We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us? Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

30

Geometry Issues of Gaze Estimation

Arantxa Villanueva, Juan J. Cerrolaza and Rafael Cabeza
Public University of Navarra
Spain

1. Introduction

Video-oculography (VOG) is a non-intrusive method used to estimate gaze. The method is based on a remote camera(s) that provides images of the eyes that are processed by a computer to estimate the point at which the subject is gazing in the area of interest. Normally, infrared (IR) light-emitting diodes (LEDs) are used in the system, as this light is not visible to humans. The objective of the light source is to increase the quality of the image and to produce reflections on the cornea. These reflections can be observed in the acquired images (see Figure 1) and represent useful features for gaze estimation.

Figure 1. Image of the eye captured by the camera. The corneal reflections are the bright dots in the image

During the last several decades, gaze tracking systems based on VOG have been employed in two main fields: interactive applications and diagnostic applications. Interactive applications permit users to control the position of the mouse on the screen and the activation of items by their gaze, thus allowing highly impaired users with controlled eye movement to interact with the computer and their environment. Diagnostic applications are devoted to eye movement analysis when executing tasks such as web page browsing or reading, which have interesting applications in the fields of psychology and market research. However, gaze tracking technologies are still not useful for a large part of society. New commercial applications, such as in video games and the automotive industry, would attract more companies and general interest in these systems, but several technical obstacles still need to be overcome. For instance, the image processing task is still problematic in outdoor scenarios, in which rapid light variations can occur. In addition, the head position constraints of these systems considerably reduce the potential applications of this technology. The accuracy of gaze tracking systems is, to a large extent, compromised by

head position since any head movement requires the system to readjust to preserve accuracy; i.e., gaze estimation accuracy can vary as the head moves.

Gaze estimation is defined as the function of converting the image processing results (image features) into gaze (gaze direction/gazed point) using a mathematical equation. The usual procedure in any gaze tracking session is to first perform a calibration of the user. The calibration consists of asking the user to fixate on a set of known points on the screen. Calibration adapts the gaze estimation function to the user's position and system configuration. The mathematical method (gaze estimation function) used determines the dependence of the system accuracy on head position, i.e. how the accuracy varies as the head moves.

This work is focused on exploring the connection between the image of the eye and gaze, and it constructs a mathematical model for the gaze estimation problem.

A more detailed description of the problem and a review of relevant works on gaze estimation are presented in Section 2. Section 3 presents the 3-D eyeball model to be used in the rest of the chapter and establishes relevant terms, variables, and definitions. Section 4 introduces some of the most relevant features of the eye image, such as glint(s), pupil centre, and pupil shape. Alternative models are constructed and evaluated based on these image features. The proposed gaze estimation model is described in Section 5. Section 6 introduces the experimental results of the model. Conclusions and future research are described Sections 7 and 8, respectively.

2. State of the art

The 3-D direction of gaze is defined as the line of sight (LoS). The point of regard (PoR) is determined as the 2-D intersection between the LoS and the area of interest, e.g. the screen. A visual fixation is defined as a stable position of the LoS (visual axis) that presents a visual angular dispersion below 1º. Hence, most gaze-tracking system designers try to achieve gaze estimation errors below 1º (see Figure 2).

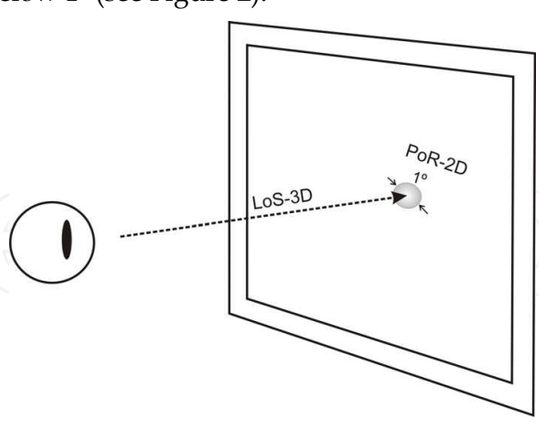


Figure 2. The LoS denotes gaze direction in 3-D space. PoR is the gazed point in 2-D. A fixation is defined as a quasi-stable position of LoS with an angular dispersion below 1º

Gaze estimation methods can be divided into two main groups: interpolation methods and geometry-based methods. Interpolation methods use general-purpose polynomial expressions, such as linear or quadratic expressions, to describe the connection between image features and gaze. These functions are based on unknown coefficients to be

determined by calibration. Although simplicity is the primary advantage of interpolation methods, the lack of control over system behaviour and system errors is a significant disadvantage. Furthermore, system accuracy can decrease as result of subject's movement from the calibration position (Cerrolaza et al., 2008) (Morimoto & Mimica, 2005).

Enhancing gaze estimation, in terms of accuracy and head movement tolerance, is one of the most sought-after objectives in gaze-estimation technology. Geometry-based methods investigate gaze estimation models that allow for free head movement, and they are based on mathematical and geometric principles. These methods provide relevant information about the systems, such as the minimal hardware required, the minimal image features and the minimal user calibration for gaze estimation purposes (hereafter, calibration refers to user calibration unless otherwise stated). Recently, remarkable studies have been published in this area. One of the most relevant (Shih & Liu, 2004) is based on a stereo system and one calibration point. (Hennessey et al., 2006) and (Guestrin & Eizenman, 2006) demonstrate geometric gaze estimation methods using a single camera and four and nine calibration points, respectively. More recent studies have been presented by (Villanueva & Cabeza, 2007) and (Guestrin & Eizenman, 2008). The aforementioned methods employ 3-D eye models and calibrated scenarios (calibrated camera, light sources, and camera positions) to estimate LoS. Interesting approaches have been carried out in non-calibrated scenarios using projective plane properties for PoR estimation as shown in (Hansen & Pece, 2005) and (Yoo & Chung, 2005).

The current work focuses on 3-D (LoS) gaze estimation models based on system geometry. Alternative models for gaze (LoS) estimation are constructed, and their properties are evaluated, including hardware used, image features, head pose constraints, and calibration requirements. This work seeks to develop a gaze estimation method with minimal hardware, allowing free head movement and minimal calibration requirements.

3. Eye Model

Figure 3 depicts the schematic system, which is composed of a subject's eye, one camera, and one infrared light source. Of all the elements of the system, the eye is the most complex one. Geometry-based gaze estimation methods employ eye models that are highly simplified to reduce the complexity introduced by consideration of physiology and eyeball kinematics. Although variations exist between the models proposed by researchers and there is no unified model, some fundamental aspects of the eyeball geometry have been agreed upon during the last few years. This model is detailed in Figure 3.

The cornea is considered a sphere with its centre at C and a corneal radius of r_c. The pupil is a circle with radius r and centre E. The pupil centre is perpendicularly connected to C at a distance of h, and both points, together with the eyeball centre A, are contained in the optical axis of the eye. The elements of the system are referenced to the camera projection centre O. The LoS can be approximated by the visual axis of the eye, the line connecting the fovea and the nodal point (approximated by C). The fovea is a small area with a diameter of about 1.2° located in the retina, and it contains a high density of cones that are responsible for high visual detail discrimination and an individual's central vision. When looking at a particular point, the eye is oriented in such a way that the observed object projects itself onto the fovea. Due to the offset of the fovea with respect to the back pole of the eyeball, an angular offset exists between the optical and visual axes, with horizontal and vertical components β (~5°) and α (2° to 3°), respectively (Guestrin & Eizenman, 2006). Several works, including the

present one, reduce this offset to just the horizontal value, i.e. β. Optical and visual axes rotate in an imaginary plane with respect to the camera when looking at different points. Once the optical axis has been determined, the 3-D position of this plane can be calculated by using Donder's and Listing's Laws and applying the "false torsion" principle stated as:

$$\sin\alpha_o = \frac{\sin\theta_0 \sin\varphi_o}{1+\cos\theta_0 \cos\varphi_o}, \tag{1}$$

where (θ_o, φ_o) are the vertical and horizontal rotation angles performed by the optical axis and α_o is the torsion angle around itself. In this manner, once the position of the plane is determined by (θ_o, φ_o, α_o), the visual axis position is calculated from the optical axis by applying the angular offset β. In the suggested eye model, the optical axis of the eye is calculated as the line connecting *C* and *E*.

