes through the 2D pupil center P_I also passes through the 3D pupil center $P_O(x_O, y_O, z_O)$. Since the 3D pupil center is on the surface of the eyeball, the ray intersects the 3D eyeball. Thus, we can obtain the equation below:

$$\begin{cases} \dfrac{x_O}{u_p - u_0} = \dfrac{y_O}{v_p - v_0} = \dfrac{z_O}{f} \\ \sqrt{(x_O - x_C)^2 + (y_O - y_C)^2 + (z_O - z_C)^2} = L, \end{cases} \tag{1}$$

where (u_0, v_0) is the principle point of the image, and f is the focal length. From Equation (1), 3D pupil center P_O can be expressed by 3D eyeball center O_C and ellipse center $P_I(u_p, v_p)$

$$P_O = f(O_C, P_I). \tag{2}$$

Let unit vector of gaze direction be $\overrightarrow{v} = \frac{P_O - O_C}{\|P_O - O_C\|}$ (see Fig.3). By substituting P_O by Equation (2), we can express unit vector of gaze direction as

$$\overrightarrow{v} = \frac{f(O_C, P_I) - O_C}{\|f(O_C, P_I) - O_C\|}. \tag{3}$$

As we mentioned previously, two unit vectors of gaze direction $\vec{v}(v_x, v_y, v_z)$ are inferred from an ellipse contour [8]. In [8], the authors use eye corners to solve the ambiguities. In our method, we can easily solve the ambiguities by judging which gaze direction is more close to the initial eyeball center. The initial eyeball center is obtained in next section.

From n frames, we obtain a set of ellipse centers $P_{Ij}(j = 0, 1, \ldots, n)$ and a set of unit gaze directions $\overrightarrow{v_j}(j = 0, 1, \ldots, n)$, obtained from [8], and $\overrightarrow{v_j'}(j = 0, 1, \ldots, n)$, obtained from Equation (3), respectively. Finally, we have the following final objective functions to solve the 3D eyeball center $O_C(x_C, y_C, z_C)$.

$$\min\left(\sum_{j=0}^{n} \left\| \overrightarrow{v_j'} - \overrightarrow{v_j} \right\|^2\right), \quad (j = 0, 1, \ldots, n) \tag{4}$$

2) INITIAL VALUE OF EYEBALL CENTER

To let Equation (4) be converged, an initial value is needed. Given an ellipse contour describing the pupil boundary on a frame (Fig. 3), its long axis is a, its ellipse center is $P_I(u_p, u_0)$ and the corresponding unit gaze direction is $\vec{v}(v_x, v_y, v_z)$. As we mentioned in advance, the true scale of pupil varies in a range, normally, the radius of a pupil ranges r from 1mm to 5 mm. By taking 0.1mm as the interval, we have a set of possible scales $r_i = \{1, 1.1, \ldots, 5\}\,(i = 1, 2, \ldots, 50)$. Then, we could have a set of possible depths of 3D pupil centers z_{Oi}.

$$z_{Oi} = \frac{r_i \cdot f}{a} \cdot (i = 1, 2, \ldots, 50) \tag{5}$$

Thus, the 3D pupil center can be:

$$P_{Oi} = \left(\frac{r_i \cdot (u_p - u_0)}{a}, \frac{r_i \cdot (v_p - v_0)}{a}, \frac{r_i \cdot f}{a}\right) \cdot (i = 1, 2, \ldots, 50) \tag{6}$$

where (u_0, v_0) is the principle point of the image, and f is the focal length.

Given L as the distance between 3D pupil center and 3D eyeball center, 3D eyeball center can be:

$$O_{Ci} = P_{Oi} - L \cdot \vec{v} \cdot (i = 1, 2, \ldots, 50) \tag{7}$$

For next frame, we obtain an ellipse as well; its ellipse center is $P_{I1}(u_{p1}, u_{01})$ and the corresponding unit gaze direction is $\overrightarrow{v_1'}$. Since the 3D eyeball center is fixed, we use the 3D eyeball center set O_{Ci} from last frame to infer the possible 3D pupil center set P_{O1i} in this frame.

$$P_{O1i} = O_{Ci} + L \cdot \overrightarrow{v_1'} \cdot (i = 1, 2, \ldots, 50) \tag{8}$$

Then, we project P_{O1i} back to 2D image as P_{1i}', the best choice of the 3D eyeball center set O_{Ci}should meet the

distance between P'_{1i} and P_{I1} is the minimum. We also use n frames to constrain the initial value of eyeball center.

$$O_C = \arg \min_{i=1,2,\ldots,50} (\sum_{j=1}^{n} \left\| P'_{ji} - P_{Ij} \right\|_2). \quad (9)$$

B. PUPIL DEFORMATION MODEL

As we mentioned, after the preprocessing, we obtain lots of possible pupil contours. To eliminate most of false candidates, we propose a pupil deformation model.

For head-mounted cameras, once the eyeball is determined, the movement of pupil is also determined. Based on the biological structure, pupil moves on the surface of the eyeball in a fixed model, so the pupil contour moves on the image also in a fixed deformation model. Since we have obtained the 3D eyeball center position in Section 3A for current situation, the deformation model of current pupil should also be known. In our method, we use a coarse deformation model to help pupil contour selection. For example, when the user is gazing at the head-mounted camera, the projection of pupil would not have a big deformation, the length of major axis and minor axis are approximately the same. In this situation, the projection point of 3D pupil center and 3D eyeball center coincide. When the user is gazing at the left or right side of camera, the size would be reduced in horizontal direction. When the user is gazing at the up or down side of camera, the size would be reduced in vertical direction. We conclude a coarse deformation as Fig. 4. The model center is the projection point of 3D eyeball center (from Section 3A), and the further away from the model center, the deformation is greater. In our case, we do not discuss the situation that 2D pupil center is very near from the model center for avoiding arbitrary elimination caused by errors.

