A Hybrid Method for Remote Eye Tracking using RGB-IR Camera

Kenta Yamagishi and Kentaro Takemura
Graduate School of Engineering, Tokai University, Hiratsuka, Japan

Keywords: Eye-tracking, RGB-IR Camera, 3D Eye Model.

Abstract: Methods for eye tracking using images can be divided largely into two categories: Methods using a near-infrared image and methods using a visible image. These images have been used independently in conventional eye-tracking methods; however, each category of methods have different advantageous features. Therefore, we propose using these images simultaneously to compensate for the weak points in each technique, and an RGB-IR camera, which can capture visible and near-infrared images simultaneously, is employed. Pupil detection can yield better results than iris detection because the eyelid often occludes the iris. On the other hand, the iris area can be used for model fitting because the iris size is constant. The model fitting can be automated at initialization; thus, the relationship between the 3D eyeball model and eye camera is solved. Additionally, the positions of the eye and gaze vectors are estimated continuously using these images for tracking. We conducted several experiments for evaluating the proposed method and confirmed its feasibility.

1 INTRODUCTION

In eye-tracking research, there are several important topics, one of which is the estimation of point-of-gaze. Many researchers have proposed various estimation methods with high accuracy(Morimoto and Mimica, 2005; Guestrin and Eizenman, 2006). Furthermore, calibration methods for computing a point-of-gaze and visual axis have also been discussed(Ohno et al., 2002; Nagamatsu et al., 2010). Recent works involve the use of an eye-tracking device in applications such as a head-mounted displays(Toyama et al., 2012) and other devices used in day-to-day activities(Kassner et al., 2014). In addition to wearable eye trackers, a remote eye tracker is also desired for daily applications. For example, a remote eye tracker can be used with digital signage, and we can measure the region of interest and the effect of the display contents.

We also have studied several eye-tracking methods including the 3D point-of-gaze(Takemura et al., 2014a) and corneal imaging(Takemura et al., 2014b). Additionally, in our previous research, we proposed pupil and iris tracking using an RGB-IR camera(Yamagishi and Takemura, 2017), and we achieved model-based tracking. However, the previous method was proposed for wearable eye tracker; hence, it is necessary to maintain the relationship between the camera and the eyeball during measurement. Therefore, we propose a method to estimate the eyeball position and the eye tracking for a remote and glint-free eye-tracking method using an RGB-IR camera. The limbus is suitable for model fitting because the size is constant. Strictly speaking, there is an individual difference in the size of human eyes, but we use a 3D model average size(Grand and Hage, 1980) for the model-based approach. Therefore, the eyeball position can be computed using the limbus. However, the eyelids often occlude part of the iris, which is a disadvantage. On the other hand, pupil size is changed by several factors such as illumination conditions and mental state, and thus it is difficult to use it in model fitting. However, pupil area can be detected easily using a near-infrared image, and occlusion rarely occurs. Therefore, visible and infrared images are used simultaneously to compensate for weak points in each image as a hybrid eye-tracking method.There is two main contribution in this research; first, we implemented a method for iris detection by converting to polar coordinates based on a detected pupil region with an RGB-IR camera for eye model fitting. In our previous research, the limbus was input manually for computing the initial relationship. Therefore, it was necessary to initialize without user operation. Second, the eye-tracker has tolerance for head movements as a remote eye-tracking. We estimated eyes position and gaze vector by detecting dual eyes using the pupils and 3D eyeball models. We conducted several experiments for evaluating the accuracy of the point-of-gaze and confirmed the feasibility of the re-

Yamagishi, K. and Takemura, K.
A Hybrid Method for Remote Eye Tracking using RGB-IR Camera.
DOI: 10.5220/0007582705910596
In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2019)*, pages 591-596
ISBN: 978-989-758-354-4

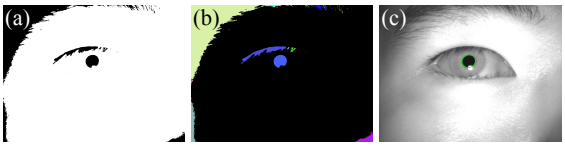

Figure 1: Blocks of binary images (a) are labeled by connected component labeling (b); the pupil area (c) is estimated using the degree of circularity.

mote eye-tracking.

