Eye gaze tracking techniques for interactive applications. Comp Vis Im Und

Article *in* Computer Vision and Image Understanding · April 2005

Impact Factor: 1.54 · DOI: 10.1016/j.cviu.2004.07.010 · Source: DBLP

Eye Gaze Tracking Techniques for Interactive Applications

Carlos H. Morimoto [a,1] Marcio R. M. Mimica [a]

[a] *Departamento de Ciência da Computação*
Universidade de São Paulo, São Paulo, Brazil
{*hitoshi,mimica*}*@ime.usp.br*

Abstract

This paper presents a review of eye gaze tracking technology and focuses on recent advancements that might facilitate its use in general computer applications. Early eye gaze tracking devices were appropriate for scientific exploration in controlled environments. Although it has been thought for long that they have the potential to become important computer input devices as well, the technology still lacks important usability requirements that hinders its applicability. We present a detailed description of the pupil-corneal reflection technique due to its claimed usability advantages, and show that this method is still not quite appropriate for general interactive applications. Finally, we present several recent techniques for remote eye gaze tracking with improved usability. These new solutions simplify or eliminate the calibration procedure and allow free head motion.

1 Introduction

Eye gaze trackers (EGTs) are devices that can estimate the direction of gaze of a person. Young and Sheena [1] present a good survey of traditional techniques for eye gaze tracking. Some more recent reviews can be found in Glenstrup and Nielsen [2] and Duchowski [3].

Early EGTs were developed for scientific exploration in controlled environments or laboratories. Eye gaze data has been used in ophthalmology, neurology, psychology and related areas to study oculomotor characteristics and

[1] We thank the support from the Conselho Nacional de Pesquisa e Desenvolvimento (CNPq) e Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e International Business Machines (IBM)

abnormalities, and their relation to cognition and mental states. There are more recent applications for research in marketing and advertising, as well as in Human Factors Engineering to evaluate computer interfaces and web sites, but they are still confined to controlled environments.

For long though EGTs have been suggested as input devices for computer interfaces [4,5], successful attempts are still limited to military applications and the development of interfaces for people with disabilities. This paper is concerned with the usability of EGTs for more general computer applications.

According to Duchowski [6], eye gaze tracking applications can be categorized as diagnostic or interactive. Diagnostic applications use eye gaze data as quantitative evidence of the user's visual and attentional processes. Interactive applications use eye gaze data to respond to or interact with the user based on the observed eye movements.

Many traditional techniques for eye gaze tracking are intrusive, i.e., they require some equipment to be put in physical contact with the user. These techniques include, for example, contact lenses, electrodes, and head mounted devices. Non-intrusive techniques (or remote techniques) are mostly vision based, i.e., they use cameras to capture images of the eye. Some camera based techniques might be somewhat intrusive if they require to be head mounted.

For diagnostic applications, where eye data can be recorded during a short experiment and processed later, the time required to setup the eye gaze tracker and the discomfort that the equipment might cause do not constitute a problem (at least in general). This is also true for a few interactive applications where the user has to depend heavily on the eye tracker to accomplish some task (i.e., there is little choice or no alternative device).

A remote eye gaze tracker (REGT) offers comfort of use, and easier and faster setup, allowing the user to use the system for longer periods than intrusive techniques. Although the accuracy of REGTs is in general lower than intrusive EGTs, they are more appropriate for use during long periods. The pupil-corneal reflection technique is commonly advertised as a remote gaze tracking system that is robust to some head motion, and easy to calibrate. Most of these claims are unfortunately not exactly true. The small head motion tolerated by such devices have considerable influence in their accuracy, therefore experiments are usually done using a chin rest or bite bar to restrict head motion, which greatly reduces the user's comfort level.

Schnipke and Todd [7] describe how hard it can be to collect reliable eye tracking data using a commercial REGT system. In order to setup their usability experiment of a popular software application, an operator with one year experience with the eye tracker was not able to make the system track 10 out of 16 subjects, and none of them wore glasses. Among their list of tribulations

there are problems related to the difficulty of setting the system up, such as controlling the illumination conditions and placing the camera such that the eyelashes do not interfere, and some intrinsic problems of the system related to calibration, dry eyes, glasses, and lag of the system when there is some head motion.

Despite the limitations of current technology, eye gaze based interactive applications have the potential to revolutionize the way we use computers. Lewis [8] discusses design issues for constructing intelligent agent-based user interfaces. The term agent has come to refer to the automation of aspects of Human-Computer Interaction (HCI). Such interfaces would be able to anticipate commands or perform actions autonomously, but in order to do so, they would require reliable means to detect the focus of attention of the user to infer the user's "intention". Vertegaal [9] discusses design issues for a similar interaction paradigm named Attentive Interfaces, and shows how it could benefit from REGTs. Other applications of eye trackers are suggested by Zhai [10], Duchowski [3], and Glenstrup and Nielsen [2]. And while we wait for a reliable, accurate, easy to operate, and low cost REGT, Edwards [11] suggests a tool that can be used to develop these eye-aware applications.

In the next section we start with a short description of the eye structure to better understand the principles and restrictions of each eye tracking technique. Section 3 describes traditional EGTs, which might be intrusive or non-intrusive. The performance of these techniques are compared, and Section 4 describes in detail the pupil-corneal reflection technique and discusses its usability to general computer applications. Section 5 describes advanced techniques for REGTs that characterize the state of the art of this technology, and Section 6 concludes the paper.

2 Human Eye Structure

Figure 1 shows the major components of the human eye. The eye has an approximately spherical shape with a radius of about 12 mm [12].

The external parts of the eye that are visible in the eye socket are the sclera (the white part of the eye), the iris (the color part of the eye), and the pupil located in the center of the iris. The cornea is a protective transparent membrane, void of blood vessels, that protrudes toward the front of the eye, and covers the iris. The iris has a circular aperture in the center, called pupil, which regulates the amount of light coming into the eye by constantly changing its size.

Behind the iris there is the lens, a biconvex multilayered structure. The shape

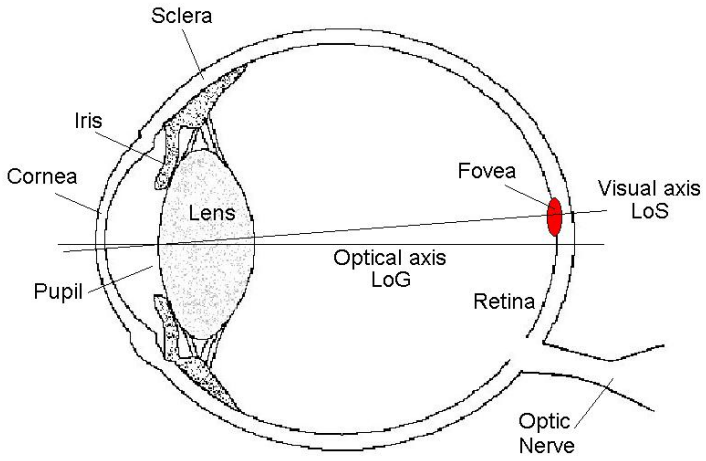


Fig. 1. Structure of the human eye

of the lens changes during accommodation, a process that allows to bring the image of an object to a sharp focus in the retina, which is a layer of photo sensitive cells located at the back of the eye. Between the cornea and the lens lies the anterior chamber which is filled with the watery aqueous humor and in the space between the lens and the retina is the transparent gelatinous vitreous body. The light that penetrates the retina has traversed the whole eye optic media, suffering reflection and refraction at each media boundary.