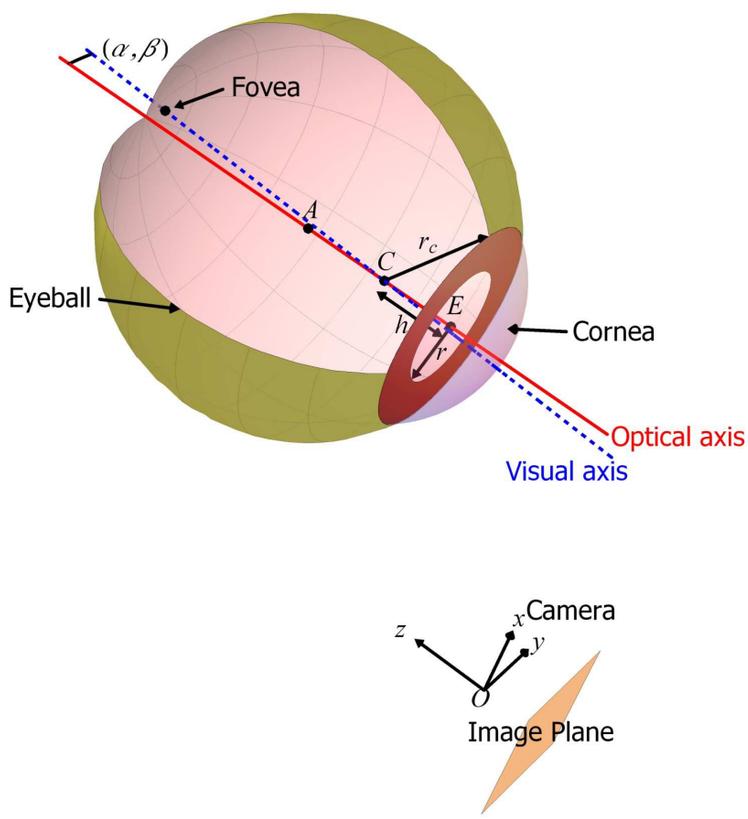


Figure 3. Eye model for gaze estimation

In the simplified model of the eyeball, the cornea is reduced to a single surface, and the aqueous humour is assumed to be homogeneous. The reflective properties of the cornea influence the position of the glint(s) in the image, while its refractive properties modify the pupil image. According to Refraction Law, corneal refraction deviates light reflected off the retina and crossing the pupil prior to reaching the VOG camera (Villanueva & Cabeza, 2008a). Figure 4 shows the 3-D pupil inside the cornea and the pupil shape after refraction. The pupil image is not the projection of the pupil but the projection of the refracted pupil.

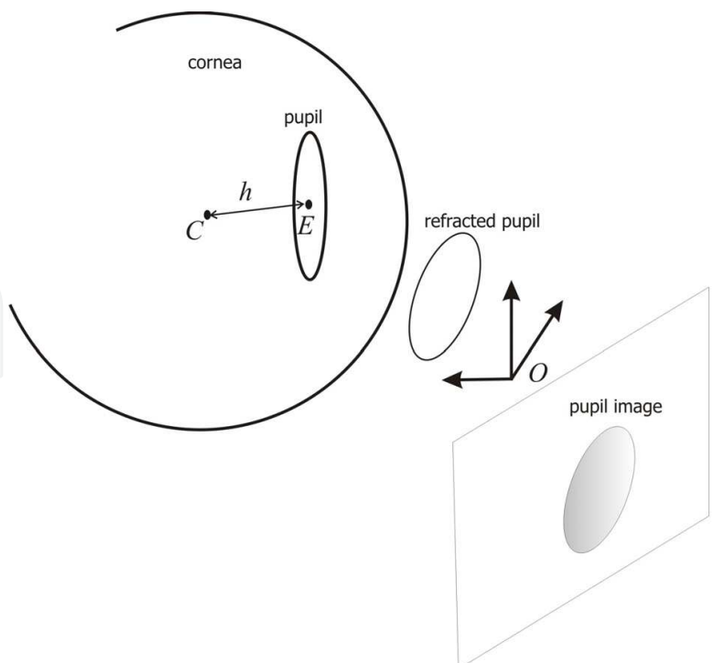


Figure 4. The pupil image is deviated due to corneal refraction when crossing the corneal sphere

Corneal refraction alters the pupil size in the image and its position with respect to the limbus, i.e. the corneal-sclera junction. Figure 5 shows the difference between the projected pupil and the real image of the pupil (refracted and projected).

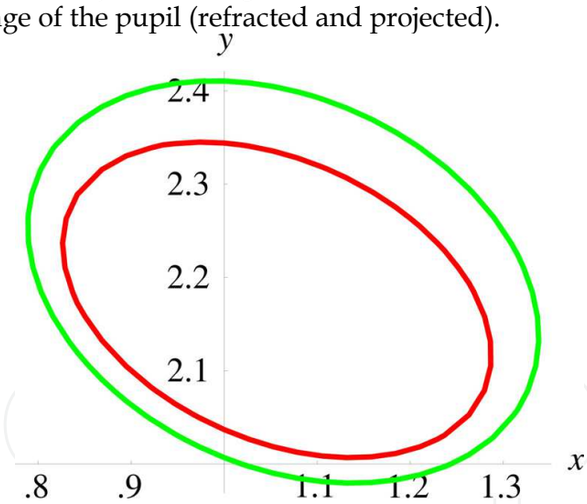


Figure 5. Comparison of the perspective projection of the pupil (smaller ellipse) and the combined refraction and projection of the pupil (larger ellipse)

4. Models for Gaze Estimation

According to the eye model described in Section 3, alternative gaze estimation methods have been proposed based on different image features. The procedure selected to accomplish the work in the simplest manner is to analyze each of the alternative features that can be extracted from the image separately. In this manner, a review of the most commonly used features employed by alternative gaze tracking systems is carried out. The

constructed models can be categorized into three groups: models based on points, models based on shapes, and hybrid models combining points and shapes. The systems of the first group are based on extracting image features which consist of single points of the image and combining them in different ways. We define a point as a specific pixel described by its row and column in the image. Thus, the following models make up this group: the model based on the centre of the pupil, the model based on the glint, the model based on multiple glints, the model based on the centre of the pupil and the glint, and the model based on the centre of the pupil and multiple glints. On the other hand, the models based on shapes involve more image information; basically, these types of systems take into account the geometrical form of the pupil in the image. This group contains one model, the model based on pupil contour. The models of the third group combine both points and shapes to sketch the system. This group contains the model based on the pupil ellipse and glint, as well as the model based on the pupil ellipse and multiple glints. Figure 6 shows a classification of the constructed models.

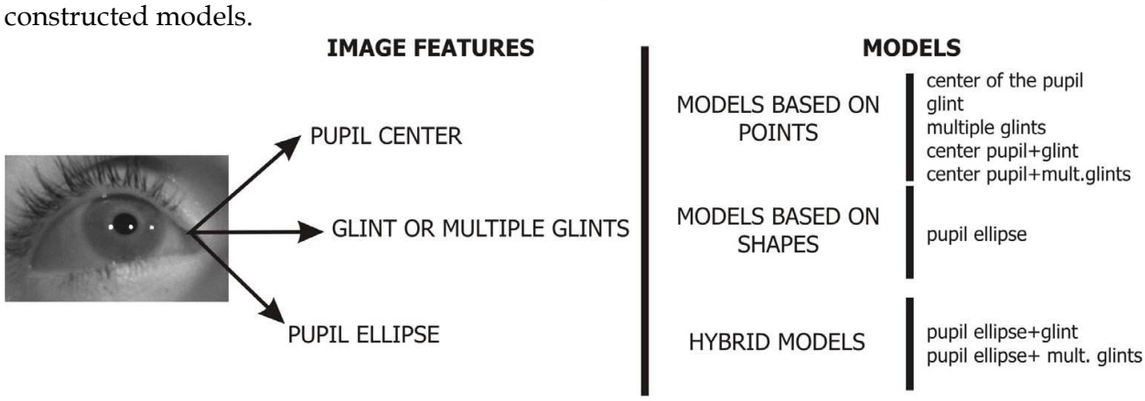


Figure 6. Model classification according to image features

Three of the most relevant features of the image of the eye are selected to be analyzed as potential features for gaze estimation models: the glint, the centre of the pupil, and the shape of the pupil.

