Novel Eye Gaze Tracking Techniques Under Natural Head Movement

Zhiwei Zhu and Qiang Ji*, *Senior Member, IEEE*

***Abstract*—Most available remote eye gaze trackers have two characteristics that hinder them being widely used as the important computer input devices for human computer interaction. First, they have to be calibrated for each user individually; second, they have low tolerance for head movement and require the users to hold their heads unnaturally still. In this paper, by exploiting the eye anatomy, we propose two novel solutions to allow natural head movement and minimize the calibration procedure to only one time for a new individual.**

The first technique is proposed to estimate the 3-D eye gaze directly. In this technique, the cornea of the eyeball is modeled as a convex mirror. Via the properties of convex mirror, a simple method is proposed to estimate the 3-D optic axis of the eye. The visual axis, which is the true 3-D gaze direction of the user, can be determined subsequently after knowing the angle deviation between the visual axis and optic axis by a simple calibration procedure. Therefore, the gaze point on an object in the scene can be obtained by simply intersecting the estimated 3-D gaze direction with the object. Different from the first technique, our second technique does not need to estimate the 3-D eye gaze directly, and the gaze point on an object is estimated from a gaze mapping function implicitly. In addition, a dynamic computational head compensation model is developed to automatically update the gaze mapping function whenever the head moves. Hence, the eye gaze can be estimated under natural head movement. Furthermore, it minimizes the calibration procedure to only one time for a new individual.

The advantage of the proposed techniques over the current state of the art eye gaze trackers is that it can estimate the eye gaze of the user accurately under natural head movement, without need to perform the gaze calibration every time before using it. Our proposed methods will improve the usability of the eye gaze tracking technology, and we believe that it represents an important step for the eye tracker to be accepted as a natural computer input device.

***Index Terms*—Eye gaze tracking, gaze estimation, human computer interaction.**

I. Introduction

EYE gaze is defined as the line of sight of a person. It represents a person's focus of attention. Eye gaze tracking has been an active research topic for many decades because of its potential usages in various applications such as Human Computer Interaction (HCI), Virtual Reality, Eye Disease Diagnosis, Human Behavior Studies, etc. For example, when a user is looking at a computer screen, the user's gaze point at the screen can be estimated via the eye gaze tracker. Hence, the eye gaze can serve as an advanced computer input [1], which is proven to be more efficient than the traditional input devices such as a mouse pointer [2]. Also, a gaze-contingent interactive graphic display application can be built [3], in which the graphic display on the screen can be controlled interactively by the eye gaze. Recently, eye gaze has also been widely used by cognitive scientists to study human beings' cognition [4], memory [5], etc.

Numerous techniques [3], [6]–[17] have been proposed to estimate the eye gaze. Earlier eye gaze trackers are fairly intrusive in that they require physical contacts with the user, such as placing a reflective white dot directly onto the eye [6] or attaching a number of electrodes around the eye [7]. In addition, most of these technologies also require the user's head to be motionless during eye tracking. With the rapid technological advancements in both video cameras and microcomputers, gaze tracking technology based on the digital video analysis of eye movements has been widely explored. Since it does not require anything attached to the user, video technology opens the most promising direction for building a nonintrusive eye gaze tracker. Various techniques [18]–[23], [3], [17] have been proposed to perform the eye gaze estimation based on eye images captured by video cameras. However, most available remote eye gaze trackers have two characteristics that prevent them from being widely used. First, they must often be calibrated repeatedly for each individual; second, they have low tolerance for head movements and require the user to hold the head uncomfortably still.

In this paper, two novel techniques are introduced to improve the existing gaze tracking techniques. First, a simple 3-D gaze tracking technique is proposed to estimate the 3-D direction of the gaze. Different from the existing 3-D techniques, the proposed 3-D gaze tracking technique can estimate the optic axis of the eye without the need to know any user-dependent parameters about the eyeball. Hence, the 3-D direction of the gaze can be estimated in a way allowing more easy implementation, improving the robustness and accuracy of the gaze estimation simultaneously. Second, a novel 2-D mapping-based gaze estimation technique is introduced to allow free head movements and minimize the calibration procedure to only one time for a new individual. A dynamic head compensation model is proposed to compensate for the head movements so that whenever the head moves, the gaze mapping function at a new 3-D head position can be updated automatically. Hence, accurate gaze information

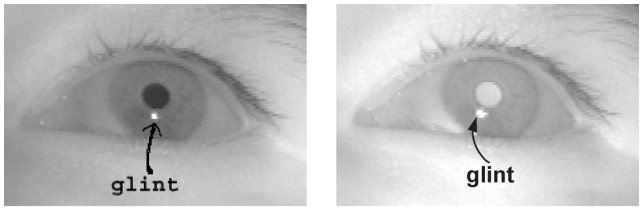
Manuscript received March 19, 2005; revised December 11, 2007. *Asterisk indicates corresponding author.*

Z. Zhu was with the Department of Electrical, Computer and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180-3590 USA. He is now with the Sarnoff Corporation, Princeton, NJ 08540 USA.

*Q. Ji is with Department of Electrical, Computer and Systems Engineering, Rensselaer Polytechnic Institute, SEC 6003, 110 8th Street, Troy, NY 12180-3590 USA (e-mail: jiq@rpi.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2007.895750



(a) (b)

Fig. 1. Eye images with glints. (a) Dark-pupil image. (b) Bright-pupil image. Glint is a small bright spot as indicated.

can always be estimated as the head moves. Therefore, by using our proposed gaze tracking techniques, a more robust, accurate, comfortable and useful eye gaze tracking system can be built.

II. Related Works

In general, most of the nonintrusive vision-based gaze tracking techniques can be classified into two groups: 2-D mapping-based gaze estimation method [3], [8]–[10] and direct 3-D gaze estimation method [11]–[17]. In the following section, each group will be discussed briefly.

A. 2-D Mapping-Based Gaze Estimation Technique

For the 2-D mapping-based gaze estimation method, the eye gaze is estimated from a calibrated gaze mapping function by inputting a set of 2-D eye movement features extracted from eye images, without knowing the 3-D direction of the gaze. Usually, the extracted 2-D eye movement features vary with the eye gaze so that the relationship between them can be encoded by a gaze mapping function. In order to obtain the gaze mapping function, an online calibration needs to be performed for each person. Unfortunately, the extracted 2-D eye movement features also vary significantly with head position; thus, the calibrated gaze mapping function is very sensitive to head motion [10]. Hence, the user has to keep his head unnaturally still in order to achieve good performance.

The pupil center cornea reflection (PCCR) technique is the most commonly used 2-D mapping-based approach for eye gaze tracking. The angle of the visual axis (or the location of the fixation point on the display surface) is calculated by tracking the relative position of the pupil center and a speck of light reflected from the cornea, technically known as the "glint" as shown in Fig. 1(a) and (b). The generation of the glint will be discussed in more detail at Section III-B.1. The accuracy of the system can be further enhanced by illuminating the eyes with near-infrared (IR) LEDs coaxial with the camera , which produces the "bright-pupil" effect as shown Fig. 1(b) and makes the video image easier to process. IR light is harmless and invisible to the user.

Several systems [24], [25], [18], [26] have been built based on the PCCR technique. Most of these systems show that if the users have the ability to keep their heads fixed, or to restrict head motion via the help of chin-rest or bite-bar, very high accuracy can be achieved in eye gaze tracking results. Specifically, the average error can be less than 1° visual angle, which corresponds to less than 10 mm in the computer screen when the subject is sitting approximately 550 mm from the computer screen. But as the head moves away from the original position where the user performed the gaze calibration, the accuracy of these gaze tracking systems drops dramatically; for example, [10] reports detailed data showing how the calibrated gaze mapping function decays as the head moves away from its original position. Jacob reports a similar fact in [25]. Jacob attempted to solve the problem by giving the user the ability to make local manual re-calibrations, which brings numerous troubles for the user. As these studies indicate, calibration is a significant problem in current remote eye tracking systems.

Most of the commercially available eye gaze tracking systems [27], [23], [28] are also built on the PCCR technique, and most of them claim that they can tolerate small head motion. For example, less than 2 square inches of head motion tolerance is claimed for the gaze tracker from *LC* technologies [23], which is still working to improve it. The ASL eye tracker [27] has the best claimed tolerance of head movement, allowing approximately one square foot of head movement. It eliminates the need for head restraint by combining a magnetic head tracker with a pan-tilt camera. However, details about how it handles head motion are not publicly known. Further, combining a magnetic head tracker with a pan-tilt camera is not only complicated but also expensive for the regular user.