There is a small but special region in the retina, known as the fovea, that concentrates most of the color sensitive cells and is responsible for the perception of the scene's fine details. The fovea is not exactly in the optical axis of the eye defined by the center of the eyeball and the center of the pupil. We will call the optical axis of the eye as the line of gaze (LoG), and the line from the fovea through the center of the pupil as the line of sight (LoS). It is the LoS and not the LoG that determines a person's visual attention. If the LoG or LoS can be estimated and there is information about the scene objects, the point of regard is computed as the intersection of the LoG (or LoS) with the nearest object of the scene. For HCI, it is reasonable to consider the monitor of the computer to be the object of analysis and the point of regard a pixel on this monitor.

Different eye models describe the optical characteristics of the human eye under different complexity levels. We will use Gullstrand's eye model [13] as an ideal eye to demonstrate some properties of EGTs. Table 1 shows the properties of the boundary surfaces that are in the light path through the cornea until the retina. The boundaries between structures are set as spherical surfaces.

Table 1
Path of a light ray using Gullstrand's eye model.

	Position (mm)	Radius (mm)	Refraction index after surface
Cornea	0	7.7	1.376
	0.5	6.8	1.336
Eyelens	3.2	5.33	1.385
	3.8	2.65	1.406
	6.6	-2.65	1.385
	7.2	-5.33	1.336
Retina	24.0	-11.5	

3 Eye Gaze Trackers

The Applied Vision Research Unit (AVRU) of the University of Derby maintains an eye movement equipment database (`http://ibs.derby.ac.uk/emed`) that might be useful for those readers interested in more detailed information about commercial eye trackers. Detailed information about several traditional eye trackers is presented by Young and Sheena [1]. In this section we briefly describe the characteristics of traditional intrusive and remote eye tracking techniques, and discuss some usability requirements for interactive computer applications.

3.1 Intrusive Eye Gaze Trackers

Intrusive eye gaze tracking techniques are in general more accurate than remote ones. Some less accurate alternatives are also less expensive. One of the most traditional methods is based on contact lenses. Robinson [14] uses a small coil (called search coil) embedded into a contact lens that is tightly fit over the sclera with a slight suction to avoid drift during fast eye movements. The user's gaze is estimated from measuring the voltage induced in the search coil by an external electro-magnetic field. Although very intrusive, the system is very accurate (approximately 0.08^o).

A less expensive technique is based on measuring skin potentials, as described by Kaufman et al.[15]. The electro-oculogram (EOG) is a very common technique for recording eye movement for clinical applications due to its technical simplicity. By placing electrodes around the eye, it is possible to measure small differences in the skin potential that correspond, among others, to eye move-

ment. This technique is also not appropriate for everyday use, and its reported accuracy is about 2^o.

Cameras or other optical devices can be used to measure the eye-position without direct contact with the user. Some camera based methods require the eye to be very close to the optical device, and therefore, must be head mounted, or the motion of the head must be restricted with the use of a chin rest of bite-bar.

3.2 Camera Based Eye Gaze Trackers

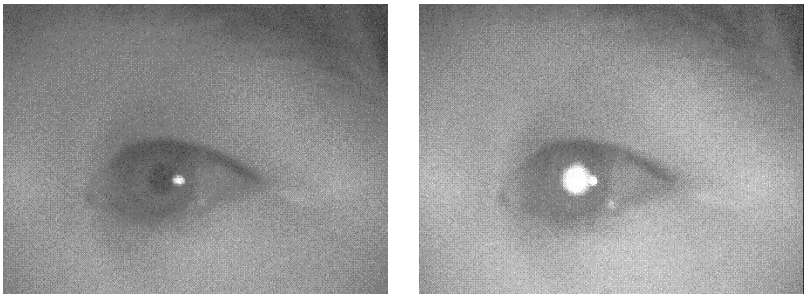
Camera based eye gaze tracking techniques rely on some properties or characteristics of the eye that can be detected and tracked by a camera or other optical or photo sensitive device. Most of these techniques have the potential to be implemented in a non-intrusive way.

The limbus and the pupil are common features used for tracking. Limbus is the boundary between the sclera and the iris. Due to the contrast of these two regions, it can be easily tracked horizontally, but because the eyelids in general cover part of the iris, limbus tracking techniques have low vertical accuracy. The pupils are harder to detect and track because of the lower contrast between the pupil-iris boundary, but pupil tracking techniques have better accuracy since they are not covered by the eyelids (except during blinking).

To enhance the contrast between the pupil and the iris, many eye trackers use an infrared (IR) light source. Because IR is not visible, the light does not distract the user. In practice, most implementations use near IR light sources with wavelength around 880nm, which is almost invisible for the human eye, but can still be detected by most commercial cameras.

Sometimes, the IR source is placed near the optical axis of the camera. Because the camera now is able to "see" the light reflected from the back of the eye, similar to the red eye effect in night photography using a bright flash light, the camera sees a bright pupil as can be seen in Figure 2b instead of a regular dark pupil, as shown in Figure 2a. Nguyen et al.[16] have conducted experiments that show the behavior of the infrared bright pupil response of human eyes, discussing some of the factors that might cause great variation of the bright pupil response between subjects.

The light source also generates a corneal reflection (CR) or glint on the cornea surface, which is clearly visible in Figure 2a, near the pupil. This glint is used as a reference point in the pupil-corneal reflection technique described in Section 4.



(a) (b)

Fig. 2. Dark and bright pupil images

Reulen et al.[17] use a variation of the limbus tracking technique with IR lighting called infrared oculography (IROG). They place IR light emitting diodes and IR light sensitive photo-transistors respectively above and bellow the eye. Several such IR pairs can be mounted on goggles or helmets, so that the limbus is always illuminated, in particular, the nasal and temporal sides of the limbus. A photo-transistor transforms the reflected IR light into a voltage. The voltage of the nasally located photo-transistors is compared to the voltage of the temporal located photo-transistors, and the resulting voltage difference is proportional to the angular deviation of the eye. They named their system IRIS eye tracker (http://www.skalar.nl), which was originally designed for clinical oculomotor diagnosis in humans. However, its high accuracy (about 2 minutes of arc) and large bandwidth allow fundamental studies of eye behavior as well, such as saccades, pursuit, vergence, etc. The IROG IRIS system is head mounted and weights about 300g.

Cornsweet and Crane [18] describe another very accurate eye tracker that uses the first and fourth Purkinje images (see Figure 3). The Purkinje images are reflections created at different layers of the eye structure. The first Purkinje image corresponds to the reflection from the external surface of the cornea. This is the brightest and easiest reflection to detect and track. Detecting the other Purkinje images require special hardware, but allow the estimation of the 3D point of regard from the 3rd and 4th Purkinje images that correspond to the relaxation of the eye lens, as described by Crane [19].

In the dual Purkinje image (DPI) eye tracker, when the eye undergoes translation, both Purkinje images move together. But when the eye rotates, the two images move through different distances and thus change their separation. This separation yields a measure of the angular orientation of the eye. The authors report an impressive accuracy of about 1 minute of arc.