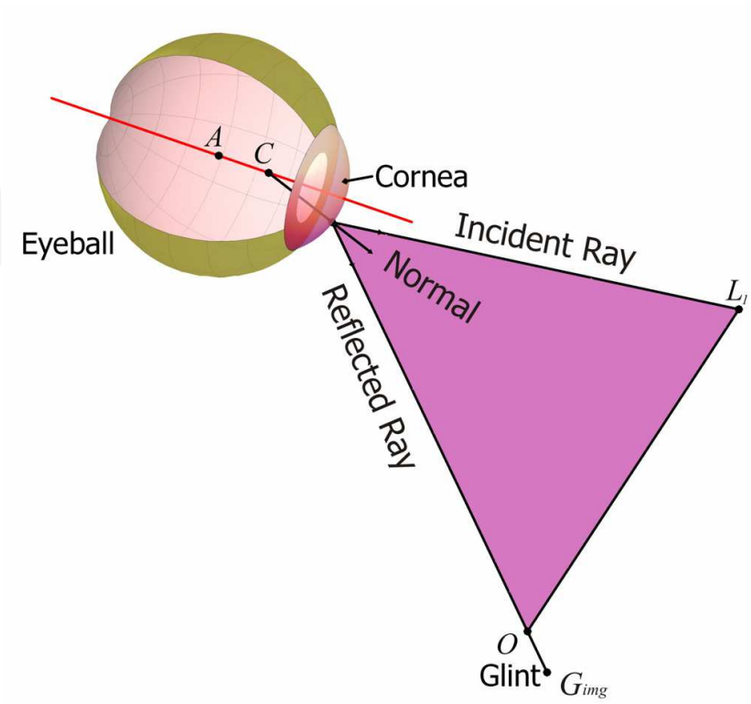
4.1 The glint

The glint or corneal reflection in the image is denoted as G_{img} and is a consequence of the IR light reflecting off of the corneal surface and reaching the camera. According to Reflection Law, given a light source denoted as L_1, the incident ray, the reflected ray, and the normal line at the point of incidence on the corneal sphere are coplanar. Consequently, the corneal centre C, the projection centre of the camera O, the glint in the image G_{1img}, and the light source L_1 are contained in the same plane (see Figure 7a).

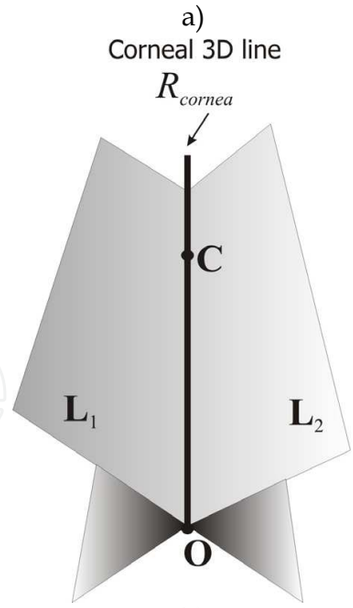
If an additional light source denoted as L_2 is introduced into the system, a new plane can be determined as a function of the new light source and glint, G_{2img}. The intersection between these planes is a 3-D line containing the projection centre of the camera O and the cornea centre C (see Figure 7b). Adding more light sources does not provide further information to determine the cornea centre, but rather just the aforementioned 3-D line.

In the system proposed by (Shih & Liu, 2004) an additional camera is introduced to the system. The combination of each camera with the pair of light sources results in a new 3-D line containing the new camera projection centre and C. The intersection between these lines determines the cornea centre C.

In a single camera model, point C can be obtained using two light sources if the corneal radius r_c is provided, as shown in studies (Guestrin & Eizenman, 2006) and (Villanueva & Cabeza, 2007).



a)



b)

Figure 7. a) Incident ray, reflected ray, and the normal line at the point of incidence are coplanar. The light source, L_1, the cornea centre, C, and the camera projection centre, O, are contained in the plane. b) If a second light source is included, two planes are created, one for each of the light sources. The intersection of the planes determines a 3D line containing the cornea centre C and the camera projection centre, O, denoted as *Rcornea*

4.2 The centre of the pupil

Although no formal proofs about the elliptical nature of the pupil image have been provided in the literature, pupil shape is normally approximated as an ellipse. The centre of the pupil shape E_{img} is generally used as a valid feature for gaze estimation purposes. This point is normally considered to be the image of point E. According to the eye model proposed, this assumption introduces two errors into the method. First, the perspective effect of the camera produces a translation in the image between the projection of point E and the centre of the pupil shape. Second, and more importantly, as mentioned in Section 3, corneal refraction produces a translation and scaling of the pupil image with respect to its projection. It could be assumed that the centre of the pupil image is the refraction of the ray starting in E. However, refraction is not a linear process and errors are introduced. The perspective effect is negligible compared to the deviation due to refraction (errors >1º can arise depending on system configuration). Recent studies (Villanueva & Cabeza, 2008a) demonstrate that errors due to refraction can be compensated for after an adequate calibration process. However, the dependence of the accuracy on variable calibration conditions should be evaluated. The work (Guestrin & Eizenman, 2006) showed a method based on the aforementioned assumption for the centre of the pupil and calibrated using 3x3 calibration with acceptable accuracies. This model makes a previous determination of the corneal centre, C. Once C has been determined, and assuming that r_c is known, E_{img} is back-projected from the image and refracted at the corneal surface (see Figure 8). The estimated pupil centre, E', is calculated as the point contained in the refracted ray at a distance h from C.

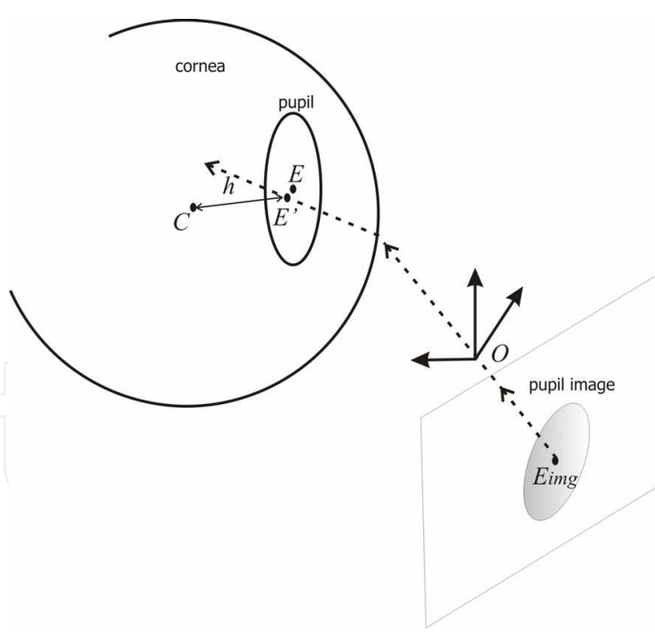


Figure 8. The approximation is based on the assumption that the pupil center is contained in the refracted line of the pupil centre in the image, i.e. $E'=E$. The error is due to the fact that corneal refraction is not linear and deviates each ray differently, hence $E' \neq E$

The system developed by (Shih & Liu, 2004) also assumes that the centre of the pupil image is the image of point E after refraction, but it uses a stereo system. According to Refraction Law, the incident ray, refracted ray, and the normal at the point of incidence are coplanar.

Consequently, if the centre of the pupil in the image E_{img} is assumed to be the projection of the ray coming from E, after refraction E (the incident ray), E_{img} (the refracted ray) and C (the normal ray) are contained in the same plane. This plane can be calculated as the plane containing C, O, and E_{img}. For each camera, a plane can be derived containing C and E (E'), the optical axis of the eye. The intersection of the two planes determines the optical axis of the eye. The method is calibrated using a single point, which makes the error due to refraction more relevant.

Regardless of the number of cameras used, both methods require a previous estimation of the corneal centre C, thus the pupil centre by itself does not provide sufficient information for optical axis estimation.

4.3 The shape of the pupil

As mentioned before, the pupil shape is the result of the refraction and projection of pupil contour. The intersection of each ray of the pupil contour with the corneal sphere suffers a deviation in its path before reaching the camera. Starting from the image, the pupil contour is sampled, and each point is back-projected from the image into 3D space. Assuming that C and r_c are known, the intersection of each ray with the corneal sphere is found, and the refracted line is calculated, using the Refraction Law equation. In this manner, we derive the distribution of pupil contour rays inside the cornea as shown in Figure 9. The pupil is contained in a plane denoted by Π that has C-E as a normal vector and is at a distance h from C. The pupil centre is calculated as the point at the centre of a circle derived from the intersections of the refracted rays with the plane Π. Alternative implementations of this method can be found in different recent works (Beymer & Flickner, 2003) (Ohno & Mukawa, 2004) (Hennesey et al., 2006) (Villanueva & Cabeza, 2007).