In summary, most of existing eye gaze systems based on the PCCR technique share two common drawbacks: first, the user has to perform certain experiments in calibrating the relationship between the gaze points and the user-dependent parameters before using the gaze tracking system; second, the user has to keep his head unnaturally still, with no significant head movement allowed.

B. Direct 3-D Gaze Estimation Technique

For the direct 3-D gaze estimation technique, the 3-D direction of the gaze is estimated so that the gaze point can be obtained by simply intersecting it with the scene. Therefore, how to estimate the 3-D gaze direction of the eye precisely is the key issue for most of these techniques. Several attempts [15], [14], [11], [12], [17] have been proposed to estimate the 3-D direction of gaze from the eye images. The direct 3-D gaze estimation technique is not constrained by the head position, and it can be used to obtain the gaze point on any object in the scene by simply intersecting it with the estimated 3-D gaze line. Therefore, the issues of gaze mapping function calibration and head movement compensation that plague the 2-D mapping-based methods can be solved nicely.

Morimoto *et al.* [15] proposed a technique to estimate the 3-D gaze direction of the eye with the use of a single calibrated camera and at least two light sources. First, the radius of the eye cornea is measured in advance for each person, using at least three light sources. A set of high order polynomial equations are derived to compute the radius and center of the cornea, but their solutions are not unique. Therefore, how to choose the correct one from the set of possible solutions is still an issue. Furthermore, no working system has been built using the proposed technique.

Ohno *et al.* [14] proposed an approximation method to estimate the 3-D eye gaze. There are several limitations for this proposed method. First, the cornea radius and the distance between the pupil and cornea center are fixed for all users although they

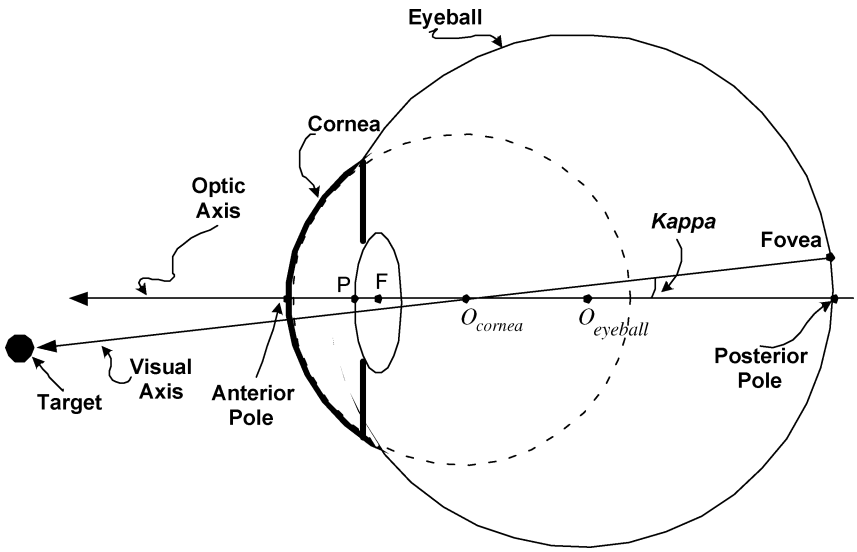


Fig. 2. Structure of the eyeball (top view of the right eye).

actually vary significantly from person to person. Second, the formulation to obtain the cornea center is based on the assumption that the virtual image of IR LED appears on the surface of the cornea. In fact, however – as shown in Section III-B.I of this paper – the virtual image of IR LED will not appear on the surface of the cornea; instead, it will appear behind the cornea surface or inside the cornea. Therefore, the calculated cornea center will be a very inaccurate approximation.

Beymer *et al.* [11] proposed another system that can estimate the 3-D gaze direction based on a complicated 3-D eyeball model with at least seven parameters. First, the 3-D eyeball model will be automatically individualized to a new user, which is achieved by fitting the 3-D eye model with a set of image features via a nonlinear estimation technique. The image features used for fitting include only the glints of the IR LEDs and the pupil edges. But as shown in Section III-B.I of this paper, the glints are the image projections of the virtual images of the IR LEDs created by the cornea, and they are not on the surface of the eye cornea, but inside the cornea. Also, the pupil edges are not on the surface of the 3-D eye model, either. Therefore, the radius of the cornea cannot be estimated based on the proposed method. Further, fitting such a complicated 3-D model with only few feature points, the solution will be unstable and very sensitive to noise.

Shih *et al.* [12] proposed a novel method to estimate 3-D gaze direction by using multiple cameras and multiple light sources. In their method, although there is no need to know the user-dependent parameters of the eye, there is an obvious limitation for the current system when stereo cameras are used. Specifically, when the user is looking at points on the line connecting the optical centers of the two cameras, the 3-D gaze direction cannot be determined uniquely.

Guestrin *et al.* [17] proposed a 3-D approach to remotely estimate 3-D gaze with the use of one or two video cameras together with multiple lights. Their method starts with the reconstruction of the optic axis of the eye. This is then followed by estimating the visual axis from the estimated optic axis through a calibration procedure. Based on a comparative analysis of different systems configurations, they theoretically show that the 3-D eye gaze can be minimally solved with either one camera and two lights, plus some subject-specific parameters or with two cameras and two lights, without any subject-specific parameters. They implemented the one camera plus two light configuration and demonstrated its accurate performance experimentally. Compared with the most of two-camera direct gaze estimation system, the one camera configuration represents a significant simplification. Their method, however, cannot accommodate large head movement.

Therefore, most of the existing 3-D gaze tracking techniques either require knowledge of several user-dependent parameters about the eye [15], [14], [11], or cannot work under certain circumstances [12] or with large head movement [17]. But in reality, these user-dependent parameters of the eyeball, such as the cornea radius and the distance between the pupil and the cornea center, are very small (normally less than 10 mm). Therefore, accurate indirect estimation techniques like the one proposed in [17] to estimate these eye parameters is a prerequisite.

III. Direct 3-D Gaze Estimation Technique

A. Structure of Human Eyeball

As shown in Fig. 2, the eyeball is made up of the segments of two spheres with different sizes placed in front of the other [29]. The *anterior*, the smaller segment, is transparent and forms about one-sixth of the eyeball, and has a radius of curvature of about 8 mm. The *posterior*, the larger segment, is opaque and forms about five-sixths of the eyeball, and has a radius of about 12 mm.

The *anterior pole* of the eye is the center of curvature of the transparent segment or *cornea*. The *posterior pole* is the center of the posterior curvature of the eyeball, and is located slightly temporal to the optical nerve. The *optic axis* is defined as a line connecting these two poles, as shown in Fig. 2. The fovea defines the center of the retina, and is a small region with highest visual acuity and color sensitivity. Since the fovea provides the sharpest and most detailed information, the eyeball is contin-

uously moving so that the light from the object of primary interest will fall on this region. Thus, another major axis, the *visual axis*, is defined as the projection of the foveal center into object space through the eye's nodal point O_{cornea} as shown in Fig. 2. Therefore, it is the visual axis that determines a person's visual attention or direction of gaze, not the optic axis. Since the fovea is a few degrees temporal to the posterior pole, the visual axis will deviate a few degrees nasally from the optic axis. The angle formed by the intersection of the visual axis and the optic axis at the nodal point is named as *angle kappa*. The angle kappa in the two eyes should have the same magnitude [29], approximately around 5°.

Pupillary axis of the eye is defined as the 3-D line connecting the center of the pupil P and the center of the cornea O_{cornea}. The pupillary axis is the best estimate of the location of the eye's optic axis; if extended through the eye, it should exit very near the anatomical posterior pole. In Fig. 2, the pupillary axis is shown as the optic axis of the eyeball. Therefore, if we can obtain the 3-D locations of the pupil center and cornea center, then the optic axis of the eye can be estimated.