Instead of using explicit geometric features such as the contours of the limbus or the pupil, an alternative approach to object pose estimation is to treat an image as a point in a high-dimensional space. Techniques using this representation are often referred to as being appearance-based or view-based. Tan et

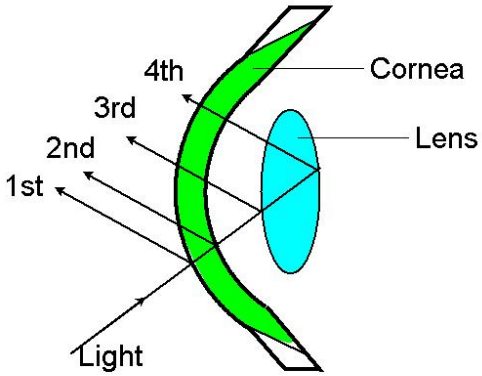


Fig. 3. Purkinje images

al.[20] use 252 images from 3 users to build an appearance-based method for estimating the gaze, and achieve an accuracy of about 0.5^o.

The method proposed by Baluja and Pomerleau [21] also does not use explicit geometric features. They describe an eye gaze tracker based on artificial neural networks (ANN). Once the eye is detected, the image of the eyes are cropped and used as input to a ANN. Training images are taken when the user is looking at a specific point on a computer monitor. In their experiments, 2000 training samples were used. Their prototype tracker runs at 15Hz, with an accuracy of about 2^o, and allows for some head movement.

3.2.1 Robustness and Accuracy

The first step for remote eye gaze estimation is the eye detection and tracking. Besides the eye, some techniques require the detection of other features. Robust and accurate eye and feature detection and tracking is fundamental to enhancing the usability of current REGTs, so the operator might be less concerned about the illumination conditions and position of the camera to have a good image of the eye, trying to avoid occlusions from eyelashes and even eye glasses.

The literature offers several techniques for detecting eyes directly, or as a sub-feature of the face. Faces can be detected from background subtraction, skin color segmentation, geometric models and templates, artificial neural networks, etc.

Direct methods for eye detection used for EGT are presented in [22–24]. Kothari and Mitchell [23] use spatial and temporal information to detect the location of the eyes. Their process starts by selecting a pool of potential candidates using gradient fields. The gradient along the iris/sclera boundary always point outward from the center of the iris, thus by accumulating along these

lines, the center of the iris can be estimated by selecting the bin with highest count. Heuristic rules and a large temporal support are used to filter erroneous pupil candidates. This method will find eye candidates that may be relatively small, but most REGT systems assume that there is a large eye in the image that is easy to segment using simple thresholds that can be adjusted by the operator.

The simple thresholding technique to segment bright or dark pupils might not be robust to variations of the ambient lighting. Besides, it requires the system's operator to find an appropriate threshold level for each session. Kim and Ramakrishna [25] and Perez et al.[26] use edge detection techniques to segment the limbus or the pupil, also requiring thresholding.

Tomono et al.[24] and Ebisawa and Satoh [22] have developed similar techniques for robust detection of the pupil and the corneal reflection. Tomono et al.[24] developed a very elaborate real-time imaging system composed of a single camera with 3 CCDs and 2 near infra red (IR) light sources. The light sources have different wavelengths (λ_1, and λ_2). The light source with wavelength λ_1 (or λ_1 for simplicity) is also polarized. λ_2 is placed near the camera optical axis, and λ_1 is placed slightly off-axis, generating bright and dark pupil images respectively. CCD3 is sensitive to λ_2 only, thus it outputs bright pupil images. CCD1 and CCD2 are sensitive to λ_1 (λ_2 is filtered out), and CCD1 also has a polarizing filter in order to receive only the diffuse light components, i.e., the corneal reflection due to λ_1 does not appear in the images from CCD1. Once the 3 images are available, the pupil is segmented from differencing and thresholding the images from CCD3 and CCD2, and the corneal reflection used for gaze estimation is segmented using the images from CCD2 and CCD1.

Observe in Figure 4c that the pupil can be easily segmented from the difference between the bright and dark pupil images, even for people wearing glasses. The thresholding operation is facilitated, and can be more easily automated. The system from Ebisawa and Satoh [22] is also based on a differential lighting scheme using two light sources, but with the same wavelength (on and off camera axis) to generate the bright/dark pupil images. The detection of the corneal reflection created by the light sources requires the use of a narrow field of view camera (long focal length) since the reflection is in general very small. In a following publication, Ebisawa [27] presents a real-time implementation of the system using custom hardware and pupil brightness stabilization for optimum detection of the pupil and the corneal reflection. This system does not use the glint from the dark pupil image, because in that particular implementation the glint from the off-axis LED moved with the zoom level and was also hard to detect due to low contrast with its surrounding neighborhood.

The eye trackers from Tomono et al.and Ebisawa were developed for eye gaze

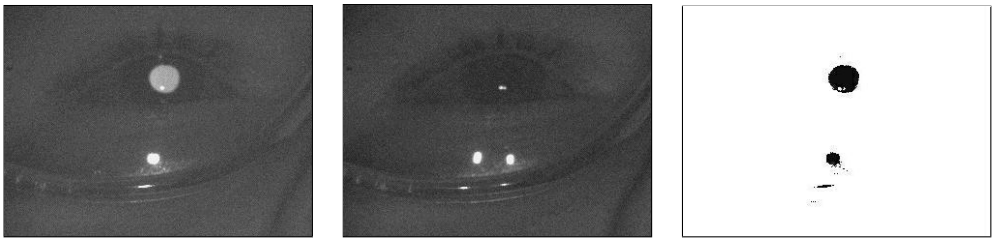


Fig. 4. a) Bright pupil; b) Dark pupil; and c) Difference pupil image

tracking in controlled (lab) environments that require high accuracy. For HCI purposes, Morimoto et al.[28], Haro et al.[29], and Zhu et al.[30] also suggested the use of the differential lighting scheme to remote eye detection and tracking. They all report good robustness to illumination changes, but for wide field of view cameras, pupil candidates must be filtered. Due to the use of active IR lighting, this technique works better indoors and even in the dark, but might not be appropriate outdoors, because sunlight contains IR and the pupils become smaller in bright environments.

The accuracy and the resolution of the direction of gaze highly depends on the accuracy of the pupil (or iris) detection. To achieve subpixel accuracy, the computation of the center of mass of the pupil region is probably the most natural way of computing its center, though its not very robust due to the presence of reflections and eyelashes.

Zhu and Yang [31] suggest ellipse fitting for subpixel iris tracking, using the edges from the limbus. The implicit representation of a conic is given by

$$a.x^2 + b.x.y + c.y^2 + d.x + e.y + f = 0$$

For a conic to be an ellipse, the following restriction: $b^2 - 4.a.c < 0$ must hold. A normalization constraint on the ellipse parameters can be imposed by the following $4.a.c - b^2 = 1$. Using this normalization, Fitzgibbon et al.[32] present a direct least-square method to compute the ellipse parameters from $N \geq 6$ points of the ellipse.

Because the contour can also be affected by outliers such as eyelashes and corneal reflections, a dual ellipse fitting mechanism as suggested by Ohno et al.[33] can further increase the robustness and accuracy of the eye gaze estimate. First, the pupil edges are used for ellipse fitting, and then, only the pupil edges that are close to the computed ellipse edges are used for the second fitting. This filters outliers out of the true pupil contour and increase the accuracy of the pupil position estimate.

Some techniques require the detection of other features, such as the corneal reflection. The detection of the CR can be done in a greedy way, by searching for a bright spot closest to the pupil center. The computation of the center of

the CR can be reliably computed as the position of the centroid of the bright spot.

A typical NTSC camera is able to generate 30 frames per second, but because the frames are interlaced, most camera based EGTs can achieve rates of up to 60 Hz. The resolution of a frame is 640×480 pixels. If the camera has a narrow field of view of about 4^o horizontally, that means that the camera can see a region of about 40mm from 600 mm (typical distance from the eye to the monitor and camera). In this particular situation, one pixel in the image would roughly correspond to 0.0625mm. Because the IROG and DPI techniques use photo sensitive cells instead of cameras to remotely detect changes in the eye position, they can detect changes faster and more accurately than regular cameras.