2 PUPIL AND IRIS DETECTION

2.1 Pupil Detection using the Degree of Circularity

Previously, we proposed hybrid eye tracking using an RGB-IR camera, and the relationship between the eye camera and eyeball was computed using the limbus. However, the limbus was selected manually. Thus, the limbus detection is implemented as an auto initialization in this paper. The details of pupil and iris detection are described below. Stable tracking of the pupil has been studied previously by many researchers(Fuhl et al., 2016) using near-infrared images. Thus, the pupil is tracked using the near-infrared image captured by the IR camera to compute the pose of the eye, and the center of the pupil is used for detecting the optical axis of the eye. The procedure for detecting the center of the pupil is described below.

1. Convert the near-infrared image to a binary image (Figure 1(a))
2. Apply the connected component labeling to the binary image (Figure 1(b))
3. Compute the degree of circularity for each divided area
4. Select the area with the highest degree of circularity as the pupil area
5. Compute the center of the pupil using elliptical approximation (Figure 1(c))

In our implementation, we calculated the degree of circularity for each labeled area to detect the pupil area, as proposed by Hausner(Haubner, 1966). The degree of circularity ψ is computed by the following equation:

$$\psi = \frac{4\pi S}{L^2}, \tag{1}$$

where S and L are the area and perimeter of the pupil, respectively. After detecting the pupil area, the pupil is tracked by searching the neighboring area of the previously known position of the pupil.

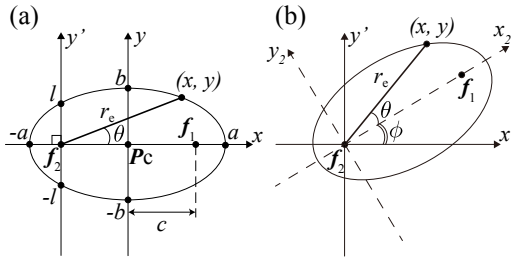

Figure 2: Deriving the equation of an ellipse.

2.2 Iris Detection using Polar Transformation

Tian et al.(Tian et al., 2012) proposed iris detection using polar transformation of the center of the pupil. This allowed the limbus to be detected as a straight line. However, the iris was assumed to be a perfect circle in the aforementioned study. In this study, we developed an improved method in which polar transformation can also be applied for iris detection when the iris is observed as an ellipse. For applying polar transformation to an ellipse, focal length $c = \sqrt{a^2 + b^2}$, eccentricity $e = c/a$ and semi-latus rectum $l = (a^2 - c^2)/a$ are required. These are computed using the pupil center $\mathbf{P}_c$, semi-major axis a and semi-minor axis b as shown in Figure 2(a). Therefore, the relationship between the rectangular coordinate system (x, y) and polar coordinate system (r_e, θ) is given as follows:

$$\begin{cases} x = f_{2x} - sr_e cos(\theta + \phi) \\ y = f_{2y} - sr_e sin(\theta + \phi), \end{cases} \tag{2}$$

where r_e and ϕ are the distance from the focal point to the elliptic curve and the slope of the ellipse, respectively, as shown in Figure 2(b), and r_e is computed using the following equation:

$$r_e = \frac{l}{1 + ecos\theta}. \quad (0 \leq \theta \leq 359) \in \mathbb{N} \tag{3}$$

Figure 3 shows the transformed eye images in the polar coordinate system, and values of each column are computed by changing scale s of Equation 2. When $s = 1$, the ellipse is normalized as $b = 1$. The pupil edge on the transformed image is detected as $s_{min} r_e$. Then, the color image is converted to the HSV color space, and the brightness image is generated. Next, we apply a bilateral filter to the brightness image, as shown in Figure 4(a). The differential images (Figure 4(b)) are computed for vertical variations, and the sum of the differential values are computed in each sr_e between 0 to 359 deg. The scale $\hat{s}$ for largest co-