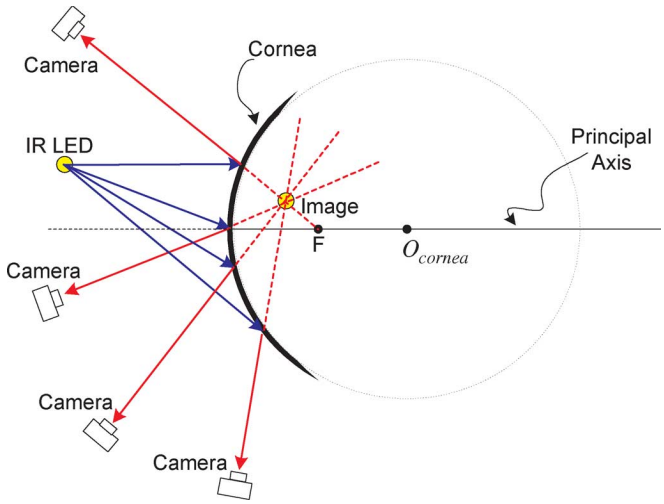


Fig. 3. Image formation of a point light source in the cornea when the cornea serves as a convex mirror.

B. Derivation of 3-D Cornea Center

1) Glint Formation in Cornea Reflection: When light passes through the eye, the boundary surface of the cornea will act like a reflective surface. Therefore, if a light source is placed in front of the eye, the reflection from the external surface of the cornea will be captured as a very bright spot in the eye image as shown in Fig. 1(a) and (b). This special bright dot is called *glint* and it is the brightest and easiest reflection to detect and track.

In order to understand the formation of the glint, the external surface of the cornea is modelled as a convex mirror with a radius R. Therefore, the eye cornea serves as a convex mirror during the process of glint formation. Specifically, the focus point F, the center of the curvature O_{cornea} and the principal axis are shown in Fig. 3. In our research, the IR LEDs are utilized as the light sources. Therefore, when an IR LED is placed in front of the eye, the cornea will produce a virtual image of the IR LED, which is located somewhere behind the cornea surface along the line that connects the cornea center and the light, as shown in the light ray diagram of the Fig. 3.

In the light ray diagram [30], the virtual image of the light image is the location in space where it appears that light diverges from. Any observer from any position who is sighting along a line at the image location will view the IR light source as a result of the reflected light. The reflection law of the convex mirrors [31] establishes that the virtual light position is only determined by the actual position of the light and by the location of the mirror, independent of the observer position. Hence, each observer sees a virtual image at the same location regardless of the observer's location. Thus, the task of determining the image location of the IR light source is to determine the location where reflected light intersects. In Fig. 3, several rays of light emanating from the IR light source are shown approaching the cornea and subsequently reflecting. Each ray is extended backwards to a point of intersection—this point of intersection of all extended reflected rays indicates the location of the virtual image of IR light source.

In our research, the cameras are the observers. Therefore, the virtual image of the IR light source created by the cornea will be shown as a glint in the image captured by the camera. If we place two cameras at different locations, each camera will capture a glint corresponding to the same virtual image of the IR light source in space as shown in Fig. 4. Therefore, in theory, with the use of two cameras, the 3-D location of the virtual image of the IR light source in space can be recovered.

2) Curvature Center of the Cornea: According to the properties of the convex mirror, an incident ray that is directed towards the center of curvature of a mirror is reflected back along its own path (since it is normally incident on the mirror). Therefore, as shown in Fig. 5, if the light ray L_1L_1' is shone directly towards the center of the curvature of the cornea O_{cornea}, it will be reflected back along its own path. Also, the virtual image of the IR light source L_1' will lie in this path. Therefore, as shown in Fig. 5, the IR light source L_1, its the virtual image L_1' and the curvature center of the cornea O_{cornea} will be co-linear.

Further, if we place another IR light source at a different place L_2 as shown in Fig. 5, then the IR light source $L2$, its the virtual image L_2' and the curvature center of the cornea O_{cornea} will lie in another line $L_2L_2'O_{\text{cornea}}$. Line $L_1L_1'O_{\text{cornea}}$ and line $L_2L_2'O_{\text{cornea}}$ will intersect at the point O_{cornea}.

As discussed in Section III-B.1, if two cameras are used, the 3-D locations of the virtual images L_1' and L_2' of the IR light sources can be obtained through 3-D reconstruction. Furthermore, the 3-D location of the IR light sources L_1 and L_2 can be obtained through the system calibration procedure discussed in Section V-A. Therefore, the 3-D location of the curvature center of cornea O_{cornea} can be obtained by intersecting the line L_1L_1' and L_2L_2' as follows:

$$\begin{cases} O_{\text{cornea}} &= L_1 + k_1(L_1 - L_1') \\ O_{\text{cornea}} &= L_2 + k_2(L_2 - L_2') \end{cases}. \tag{1}$$

Note that when more than two IR light sources are available, a set of equations can be obtained, which can lead to a more robust estimation of the 3-D location of cornea center.

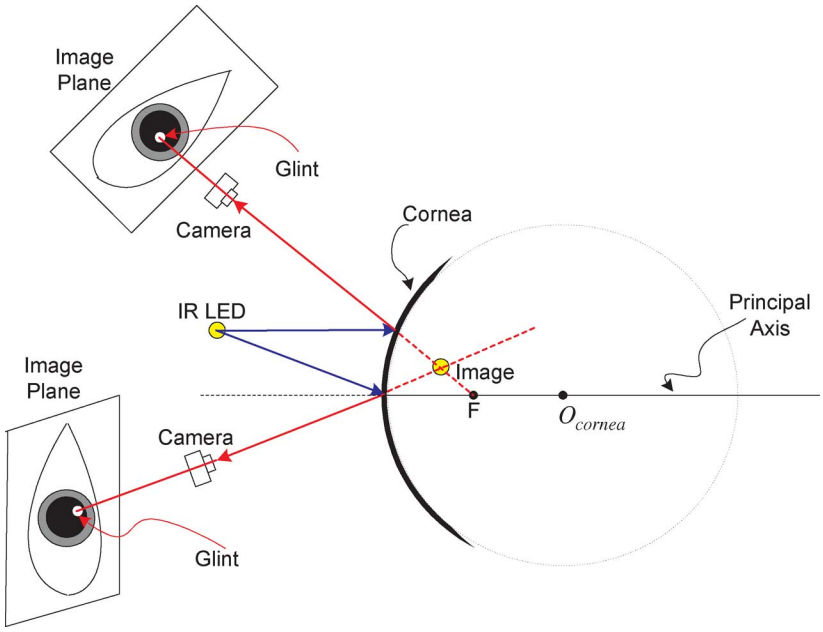


Fig. 4. Ray diagram of the virtual image of the IR light source in front of the cameras.

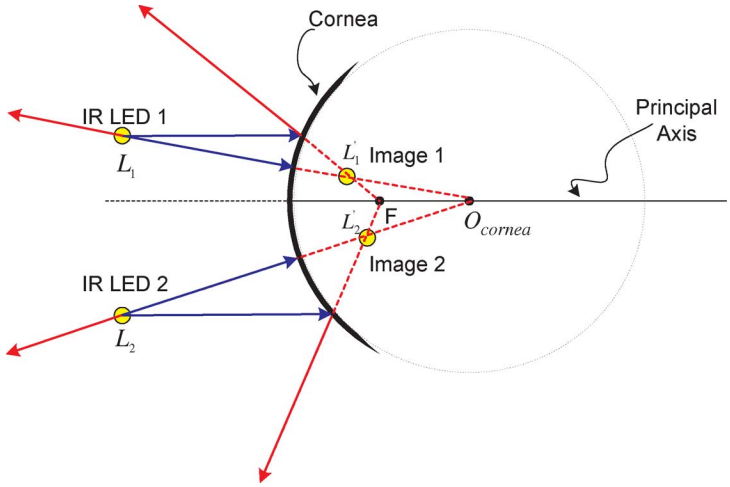


Fig. 5. Ray diagram of two IR light sources in front of the cornea.

C. Computation of 3-D Gaze Direction

1) Estimation of Optic Axis: As discussed earlier, the pupillary axis is the best approximation of the optic axis of the eye. Therefore, after the 3-D location of the pupil center P is extracted, the optic axis V_p of the eye can be estimated by connecting the 3-D pupil center P with cornea center O_{cornea} as follows:

$$V_p = O_{\text{cornea}} + k(P - O_{\text{cornea}}). \tag{2}$$

However, due to the refraction index difference between the air and the aqueous humor inside the cornea, as shown in the geometric ray diagram of the refraction at spherical surface in Fig. 6, we can see that it is the virtual image of the pupil, not the pupil itself being observed from a camera. On the other hand, the pupil center is located in the optic axis of the eye because we assume that the pupillary axis approximates the optic axis of the eye. Therefore, according to the refraction law at spherical surfaces [30], two light rays as shown in Fig. 6 can be used to locate the virtual image of the pupil center, which is still in the optic axis of the eye due to the symmetry of the pupil. As a result, the virtual image of the pupil center P' and the cornea center can be used to estimate the optic axis of the eye directly. Following the same principle as for the virtual position of the light, the virtual image of the pupil is also independent of the camera position. Hence, given two images of the same virtual pupil, its 3-D position can be estimated through a 3-D triangulation.