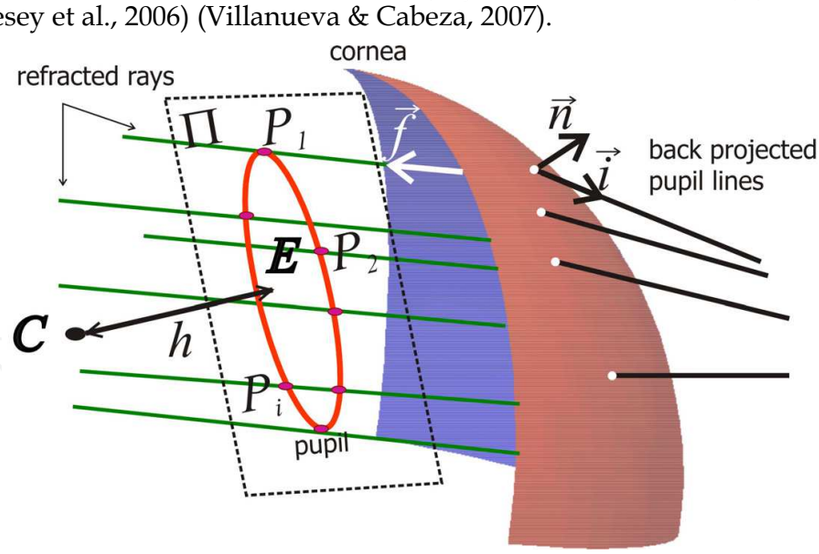


Figure 9. Cornea and pupil after refraction. E is the centre of a circumference formed by the intersections of the plane Π with the refracted lines. The plane Π is perpendicular to (C-E) and the distance between pupil and cornea centres is h

According to the proposed eye model, this method introduces corneal refraction in the method, hence no approximations are assumed, and zero error is produced. As in the previous section, to apply the model based on the shape of the pupil, knowledge about the cornea centre is required, and thus the pupil shape by itself does not permit optical axis estimation.

5. Proposed Gaze Estimation Model

Based on the above analysis, it is concluded that, models based on points and shapes do not allow for optical axis estimation. To estimate the pupil centre of the eye, *E*, a previous estimation of the corneal centre *C* is needed. Since the corneal centre information is provided by the glints and the pupil centre *E* is given by the pupil image, a hybrid model is needed for optical axis estimation in a head pose variant scenario, i.e. a model combining glints and the pupil as image features.

As shown in Section 4.1, two light sources are needed to estimate *C*. The proposed model attempts to minimise the hardware used, and although the method using the stereo system provides a geometric solution to estimate the optical axis (Shih & Liu, 2004), the proposed model focuses on a single camera system. The proposed framework uses a single calibrated camera and two light sources L_1 and L_2 with known positions with respect to the camera projection centre. According to the aforementioned analysis, the optical axis estimation is performed in two steps.

5.1 Center of the cornea estimation

According to the analysis presented in Section 4.1, the following equations can be stated:

- As shown in Figure 7a, given two light sources, L_1 and L_2, two corneal reflections will be generated in the camera, named G_{1img} and G_{2img}, respectively. Each light source, L_i, will define a plane, $\prod_{Li}$, containing the projection centre of the camera *O*. These planes can be defined as:

$$\prod_{L_1} = L_1 \times G_{1img}; \prod_{L_2} = L_2 \times G_{2img}. \tag{2}$$

- The cornea centre C is contained in both planes, which can be stated as:

$$C \cdot \prod_{L_1} = C \cdot \prod_{L_2} = 0. \tag{3}$$

- The Reflection Law can be applied for each one of the light sources as:

$$r_i = 2(n_i \cdot l_i)n_i - l_i, \; i=1,2, \tag{4}$$

where r_i is the unit vector in the G_{1img} direction, l_i is the unit vector in the (L_i - G_i) direction, and n_i is the normal vector at the point of incidence in the (G_i - *C*) direction. G_i denotes the incidence point at the corneal surface.

Assuming that r_c is known, G_i can be expressed as a function of *C* from:

$$|G_i - C| = r_c, \; i=1,2. \tag{5}$$

By solving equations (3) to (5), the cornea centre is determined numerically (Villanueva & Cabeza 2008b). This method requires knowledge of the corneal radius r_c to be obtained by means of calibration.

5.2 Center of the pupil estimation

Assuming that the cornea parameters C and r_c are known as calculated in previous section, the following equations can be stated:

- The pupil contour is sampled, and each point is back-projected from the image into 3-D space. The intersection of each ray with the corneal sphere is found, and the refracted line is calculated using the Refraction Law equation. Given a point k (k=1..n) of the pupil contour, the refracted ray can be calculated as:

$$f_k = \left[\frac{n_a}{n_b}\right]\left[i_k - \left((i_k \cdot n_k) + \sqrt{\left(\frac{n_a}{n_b}\right)^2 - 1 + (i_k \cdot n_k)^2}\right) n_k\right], k=1,..n, \tag{6}$$

  where n_a and n_b are the refractive indices of air and the aqueous humor in contact with the back surface of the cornea, f_k and i_k represent the refracted light inside the cornea and the incident light directions, respectively, and n_k is the surface normal vector at the point of incidence (see Figure 9).
- The pupil is contained in a plane Π having (E - C) as the normal vector of the plane at a distance h from C. Given a 3-D point, (x, y, z), with respect to the camera, the plane Π can be formulated by:

$$\frac{(E-C)}{h} \cdot [(x, y, z) - C] + h = 0. \tag{7}$$

- After Π has been defined, the intersections of the plane with the refracted lines derive a set of points, P_k , k=1..n, that represent the contour of a circumference whose centre is E. If $|P_k - E|$ represents the distance between P_k and E, this statement can be expressed as follows:

$$|P_i - E| = |P_j - E| \quad \text{where } i \neq j, (i,j=1..n). \tag{8}$$

- The pupil centre is determined numerically by solving equation (8) to find the global optimum (Villanueva & Cabeza, 2008b). The method requires knowledge of the distance between corneal and pupil centers h to be obtained by means of calibration.

5.3 LoS estimation

Once the optical axis of the eye is determined as the line connecting C and E, gaze direction (LoS) is estimated by calculating the visual axis. Knowing the optical axis in 3-D permits calculation of the rotation angles (θ_o, φ_o) of this line with respect to the camera; thus, the additional torsion α_o is calculated by means of (1). Defining the visual axis direction (for the left eye) with respect to C as ($-\sin\beta$, 0, $\cos\beta$) permits us to calculate the LoS direction with respect to the camera by means of the Euclidean coordinate transformation expressed as:

$$C + R_{\alpha_o} R_{\theta_o \varphi_o} .(-\sin\beta, 0, \cos\beta)^T, \tag{9}$$

where $R_{\theta_o \varphi_o}$ is the rotation matrix calculated as a function of the vertical and horizontal rotations of the vector ($C-E$) with respect to the camera coordinate system, and R_{α_o} represents the rotation matrix of the torsion around the optical axis needed to deduce the final eye orientation. The computation of the PoR as the intersection of the gaze line with the screen plane can be calculated if the screen position is known. This method requires knowledge of β, which is obtained by means of calibration.

5.4 Calibration

The model just described requires specific subject's parameters to determine gaze direction. These parameters, i.e., h, r_c, and β, are obtained using calibration. The subject is asked to observe known points on the screen, and the model parameters are adjusted in order to minimise the error for the calibration points. The equations used for LoS estimation ((3) to (5) and (8)) are also used for calibration purposes. It can be theoretically demonstrated that for the proposed method, one calibration point is sufficient to obtain the necessary model parameters, i.e. h, r_c and β (Villanueva and Cabeza, 2008b).

6. Experiments

Ten users were selected to test the model. The working distance was selected as between 400 and 500 mm from the camera. The subjects had little or no experience with the system. They were asked to fixate on each test point for a period of time. Figure 10 shows the selected fixation marks uniformly distributed through the gazing area, the positions of which are known with respect to the camera. The position in mm for each point is shown.

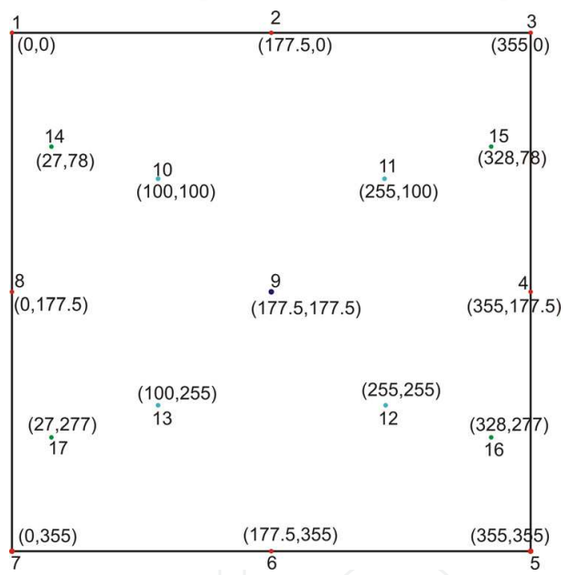


Figure 10. Test sheet

The errors obtained are compared with the limit value of 1º of the visual angle, which is a system performance indicator. The users selected the eye with which they felt more comfortable. They were allowed to move the head between fixation points and could take breaks during the experiment. However, they were asked to maintain a fixed head position during each test point (ten images). The constructed model presents the following requirements:

- The camera must be calibrated.
- Light sources and screen positions must be known with respect to the camera
- The subject eyeball parameters r_c, β, and h must be calibrated.