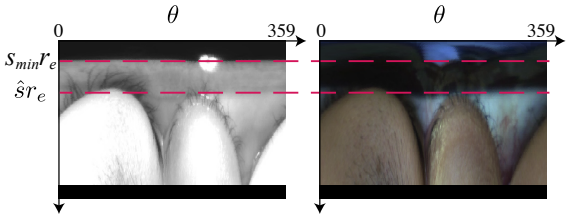

Figure 3: Transformed eye images in the polar coordinate system for detecting iris and pupil areas.

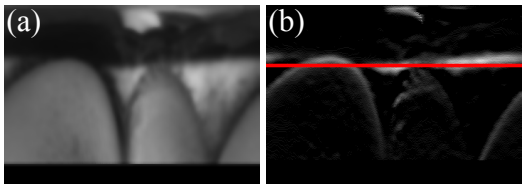

Figure 4: Brightness (a) of the eye image in HSV color space, and the differential image (b).

lumn sr_e is given by

$$\hat{s} = \underset{s\in\mathbb{N}}{argmax}(\sum_{\theta}(f(sr_e,\theta) - f(sr_e - 1,\theta))), \quad (4)$$

where $f(sr_e,\theta)$ is the brightness value for pixel (sr_e,θ). The red line shows the largest column, and the limbus is estimated as shown in Figure 4(b).

Theoretically, the edges of the limbus are extracted as straight lines. However, in reality, some distortion exists. The distortion is caused mainly by the difference between the centers of the iris and pupil. The center of the iris is different from the center of the pupil, as has been reported(Wilson et al., 1992). Thus, it is necessary to detect the limbus in the polar coordinate system for accurate detection of the iris. Consequently, recalculation is required for estimating the optimal limbus. Binary images (Figure 5(a)) are converted by using the following threshold:

$$T = \frac{\sum_\theta f(\hat{s}r_e,\theta)}{\sum_\theta \chi(\hat{s}r_e,\theta)}, \quad \chi(\hat{s}r_e,\theta) = \begin{cases} 1 & (f(\hat{s}r_e,\theta) \neq 0) \\ 0 & (f(\hat{s}r_e,\theta) = 0). \end{cases} \quad (5)$$

The connected components are searched around $\hat{s}r_e$, and the points of the connected components are projected by the input image, as shown in Figure 5(b), and the iris is detected based on these points using the least-square method and RANSAC, as shown in Figure 5(c). Figure 6 shows the comparison of images generated by polar transformation of a perfect circle and of an ellipse. It is observed that the edge of the iris has high linearity even when an ellipse is subject to polar transformation, and the proposed method can fit limbus well.

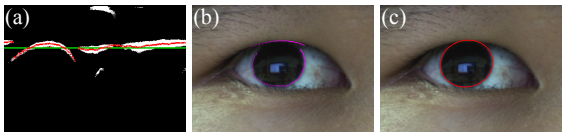

Figure 5: The binary image (a) for searching connected components, the image (b) that the points on the components are projected on, and the detected iris area (c) using the least-square method.

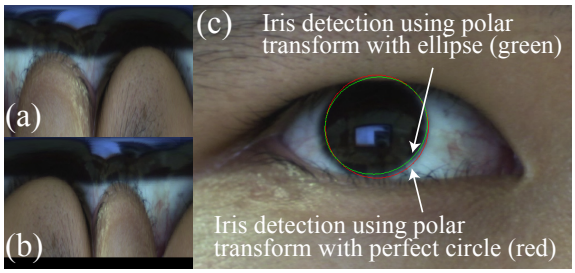

Figure 6: Comparative results between using a perfect circle and an ellipse.