3.3 Calibration and Head Motion

So far we have basically just described what is measured in each technique to estimate the direction of gaze. These measurements, such as pupil position, limbus position, skin potentials, relative position of Purkinje images, and so on, must be translated to eye orientation. A calibration procedure is required to compute the mapping between the measurements and the eye orientation. Besides the eye orientation, the accommodation of the lens could be measured for 3D eye tracking.

A typical calibration procedure presents the user a set of visual targets that the user has to look at while the corresponding measurement is taken. From these correspondences, a mapping or calibration function can be computed. Ideally this function should be linear over a wide visual angle. Cornsweet and Crane [18] show that the DPI technique has good linearity within 10^o in diameter.

For the traditional pupil and limbus tracking techniques, the position of the center of the pupil or iris must be mapped to the visual targets. Since the eye position varies with the head position, the head should remain still during and after the calibration. One way to compensate for small head motion, is to consider the pupil/iris position relative to the eye socket, or some reliable fixed point on the user's face. Therefore the mapping is computed using the vector from the reference point to the pupil/iris center. For the pupil-corneal reflection technique, the CR is used as the reference point.

The appearance-based and ANN techniques "learn" the calibration from a large set of images, and generalize this mapping to other users, i.e., they have the advantage of not requiring a per user calibration once they are trained, but because the image of the eye also changes with head positions (and illumination conditions), these techniques are also sensitive to head motion.

Table 2
Characteristics of traditional EGTs.

Technique	Accuracy	Comments
Contact Lens	1'	very intrusive, but fast and accurate
EOG	2^o	intrusive, but simple and low cost
IROG	2'	head mounted, limbus tracking
DPI	1'	not intrusive, but requires bite bar
Limbus Tracking	1^o	camera based, lower vertical accuracy
Pupil Tracking	1^o	camera based, hard to detect the pupil
Pupil-Glint	1^o	camera based, tolerate some head motion
Image-based	0.5^o - 2^o	camera based, requires training

3.4 Eye Gaze Tracker Usability Requirements

Table 2 summarizes some characteristics of traditional eye gaze trackers. Besides the accuracy, there are several usability requirements that an EGT should satisfy. According to Scott and Findlay [34] and Hallett [35] the ideal EGT would:

(1) Offer an unobstructed field of view with good access to the face and head;
(2) Make no contact with the subject;
(3) Meet the practical challenge of being capable of artificially stabilizing the retinal image if necessary ;
(4) Possess an accuracy of at least one percent or a few minutes of arc;
(5) Offer a resolution of 1 minute of arc.sec^{-1}, and thus be capable of detecting the smallest changes in eye position;
(6) Offer a wide dynamic range of one minute to 45^o for eye position and one minute arc.sec^{-1} to 800.sec^{-1} for eye velocity
(7) Offer good temporal dynamics and speed of response
(8) Possess a real-time response
(9) Measure all three degrees of angular rotation and be insensitive to ocular translation
(10) Be easily extended to binocular recording
(11) Be compatible with head and body recordings
(12) Be easy to use on a variety of subjects

The above list includes several technical requirements for speed, accuracy and resolution, some particular requirements for laboratory use, and other more general requirements for ease of use. To build eye-aware applications for the general public, many of these requirements are too restrictive, thus they can be "relaxed", if not completely ignored, such as requirement 3.

To be applied in general computer interfaces, an ideal eye tracker should:

(1) be accurate, i.e., precise to minutes of arc;
(2) be reliable, i.e., have constant, repetitive behavior;
(3) be robust, i.e, should work under different conditions, such as indoors/outdoors, for people with glasses and contact lenses, etc;
(4) be non intrusive, i.e., cause no harm or discomfort;
(5) allow for free head motion;
(6) not require calibration, i.e., instant setup;
(7) have real-time response;

In short, it should work anywhere, for everyone, all the time, with any application, without the need for setup, and should cause the user no harm or discomfort.

Of course no technique satisfies these usability requirements, but the pupil-corneal reflection technique offers some advantages over the other available alternatives. Although the performance of this method can benefit from the use of a chin rest or bite-bar, many manufacturers such as ASL[36], LC Technologies [37], and SMI [38], claim that their pupil-corneal reflection eye trackers can tolerate small head motion, typically within one cubic feet, after a simple calibration procedure, and can achieve an accuracy between 0.5^o and 1^o.

Although we do not have access to their proprietary software, the next section presents some algorithms and methods that can be used to implement a pupil-corneal reflection EGT, and discusses those usability claims.

4 Pupil-Corneal Reflection Technique

Due to its simplicity and reasonable accuracy, many REGTs today are based on the corneal reflection technique [36,39,4,37,28,38]. Figure 5 shows a typical pupil-corneal reflection setup. Similar to the DPI technique, it also requires an IR light source to generate the Purkinje images, but only the first Purkinje image, or corneal reflection, needs to be detected and tracked.

Assuming that the eye is a sphere that only rotates around its center, and that the camera and light source are fixed, the position of the CR does not move with the eye rotation, and therefore can be used as a reference point. The center of the pupil (or iris) and the CR defines a vector in the image. This vector can be easily mapped to screen coordinates on a computer monitor after a calibration procedure, and used, for example, to control the cursor on a graphical user interface. This is an easy way to assess the quality of the calibration over the monitor.

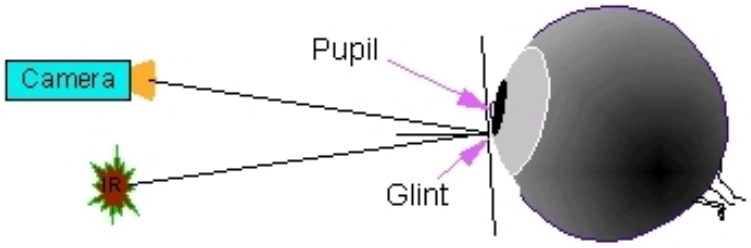


Fig. 5. Pupil-corneal reflection technique

A calibration procedure is required to compute the mapping from the pupil-glint vector to monitor screen coordinates. In general, the user is asked to look at several points on the computer screen, one point at a time, and press a button. Morimoto et al.[28] use 9 points for calibration and a second order polynomial calibration function. The polynomial is defined as:

$$s_x = a_0 + a_1 x + a_2 y + a_3 xy + a_4 x^2 + a_5 y^2$$

$$s_y = b_0 + b_1 x + b_2 y + b_3 xy + b_4 x^2 + b_5 y^2$$

where (s_x, s_y) are screen coordinates and (x, y) is the pupil-corneal reflection vector. The parameters a_0 to a_5 and b_0 to b_5 are the unknowns. Since each calibration point defines 2 equations, the system is over constrained with 12 unknowns and 18 equations, and can be solved using least squares. Actually, because the set of parameters a and b are independent, it can be solved as 2 systems of 6 unknowns and 9 equations.

Fitting even higher order polynomials has been shown to increase the accuracy of the system [40], but the second order requires less calibration points and provides a good approximation. Simpler linear models has also been successfully used.

The use of the corneal reflection as reference allows for small head motion because the CR follows the head motion, and the calibration nicely handles the offset due to the difference of the LoG and LoS, imperfections of the cornea, position of the camera relative to the computer screen, etc.