The images were captured with a calibrated Hamamatsu C5999 camera and digitised by a Matrox Meteor card with a resolution of 640x480 (RS-170). The LEDs used for lighting have a spectrum centred at 850 nm. The whole system is controlled by a dual processor Pentium

system at 1.7 GHz with 256 MB of RAM. It has been demonstrated theoretically that system accuracy can be considerably influenced by a non-accurate glint position estimation. To reduce this effect, the number of light sources can be increased, thus compensating for the error by averaging the value of the cornea centre (Villanueva & Cabeza, 2007). Four LEDs were selected to produce the needed multiple glints. They were located in the lower part, and their positions with respect to the camera were calculated, which considerably reduced the possibility of misleading partial occlusions of the glints by eyelids when looking at different points of the screen because, in this way, the glints in the image appear in the lower half of the pupil.

6.1 Gaze Estimation

Once the hardware was defined, and in order to apply the constructed model based on the shape of the pupil and the glint positions, some individual subject eyeball characteristics needed to be calculated (r_c, β, and h). To this end, a calibration was performed. The constructed model based on multiple glints and pupil shape theoretically permits determination of these data by means of a single calibration mark and applying the model already described in Section 3. Given the PoR as the intersection of the screen and the LoS, model equations (2)-(4) and (6)-(8) can be applied to find the global optima for the parameters r_c, h, and β that minimise the difference between the model output and the real mark position. Figure 11 shows the steps for the subject calibration.

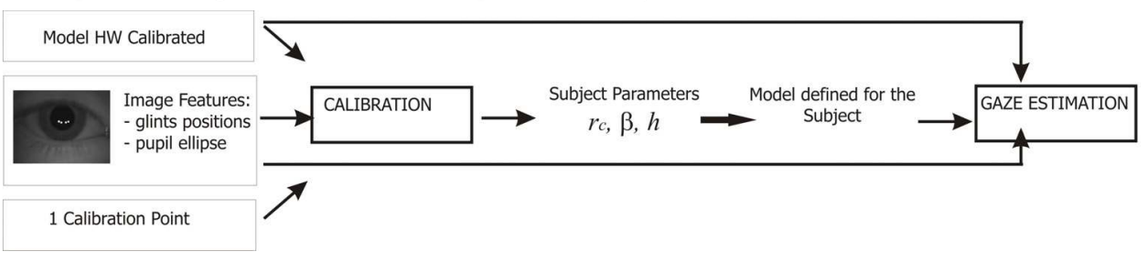


Figure 11. The individual calibration permits us to extract the physical parameters of the subject's eyeball using one calibration point, the captured image, and gaze estimation model information

In theory, one calibration point is sufficient to estimate a subject's personal parameters (Villanueva & Cabeza, 2008b). However, in practice, and to increase confidence in the obtained values, three fixations were performed for each subject, and the mean values were used for the eye parameters in the experiment. For each subject, the three points with lower variances in the extracted glint positions were selected for calibration. Each point among the three permits estimation of values for h, β, and r_c (Villanueva & Cabeza, 2007). Once the system and the subject were calibrated, the performance of the model was tested for two, three, and four LEDs. Figure 12 shows the obtained results. For each user, black dots represent the real values for the fixations. The darkest marks are the gaze points estimated using four LEDs, whereas the lighter marks represent the gaze points estimated using three LEDs. Finally, the medium grey marks are the estimations performed using two LEDs.

Corneal refraction effects are more important as eye rotation increases. In Figure 12, no significant differences can be found between the error for corner points and other points. If the model did not adequately take refraction into account, higher errors would be expected

for the corners. However, the accuracy does not depend on eye rotation, and the model is not affected by an increase in the refraction effect, since this is compensated for.

Table 1 shows a quantitative evaluation of the model competence for two, three, and four LEDs. For each subject, the average error for the 17 fixation marks was calculated in visual degrees since this is the most significant measurement of the model performance. It is clear that the model with four LEDs has the smallest errors. On average, the model with two LEDs has an error of 1.08º, the model with three LEDs 0.89º, and the model with four 0.75º. Therefore, on average, the models with three and four LEDs render acceptable accuracy values. As expected, an increase in the number of light sources results in an improvement of the system accuracy.

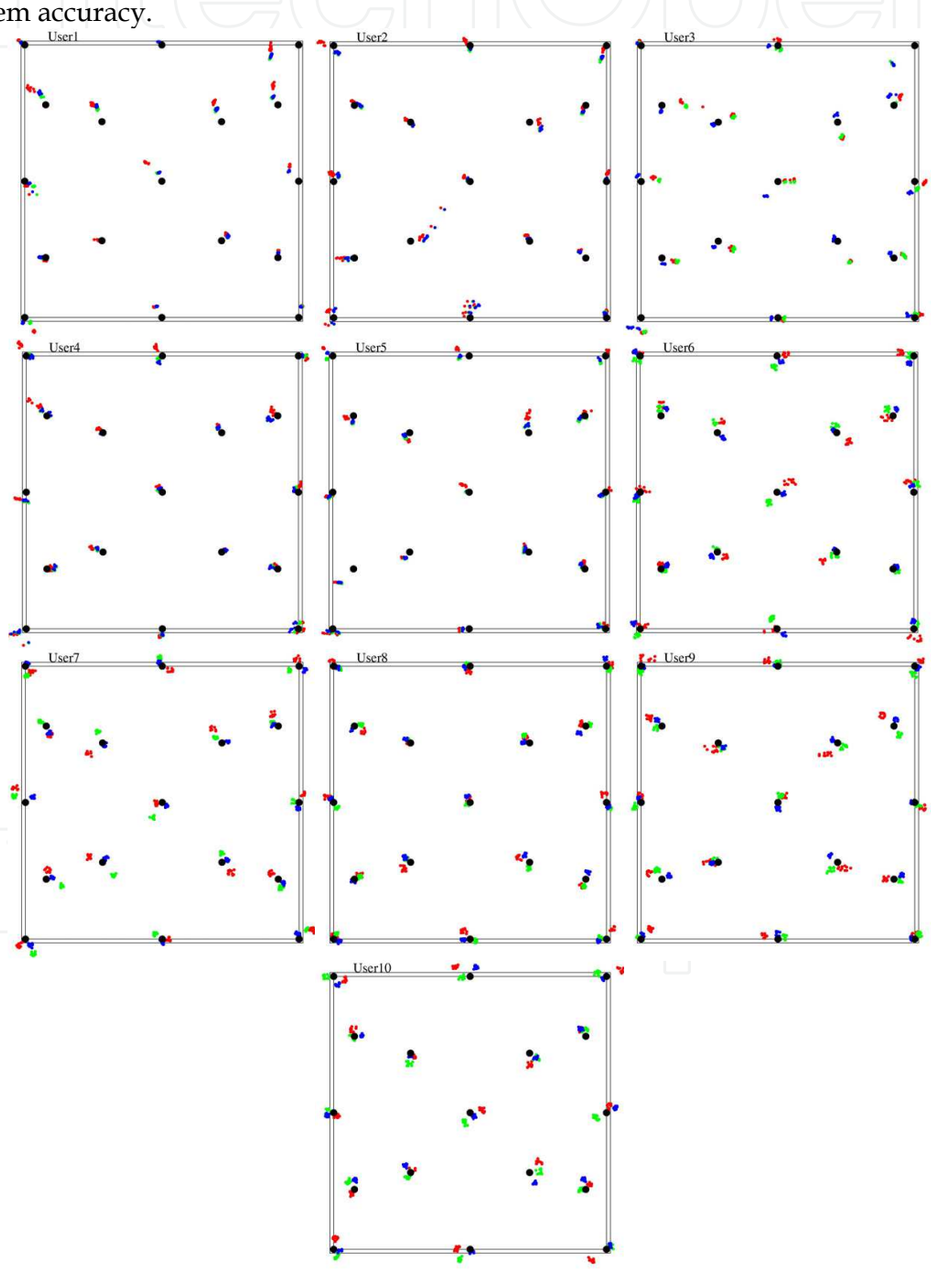


Figure 12. Results obtained by the model for users 1 to 10

Subject	1	2	3	4	5	6	7	8	9	10	Mean
2 LEDs	1.47	0.85	1.46	0.90	0.92	0.97	1.24	0.78	1.19	1.06	1.08
3 LEDs	1.06	0.80	1.35	0.58	0.75	0.78	1.20	0.79	0.74	0.86	0.89
4 LEDs	1.04	0.76	1.01	0.62	0.72	0.71	0.62	0.65	0.59	0.80	0.75