2.3 3D Eyeball Model Fitting

The 3D model of the eye consists of the corneal and eye spheres, as shown in Figure 7. The parameters of this model have been determined by anatomical data(Snell and Lemp, 1997), as shown in Table 1. First, we the calculated depth and attitude of the eye using Nishino's method(Nishino and Nayar, 2006). This method is important, so it is briefly described in section 2.3.1. Next, we estimated the center of the eyeball for both eyes and the distance of both eyes.

2.3.1 Estimating the Pose of the Eyes

First, the initial position between the camera center and the eye is computed using the model. Then, the iris, along with both corners of the eye, are registered on the color image, and the center of the iris (i_{Lx}, i_{Ly}) is drawn on the color image using an elliptical approximation. The 3D center of the limbus **P** can then be computed as:

$$\mathbf{P} = \left[(d+f)\frac{i_{Lx}-cx}{f}, (d+f)\frac{i_{Ly}-cy}{f}, (d+f)\right]^T, \quad (6)$$

where (cx,cy) and f are camera parameters (the optical centers in the image coordinates and the focal length, respectively). The distance d between the center of the limbus and the image plane is calculated as

$$d = f\frac{r_L}{r_{max}}, \quad (7)$$

where r_{max} is the major radius of the ellipse. The optical axis of the eye is given by

$$\mathbf{g} = [sin(\tau)sin(\phi), -sin(\tau)cos(\phi), -sin(\tau)]^T, \quad (8)$$

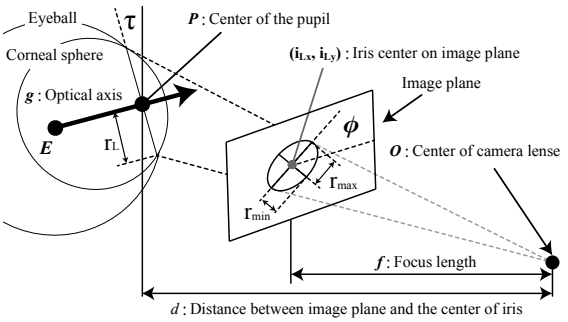

Figure 7: Geometric 3D eye model consisting of the eye sphere and the corneal sphere for tracking the pupil and iris.

Table 1: Anatomical parameters for 3D eye model.

Description	Def.	Parameter
Radius of the eye sphere	r_E	12.1 mm
Radius of the corneal sphere	r_C	7.78 mm
Radius of the limbus	r_L	5.6 mm
Distance from the cornea's center to the eyeball's center	d_{CE}	5.2 mm
Distance from the pupil's center to the cornea's center	d_{PC}	5.6 mm

where ϕ is the rotation of the limbus ellipse and τ is the angle of the tilt of the limbus plane with respect to the image plane. The estimated ellipse corresponds to two poses of the eye, such as upward and downward attitude, and the pose cannot be determined uniquely. Thus, the camera position relative to the face is used for determination.

2.3.2 Estimating the Distance between Eyeballs

The center of the cornea **C** and the center of the eye **E** are determined using the following equations, respectively:

$$\mathbf{C} = -d_{PC}\frac{\mathbf{g}}{||\mathbf{g}||} + \mathbf{P}, \tag{9}$$

$$\mathbf{E} = -d_{CE}\frac{\mathbf{g}}{||\mathbf{g}||} + \mathbf{C}. \tag{10}$$

The centers of the eyeballs $\mathbf{E}_L$ and $\mathbf{E}_R$ are computed for each eye, and the distance d_{EE} between the eyeballs is calculated. These parameters are used for the position tracking mentioned in section 3 as a constraint.