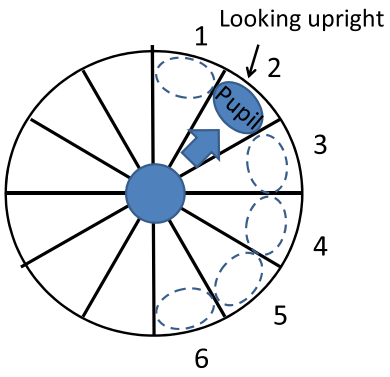


FIGURE 4. Pupil deformation model.

Due to symmetry, we only divide 180 degrees into 6 directions. For example, when the 2D pupil center is in area 2 (looking upright), the deviation angle of ellipse is very different from other areas. According to this principle, for all possible pupil contours in area 2, we eliminate those that deviation angle dose not satisfy the area 2 or the neighbor areas (area 1, 3). One sample frame is shown in Fig. 5. As you can see, there are many possible pupil contours in left figure; most of them are caused by eyelash, it is hard to eliminate them by traditional image-based methods. But after

FIGURE 5. One sample frame after deformation model.

deformation model, we eliminate almost all; the small bold white circle is the projection point of calibrated 3D eyeball center.

C. THE EDGE GRADIENTS OF A CIRCULAR PATTERN

As we stick to a coarse elimination, some possible pupil contours still remain. To select the best pupil contour, we use the contour gradient as well as internal pixel values of pupil.

As Fig. 6 shows, given the unit distance vector d from 2D pupil center C to contour edge point E, and the gradient of point E is g. For a contour, we compute the vector product of all edge points. Considering there are m edge points in a contour, so we can get average vector product below:

$$\sum_{i=0}^{m} (d_i \cdot g_i)/m. \quad (10)$$

In theory, the average of vector product of pupil should be maximum than any other contours. Thus, for remaining n possible pupil contours, we use equation (11) to select the best contour.

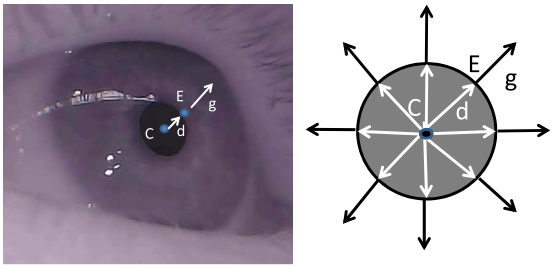


FIGURE 6. Vector product by the gradients of circular pattern.

$$\text{argmax} \frac{\{(\omega_j \sum_{m}^{i=0} d_{ji} \cdot g_{ji})/m\}}{j - 0, 1, \ldots, n}, \quad (11)$$

where $\omega_j = \text{Inv}(\text{Gauss}\left(I_{Cj}\right))/125$. ω_j is a weight considering the internal values inside the pupil should be low. Gauss $\left(I_{Cj}\right)$ is the ellipse center value after a Gauss filter, and the size of Gauss window is 21. Inv(Gauss $\left(I_{Cj}\right)$) makes the lower value higher. Therefore, ω_j ensures that ellipse with lower pixel values inside can have a higher weight.

D. PUPIL FITTING WITH THREE PARAMETERS

The last step on selected pupil contour is pupil fitting, the usual way is using ellipse fitting with five parameters. This fitting method is based on image information only, which is not robust against outliers.

Our proposed fitting algorithm is based on the eye model. In 3D space, pupil is nearly a circle. Based on the 3D pupil contour, the pupil coordinate system is built. In the pupil coordinate system, the origin is at the 3D eyeball center, the Z axis passes through the pupil center perpendicular to the pupil plane (Fig. 2).

Given the 3D pupil edge points$P_p(r \cdot \cos(t), r \cdot \sin(t), L)$, r is the real pupil radius, which is unknown. L is the distance between eyeball center and 3D pupil center P_0, and t is a parameter. By coordinate transformation, the pupil edge points (P_p) are then transformed to the camera coordinate system (P_c).

AssumingT(T1,T2,T3), Pc(xpc,ypc,zpc),

$$R = \begin{bmatrix} R_1 & R_2 & R_3 \\ R_4 & R_5 & R_6 \\ R_7 & R_8 & R_9 \end{bmatrix} \text{R=R1R2R3R4R5R6R7R8R9.}$$

$$\text{Pe} \begin{cases} x_{pc} = R_1 * r * \cos(t) + R_2 * r * \sin(t) + R_3 * L + T_1 \\ y_{pc} = R_4 * r * \cos(t) + R_5 * r * \sin(t) + R_6 * L + T_2 \\ z_{pc} = R_7 * r * \cos(t) + R_8 * r * \sin(t) + R_9 * L + T_3, \end{cases} \tag{12}$$

where the translation T is the 3D eyeball center position. As the pupil rotates on the eyeball surface, the roll angle does not change relative to the camera. Hence, the Rotation R can be expressed using roll, pitch, and yaw, where the roll is zero.