Since the fovea is invisible from the captured eye images, the visual axis of the eye cannot be estimated directly. Without knowing the visual axis of the eye, the user's fixation point in the 3-D space still cannot be determined. However, the deviation angle *kappa* between the visual axis and the optic axis of the eye is constant for each person.

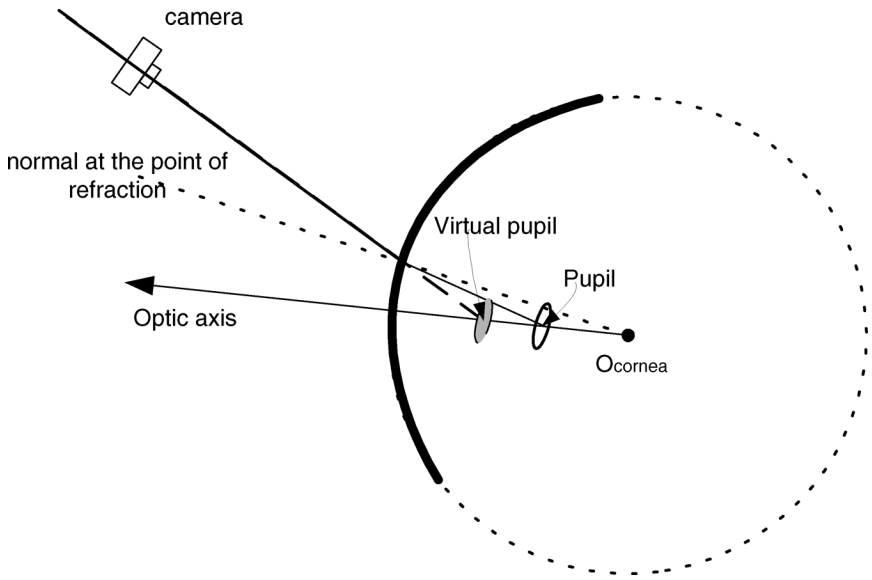


Fig. 6. Ray diagram of virtual pupil image under cornea-air surface refraction.

It is assumed that there is a 3-D coordinate system attached to the eyeball, whose Z axis is the principal axis of the eyeball and X axis is parallel to the horizontal plane in 3-D space. In addition, under the pan (left-right) or tilt (up-down) movement of the head, the X axis of the eye will be always parallel to the horizontal plane. Under the swing (in-plane) movement of the head, the eye will involuntarily perform a torsional movement around the Z axis at the inverse direction of the head swing to keep the X axis parallel to the horizontal plane. However, the eye can only perform small torsional movement. Therefore, if the head swing angle is too large, the X axis of the eye will not be parallel to the horizontal plane.

But normally, large head swing movements are rarely happened, hence, during the eye movements, it can be assumed that the X axis of the eye will be always parallel to the horizontal plane. Therefore, if the deviation angle *kappa* is known, then the visual axis can be computed from the estimated optic axis easily. In the following, a technique is proposed to estimate the deviation angle *kappa* accurately.

2) Compensation of the Angle Deviation Between Visual Axis and Optic Axis: When a user is looking at a known point S in the screen, the 3-D location of the screen point S can be known in that the screen is calibrated. At the same time, the 3-D location of the cornea center O_{cornea} and the 3-D location of the virtual pupil center P' can be computed from the eye images via the proposed technique discussed above. Therefore, the direction of visual axis $\overrightarrow{V_v}$ and the direction of optic axis $\overrightarrow{V_p}$ can be computed as follows:

$$\begin{cases} \overrightarrow{V_v} = \frac{(S-O_{\text{cornea}})}{\|S-O_{\text{cornea}}\|} \\ \overrightarrow{V_p} = \frac{(P'-O_{\text{cornea}})}{\|P'-O_{\text{cornea}}\|} \end{cases}. \tag{3}$$

In addition, let's represent the relationship between the visual axis and the optic axis as follows:

$$\overrightarrow{V_v} = M\overrightarrow{V_p} \tag{4}$$

where M is a 3×3 rotation matrix and it is constructed from the deviation angles between the vectors $\overrightarrow{V_v}$ and $\overrightarrow{V_p}$, or the deviation angle *kappa*. Once the rotation matrix M is estimated, then the 3-D visual axis can be estimated from the extracted 3-D optic axis. Therefore, instead of estimating the deviation angle *kappa* directly to obtain the relationship between the visual axis and the optic axis, it can be encoded through the rotation matrix M implicitly. In addition, the rotation matrix M can be estimated by a simple calibration as follows.

During the calibration, the user is asked to look at a set of k pre-defined point S_i $(i = 1, \ldots, k)$ in the screen, where $k = 9$ during the experiment. After the calibration is done, a set of k pairs of vectors $\overrightarrow{V_v}$ and $\overrightarrow{V_p}$ are obtained via (3). In addition, since the rotation matrix M is an orthonormal matrix, (4) can be represented as

$$\overrightarrow{V_p} = M^T\overrightarrow{V_v}. \tag{5}$$

Therefore, according to (4) and (5), one pair of vectors $\overrightarrow{V_v}$ and $\overrightarrow{V_p}$ can give 6 linear equations so that two screen points are enough to estimate the 3×3 rotation matrix M.

Once the rotation matrix M is estimated, the visual axis of the eye can be estimated from the computed optic axis $\overrightarrow{V_p}$ through (4). Finally, an accurate point of regard of the user can be computed by intersecting the estimated 3-D visual axis with any object in the scene.

IV. 2-D Mapping-Based Gaze Estimation Technique

A. Specific Eye Gaze Mapping Function

Most commercially available remote eye gaze trackers are built from the PCCR technique. Specifically, the PCCR-based technique consists of two major components: pupil-glint vector extraction and gaze mapping function acquisition. Both components will be discussed briefly as follows.

1) Pupil-Glint Vector Extraction: Gaze estimation starts with pupil-glint vector extraction. After grabbing the eye image from

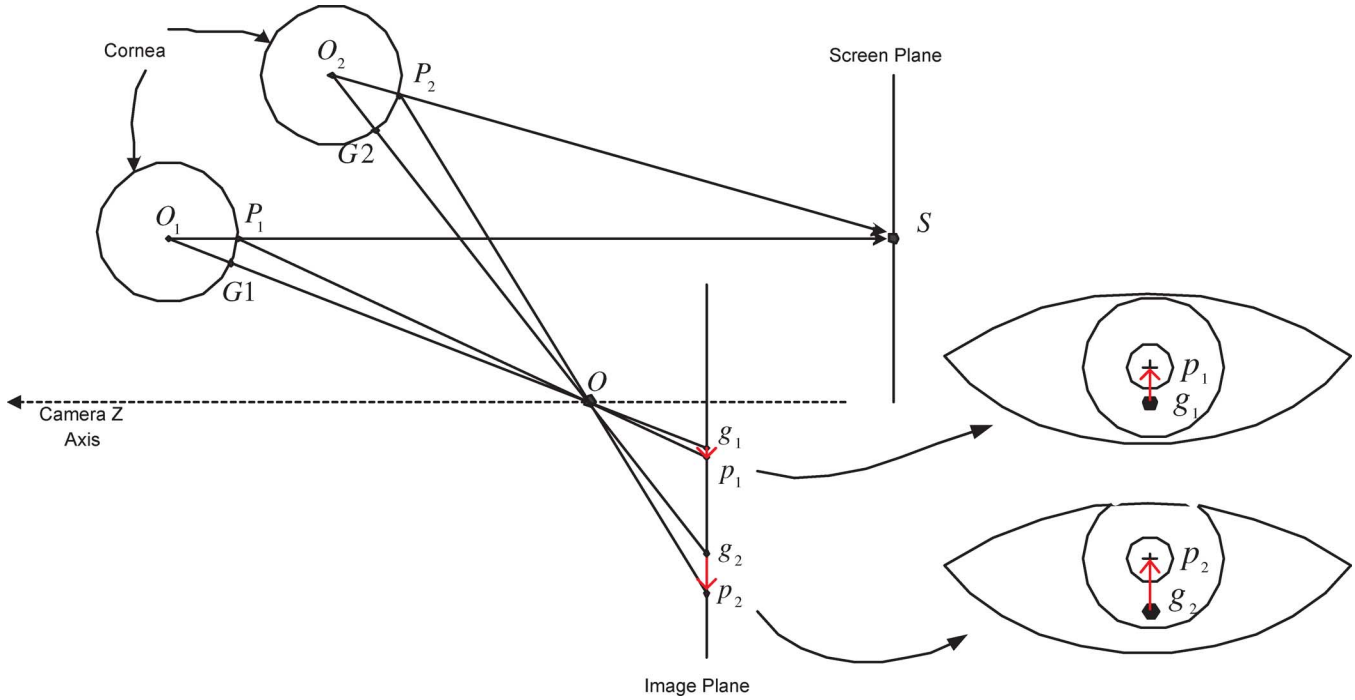


Fig. 7. Pupil and glint image formations when eyes are located at different positions while gazing at the same screen point (side view) and the captured eye images are shown on the right side.

the camera, computer vision techniques [3], [32] are proposed to extract the pupil center and the glint center robustly and accurately. The pupil center and the glint center are connected to form a 2-D pupil-glint vector v as shown in Fig. 7.