3 3D POSITION TRACKING FOR EYES

When we implement our algorithm in a head-mounted eye tracker, the center of the cornea calculated in the initialization is used for tracking. However, the eye position is changed when the remote eye tracker is employed, and thus a continuous position estimation is required. Therefore, the pupil and iris are used for estimating the pose and the position of the eye, respectively. The details of the continuous position estimation are as follows. First, the pupil region is extracted, and the optical axis **g** is estimated using the pupil ellipse's elevation mentioned in section 2. The vectors from the camera lens to the pupil centers are defined as $\mathbf{V}_{PR}$ and $\mathbf{V}_{PL}$, and the vector passing through the centers of the pupils are defined as $\mathbf{P}_{IR}$ and $\mathbf{P}_{IL}$ and are observed approximately 0.55 mm ahead of the real pupil position by the refraction of the cornea as shown in Figure 8. Furthermore, the eyeball centers $\mathbf{E}_R$ and $\mathbf{E}_L$ are computed using the anatomical parameters d_{IP}, d_{PC}, and d_{CE}. The vectors $\mathbf{V}_{PR}$ and $\mathbf{V}_{PL}$ are shifted parallel by d_{IP} as shown in Figure 9. The shifted origins are defined as $\mathbf{O}_L$ and $\mathbf{O}_R$ by the following equations,

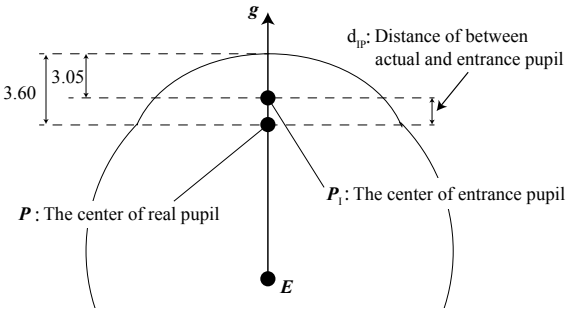

Figure 8: The entrance and exit pupil of the eye.

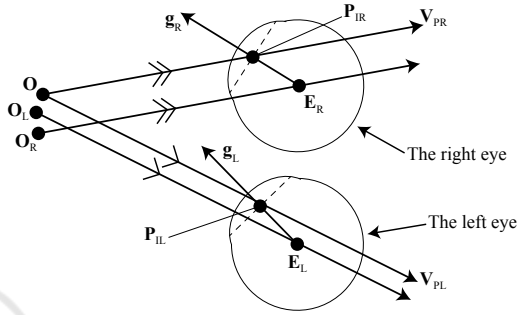

Figure 9: Geometrical constraint of the eyeball position.

$$\mathbf{O}_R = \mathbf{O} - \frac{\mathbf{g}_R}{||\mathbf{g}_R||}(d_{IP} + d_{PC} + d_{CE}), \tag{11}$$

$$\mathbf{O}_L = \mathbf{O} - \frac{\mathbf{g}_L}{||\mathbf{g}_L||}(d_{IP} + d_{PC} + d_{CE}), \tag{12}$$

and the actual positions of the eyes are located on two vectors defined as follows:

$$\mathbf{E}_R = \mathbf{O}_R + t\mathbf{V}_{PR}, \tag{13}$$

$$\mathbf{E}_L = \mathbf{O}_L + \mu\mathbf{V}_{PL}. \tag{14}$$

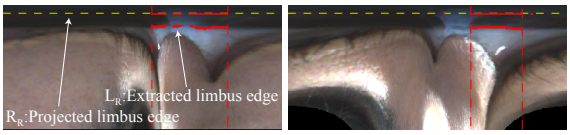

Figure 10: Re-projected limbus R and extracted limbus L are displayed on transformed images for right and left eyes.

The eye position cannot be calculated by closed-form solution. However, when the distance between the eyes is defined as d_{EE}, the distance is used as a criteria for estimating the position. Additionally, the positions of the eyes are located on these vectors, and thus, the searching area is restricted. When the candidates for the eye positions are selected, the re-projected limbus are matched to the actual limbus on the image. Figure 10 shows the extracted limbus and the re-projected limbus from candidates for both eyes. Limbus edges that have the brightness higher than threshold are used for evaluation. The coordinates of the extracted limbus edges from the image are $\mathbf{L}_R$ and $\mathbf{L}_L$ in polar coordinates, and the polar coordinates of the re-projected limbus are $\mathbf{R}_R$ and $\mathbf{R}_L$. The positions of the eyes are estimated using the following evaluation formula:

$$\{\hat{t},\hat{\mu}\} = \underset{t,\mu}{argmin}\{\|\mathbf{L}_R - \mathbf{R}_R\| + \|\mathbf{L}_L - \mathbf{R}_L\|\}. \quad (15)$$

The computed optimal parameters $\hat{t}$ and $\hat{\mu}$ are substituted for equation 13 and 14, respectively. Finally, the eyeball centers can be estimated in each frame.