The pupil edge points in the camera coordinate system (P_c) are then projected back to a 2D image. The image points $P_i(u, v)$ Ipu,vcan be expressed as

$$\text{u} = \text{x}pc * \text{f}zpc + \text{u0}v = \text{y}pc * \text{f}zpc + \text{v0}, \quad \text{u}v1 = \text{M}xcyczc \tag{13}$$

where $M(u0, v0, f)M(u_0, v_0, f)$ is the internal camera parameter determined using a chessboard calibration. Substituting (12) into (13), we can express (13) as

$$\begin{aligned} \text{u} &= \text{f(sint,cost,pitch,yaw,R,T,a,b,L,M,Te)} \\ \text{v} &= \text{g(sint,cost,pitch,yaw,R,T,a,b,L,M,Te)} \\ &\times \begin{cases} u = f(\sin(t), \cos(t), R, T, r, L, M) \\ v = g(\sin(t), \cos(t), R, T, r, L, M), \end{cases} \end{aligned} \tag{14}$$

The equations are too long; hence, we did not include them here. Obviously, they are linear equations. The values for sin (t) and cos (t) can be easily obtained from (14)

$$\begin{cases} \sin(t) = h(u, v, R, T, r, L, M) \\ \cos(t) = k(u, v, R, T, r, L, M), \end{cases}$$
$$\text{sint} = \text{h(u,v,pitch,yaw,R,T,a,b,L,M,Te)}$$
$$\text{cost} = \text{k(u,v,pitch,yaw,R,T,a,b,L,M,Te)} \tag{15}$$

Using the relation, $\sin2(\text{t}) + \cos2(\text{t}) = 1$, from equation (15) we can have the following equation.

$$\Psi(u, v, R, T, r, L, M) = 0,$$
$$\Psi\text{u,v,pitch,yaw,R,T,a,b,L,M,Te} = 0 \tag{16}$$

where R is unknown, and the rotation matrix R can be expressed in terms of roll, pitch, and yaw, when the roll is zero, as mentioned previously. Besides, pupil scale r is also unknown.

As you can see, there are only three unknowns in equation (16). Because a set of 2D pupil edge points can be detected on the image, a set of over-constraint equations (16) can be obtained from one image. The Levenberg–Marquardt algorithm [24] is used to solve the equation set. After solving this equation set, we project pupil edge points $P_p(r \cdot \cos(t), r \cdot \sin(t), L)$ back to 2D images to find the final pupil contour.

IV. EVALUATION

To verify the effects of the proposed method, a series of experiments were conducted. Since the real 3D eyeball center position for each person is immeasurable, a simulation is conducted. Then, we conduct a simulation for our proposed pupil fitting algorithm. Finally, we test our proposed pupil detection method on public datasets: LPW. The following shows the details of experiments and results.

A. SIMULATION FOR EYEBALL CENTER CALIBRATION

We assume a 3D eyeball center position and ten gaze directions, so we can infer ten 3D pupil centers. Then we project these 3D pupil centers to 2D images. Thus, under this assumption, we have ten gaze directions and ten 2D pupil centers. By these synthetic data, given a probable initial value, we use our proposed objective function to solve it. The result is shown in Table 1. As you can see, the calculated values are consistent with the theoretical values if the initial value is quite rational. It also shows sometimes the convergent failure happens. We found the objective function has a high fault-tolerance on Z axis.

TABLE 1. Eyeball calibration validation.

Initial values (mm)	Final values (mm)	Ground truth (mm)
(-1.00, -4.00, 29.00)	(-1.999, -3.001, 28.006)	(-2.00, -3.00, 28.00)
(0.00, -5.00, 25.00)	(-1.999, -3.002, 28.002)	(-2.00, -3.00, 28.00)
(-1.00, -4.00, 29.00)	(-0.001, -5.999, 29.998)	(0.00, -6.00, 30.00)
(0.00, -5.00, 25.00)	(0.000, -6.000, 30.000)	(0.00, -6.00, 30.00)
(1.00, 0.00, 27.00)	(-2.297,-4.402,23.349)	(-2.30, -4.40, 23.34)
(-5.00, -6.00, 19.00)	(-2.301, -4.400, 23.339)	(-2.30, -4.40, 23.34)
(3.00, -8.00, 28.00)	(-2.299, -4.401, 23.346)	(-2.30, -4.40, 23.34)
(2.00, 0.00, 30.00)	(0.00, 0.00, 30.002)	(0.00, 0.0, 30.00)
(3.00, 0.00, 30.00)	Failed	(0.00, 0.0, 30.00)
(0.00, 4.00, 30.00)	(0.00, 0.00, 30.004)	(0.00, 0.0, 30.00)
(0.00, 5.00, 30.00)	Failed	(0.00, 0.0, 30.00)
(0.00, 0.00, 22.00)	(0.00, 0.00, 30.000)	(0.00, 0.0, 30.00)
(0.00, 0.00, 20.00)	(0.00, 0.00, 29.998)	(0.00, 0.0, 30.00)

To take into account observation noise, we add noise to 2D pupil centers and 3D gaze directions separately, and evaluate the final performance. We assume the 3D eyeball center position is *(0.0, 0.0, 30.0) mm*, then randomly generate ten gaze directions in 80 degrees view, the corresponding 2D