2) Specific Gaze Mapping Function Acquisition: After obtaining the pupil-glint vectors, a calibration procedure is proposed to acquire a specific gaze mapping function that will map the extracted pupil-glint vector to the user's fixation point in the screen at the current head position. The extracted pupil-glint vector v is represented as (v_x, v_y) and the screen gaze point S_s is represented by $(x_{\text{gaze}}, y_{\text{gaze}})$ in the screen coordinate system. The specific gaze mapping function $S_s = f(v)$ can be modelled by the following nonlinear equations [23]:

$$\left.\begin{aligned} x_{\text{gaze}} &= a_0 + a_1 * v_x + a_2 * v_y + a_3 * v_x * v_y \\ y_{\text{gaze}} &= b_0 + b_1 * v_x + b_2 * v_y + b_3 * {v_y}^2 \end{aligned}\right\} \quad (6)$$

where the $(a_3 * v_x * v_y)$ term in x_{gaze} and $(b_3 * {v_y}^2)$ term in y_{gaze} amount to scale factors on v_x and v_y such that the gains of v_x on x_{gaze} and v_y on y_{gaze} vary as a function of how high (v_y) on the screen the user is looking. Within the range of the computer screen, these two nonlinear terms accommodate most of the nonlinearity associated with the tilt of the screen (In reality, the camera is mounted below the monitor and a significant nonlinearity is introduced by the screen being tilted with respect to the camera Z axis). The $(b_3 * {v_y}^2)$ term in y_{gaze} also accommodates flattening of the corneal surface toward the edges, which is typical in the human eye.

The coefficients a_0, a_1, a_2, a_3 and b_0, b_1, b_2, b_3 are estimated from a set of pairs of pupil-glint vectors and the corresponding screen gaze points. These pairs are collected in a calibration procedure. During the calibration, the user is required to visually follow a shining dot as it displays at several predefined locations on the computer screen. In addition, the subject must keep his head as still as possible.

If the user does not move his head significantly after the gaze calibration, the calibrated gaze mapping function can be used to estimate the user's gaze point on the screen with high accuracy, based on the extracted pupil-glint vector. But when the user moves his head away from the position where the gaze calibration is performed, the calibrated gaze mapping function will fail to estimate the gaze point because of the pupil-glint vector changes caused by the head movements. In the following section, head movement effect on the pupil-glint vector will be illustrated.

B. Head Motion Effect on Pupil-Glint Vector

Fig. 7 shows the ray diagram of the pupil-glint vector generation in the image when an eye is located at two different 3-D positions O_1 and O_2 in front of the camera due to head movement. For simplicity, the eye is represented by a cornea, the cornea is modelled as a convex mirror, and the IR light source used to generate the glint is located at O. In addition, the eye cornea is further represented by a virtual sphere whose surface goes through the virtual pupil center P, which functions exactly as a real eye cornea. All of which are applicable to subsequent figures in this paper. Assume that the origin of the camera is located at O, p_1 and p_2 are the pupil centers and g_1 and g_2 are the glint centers generated in the image. Further, at both positions, the user is looking at the same point of the computer screen S. According to the light ray diagram shown in Fig. 7, the generated pupil-glint vectors $\overrightarrow{g_1p_1}$ and $\overrightarrow{g_2p_2}$ will be significantly different in the images, as shown in Fig. 7. Two factors are responsible for this pupil-glint vector difference: first, the eyes are at different positions in front of the camera; second, in order to look at the

same screen point, eyes at different positions rotate themselves differently.

The eye will move as the head moves. Therefore, when the user is gazing at a fixed point on the screen while moving his head in front of the camera, a set of pupil-glint vectors in the image will be generated. These pupil-glint vectors are significantly different from each other. If uncorrected, inaccurate gaze points will be estimated after inputting them into a calibrated specific gaze mapping function obtained at a fixed head position.

Therefore, the head movement effects on these pupil-glint vectors must be eliminated in order to utilize the specific gaze mapping function to estimate the screen gaze points correctly. In the following section, a technique is proposed to eliminate the head movement effects from these pupil-glint vectors. With this technique, accurate gaze screen points can be estimated under natural head movement.

C. Dynamic Head Compensation Model

1) Approach Overview: The first step of our technique is to find a specific gaze mapping function f_{O_1} between the pupil-glint vector v_1 and the screen coordinate S at a reference 3-D eye position O_1. This is usually achieved via a gaze calibration procedure using (6). The function f_{O_1} can be expressed as follows:

$$S = f_{O_1}(v_1). \tag{7}$$

Assume that when the eye moves to a new position O_2 as the head moves, a pupil-glint vector v_2 will be generated in the image while the user is looking at the same screen point S. When O_2 is far from O_1, v_2 will be significantly different from v_1. Therefore, v_2 cannot be used as the input of the gaze mapping function f_{O_1} to estimate the screen gaze point due to the changes of the pupil-glint vector caused by the head movement. If the changes of the pupil-glint vector v_2 caused by the head movement can be eliminated, then a corrected pupil-glint vector v_2' will be obtained. Ideally, this corrected pupil-glint vector v_2' is the generated pupil-glint vector v_1 of the eye at the reference position O_1 when gazing at the same screen point S. Therefore, this is equivalent to finding a head mapping function g between two different pupil-glint vectors at two different head positions when still gazing at the same screen point. This mapping function g can be written as follows:

$$v_2' = g(v_2, O_2, O_1) \tag{8}$$

where v_2' is the equivalent measurement of v_1 with respect to the initial reference head position O_1. Therefore, the screen gaze point can be estimated accurately from the pupil-glint vector v_2' via the specific gaze mapping function f_{O_1} as follows:

$$S = f_{O_1}(g(v_2, O_2, O_1)) = F(v_2, O_2) \tag{9}$$

where the function F can be called as a generalized gaze mapping function that explicitly accounts for the head movement. It provides the gaze mapping function dynamically for a new eye position O_2.

Via the proposed technique, whenever the head moves, a gaze mapping function at each new 3-D eye position can be updated automatically; therefore, the issue of the head movement can be solved nicely.

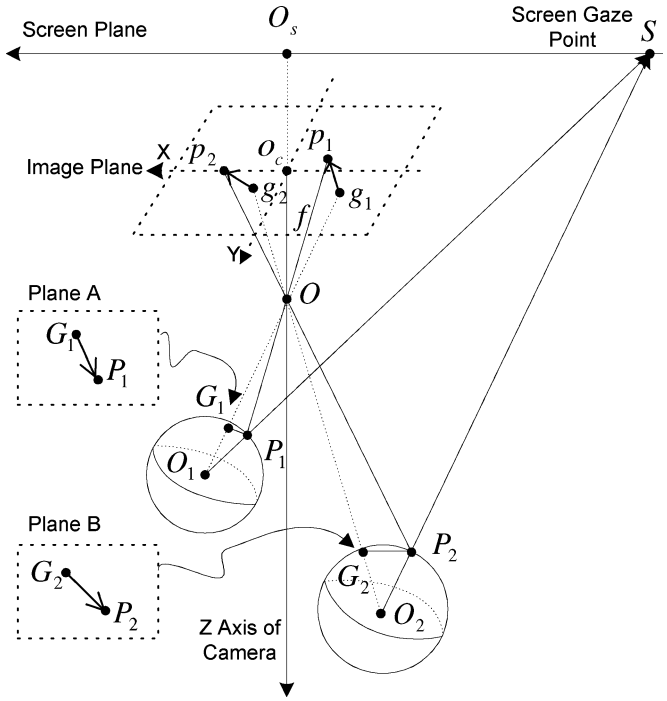


Fig. 8. Pupil and glint image formation when the eye is located at different positions in front of the camera.