4 EVALUATION EXPERIMENT

We performed a experiment for confirming the tolerance of head movements by measuring the accuracy of the point-of-gaze. The experiment is conducted under the experimental setup. Nine(3×3) markers are displayed at 480 mm × 240 mm intervals on the 27-inch monitor. The camera was calibrated using a circular grid in advance, and the intrinsic parameters were known. For estimating the point-of-gaze, the geometric relationship between the camera and the monitor are required. Therefore, retro-reflective markers were attached to the camera and the monitor, and the geometric relationship was solved by the motion capture system as the hardware calibration. Furthermore, the visual axis was calculated from the optical axis determined using the 3D eyeball model using the angle calculated by the one-point calibration (user calibration), and the point-of-gaze was computed as an intersection between the visual axis and the monitor. Research participants stared at nine points in each experiment, and ten images were captured at each point for calculating the average error.

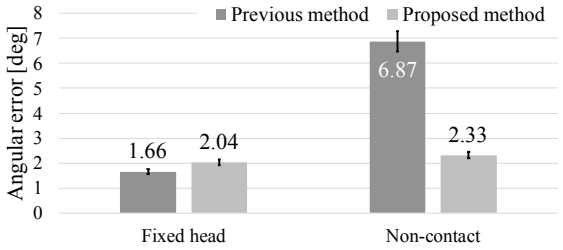

Figure 11: The average error of the estimated point-of-gaze with and without the chin rest

There were ten research participants aged 20–23 years in this experiment. Figure 11 shows the average error and the standard error of the estimated point-of-gaze with and without the chin rest, when our previous method(Yamagishi and Takemura, 2017) and the proposed method are applied. The error of the estimated point-of-gaze was increased without the chin rest in our previous method. In contrast, the computed accuracies under the two experimental conditions are very similar in the proposed method, and we confirmed that the proposed method has a tolerance for head movements.

5 DISCUSSION

In this section, the elevation of the camera is discussed through additional experiment. We prepared three camera conditions, large to small elevation, for evaluating the influence of the camera setting. When the pupil is detected as an ellipse, the major and minor axes are used for detecting the pose of the eye. However, the inverse cosine is used for estimating the pose as per the following equation:

$$\tau = cos^{-1}\left(\frac{r_{min}}{r_{max}}\right). \quad (16)$$

Thus, when an ellipse is closed to a perfect circle, the estimation error could might be increased. Therefore, the elevation angles of the cameras were 20, 30 and, 40 deg, and the point-of-gaze was estimated with our proposed method. In this experiment, the head is fixed with the chin rest while estimating the point-of-gaze. Figure 13 shows the average error and the standard error of the estimated point-of-gaze as the comparative result. As shown by these results, accuracy is influenced by the elevation of the camera. When the point-of-gaze is calculated, the large elevation angle works better than the small angle. Figure 12 shows how the value of the inverse cosine depends on the ratio between the major axis and minor axis. When the ratio is close to one, the value of the inverse cosine varies extremely. As a result, we can understand that a large

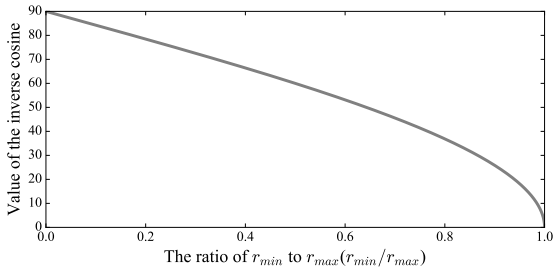

Figure 12: Variance of the inverse cosine.