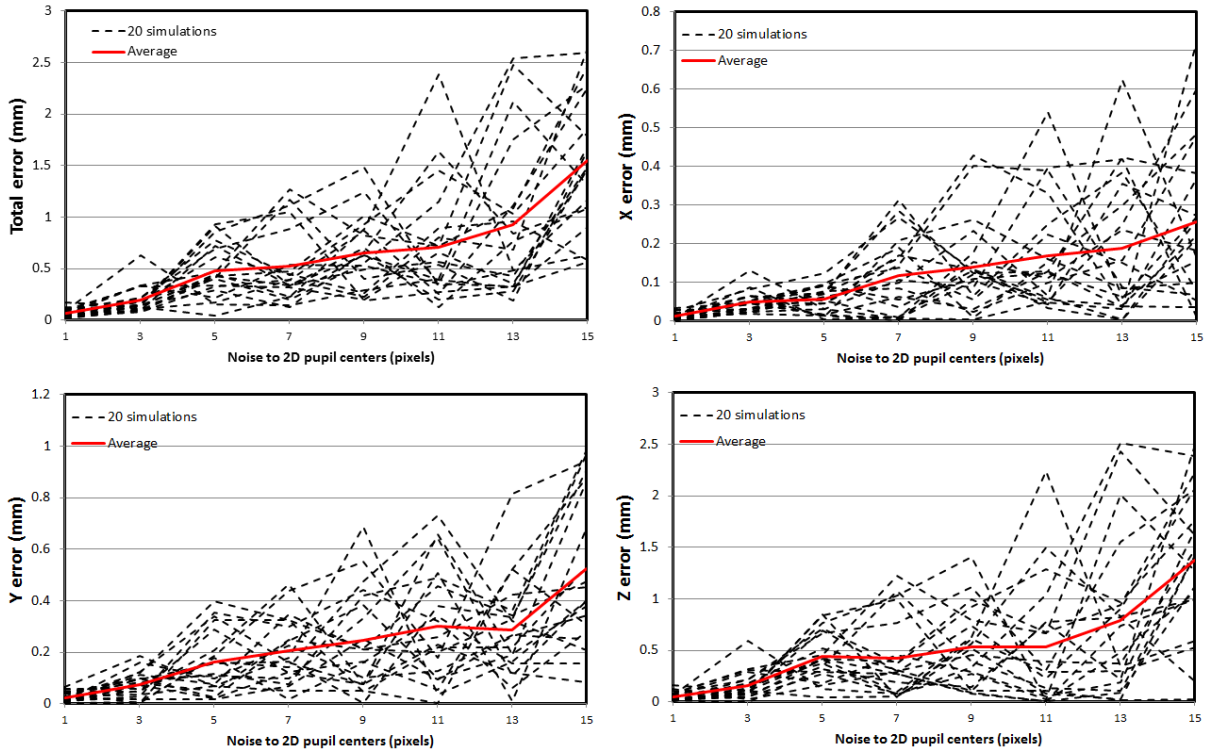


FIGURE 7. Eyeball center calibration error with noise to 2D pupil centers.

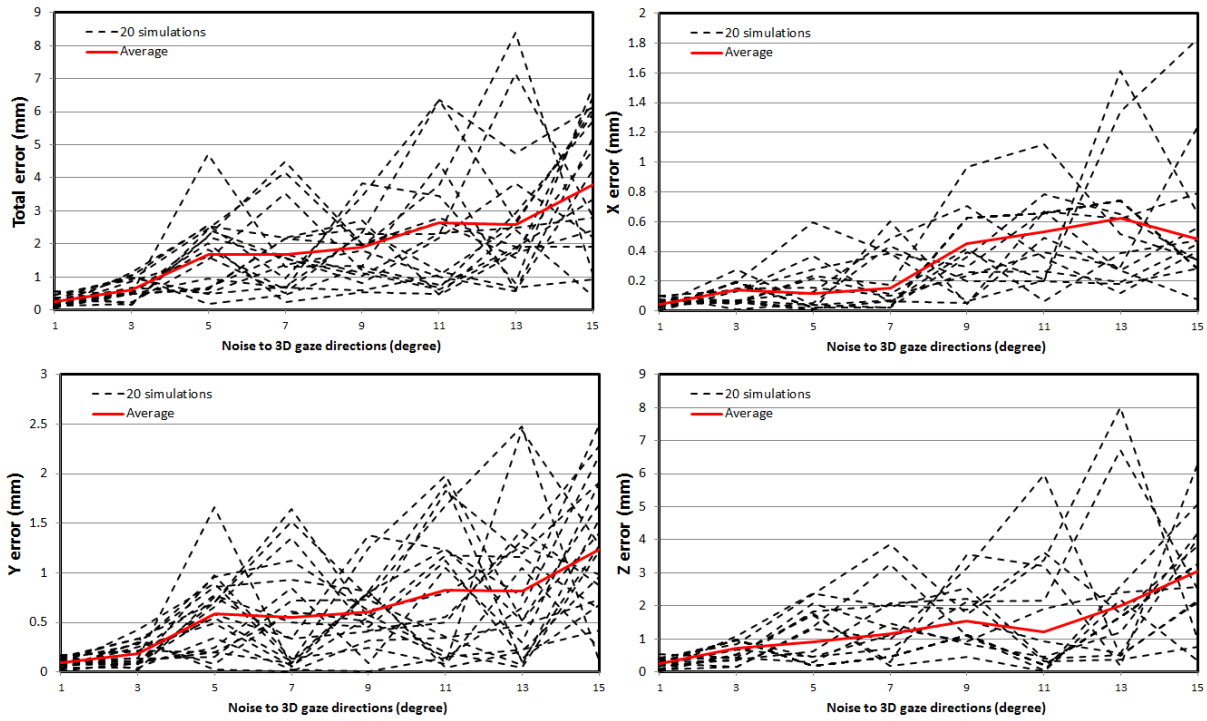


FIGURE 8. Eyeball center calibration error with noise to 3D gaze directions.

pupil centers are calculated based on 3D eyeball center and gaze directions.