2) Image Projection of Pupil-Glint Vector: In this section, we show how to find the head mapping function g. Fig. 8 shows the process of the pupil-glint vector formation in the image for an eye in front of the camera. When the eye is located at two different positions O_1 and O_2 while still gazing at the same screen point S, two different pupil-glint vectors $\overrightarrow{g_1p_1}$ and $\overrightarrow{g_2p_2}$ are generated in the image. Further, as shown in Fig. 8, a plane A parallel to the image plane that goes through the point P_1 will intersect the line O_1O at G_1.[1] Another plane B parallel to the image plane that goes through the point P_2 will intersect the line O_2O at G_2.[2] Therefore, $\overrightarrow{g_1p_1}$ is the projection of the vector $\overrightarrow{G_1P_1}$ and $\overrightarrow{g_2p_2}$ is the projection of the vector $\overrightarrow{G_2P_2}$ in the image plane. Because plane A, plane B and the image plane are parallel, the vectors $\overrightarrow{g_1p_1}$, $\overrightarrow{g_2p_2}$, $\overrightarrow{G_1P_1}$ and $\overrightarrow{G_2P_2}$ can be represented as 2-D vectors in the X–Y plane of the camera coordinate system.

Assume that in the camera coordinate system, the 3-D virtual pupil centers P_1 and P_2 are represented as (x_1, y_1, z_1) and (x_2, y_2, z_2), the glint centers g_1 and g_2 are represented as $(x_{g_1}, y_{g_1}, -f)$ and $(x_{g_2}, y_{g_2}, -f)$, where f is focus length of the camera, and the screen gaze point S is represented by (x_s, y_s, z_s). Via the pinhole camera model, the image projection of the pupil-glint vectors can be expressed as follows:

$$\overrightarrow{g_1p_1} = -\frac{f}{z_1} * \overrightarrow{G_1P_1} \tag{10}$$

$$\overrightarrow{g_2p_2} = -\frac{f}{z_2} * \overrightarrow{G_2P_2}. \tag{11}$$

[1] G_1 is not the actual virtual image of the IR light source.

[2] G_2 is not the actual virtual image of the IR light source.

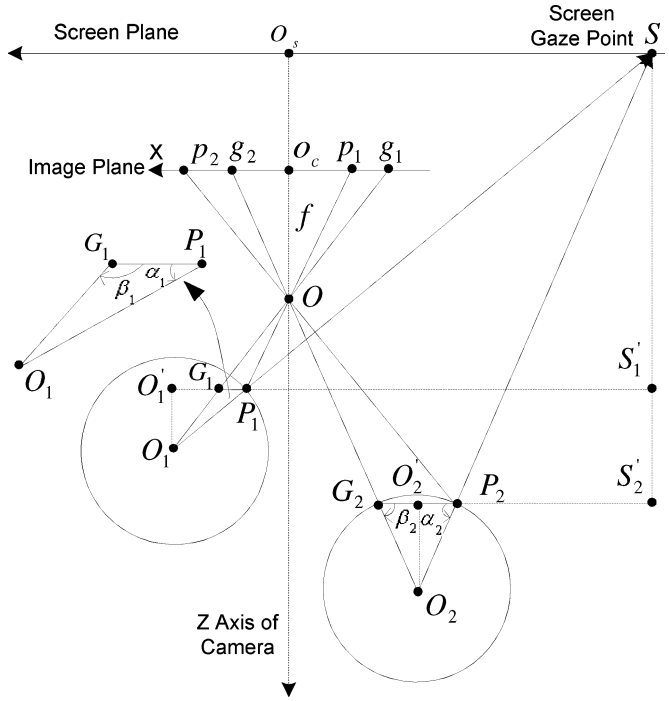


Fig. 9. Pupil and glint image formation when the eye is located at different positions in front of the camera (top-down view).

Assume that the pupil-glint vectors $\overrightarrow{g_1p_1}$ and $\overrightarrow{g_2p_2}$ are represented as (v_{x1}, v_{y1}) and (v_{x2}, v_{y2}) respectively, and the vectors $\overrightarrow{G_1P_1}$ and $\overrightarrow{G_2P_2}$ are represented as (V_{x1}, V_{y1}) and (V_{x2}, V_{y2}) respectively. Therefore, the following equation can be derived by combining the (10) and (11):

$$v_{x1} = \frac{V_{x1}}{V_{x2}} * \frac{z_2}{z_1} * v_{x2} \tag{12}$$

$$v_{y1} = \frac{V_{y1}}{V_{y2}} * \frac{z_2}{z_1} * v_{y2}. \tag{13}$$

The above two equations describe how the pupil-glint vector changes as the head moves in front of the camera. Also, based on the above equations, it is obvious that each component of the pupil-glint vector can be mapped individually. Therefore, (12) for the X component of the pupil-glint vector will be derived first as follows.

3) Case One: The Cornea Center and the Pupil Center Lie on the Camera's X–Z Plane: Fig. 9 shows the ray diagram of the pupil-glint vector formation when the cornea center and pupil center of an eye happen to lie on the $X-Z$ plane of the camera coordinate system. Therefore, either the generated pupil-glint vectors $\overrightarrow{p_1g_1}$ and $\overrightarrow{p_2g_2}$ or the vectors $\overrightarrow{P_1G_1}$ and $\overrightarrow{P_2G_2}$ can be represented as one dimensional vectors, specifically, $\overrightarrow{p_1g_1} = v_{x1}$, $\overrightarrow{p_2g_2} = v_{x2}$, $\overrightarrow{P_1G_1} = V_{x1}$ and $\overrightarrow{P_2G_2} = V_{x2}$.

According to Fig. 9, the vectors $\overrightarrow{G_1P_1}$ and $\overrightarrow{G_2P_2}$ can be represented as follows:

$$\overrightarrow{G_1P_1} = \overrightarrow{G_1O_1'} + \overrightarrow{O_1'P_1} \tag{14}$$

$$\overrightarrow{G_2P_2} = \overrightarrow{G_2O_2'} + \overrightarrow{O_2'P_2}. \tag{15}$$

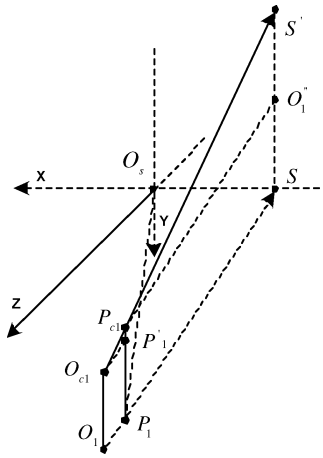


Fig. 10. Projection into camera's $X-Z$ plane.

For simplicity, r_1 is used to represent the length of $\overrightarrow{O_1P_1}$, r_2 is used to represent the length of $\overrightarrow{O_2P_2}$, $\angle G_1P_1O_1$ is represented as α_1, $\angle G_2P_2O_2$ is represented as α_2, $\angle P_1G_1O_1$ is represented as β_1 and $\angle P_2G_2O_2$ is represented as β_2. According to the geometries shown in Fig. 9, the vectors $\overrightarrow{G_1P_1}$ and $\overrightarrow{G_2P_2}$ can be further achieved as follows:

$$\overrightarrow{G_1P_1} = -\frac{r_1 * \sin(\alpha_1)}{\tan(\beta_1)} - r_1 * \cos(\alpha_1) \tag{16}$$

$$\overrightarrow{G_2P_2} = -\frac{r_2 * \sin(\alpha_2)}{\tan(\beta 2)} - r_2 * \cos(\alpha_2). \tag{17}$$

As shown in Fig. 9, line G_1P_1 and line G_2P_2 are parallel to the X axis of the camera. Therefore, $\tan(\beta_1)$ and $\tan(\beta_2)$ can be obtained from the rectangles g_1o_cO and g_2o_cO individually as follows:

$$\tan(\beta_1) = \frac{f}{\overrightarrow{o_cg_1}} \tag{18}$$

$$\tan(\beta_2) = \frac{f}{\overrightarrow{o_cg_2}}. \tag{19}$$

In the above equation, g_1 and g_2 are the glints in the image, and o_c is the principal point of the camera. For simplicity, we choose x_{g_1} to represent $\overrightarrow{o_cg_1}$ and x_{g_2} to represent $\overrightarrow{o_cg_2}$. Therefore, after detecting the glints in the image, $\tan(\beta_1)$ and $\tan(\beta_2)$ can be obtained accurately.