There are several problems with this simple model though. Unfortunately the calibration mapping decays as the head moves away from its original position, and as mentioned by Schnipke and Todd [7], the calibration is one of the worst problems in current REGTs because it requires the operator to adjust several system parameters such as illumination conditions and relative position of the user, monitor, and camera.

The use of the differential lighting scheme may facilitate the system setup

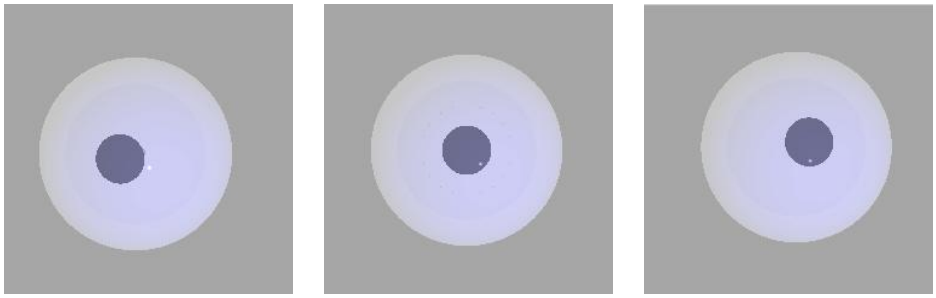


Fig. 6. Ray traced images using Gullstrand's eye model

and make it more robust to illumination changes, but the calibration is still a problem. To test the assumption that the calibration holds for small head motion, we have generated several synthetic images of the eye model, rendered using ray tracing. Figure 6 shows 3 such images of the eye. We are mostly interested in the pupil center and corneal reflection. The images show the cornea sphere inside the sclera, and the pupil as a disc inside the cornea. The rotation is done around the center of the eye ball (i.e., not the center of the cornea). Because the indexes of refraction of the cornea and the aqueous humor are very similar (see Table 1), only the external boundary of the cornea is considered.

For this experiment, the optical center of the camera is assumed to be the center of coordinates. The (x, y) coordinates correspond to the left and up directions, respectively. The z-axis defines the optical axis, in a camera centered right-handed coordinate system. The camera's vertical field of view was set to 3.5^o.

Let $P_0 = (0, 270, 600)$ be the eye position, and the computer screen be defined by the rectangle in the xy plane, with upper left coordinate (183, 274) and lower right coordinate (-183, 0), which roughly corresponds to a 18" monitor. All coordinates are in millimeters. Figure 7 shows the relative position of the eye with respect to the camera and monitor.

To calibrate the system, the screen was divided into a 3×3 grid (see Figure 7). Thus, 9 images of the eye were rendered looking at the center of each grid element. A refined version of the system presented by Morimoto et al.[28] was used for image processing, calibration and gaze estimation.

To test the calibration and estimate the error without head motion, each grid element was further divided into a 3×3 grid, and images of the eye looking at each of these new locations were generated. Figure 8 shows the calibration errors along the screen for the eye positioned at P_0. The error is defined as the distance from the estimated gaze position and the true grid position. The average error over the whole computer screen is about 8mm, or about 0.8^o of visual angle. Notice that the error is not uniform along the screen. Jacob

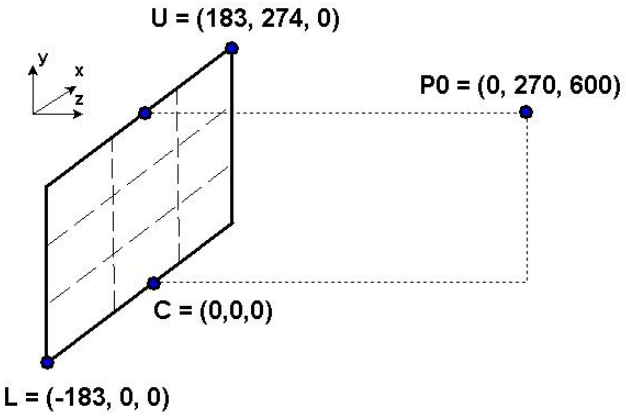


Fig. 7. Camera and monitor setup

reports a similar fact in [41], and solved the problem by giving the user the possibility of making local manual re-calibration. The user should move the cursor with the mouse to the area needing recalibration, and click on the region while looking at the cursor.

To test the influence of head motion in the accuracy of the calibration function, the eye was moved to different positions to simulate head motion. The eye was translated to 3 different positions along the 3 axis, corresponding to displacements of 50mm, 100mm, and 150mm from the original position. For each position, 81 new images of the eye, corresponding to gazing at the 81 grid elements on the computer screen, were rendered and given as input to the REGT.

Translating the eye along the x-axis resulted in small variations in the average errors. Figure 9 compares the average errors at each of the nine screen grid coordinates for the eye at P_0 and the eye positioned at (-100, 270, 600). Notice that when the eye is moved to a new position, the camera direction was changed to keep the eye centered in its view, but the same calibration parameters are used. The average error is 9.92 mm for this new eye position.

Figure 10 shows the comparison for the eye at P_0 and at (0, 270, 700), a translation in z. The average errors increases to 40.56 mm in this position, showing that this calibration function is more sensitive to eye movements along the z-axis.

The average error for the eye positioned at (0, 170, 600) is 21.76 mm, which shows that the technique is not as robust for movements along the y-axis as it is for horizontal translations, though not as bad as for translations along the z-axis. These results are in accordance with our personal experience, since small left-right or up-down rotations of the head (which are the most common motions) do not affect the calibration much. Moving the head up-down (along

the y-axis) is uncommon, and moving the head closer to the monitor really affects the calibration.

White et al.[42] use a simple linear model with independent components, and mention that in practice, higher order polynomial do not provide better calibration. We have also tested a simpler linear model with 6 calibration parameters instead of 12, and noticed that the calibration is better near the monitor edges for the more complex calibration model, but in our practical experiments, this refinement is not noticeable.

Fitting a polynomial function over the whole monitor screen is also not a requirement. For example, Zhu and Yang [31] construct a 2D linear mapping from the vector between the eye corner and the iris center to the gaze angle. After calibration, the gaze direction is computed by interpolation. For example, suppose the gaze angle and the eye corner to iris vector used for calibration in points P_1 and P_2 are respectively $\{(\alpha_1, \beta_1), (x_1, y_1)\}$ and $\{(\alpha_2, \beta_2), (x_2, y_2)\}$.