We add noise to 2D pupil centers by (1, 3, 5, 7, 9, 11, 13, 15) Gaussian pixel errors separately, and 20 times are conducted. The calculated eyeball center error is shown in Fig. 7. The experiment shows that the average error from ground truth is 1.5 mm with 15 Gaussian pixel errors to 2D pupil centers, and mostly is from Z axis.

Then we add noise to 3D gaze directions by (1, 3, 5, 7, 9, 11, 13, 15) Gaussian degree errors separately, and 20 times are conducted. The calculated eyeball center error is shown in Fig. 8. The experiment shows that the average error is 4 mm with 15 Gaussian pixel errors to 3D gaze directions, and mostly is from Z axis.

Referring to the average biological value, we take the distance between eyeball center and pupil center as $L = 10.5mm$; normally, this parameter is in a range $\pm 1mm$. To study the sensibility of this parameter, a simulation is conducted. We assume the 3D eyeball center position is *(0.0, 0.0, 30.0) mm*, then randomly generate ten gaze directions, with the supposition of a $L = \{9.5, 10, 11, 11.5\}mm$, the corresponding 2D pupil centers are calculated. Then we still calibrate the eyeball center with our supposition of a $L = 10.5mm$, the errors on 3D eyeball center are shown in Table 2. It concludes that the sensitivity of L would only influence the Z coordinate. We think the reason is that L actually determines the scale of eyeball model, and the error is finally reflected on the Z coordinate.

TABLE 2. Sensibility with the supposition of L.

L	11mm	10mm	11.5mm	9.5mm
Total error from ground truth	1.36mm	1.50mm	2.61mm	3.16mm
X error from ground truth	2.72e-05mm	1.49e-05mm	1.10e-4 mm	1.90e-05 mm
Y error from ground truth	6.32e-05 mm	3.77e-05mm	2.40e-4 mm	4.98e-05 mm
Z error from ground truth	1.36mm	1.50mm	2.61mm	3.16mm

B. SIMULATION FOR PUPIL FITTING

In last section, we discussed the 3D eyeball center error. Thus, in this section, we provide a simulation of the effects of 3D eyeball center error on the accuracy of proposed pupil fitting.

At the camera coordinate system, assuming that the 3D eyeball center position and the pupil pose are known, the pupil contour, which is projected onto the image, can also be determined. The contour on the image represents the pupil location. Thus, these pupil edge points can be considered as the ground truth.

The error is added to the ground truth of 3D eyeball center position. Using the proposed pupil fitting to fit the pupil based on the simulated edge points, the final fitted pupil center error is shown in Fig.9. It shows that if the error on X or Y coordinate of 3D eyeball center is under 2.5mm, the final fitted pupil center error is up to an error of five pixels. It also shows that the final fitted pupil center error has a high fault-tolerance on Z axis. As the simulations in last section shows, the 3D eyeball error appears mostly in Z coordinate, while this section proves that it would not have much influence on the final pupil fitting result.

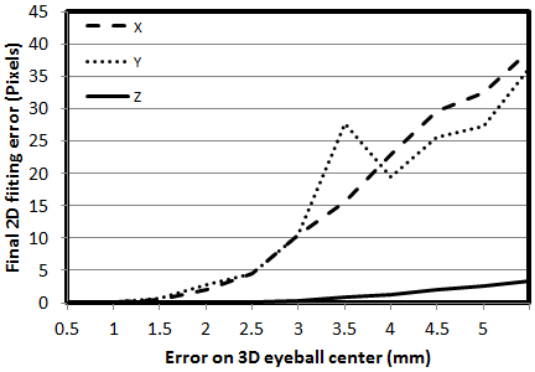


FIGURE 9. Proposed pupil fitting error with the error on 3D eyeball center.

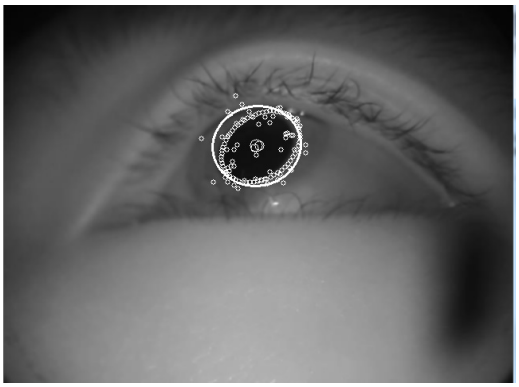


FIGURE 10. One simulation frame for pupil fitting.

Next we provide a simulation of the effects of edge point errors and outliers on the accuracy of fitting using traditional ellipse fitting and the proposed pupil fitting.

To validate the gaze direction against edge point errors, errors from 1 to 15 Gaussian pixel were added to the ground truth of pupil edge points. Based on the ground truth, the pupil fitting was performed using our method and traditional ellipse fitting. As Fig. 10 shows one simulation frame, the dotted line represents the contour by our proposed pupil fitting algorithm; the solid line represents the contour by ellipse fitting. The simulation shows that outliers have a major impact on ellipse fitting because ellipse fitting would take every point into consideration, while our pupil fitting is under the supervision of pupil projection. Since we found that sometimes even the fitted contour by ellipse fitting is very inaccurate, the fitted center is still close to the ground truth, at this time, we compare the gaze direction between two methods; the gaze direction is computed based on the fitting contour. The result is shown in Fig. 11. The error of our method increases slowly, around 2 degrees with 15 Gaussian pixel errors while the traditional algorithm has 16 degrees error.