Further, $\sin(\alpha_1)$, $\cos(\alpha_1)$, $\sin(\alpha_2)$ and $\cos(\alpha_2)$ can be obtained from the geometries of the rectangles P_1SS_1' and P_2SS_2' directly. Therefore, (16) and (17) can be derived as follows:

$$V_{x1} = r_1 * \frac{(z_s - z_1) * x_{g_1}}{P_1S * f} + r_1 * \frac{(x_s - x_1)}{P_1S} \tag{20}$$

$$V_{x2} = r_2 * \frac{(z_s - z_2) * x_{g_2}}{P_2S * f} + r_2 * \frac{(x_s - x_2)}{P_2S}. \tag{21}$$

4) Case Two: The Cornea Center and the Pupil Center do Not Lie on the Camera's X–Z Plane: In fact, the cornea center and the pupil center do not always lie on the camera's $X-Z$ plane. However, we can obtain the ray diagram shown in Fig. 9 by projecting the ray diagram in Fig. 8 into $X-Z$ plane along the Y axis of the camera's coordinate system. Therefore, as shown in Fig. 10, point P_1 is the projection of the pupil center P_{c1}, point

O_1 is the projection of the cornea center O_{c1}, and point S is also the projection of the screen gaze point S' in the $X-Z$ plane. Starting from O_{c1}, a parallel line $O_{c1}P_1{}'$ of line O_1P_1 intersects with line $P_{c1}P_1$ at $P_1{}'$. Also starting from P_{c1}, a parallel line $P_{c1}O_1{}''$ of line P_1S intersects with line SS' at $O_1{}''$.

Because $O_{c1}P_{c1}$ represents the distance r between the pupil center to the cornea center, which will not change as the eyeball rotates, O_1P_1 can be derived as follows:

$$r_1 = O_1P_1 = r * \frac{P_1S}{\sqrt{P_1S^2 + (y_1 - y_s)^2}}. \tag{22}$$

Therefore, when the eye moves to a new location O_2 as shown in Fig. 9, O_2P_2 can be represented as follows:

$$r_2 = O_2P_2 = r * \frac{P_2S}{\sqrt{P_2S^2 + (y_2 - y_s)^2}}. \tag{23}$$

After substituting the formulations of r_1 and r_2 into (20) and (21), we can obtain V_{x1}/V_{x2} as follows:

$$\frac{V_{x1}}{V_{x2}} = d * \frac{[(z_s - z_1) * x_{g_1} + (x_s - x_1) * f]}{[(z_s - z_2) * x_{g_2} + (x_s - x_2) * f]} \tag{24}$$

where d is set as follows:

$$d = \frac{\sqrt{(z_2 - z_s)^2 + (x_2 - x_s)^2 + (y_2 - y_s)^2}}{\sqrt{(z_1 - z_s)^2 + (x_1 - x_s)^2 + (y_1 - y_s)^2}}.$$

As a result, (12) and (13) can be finally obtained as follows:

$$v_{x1} = d * \frac{[(z_s - z_1) * x_{g_1} + (x_s - x_1) * f]}{[(z_s - z_2) * x_{g_2} + (x_s - x_2) * f]} * \frac{z_2}{z_1} * v_{x2} \tag{25}$$

$$v_{y1} = d * \frac{[(z_s - z_1) * y_{g_1} + (y_s - y_1) * f]}{[(z_s - z_2) * y_{g_2} + (y_s - y_2) * f]} * \frac{z_2}{z_1} * v_{y2}. \tag{26}$$

The above equations constitute the head mapping function g between the pupil-glint vectors of the eyes at different positions in front of the camera, while gazing at the same screen point.

5) Iterative Algorithm for Gaze Estimation: Equations (25) and (26) require the knowledge of gaze point $S = (x_s, y_s, z_s)$ on the screen. However, the gaze point S is the one that needs to be estimated. As a result, the gaze point S is also a variable of the head mapping function g, which can be further expressed as follows:

$$v_2{}' = g(v_2, P_2, P_1, S). \tag{27}$$

Assume that a specific gaze mapping function f_{P_1} is known via the calibration procedure described in Section IV-A.2. Therefore, after integrating the head mapping function g into the specific gaze mapping function f_{P_1} via (9), the generalized gaze mapping function F can be recursively rewritten as follows:

$$S = F(v_2, P_2, S). \tag{28}$$

Given the extracted pupil-glint vector v_2 from the eye image and the new location P_2 that the eye has moved to, (28) becomes a recursive function. An iterative solution is proposed to solve it.

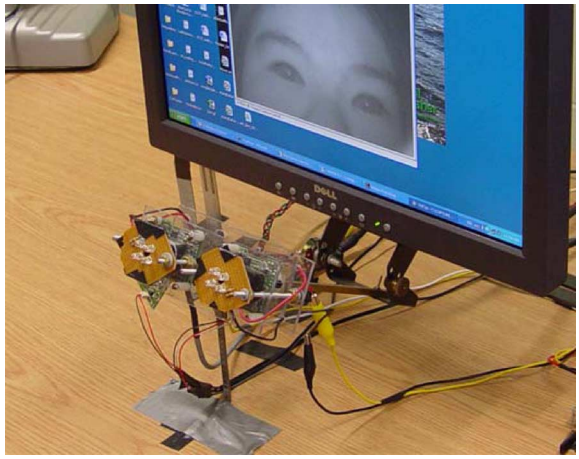


Fig. 11. Configuration of the gaze tracking system.

First, the screen center S_0 is chosen as an initial gaze point, then a corrected pupil-glint vector $v_2{}'$ can be obtained from the detected pupil-glint vector v_2 via the head mapping function g. By inputting the corrected pupil-glint vector $v_2{}'$ into the specific gaze mapping function f_{P_1}, a new screen gaze point S' can be estimated. S' is further used to compute a new corrected pupil-glint vector $v_2{}'$. The loop continues until the estimated screen gaze point S' does not change any more. Usually, the whole iteration process will converge in less than 5 iterations, which makes the real-time implementation possible.

V. Experiment Results

A. System Setup

Our system consists of two cameras mounted under the computer monitor, as shown in Fig. 11. An IR illuminator is mounted around the center of the camera lens to produce the corneal glint in the eye image.

Before the usage of the system, two steps are performed to calibrate the system. The first step is to obtain the parameters of the stereo camera system, which is obtained through camera calibration [33]. Once the stereo camera system is calibrated, given any point P_i in front of it, the 3-D position $(x_i\ y_i\ z_i)^T$ of P_i can be reconstructed from the image points of P_i in both cameras. The second step is to obtain the 3-D positions of the IR LEDs and the computer screen in the stereo camera system. Since the IR LEDs and the computer screen are located behind the view-field of the stereo camera system, they cannot be observed directly by the stereo camera system. Therefore, similar to [11], [12], a planar mirror with a set of fiducial markers attached to the mirror surface is utilized. With the help of the planar mirror, the virtual images of the IR LEDs and the computer screen reflected by the mirror can be observed by the stereo camera system. Thus, the 3-D locations of the IR LEDs and the computer screen can be calibrated after obtaining the 3-D locations of their virtual images. In the following sections, experiment results of both gaze tracking techniques will be reported.

B. Performance of 3-D Gaze Tracking Technique

1) Gaze Estimation Accuracy: Once the system is calibrated and the angle deviation between the visual axis and the optic

TABLE I
GAZE ESTIMATION ACCURACY FOR THE FIRST SUBJECT

Session	Horizontal accuracy (λ, σ)	Vertical accuracy (λ, σ)	Distance to the camera
1	5.02±2.03 mm ($0.72^\circ \pm 0.29^\circ$)	6.40±2.35 mm ($0.92^\circ \pm 0.34^\circ$)	280 mm
2	7.20±2.81 mm ($0.92^\circ \pm 0.40^\circ$)	9.63±3.02 mm ($1.22^\circ \pm 0.43^\circ$)	320 mm
3	9.74±3.23 mm ($1.24^\circ \pm 0.46^\circ$)	13.24±3.69 mm ($1.68^\circ \pm 0.53^\circ$)	370 mm
4	12.47±3.78 mm ($1.37^\circ \pm 0.54^\circ$)	17.30±4.35 mm ($1.90^\circ \pm 0.62^\circ$)	390 mm
5	19.60±5.06 mm ($1.97^\circ \pm 0.73^\circ$)	24.32±6.97 mm ($2.45^\circ \pm 0.99^\circ$)	440 mm

axis for a new user is obtained, his gaze point on the screen can be determined by intersecting the estimated 3-D visual axis of the eye with the computer screen. In order to test the accuracy of the gaze tracking system, seven users were involved in the experiments and none of them wears glasses.