Table 1. Error quantification (degree) of the final model using 2, 3, and 4 LEDs for ten users

6.2 Image Feature Detection

The proposed gaze estimation model is based on a single camera and multiple light sources. The pupil shape and glints positions are used as working features for image data. As mentioned before, the exact determination of glints positions considerably influences system accuracy as studied in (Villanueva & Cabeza, 2008b). In the same manner, pupil contour estimation can be critical for gaze estimation. This task is frequently obviated in many gaze tracking systems devoted to gaze estimation, but we consider it important to include the following recent study in this chapter.

The acquired images (see Figure 1) are processed to extract the pupil and glints. The pupil is detected as a blob matching specific characteristics of size, grey level and compactness. These conditions depend to a large extent on the ambient light level in the room and the optics of the camera. According to the hardware configuration, it can be assumed that the glints are located in a region close to the previously estimated pupil blob. Moreover, the glints should also verify selected assumptions about size, compactness, brightness, and relative position in the image. All of these requirements permit determination of the, approximate glint and pupil position in the image. Figure 13 shows the steps used to determine pupil area. First, a rough segmentation of the glints and pupil area is carried out; second, the reduced region of interest allows for a more careful determination of the grey values in the pupil area to be used for thresholding.

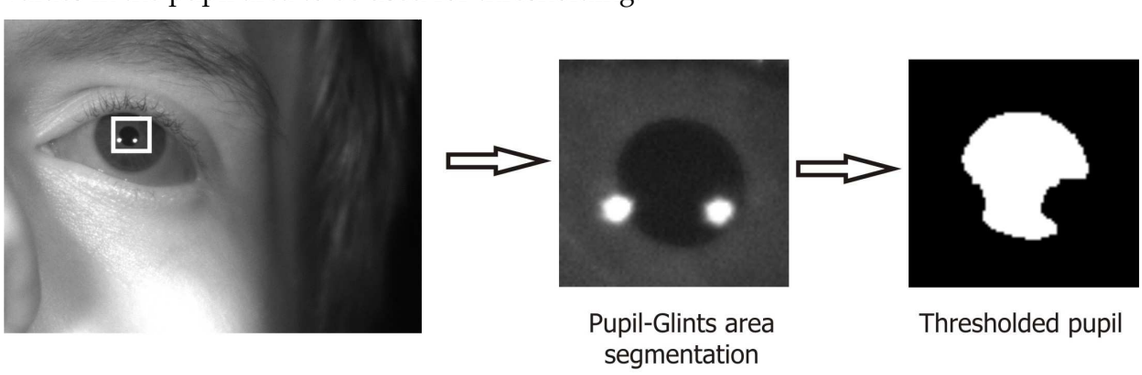


Figure 13. Pupil and glint area extraction from the image and thresholding

Once the pupil blob has been identified, two alternative methods have been compared to determine pupil contour: i) detection using grey level image and ii) ellipse detection in the thresholded image.

The first method uses a Canny edge detector. This is known to many as the optimal edge detector, and it combines Gaussian smoothing with border detection. Once the gradient is calculated, each value is analysed at each pixel by looking in the gradient direction and

eliminating those pixels not belonging to the border by establishing selected threshold values. In this manner, thinner lines are obtained for image contours.

For the second method, called the Thresholding method, the binarised (black and white) pupil is used to determine the pupil contour. Once the pupil area has been thresholded, the image is sampled by rows and columns detecting the white-to-black changes. Consequently, the pupil contour points are determined.

Figure 14 shows the pupil-glint image together with the detected contour points. Box and cross shape points represent contour points calculated using the Canny and Thresholding methods, respectively. In both methods, the influence of glints is removed by removing the points of the pupil contour that are in contact with any of the glints, since the glint positions are known beforehand.

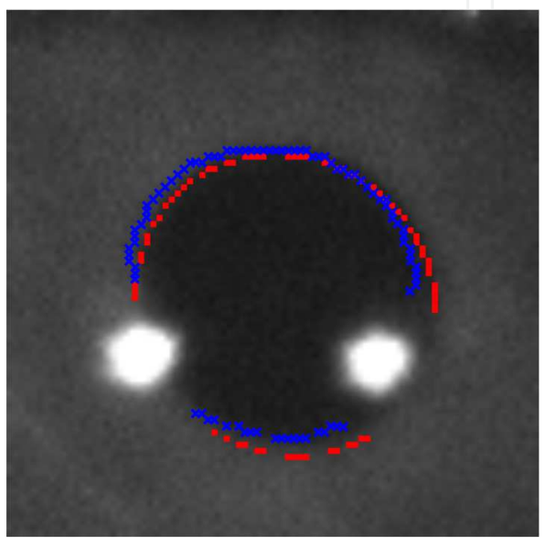


Figure 14. Comparison between Canny and Thresholding edge-detection methods for pupil contour points detection. Box shape points represent the points obtained using a Canny detector, while the cross shape points show the points obtained using the Thresholding method

As expected, the two methods present inaccuracies in the determination of contour points. Inaccurate contour detection introduces errors in gaze estimation. The objective of the analysis is first to evaluate the reliability of the two methods and second to obtain an approximate value of the existing uncertainty in detection of pupil contour points. The reliability of the methods is measured as an indicator of the goodness of fit. To this end, the processed results of images from the same gazed point are compared.

Five new subjects took part in this experiment. They were asked to observe a grid of 3x3 points on the screen. Thirty images were acquired for each one of the grid points. The estimated contours were fitted to an ellipse, and the ellipse parameters, i.e. centre, semi-axes and orientation, were calculated. Ellipse features were used as inputs of a filtering process that selected 18 images from the 30 acquired for each gazed point using the Mahalanobis Distance criteria. This process permits elimination of artifacts or outliers from the sample. In order to estimate the uncertainty of the contour detection procedure, the pupil centre of gravity is calculated as a function of the contour points by both the Canny and Thresholding methods. Thus, the statistical distribution of the centre of gravity can be calculated for each

gazed point by means of the two alternative contour detection methods. In order to compare data from different points, users, and methods, the Coefficient of Variation (CV) is calculated as the ratio between the standard deviation and the absolute value of the mean. Low CVs indicate good stability. Data from all subjects and points are evaluated, and the results show that the two methods have low CVs for pupil centre of gravity. Using the Thresholding method, a mean value of 1% is obtained for the CV, while a 2% is obtained for the Canny detector. It is concluded that both methods present and acceptable reliability. Moreover, the results show low variability regarding CV between users and screen points.

In order to evaluate the influence of contour detection errors on the calculated gaze position a previous estimation of the existing indetermination is necessary. Pupil contour can vary when looking at the same point due to alternative reasons, such as eye micro-movement, head movement or image processing algorithms errors. The objective of this study is to estimate the indetermination derived from the image processing method used. Employing the same data the difference between Thresholding and Canny methods for pupil centre of gravity is calculated for each acquired image. The obtained mean difference is ~1.1 pixels for both, x and y coordinates of the pupil centre of gravity. Consequently, a $1.1\sqrt{N}$ pixels of difference can be expected for pupil contour points, where N is the number of contour points used to calculate the pupil centre of gravity. It is assumed that this error is due to image processing algorithms inaccuracies.

Nevertheless, to evaluate the effect of image processing inaccuracies properly, gaze estimation data are needed. Forty pupil contour point positions of the pupil are corrupted in a simulated environment using Gaussian noise with alternative standard deviation values in the range of 2 to 12 pixels. Glint data are preserved with no corruption, and the effect of the pupil contour is measured on its own. For each grid point, ten noisy iterations are carried out, and the estimated gaze points are calculated. Thus, the error with respect to the real gaze point can be computed. In Figure 15, the 3x3 grid of points is represented with the estimations obtained for a standard deviation of 8 pixels.