C. SIMULATIONS ON CIRCULAR PATTERN

In real situations, the pupil is always an ellipse on images. Thus, to validate the proposed Circular Pattern can be used

TABLE 3. Performance comparison of state-of-the-art methods in terms of detection rate up to an error of five pixels.

LPW samples	SET (%)	Starburst (%)	Swirski (%)	ExCuSe (%)	ElSe (%)	Pupil Labs (%)	Proposed (%)
1	56.86	39.79	84.48	63.53	87.95	65.58	***90.22***
2	48.68	19.70	41.58	29.90	69.87	24.72	***76.97***
3	27.55	6.75	31.43	34.83	57.50	21.82	***58.28***
4	7.7	9.27	16.87	25.38	37.42	14.53	***46.15***
5	6.75	0.00	8.38	19.08	22.95	13.15	***26.95***
6	11.10	13.30	63.48	53.44	84.10	38.72	***87.05***
7	43.55	7.65	66.17	66.48	73.60	68.43	***81.60***
8	42.17	34.32	77.68	75.32	81.00	64.22	***86.58***
9	35.65	30.90	56.40	60.42	61.97	45.20	***62.03***
10	10.42	3.65	71.23	59.00	72.65	41.62	***80.10***
11	31.07	18.10	31.58	49.52	71.48	9.45	***73.40***
12	54.92	24.10	71.82	72.58	89.73	49.58	***90.35***
13	14.75	16.52	27.03	45.04	51.51	16.68	***52.94***
14	30.25	23.50	***76.07***	58.60	70.50	57.52	76.02
15	27.17	8.15	37.80	44.83	53.95	43.78	***54.60***
16	23.24	17.24	74.11	72.73	82.13	82.01	***85.10***
17	20.90	2.42	68.88	42.10	***72.97***	47.52	69.70
18	50.67	33.48	61.18	66.25	78.57	48.73	***80.65***
19	11.97	3.45	24.87	21.88	54.05	2.60	***68.02***
20	19.83	16.58	41.40	11.72	83.35	0.98	***90.17***
21	30.43	25.63	55.90	47.45	88.92	28.18	***89.65***
22	41.60	11.85	6.48	31.23	70.00	0.83	***71.03***

The best performance on each data set is shown in red.

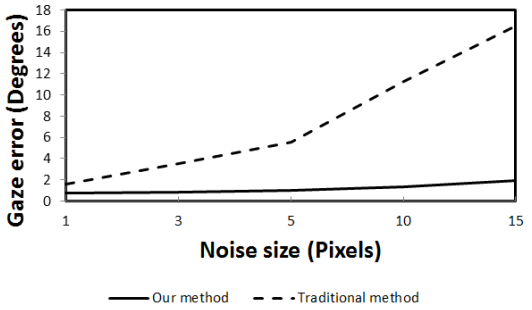


FIGURE 11. Average gaze error at different Gaussian noises.

FIGURE 12. Assume a black ellipse as pupil.

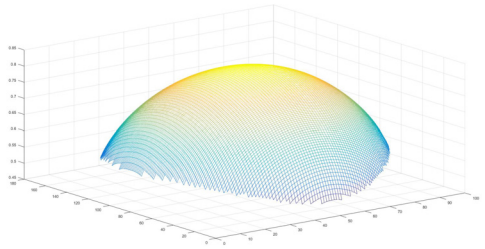


FIGURE 13. Average vector product for each point inside the ellipse.

in these situations, a simulation is conducted. We assume a black ellipse as a pupil (Fig. 12). Using Equation (10) to compute the average vector product for each point inside the ellipse (Fig. 13, 14), the result shows that the value of ellipse center is maximum; it proves the proposed circular pattern is enabled even for ellipse.

D. VALIDATION ON LPW

The recent Labeled Pupils in the Wild (LPW) dataset [25] contains 66 high-quality eye region videos that were recorded from 22 participants using a dark-pupil head-mounted eye tracker from Pupil Labs [17]. Each video in the dataset consists of about 2000 frames with a resolution of 640×480 pixels. The dataset is one order of magnitude larger than other datasets and covers a wide range of realistic indoor and outdoor illumination conditions, includes participants wearing glasses and eye make-up, as well as different ethnicities with variable skin tones, eye colors, and face shapes. According to the evaluation in [9], ElSe performs best on this dataset.

We compared our pupil detection method to six state-of-the-art approaches on LPW. The competitor algorithms are SET [18], Starburst [15], Swirski [16], ExCuSe [19], ElSe [20], and Pupil Labs [17]. Their results are taken from a review [9], we conducted some repeated trials with ElSe to verify their report, and their report is credible as the results

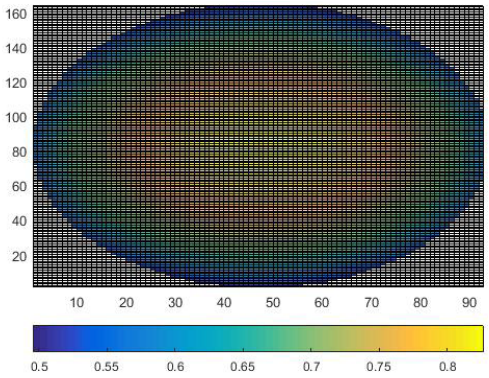


FIGURE 14. Average vector product for each point inside the ellipse (XY plane).