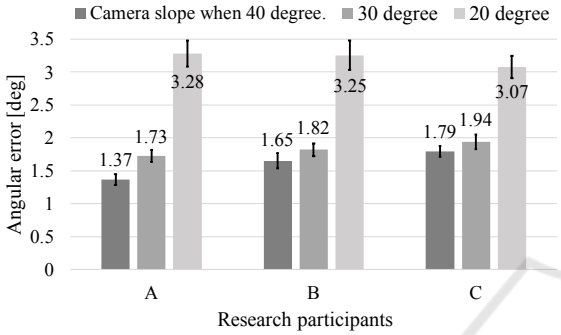

Figure 13: The average error of the estimated point-of-gaze under three camera conditions.

elevation is required. Currently, it is difficult to measure the eye from a low angle with high accuracy, but a low angle is needed for various applications. Therefore, this problem should be solved in future work.

6 CONCLUSIONS

In this research, we proposed a remote eye-tracking method, and the eyeball positions and gaze vectors were estimated using an RGB-IR camera. The IR image and visible image were simultaneously used to compensate for weak points. The pupil was tracked using the near-infrared image for estimating the pose of the eye. Additionally, the position was estimated using the limbus in the remote eye-tracking method. We conducted several experiments and confirmed the feasibility of the proposed method as a remote eye tracker. In future work, we will solve the problem of measurement range that is discussed above, and the method will be used for digital signage.

REFERENCES

Fuhl, W. et al. (2016). 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, pages 123–130.

Grand, Y. L. and Hage, S. G. E. (1980). *Physuological Optics*. Springer, Berlin.

Guestrin, E. D. and Eizenman, M. (2006). General theory of remote gaze estimation using the pupil center and corneal reflections. *IEEE Transactions on Biomedical Engineering*, 53(6):1124–1133.

Haubner, H. H. (1966). Characterization of the powder particle shape. *Planseeber Pulvermetall*, 14(2):75–84.

Kassner, M. et al. (2014). Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, UbiComp '14 Adjunct, pages 1151–1160.

Morimoto, C. H. and Mimica, M. R. M. (2005). Eye gaze tracking techniques for interactive applications. *Comput. Vis. Image Underst.*, 98(1):4–24.

Nagamatsu, T. et al. (2010). User-calibration-free gaze tracking with estimation of the horizontal angles between the visual and the optical axes of both eyes. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, ETRA '10, pages 251–254.

Nishino, K. and Nayar, S. K. (2006). Corneal imaging system: Environment from eyes.

Ohno, T. et al. (2002). Freegaze: A gaze tracking system for everyday gaze interaction. In *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications*, ETRA '02, pages 125–132.

Snell, R. S. and Lemp, M. A. (1997). *Clinical Anatomy of the Eye*. Blackwell Publishing. Malden, 2nd edition.

Takemura, K. et al. (2014a). Estimating 3-d point-of-regard in a real environment using a head-mounted eye-tracking system. *IEEE Transactions Human-Machine Systems*, 44(4):531–536.

Takemura, K. et al. (2014b). Estimation of a focused object using a corneal surface image for eye-based interaction. *Journal of Eye Movement Research*, 7(3)(4):1–9.

Tian, Q. et al. (2012). Iris boundary localization method based on polar coordinate. *Procedia Engineering*, 29:2194 – 2198.

Toyama, T., Kieninger, T., Shafait, F., and Dengel, A. (2012). Gaze guided object recognition using a head-mounted eye tracker. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ETRA '12, pages 91–98.

Wilson, M. et al. (1992). Change of pupil centration with change of illumination and pupil size. In *Optometry and Vision Science*, volume 69, pages 129–136.

Yamagishi, K. and Takemura, K. (2017). A hybrid eye-tracking method using a multispectral camera. In *Proceedings of 2017 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2017*, pages 1529–1534.