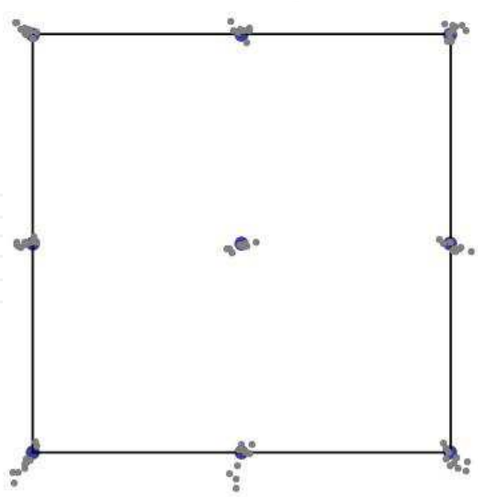


Figure 15. 3x3 grid of points and the estimated gaze positions when 8 pixels of noise is introduced

Figure 16 shows the distribution of the error for the full grid and for the different deviation values. On the horizontal axis, the error is shown in visual degrees. In addition, the

threshold at 1º is also indicated, since errors below this limit are acceptable performance indicators. From the figure it is observed that 8 pixels of noise produce a mean error above the acceptable limit of 1º.

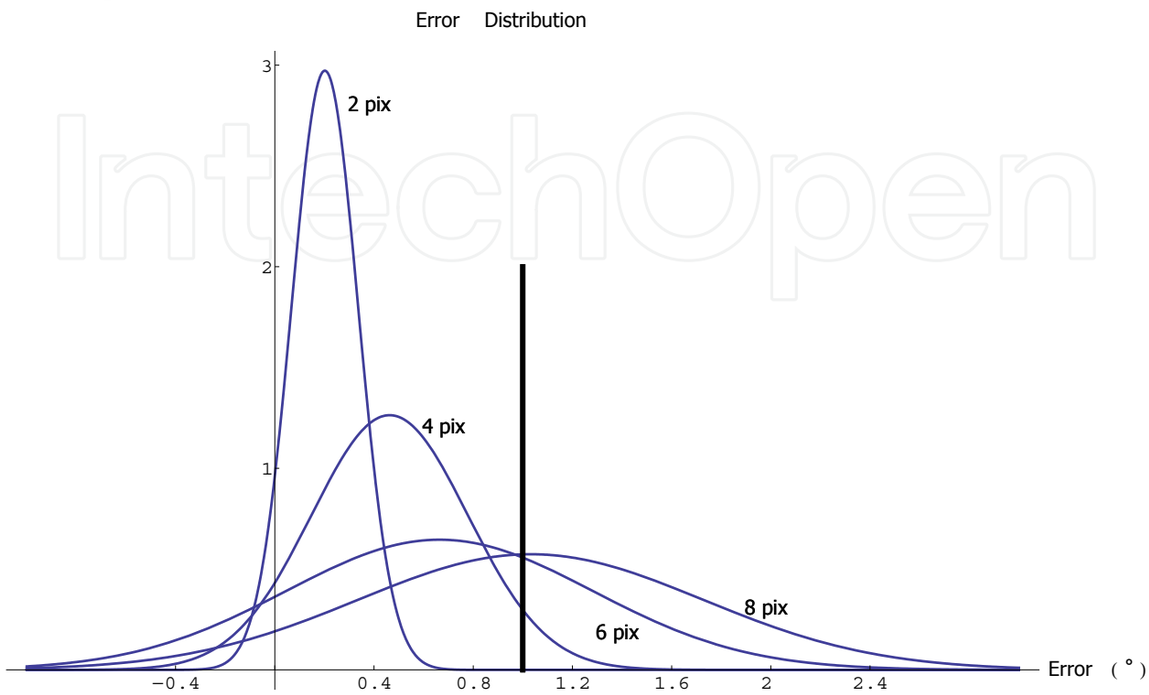


Figure 16. Error distribution for the 3x3 grid of points for different noise levels

However, the distribution can change slightly if the error is computed as the difference between the real gaze point and the averaged estimations. Figure 17 shows the same grid from Figure 15, but only shows the real gaze point and the average of the estimations. The error is clearly reduced.

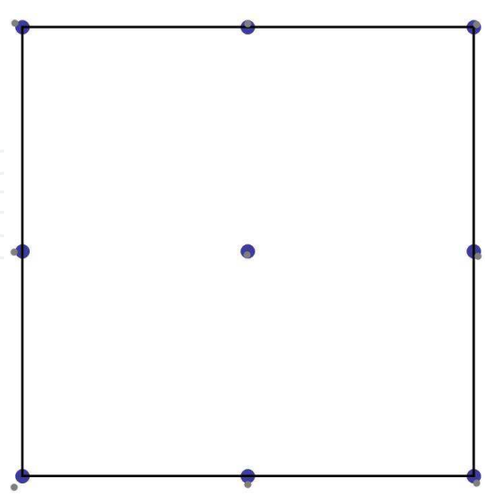


Figure 17. 3x3 grid of points and the estimated averaged gaze positions when 8 pixels of noise is introduced

In the same manner, the distribution of the error changes if this method is used, since the error is partially compensated by averaging. Thus, the average behaviour of the model is acceptable for deviations of 8 and 10 pixels (see Figure 18). Nevertheless, eye tracking system speed and our application requirements would determine the error computation method to a large extent.

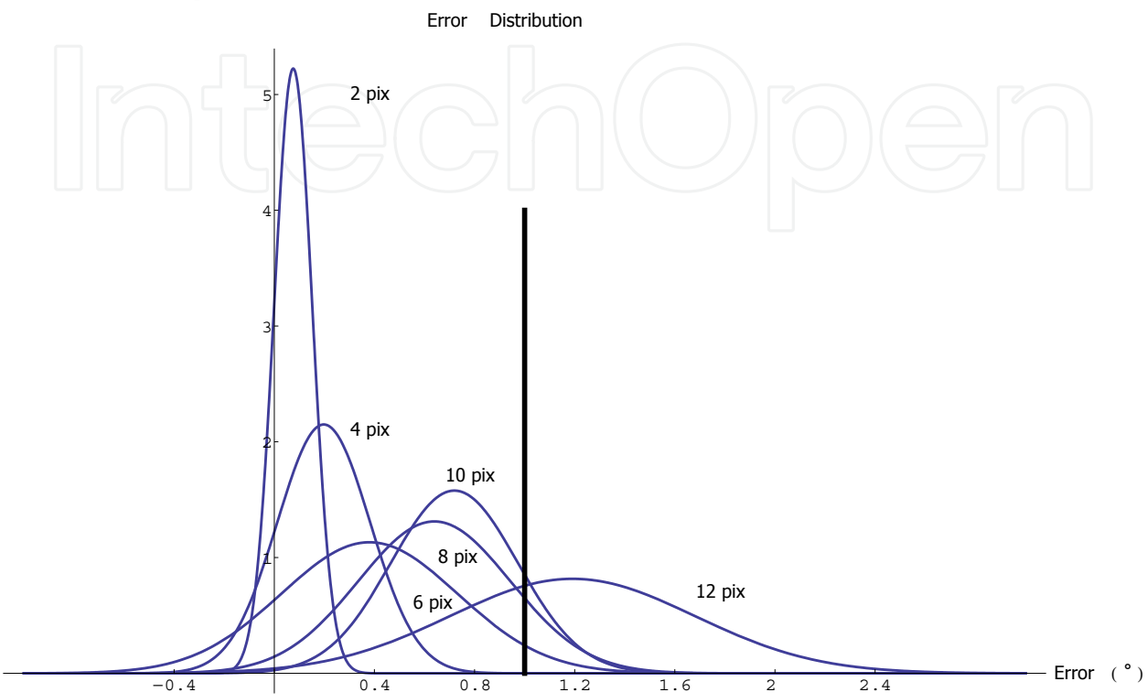


Figure 18. Error distribution for the 3x3 grid of points for different noise levels when averaged estimations are used for each point

7. Conclusions

A geometry-based model for gaze estimation has been constructed and evaluated. Alternative models have been proposed based on different image features. The results show that a hybrid model is needed for LoS estimation in a free head pose scenario, since glints and pupil separately do not provide sufficient information for optical axis estimation. Glints have been shown to be connected to the corneal centre position. Once the cornea centre has been determined, pupil shape can be used to estimate the pupil centre. The line connecting the pupil centre and the cornea centre is determined as the optical axis of the eye. Thus, the LoS is deduced using Listing's and Donder's Laws.

Cornea centre estimation is significantly sensitive to glint position inaccuracies. Hence, the number of light sources has been increased to obtain an averaged cornea centre and to compensate for the error. The obtained results show a better performance as the number of light sources is increased from two to four.