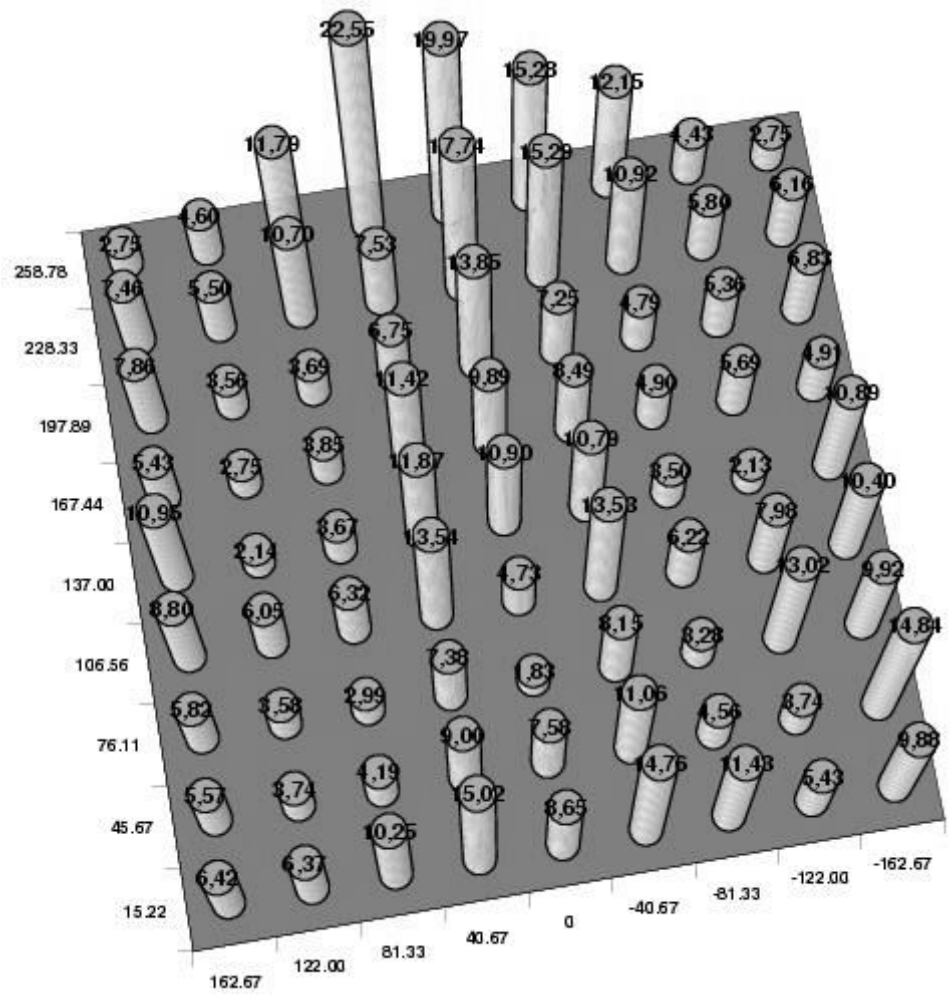


Fig. 8. Distribution of errors (in mm) along the screen for eye at P_0.

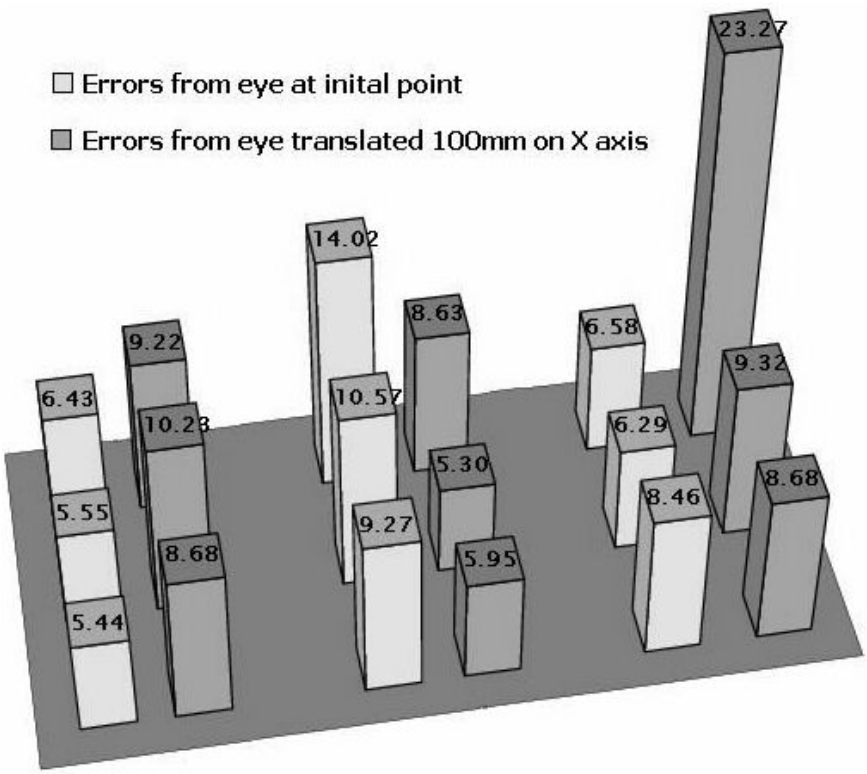


Fig. 9. Average screen grid errors (in mm) to eye at P_0 and at (-100, 270, 600).

Then after the measurement of a corner-iris vector (x, y) is taken, the gaze angle is computed as follows:

$$\alpha = \alpha_1 + \frac{x - x_1}{x_2 - x_1}(\alpha_2 - \alpha_1)$$

$$\beta = \beta_1 + \frac{y - y_1}{y_2 - y_1}(\beta_2 - \beta_1)$$

They report an average error of about 1.1^o using subpixel accuracy for tracking the eye corner and iris-center, and about 3.3^o using pixel level accuracy.

5 Advanced REGT

Advanced REGTs that are being researched today basically try to eliminate two problems, the need of calibration per user session, and the large restriction on head motion.

A simple method for estimating eye gaze without calibration and allowing free head motion was suggested by Morimoto et al.[43]. They use one single camera and 2 IR light sources, one light to generate the bright pupil image

and the second to generate the dark pupil image. Because the cornea surface can be modeled as a spherical convex mirror, assuming paraxial rays from the light sources when reflected by the mirror (cornea), it is possible to compute the center of the cornea in 3D. This requires the calibration of the camera with respect to the monitor and light positions, and a model of the user's eye. From the center of the cornea, they also estimate the 3D position of the pupil, and the gaze direction is defined as the 3D vector from the cornea center to the pupil center. Experimental results show an accuracy of about 3^o using synthetic images.

Another interesting method is described by Yoo et al.[44]. They use 4 LEDs around the monitor screen to project these corners on the corneal surface (see Figure 11). A 5th LED is placed near the CCD camera lens to create a bright pupil image, and help segmenting the pupil. They assume the cornea is flat, so when the user is looking at the monitor, the center of the pupil will appear within the polygon defined by the 4 corneal reflections from the light sources on the monitor. Using the invariance property of cross ratios under perspective, they compute the point of regard very efficiently with an accuracy of about 2^o. One great advantage of this method is that it does not require camera calibration.

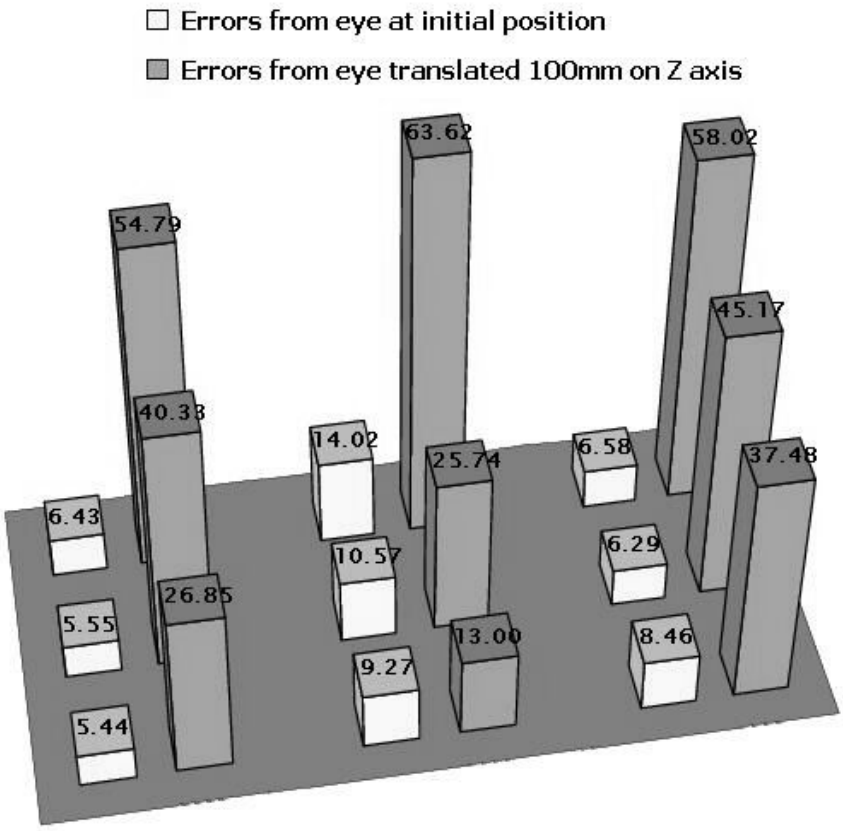


Fig. 10. Average screen grid errors (in mm) to eye at P_0 and at (0, 270, 700).