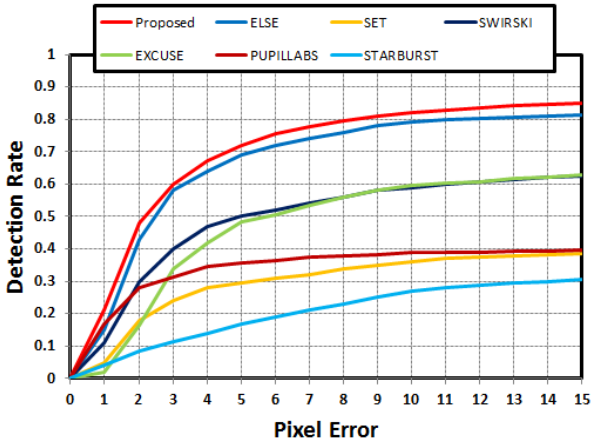


FIGURE 15. Average detection rates at different pixel distances for LPW.

coincide. Our method was applied with one fixed parameter setting to all videos. Some detection samples are shown in Fig. 1. Table 3 summarizes the performance results for each of the competing algorithms in terms of the detection rate up to an error of five pixels. In addition, Fig. 15 presents the performance of proposed method and its competitors in terms of the detection rate for different pixel error rates (0-15 pixels). On 20 out 22 data sets, proposed method clearly outperformed the other state-of-the-art algorithms, being thus the most promising approach toward robust pupil detection in heavily noisy eye images. At a pixel error of 5, the proposed method shows a detection rate of 72.62%. And the average detection rates of proposed method at different pixel errors show that the proposed method is better than other methods.

V. CONCLUSION

In this paper, we have proposed an eye-model based pupil detection method for head-mounted eye tracking in the wild. Our method relies not only on image appearance to pursue the shape and gradient variation of the pupil contour, but also on structure principle to explore the mechanism of pupil projection. In order to complement the pupil projection information, an eyeball center calibration method is proposed to build an eye model. Then, by utilizing the deformation model of pupils under head-mounted cameras and the edge gradients of a circular pattern, we find the best fitting ellipse describing the pupil boundary. Last, an eye-model based pupil fitting algorithm with only three parameters is proposed to fine-tune the final pupil contour. The proposed method extracts the geometry-appearance information, effectively boosting the performance of pupil detection. Experimental results show that this method outperforms the state-of-the-art ones.

Although we have introduced the projection information into pupil detection, it is just a cursory attempt. We believe it still has a great potential to improve the accuracy. We would keep focusing on this part of research.

REFERENCES

[1] S. Goni, J. Echeto, A. Villanueva, and R. Cabeza, "Robust algorithm for pupil-glint vector detection in a video-oculography eyetracking system," in *Proc. 17th Int. Conf. Pattern Recognit. (ICPR)*, vol. 4. Aug. 2004, pp. 941–944.

[2] L. Lin, L. Pan, L. Wei, and L. Yu, "A robust and accurate detection of pupil images," in *Proc. 3rd Int. Conf. Biomed. Eng. Inf. (BMEI)*, vol. 1. Oct. 2010, pp. 70–74.

[3] S. K. Schnipke and M. W. Todd, "Trials and tribulations of using an eye-tracking system," in *Proc. Extended Abstracts Hum. Factors Comput. Syst. (CHI)*, Apr. 2000, pp. 273–274.

[4] C. Braunagel, E. Kasneci, W. Stolzmann, and W. Rosenstiel, "Driver-activity recognition in the context of conditionally autonomous driving," in *Proc. IEEE 18th Int. Conf. Intell. Transp. Syst. (ITSC)*, Sep. 2015, pp. 1652–1657.

[5] X. Liu, F. Xu, and K. Fujimura, "Real-time eye detection and tracking for driver observation under various light conditions," in *Proc. IEEE Intell. Vehicle Symp.*, vol. 2. Jun. 2002, pp. 344–351.

[6] K. Sippel *et al.*, "Binocular glaucomatous visual field loss and its impact on visual exploration—A supermarket study," *PLoS ONE*, vol. 9, no. 8, p. e106089, 2014.

[7] Y. Sugano and A. Bulling, "Self-calibrating head-mounted eye trackers using egocentric visual saliency," in *Proc. 28th Annu. ACM Symp. User Interface Softw. Technol.*, Nov. 2015, pp. 363–372.

[8] J. Wang, E. Sung, and R. Venkateswarlu, "Eye gaze estimation from a single image of one eye," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, Oct. 2003, pp. 136–143.

[9] W. Fuhl, M. Tonsen, A. Bulling, and E. Kasneci, "Pupil detection for head-mounted eye tracking in the wild: An evaluation of the state of the art," *Mach. Vis. Appl.*, vol. 27, no. 8, pp. 1275–1288, 2016.

[10] J. Li and S. Li, "Gaze estimation from color image based on the eye model with known head pose," *IEEE Trans. Human-Mach. Syst.*, vol. 46, no. 3, pp. 414–423, Jun. 2016.

[11] X. Xiong, Z. Liu, Q. Cai, and Z. Zhang, "Eye gaze tracking using an RGBD camera: A comparison with a RGB solution," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput., Adjunct Publication*, Sep. 2014, pp. 1113–1121.

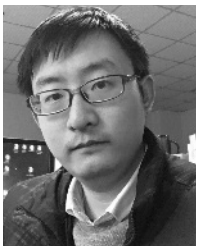
[12] J. Chen and Q. Ji, "3D gaze estimation with a single camera without IR illumination," in *Proc. 19th Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2008, pp. 1–4.