The proposed model requires the system hardware, such as the camera, light sources, and screen, to be calibrated beforehand. In addition, individual parameters such as, r_c, h, and β are required for gaze estimation. Theoretically, one calibration point is sufficient to obtain the aforementioned subject's parameters.

8. Future directions

Important obstacles need to be overcome in order to make eye tracking technology more accessible. Light variations make the image analysis more difficult, especially in outdoor settings in which light can change rapidly. New image acquisition devices, such as the silicon retina, can help solve this issue. This device produces data from those pixels for which relative brightness variations occur; however, no response is given if the overall lighting changes. Interesting attempts have been carried out using the silicon retina for eye tracking (Delbrück et al., 2008), but their production is still a difficult process and the devices that are currently available are low-resolution. Glasses and contact lenses can also introduce problems to the image processing task. Glasses can produce undesired glints in the image, while contact lenses can modify the pupil shape in the image. Most systems generally permit users to wear glasses and contact lenses; however, problems have been reported for specific subjects.

Another important area of research in the eye-gaze tracking field is gaze estimation. The connection between the image and gaze determines to a large extent the system accuracy and head movement constraints. In order to make the technology more accessible, accurate systems with head movement tolerance are needed, as head movement tolerance is a desired requirement for most eye tracking applications. Geometry-based models can provide useful information for constructing more accurate and versatile gaze tracking systems. Building models based on geometric principles results in valuable *a priori* data about possibly less accurate screen areas or system error sensitivity with respect to head movement. Geometric models can also contribute to a reduction of calibration requirements or to an increase in the effectiveness of calibration procedures. A great deal of effort has recently been put into this field, but a lot of work is still needed.

The eye model used in most of the papers devoted to gaze estimation is incomplete. Recently, a Matlab implementation for the eye model has been presented (Böhme et al., 2008) valid for gaze estimation studies. The eyeball model used in this work has been used in many works in recent years. Most researchers agree that the simplest possible eyeball model is needed and that model inaccuracies are partially compensated for during the calibration process. However, it is worth exploring additional eyeball properties, such as corneal ellipsoidal shape, and evaluating their influence in system accuracy in order to reach a proper trade-off between model complexity and accuracy. Moreover, most gaze estimation models are restricted to single-eye systems. One straightforward step would be to introduce the additional eyeball into the model. The geometrical analysis of binocular gaze estimation can considerably improve system performance and robustness.

One important issue to solve for most geometry-based gaze estimation systems is the necessity of hardware calibration. Many of these models require some extent of camera, light source, and screen position knowledge, which is not a trivial step. Although camera calibration has been widely studied, knowing the positions of light sources and the screen with respect to the camera is still a problem that needs to be solved to make this technology accessible. One possible solution is to provide the system in a closed form solution in which the camera and light sources are encapsulated together with the screen in a dedicated structure. However, the drawbacks of this solution are that it would increase the price and reduce the accessibility of the technology. Reducing hardware calibration requirements is a highly desirable task that would facilitate the employment of this technology.

Similarly, construction of eye tracking devices using off-the-shelf elements is a very interesting line of research. Thus, using webcams for eye-gaze tracking represents a goal of the field. This type of system would make eye tracking technology independent of expensive and complex hardware, but new image processing algorithms would be needed. Moreover, it is still difficult to achieve equivalent performance (resolution) with this kind of technology, since, e.g., the wider field of view of a webcam does not permit detection of the pupil with the same accuracy as a dedicated camera and optics. One possible solution to compensate for the reduction in accuracy would be to use adaptive user interfaces, such as zooming interfaces (Hansen et al., 2008).

9. References

Beymer, D. & Flickner, M. (2003). Eye gaze tracking using an active stereo head. *Proceedings of the 2003 conference on Computer Vision and Pattern Recognition*, vol. 02, p. 451, Los Alamitos, CA, USA. 2003, IEEE Computer Society.

Böhme, M., Dorr, M., Graw, M., Martinetz, M. & Barth, E. (2008). A software framework for simulating eye trackers. *Proceedings of the 2008 symposium on Eye tracking research and applications.*, pp. 251-258, Savannah, March, 2008, ACM Press, New York, NY, USA.

Cerrolaza, J.J., Villanueva, A. & Cabeza, R. (2008). Taxonomic study of polynomial regressions applied to the calibration of video-oculographic systems. *Proceedings of the 2008 symposium on Eye tracking research and applications.*, pp. 259-266, Savannah, March, 2008, ACM Press, New York, NY, USA.

Delbrück, T., Grover, D., Gisler, D., Lichtsteiner, P., Ersbøll, B. & Tureczek, A. (2008). D5.5 Silicon retina and other novel approaches for eye tracking. *Communication by Gaze Interaction* (COGAIN), IST- 2003-511598: Deliverable 5.5.

Guestrin, E. & Eizenman, M. (2006). General theory of remote gaze estimation using the pupil center and corneal reflections. *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 6, pp. 1124–1133, 2006.

Guestrin, E. & Eizenman, M. (2008). Remote point-of-gaze estimation requiring a single-point. *Proceedings of the 2008 symposium on Eye tracking research and applications.*, pp. 267-274, Savannah, March, 2008, ACM Press, New York, NY, USA.

Hansen, D.W. & Pece, A.E.C. (2005). Eye tracking in the wild. *Computer Vision and Image Understanding*, vol. 98, no. 1, pp. 155–181, 2005.

Hansen, D.W., Skovsgaard, H.H., Hansen, J.P. and Møllenbach, E. (2008). Noise tolerant selection by gaze-controlled pan and zoom in 3D. *Proceedings of the 2008 symposium on Eye tracking research and applications.*, pp. 205-212, Savannah, March, 2008, ACM Press, New York, NY, USA.

Hennessey, C., Noureddin, B. & Lawrence, P. (2006). A single camera eye-gaze tracking system with free head motion *Proceedings of the 2006 symposium on Eye tracking research and applications.* pp. 87-94, San Diego, March 2006, ACM Press, New York, NY, USA.

Morimoto, C.H. & Mimica, M.R.M. (2005). Eye gaze tracking techniques applications, *Computer Vision and Image Understanding* 98 (1) (2005) pp. 4–24.

Ohno, T. & Mukawa, N. (2004). A free-head, simple calibration, gaze tracking system that enables gaze-based interaction. *Proceedings of the 2004 symposium on Eye tracking research and applications.* pp. 115-122, San Antonio (TX), March 2004, ACM Press, New York, NY, USA.

Shih, S.W. & Liu, J. (2004). A novel approach to 3-D gaze tracking using stereo cameras, *IEEE Transactions Systems Man and Cybernetics Part-B*, vol. 34, no. 1, February 2004, pp. 234–245.

Villanueva, A. & Cabeza, R. (2007). Models for gaze tracking systems. Journal on Image and Video Processing. Volume 2007, Issue 3 (November 2007) Article No. 4, 2007, ISSN:1687-5176.

Villanueva, A. & Cabeza, R. (2008a). Evaluation of Corneal Refraction in a Model of a Gaze Tracking System. *IEEE Transactions on Biomedical Engineering.* (in press). Urbana IL, USA.

Villanueva, A. & Cabeza, R. (2008b). A Novel Gaze Estimation System with One Calibration Point. *IEEE Transactions on Systems, Man and Cybernetics* Part-B, vol. 38, no. 4, August 2008, pp. 1123-1138.

Yoo, D.H. & Chung, M. J. (2005). A novel non-intrusive eye gaze estimation using cross-ratio under large head motion, *Computer Vision and Image Understanding*, vol. 98, no. 1, pp. 25–51, 2005.

Advances in Human Computer Interaction

Edited by Shane Pinder

ISBN 978-953-7619-15-2

Hard cover, 600 pages

Publisher InTech

Published online 01, October, 2008

Published in print edition October, 2008

In these 34 chapters, we survey the broad disciplines that loosely inhabit the study and practice of human-computer interaction. Our authors are passionate advocates of innovative applications, novel approaches, and modern advances in this exciting and developing field. It is our wish that the reader consider not only what our authors have written and the experimentation they have described, but also the examples they have set.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Arantxa Villanueva, Juan J. Cerrolaza and Rafael Cabeza (2008). Geometry Issues of Gaze Estimation, Advances in Human Computer Interaction, Shane Pinder (Ed.), ISBN: 978-953-7619-15-2, InTech, Available from:
http://www.intechopen.com/books/advances_in_human_computer_interaction/geometry_issues_of_gaze_estimation

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821