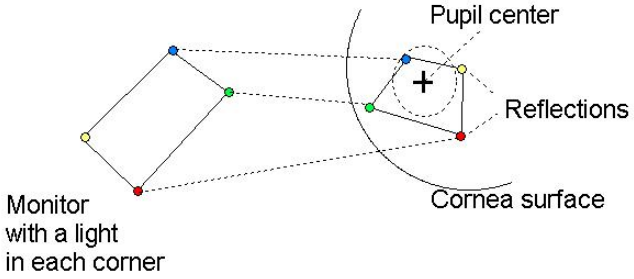


Fig. 11. Projection of light sources around the surface of the monitor

Other systems use information about the face to estimate gaze. As mentioned in [45], though not implemented, any 2D gaze tracking system could be extended to 3D if the absolute position of the eye can be determined. Wang and Sung [46] and Newman et al.[47] give examples of systems that first compute the face pose in 3D, and then compute the eye gaze. Newman et al.[47] locate the 3D position of the corners of the eye from stereo, and computes the 3D *LoG* from the orientation of the eyeball. Some of the eye parameters have to be trained, per person. The system runs in real time, but the accuracy is low, about 5^o. Wang and Sung [46] also combine a face pose estimation system with a narrow FOV camera to compute the gaze direction. They assume that the iris contour is a circle to estimate its normal direction in 3D. To compute the point of regard using real images, a second image of the eye from a different position is used. Their tests using synthetic images and real images from 3 subjects show that the accuracy of the system is bellow 1^o, which is very good considering that they do not use an eye model and do not compensate for the foveal offset.

Beymer and Flickner [48] use a separate stereo system to detect the face in 3D. Once a face is detected, this information is used to steer a second narrow field of view stereo pair of cameras. The large pupil images are then used to fit projected model features from their 3D eye model to detected image features. A one-time calibration per user is required to estimate intrinsic parameters of the eye, such as the radius of the cornea and the angular offset of the LoS. The authors report a 0.6^o accuracy in the gaze direction, for one person at a distance of 622mm from the monitor.

Shih and Liu [49] do not have a system to position the narrow field of view stereo cameras, but similar to Beymer and Flickner, their method is based on a simplified eye model. They use multiple cameras and multiple point light sources to estimate the optical axis of the eye. Using the simplified eye model of Le Grand, they show that using 2 calibrated cameras and at least 2 point light sources with known positions, it is possible to compute the *LoG*. The offset of the *LoS* to the *LoG* can be obtained from a one-time calibration

Table 3
Comparison and brief description of methods that allows free head motion

Author	year	f/s	accuracy	desc
Beymer and Flickner	2003	20	0.6^o	3D face tracking + 3D gaze
Morimoto et al.	2002	na	3^o	Single camera and at least 2 lights
Newman et al.	2000	30	$> 5^o$	3D face track + 2D eye gaze
Shih and Liu	2003	30	$< 1^o$	Two cameras and at least 2 lights
Wang and Sung	2002	na	$< 1^o$	3D face pose + 2D eye gaze
Yoo et al.	2002	na	2^o	Single camera and 5 lights

procedure per user, and usually takes 2-3 seconds. In their implementation they use 3 IR LEDs, and process 30 frames a second with an accuracy of under 1^o.

Table 3 shows a comparison of the speed and accuracy between the methods presented in this section that allow free head motion. All these methods are quite recent. The most accurate and fast ones are from Beymer and Flickner [48] and Shih and Liu [49]. Both these systems require a one time calibration per user but because they require system calibration between independent parts such as the stereo system and monitor, we do not believe they are quite ready for wide spread use. This is somewhat true also for the system presented by Morimoto et al.[43], since they also require calibration, although for a single camera. The most promising technique, at least in terms of easy of use, seems to be the one presented by Yoo et al.[44], despite their lower accuracy.

We did not consider head mounted eye trackers that compensate for head motion in this section. All the advanced methods presented are non-intrusive, and none of them requires calibration per session, although some of them require one calibration per user. Unfortunately quite a few also require camera calibration, which might be even harder to achieve for inexperienced users.

6 Conclusion

We have described the state of the art of remote eye gaze trackers (REGTs), and showed that two of the major usability concerns of current REGT technology, the requirements for constant system calibration and very limited head motion, are being answered by the latest generation of REGTs. We have also described traditional methods used for eye gaze tracking, and focussed our discussion on the pupil-corneal reflection technique. From this review, we showed that eye trackers were laboratory instruments, and as such, many intrusive techniques could be tolerated. But for the development of general computer

eye-aware computer applications, new usability requirements must be satisfied.

Acknowledgments

We would like to thank Arnon Amir, Myron Flickner and Dave Koons for the many meaningful suggestions and contributions to our work.

References

[1] L. Young, D. Sheena, Methods & designs: Survey of eye movement recording methods, Behavioral Research Methods & Instrumentation 7 (5) (1975) 397–429.

[2] A. Glenstrup, T. Engell-Nielsen, Eye controlled media: Present and future state, Master's thesis, University of Copenhagen DIKU (Institute of Computer Science), Universitetsparken 1 DK-2100 Denmark (June 1995).

[3] A. T. Duchowski, Eye Tracking Methodology: Theory and Practice, Springer Verlag, 2003.

[4] T. Hutchinson, K. W. Jr., K. Reichert, L. Frey, Human-computer interaction using eye-gaze input, IEEE Transactions on Systems, Man, and Cybernetics 19 (1989) 1527–1533.

[5] R. Jacob, What you look at is what you get, IEEE Computer 26 (7) (1993) 65–66.

[6] A. T. Duchowski, A breadth-first survey of eye tracking applications, Behavioral Research Methods, Instruments, and Computers (2002) 1–16.

[7] S. K. Schnipke, M. W. Todd, Trials and tribulations of using an eye-tracking system, in: Proc. ACM SIGCHI - Human Factors in Computing Systems Conference, 2000, pp. 273–274.

[8] M. Lewis, Designing for human agent interaction, AI Magazine 19 (2) (1998) 67–78.

[9] R. Vertegaal, Designing attentive interfaces, in: Proc. of the Eye Tracking Research & Applications Symposium, New Orleans, LA, 2002, pp. 23–30.

[10] S. Zhai, What's in the eyes for attentive input, Communications of the ACM 46 (3) (2003) 34–39.

[11] G. Edwards, A tool for creating eye-aware applications that adapt to changes in user behavior, in: Proc. of ASSETS 98, Marina del Rey, CA, 1998.

[12] G. Wyszecki, W. Stiles, Color Science: Concepts and Methods, Quantitative Data and Formulae, John Wiley & Sons, New York, 1982.