[13] L. Jianfeng and L. Shigang, "Eye-model-based gaze estimation by RGB-D camera," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2014, pp. 592–596.

[14] J. Li and S. Li, "Two-phase approach—Calibration and iris contour estimation—For gaze tracking of head-mounted eye camera," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 3136–3140.

[15] D. Li, D. Winfield, and D. J. Parkhurst, "Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops (CVPR Workshops)*, Jun. 2005, p. 79.

[16] L. Świrski, A. Bulling, and N. Dodgson, "Robust real-time pupil tracking in highly off-axis images," in *Proc. Symp. Eye Tracking Res. Appl.*, Mar. 2012, pp. 173–176.
[17] M. Kassner, W. Patera, and A. Bulling, "Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput., Adjunct Publication*, Sep. 2014, pp. 1151–1160.
[18] A.-H. Javadi, Z. Hakimi, M. Barati, V. Walsh, and L. Tcheang, "SET: A pupil detection method using sinusoidal approximation," *Frontiers Neuroeng.*, vol. 8, Apr. 2015, Art. no. 4.
[19] W. Fuhl, T. Kübler, K. Sippel, W. Rosenstiel, and E. Kasneci, "ExCuSe: Robust pupil detection in real-world scenarios," in *Proc. Int. Conf. Comput. Anal. Images Patterns*, Sep. 2015, pp. 39–51.
[20] W. Fuhl, T. C. Santini, T. Kübler, and E. Kasneci, "ElSe: Ellipse selection for robust pupil detection in real-world environments," in *Proc. 9th Biennial ACM Symp. Eye Tracking Res. Appl.*, Mar. 2016, pp. 123–130.
[21] D. Zhu, S. T. Moore, and T. Raphan, "Robust pupil center detection using a curvature algorithm," *Comput. Methods Programs Biomed.*, vol. 59, no. 3, pp. 145–157, 1999.
[22] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.
[23] E. D. Guestrin and M. Eizenman, "General theory of remote gaze estimation using the pupil center and corneal reflections," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 6, pp. 1124–1133, Jun. 2006.
[24] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *J. Soc. Ind. Appl. Math.*, vol. 11, no. 2, pp. 431–441, 1963.
[25] M. Tonsen, X. Zhang, Y. Sugano, and A. Bulling, "Labelled pupils in the wild: A dataset for studying pupil detection in unconstrained environments," in *Proc. 9th Biennial ACM Symp. Eye Tracking Res. Appl.*, Mar. 2016, pp. 139–142.
[26] A. Villanueva and R. Cabeza, "A novel gaze estimation system with one calibration point," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 38, no. 4, pp. 1123–1138, Aug. 2008.
[27] L. Świrski and N. A. Dodgson, "A fully-automatic, temporal approach to single camera, glint-free 3D eye model fitting," in *Proc. ECEM*, 2013, pp. 1–10.
[28] A. Villanueva and R. Cabeza, "Models for gaze tracking systems," *EURASIP J. Image Video Process.*, vol. 2007, 2007, Art. no. 023570. [Online]. Available: https://doi.org/10.1155/2007/23570
[29] [Online]. Available: https://docs.pupil-labs.com/#pupil-detection



JIANFENG LI (M'17) received the B.S. degree in computer science and technology from the Xi'an University of Technology, China, in 2010, the M.S. degree from the College of Computer Science, Sichuan University, China, in 2013, and the Ph.D. degree in information engineering from Tottori University, Japan, in 2017. His research interests include computer/robot vision, SLAM, and gaze estimation.

SHIGANG LI (M'93) received the B.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 1985, and the M.S. and Ph.D. degrees from Osaka University, Japan, in 1990 and 1993, respectively. He was with Osaka University as a Research Associate in 1993. He became an Associate Professor with the Faculty of Information Sciences, Hiroshima City University, Japan, in 1995. In 2001, he moved to the Faculty of Engineering, Iwate University, Japan. He became a Professor with the Graduate School of Engineering, Tottori University, Japan, in 2007.Since 2015, he has been a Professor with the Graduate School of Information Sciences, Hiroshima City University. His research interests include computer/robot vision, intelligent transportation system, and brain–computer interface.

TONG CHEN (M'14) received the B.E. degree in communication engineering from the Second Artillery Command Institute, China, in 2002, the M.Sc. degree in digital sound and vision processing from the University of Wales, U.K., in 2006, and the Ph.D. degree from Cranfield University, U.K., in 2013. He is currently a Professor with Southwest University, China. His research interests include image/signal processing, hyperspectral imaging, and affective computing.

YIGUANG LIU (SM'18) received the M.S. degree from Peking University, Beijing, China, in 1998, and the Ph.D. degree from Sichuan University, Chengdu, China, in 2004. He was a Research Fellow, a Visiting Professor, and a Senior Research Scholar with the National University of Singapore, Singapore, Imperial College London, London, U.K., and Michigan State University, East Lansing, MI, USA, respectively. He was chosen into the program for new century excellent talents of MOE in 2008 and chosen as a Scientific and Technical Leader in Sichuan Province in 2010. He is currently the Director of the Vision and Image Processing Laboratory and a Professor with the School of Computer Science, Sichuan University. He has co-authored over 100 international journal and conference papers and a chapter of the book *Computational Intelligence and Its Applications* (Imperial College Press, 2011). His research interests include computer vision and image processing, pattern recognition, and computational intelligence. He is a Reviewer of the Mathematical Reviews of the American Mathematical Society.