[13] R. Longhurst, Geometrical and Physical Optics, 3rd Edition, John Wiley & Sons, New York, 1974.

[14] D. A. Robinson, A method of measuring eye movements using a scleral search coil in a magnetic field, IEEE Transactions on Biomedical Engineering 10 (1963) 137–145.

[15] A. Kaufman, A. Bandopadhay, B. Shaviv, An eye tracking computer user interface, in: Proc. of the Research Frontier in Virtual Reality Workshop, IEEE Computer Society Press, 1993, pp. 78–84.

[16] K. Nguyen, C. Wagner, D. Koons, M. Flickner, Differences in the infrared bright pupil response of human eyes, in: Proc. of the Eye Tracking Research & Applications Symposium, New Orleans, LA, 2002.

[17] J. Reulen, j.T. Marcus, D. Koops, F. de Vries, G. Tiesinga, K. Boshuizen, J. Bos, Precise recording of eye movement: the iris technique, part 1., Med Biol Eng Comput 26 (1) (1988) 20–26.

[18] T. Cornsweet, H. Crane, Accurate two-dimensional eye tracker using first and fourth purkinje images, Journal of the Optical Society of America 63 (8) (1973) 921–928.

[19] H. Crane, C. Steele, Accurate three-dimensional eyetracker, Journal of the Optical Society of America 17 (5) (1978) 691–705.

[20] K. Tan, D. Kriegman, H. Ahuja, Appearance based eye gaze estimation, in: Proc. of the IEEE Workshop on Applications of Computer Vision - WACV02, 2002, pp. 191–195.

[21] S. Baluja, D. Pomerleau, Non-intrusive gaze tracking using artificial neural networks, Tech. Rep. CMU-CS-94-102, School of Computer Science, CMU, CMU Pittsburgh, Pennsylvania 15213 (January 1994).

[22] Y. Ebisawa, S. Satoh, Effectiveness of pupil area detection technique using two light sources and image difference method, in: A. Szeto, R. Rangayan (Eds.), Proceedings of the 15th Annual Int. Conf. of the IEEE Eng. in Medicine and Biology Society, San Diego, CA, 1993, pp. 1268–1269.

[23] R. Kothari, J. Mitchell, Detection of eye locations in unconstrained visual images, in: Proc. of the International Conference on Image Processing, Vol. I, Lausanne, Switzerland, 1996, pp. 519–522.

[24] A. Tomono, M. Iida, Y. Kobayashi, A tv camera system which extracts feature points for non-contact eye movement detection, in: Proceedings of the SPIE Optics, Illumination, and Image Sensing for Machine Vision IV, Vol. 1194, 1989, pp. 2–12.

[25] K. Kim, R. Ramakrishna, Vision based eye gaze tracking for human computer interface, in: Proc. of the IEEE International Conf. on Systems, Man and Cybernetics, Tokyo, Japan, 1999.

[26] A. Prez, M.L.Crdoba, A. Garca, R. Mndez, M. Muoz, J. Pedraza, F. Snchez, A precise eye-gaze detection and tracking system, in: Proc. of the 11th International Conference in Central Europe of Computer Graphics, Visualization and Computer Vision'2003, Plzen, Czech Republic, 2003.

[27] Y. Ebisawa, Unconstrained pupil detection technique using two light sources and the image difference method, Visualization and Intelligent Design in engineering and architecture (1995) 79–89.

[28] C. Morimoto, D. Koons, A. Amir, M. Flickner, Pupil detection and tracking using multiple light sources, Image and Vision Computing 18 (4) (2000) 331–336.

[29] A. Haro, M. Flickner, I. Essa, Detecting and tracking eyes by using their physiological properties, dynamics, and appearance, in: Proc. of CVPR 2000, 2000, pp. 163–168.

[30] Z. Zhu, K. Fujimura, Q. Ji, Real time eye detectin and tracking under various light conditions, in: Proc. of the Eye Tracking Research & Applications Symposium, New Orleans, LA, 2002, pp. 25–27.

[31] J. Zhu, J. Yang, Subpixel eye gaze tracking, in: Proc. of the 5th IEEE International Conference on Automatic Face and Gesture Recognition, Washington D.C., 2002, pp. 131–136.

[32] A. Fitzgibbon, M. Pilu, R. B. Fisher, Direct least squares fitting of ellipses, IEEE Transactions on Pattern Analysis and Machine Intelligence 21 (5) (1999) 476–480.

[33] T. Ohno, N. Mukawa, A. Yoshikawa, Freegaze: a gaze tracking system for everyday gaze interaction, in: Proc. of the Eye Tracking Research & Applications Symposium, New Orleans, LA, 2002, pp. 125–132.

[34] D. Scott, J. Findlay, Visual search, eye movements and display units, Human factor report, University of Durham (1993).

[35] P. Hallett, Eye movements, Wiley, New York, 1986, Ch. 10, pp. 25–28.

[36] ASL, http://www.a-s-l.com/gazetracker.htm.

[37] LCTech, The eyegaze development system, http://www.eyegaze.com.

[38] S. M. I. Inc, Eyelink gaze tracking, http://www.smi.de.

[39] Y. Ebisawa, M. Ohtani, A. Sugioka, Proposal of a zoom and focus control method using an ultrasonic distance-meter for video-based eye-gaze detection under free-hand condition, in: Proceedings of the 18th Annual Int. Conf. of the IEEE Eng. in Medicine and Biology Society, 1996.

[40] Z. Cherif, A. Nait-Ali, J. Motsch, M. Krebs, An adaptive calibration of an infrared light device used for gaze tracking, in: Proc. of the IEEE Instrumentation and Measurement Technology Conference, Anchorage, AK, 2002, pp. 1029–1033.

[41] R. Jacob, Eye-movement-based human-computer interaction techniques: Toward non-command interfaces, Vol. 4, Ablex Publishing Corporation, Norwood, NJ, 1993, Ch. 6, pp. 151–190.

[42] K. W. Jr., T. Hutchinson, J. M. Carley, Spatially dynamic calibration of an eye-tracking system, IEEE Transactions on Systems, Man, and Cybernetics 23 (4) (1993) 1162–1168.

[43] C. Morimoto, A.Amir, M. Flickner, Detecting eye position and gaze from a single camera and 2 light sources, in: Proc. of the International Conference on Pattern Recognition, Quebec, Canada, 2002.

[44] D. Yoo, J. Kim, B. Lee, M. Chung, Non contact eye gaze tracking system by mapping of corneal reflections, in: Proceedings Int. Conf. on Automatic Face and Gesture Recognition, Washington D.C., 2002, pp. 94–99.

[45] C. Collet, A. Finkel, R. Gherbi, Capre: a gaze tracking system in human machine interaction, Journal of advanced computational intelligence 2 (3) (1998) 77–81.

[46] J. Wang, E. Sung, Study on eye gaze estimation, IEEE Transactions on systems, man, and cybernetics - PART B 32 (3) (2002) 332–350.

[47] R. Newman, Y. Matsumoto, S. Rougeaux, A. Zelinsky, Real time stereo tracking for head pose and gaze estimation, in: Proceedings Int. Conf. on Automatic Face and Gesture Recognition, Grenoble, France, 2000.

[48] D. Beymer, M. Flickner, Eye gaze tracking using an active stereo head, in: Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, Vol. II, Madison, WI, 2003, pp. 451–458.

[49] S. Shih, J. Liu, A novel approach to 3d gaze tracking using stereo cameras, IEEE Transactions on systems, man, and cybernetics - PART B (2003) 1–12.