Personal calibration is needed for each user before using our gaze tracking system in order to obtain the angle deviation of the visual axis and the optic axis. The calibration procedure described in Section III-C.2 is very fast and only lasts for less than 5 s. Once the calibration is done, the user does not need to do the calibration any more if he wants to use the system later.

During the experiments, a marker will display at nine fixed locations in the screen randomly, and the user is asked to gaze at the marker when it appears at each location. The experiment contains five 1-minute sessions. At each session, the user is required to position his head at a different position purposely. Table I summarizes the computed gaze estimation accuracy for the first subject, where the last column represents the average distance from the user to the camera during each session. As shown in Table I, the accuracy of the gaze tracker (mean λ and standard deviation σ) significantly depends on the user's distance to the camera. Normally, as the user moves closer to the camera, the gaze accuracy will increase dramatically. This is because the resolution of the eye image increases as the user moves closer to the camera. Also, the vertical accuracy is lower than the horizontal accuracy due to lower vertical image resolution (480 pixels) as compared to horizontal resolution (640 pixels). Besides image resolution, the short focus length with our current camera further limits the range of the subject to the camera. The gaze accuracy can be improved with a zoomable lens and a high resolution camera.

Table II summarizes the computed average gaze estimation accuracy for all the seven subjects during the experiments. During the experiments, the allowed head movement volume is around 200 mm in the X, Y, and Z directions respectively centered at approximately 350 mm to the camera. On the other hand, the head movement volume is mostly limited by the distance from the user to the camera. When the user moves closer to the camera, the allowed head movement volume along the X and Y directions will get smaller, but with a higher gaze accuracy. When the user moves away from the camera, the allowed head movement volume along the X and Y directions will become larger, but the gaze accuracy will decrease significantly. Therefore, when the user has the largest allowed head movement, which is around 200 mm along the X and Y directions, it will also produce the largest gaze estimation error, which is around 2° normally.

TABLE II
AVERAGE GAZE ESTIMATION ACCURACY FOR SEVEN SUBJECTS

Subject	Horizontal accuracy	Vertical accuracy
1	1.24°	1.63°
2	1.28°	1.70°
3	1.33°	1.74°
4	1.39°	1.79°
5	1.43°	1.87°
6	1.66°	2.05°
7	1.97°	2.32°

In summary, within the allowed head movement volume, the average horizontal angular gaze accuracy is 1.47° and the average vertical angular gaze accuracy is 1.87° for all these seven users, which is acceptable for many Human Computer Interaction (HCI) applications, allowing natural head movement.

2) Comparison With Other Systems: Table III shows the comparison of accuracy and allowable head movements among several practically working gaze tracking systems that allow natural head movements. In addition, all of these systems were built recently and require only a very simple personal calibration instead of a tedious gaze mapping function calibration. For simplicity, only the depth or Z direction of the allowed head movement is illustrated, as shown in the second column of Table III. We can see that our proposed technique can provide a competitive gaze accuracy as well as a large head movement volume with only one stereo camera system and without the help of a face tracking system. Therefore, it represents the state of the art in the gaze tracking research under natural head movements.

C. Performance of 2-D Mapping Based Gaze Tracking Technique

1) Head Compensation Model Validation: For the proposed 2-D mapping-based gaze tracking technique, (25) and (26) of the head mapping function g are validated first by the following experiments.

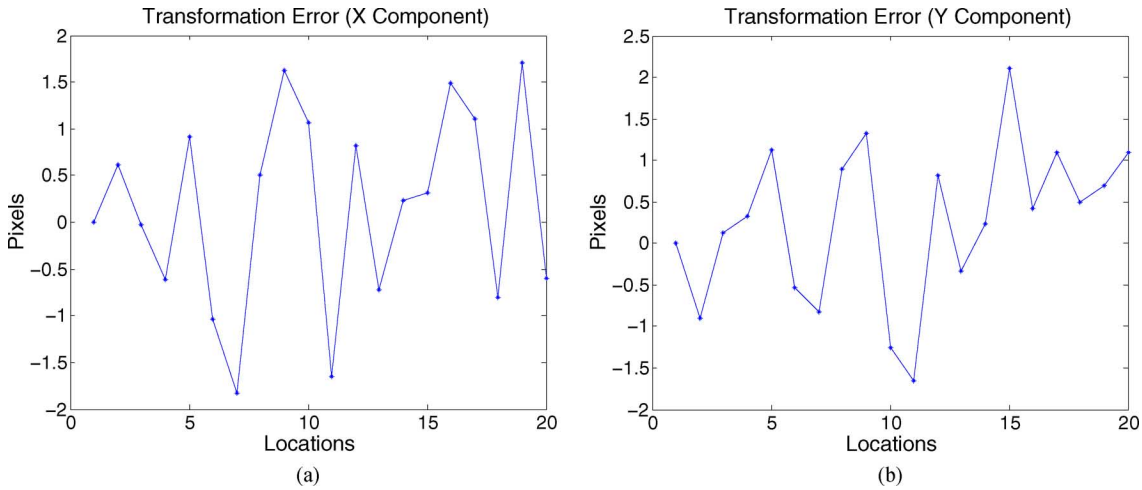
A screen point $S_c = (132.75, -226.00, -135.00)$ is chosen as the gaze point. The user gazes at this point from twenty different locations in front of the camera; at each location, the pupil-glint vector and the 3-D pupil center are collected. The 3-D pupil centers and the pupil-glint vectors of the first two samples P_1, P_2 are shown in Table IV, where P_1 serves as the

TABLE III
COMPARISON WITH OTHER SYSTEMS

Methods	Head movement volume (Z)	Average accuracy	Features
[12]	< 70 mm	0.8°	1 stereo cameras, eye tracking
[11]	N/A, but > 70 mm	0.6° (only one person)	2 stereo cameras, face & eye tracking
Ours	around 200 mm	1.6°	1 stereo cameras, eye tracking
[3]	around 500 mm	5°	single camera, eye tracking
[17]	around 40 mm	0.9°	single camera and two lights

TABLE IV
PUPIL-GLINT VECTOR COMPARISON AT DIFFERENT EYE LOCATIONS

3D pupil position (mm)	2D pupil-glint vector (pixel)	Transformed pupil-glint vector (pixel)
P_1(5.25, 15.56, 331.55)	(9.65, -16.62)	(9.65, -16.62)
P_2(-8.13, 32.29, 361.63)	(7.17, -13.33)	(8.75, -16.01)



(a) (b)

Fig. 12. Pupil-glint vector transformation errors. (a) Transformation error on the X component of the pupil-glint vector. (b) Transformation error on the Y component of the pupil-glint vector.

reference position. The second column indicates the original pupil-glint vectors, while the third column indicates the transformed pupil-glint vectors by the head compensation model. The difference between the transformed pupil-glint vector of P_2 and the reference pupil-glint vector at P_1 is defined as the transformation error.

Fig. 12 illustrates the transformation errors for all these twenty samples. It is observed that the average transformation error is only around 1 pixel, which validates our proposed head compensation model.

2) Gaze Estimation Accuracy: In order to test the accuracy of the gaze tracking system, seven users were asked to participate in the experiment.

For the first user, the gaze mapping function calibration was performed when the user was sitting approximately 330 mm from the camera. After the calibration, the user was asked to stand up for a while. Then, the user was asked to sit approximately 360 mm from the camera and follow a shining object that would display at 12 different pre-specified positions across the screen. The user was asked to reposition his head to a different position before the shining object moved to the next position. Fig. 13 displays the error between the estimated gaze points and the actual gaze points. The average horizontal error is around 4.41 mm in the screen, which corresponds to around 0.51° angular accuracy. The average vertical error is around 6.62 mm in the screen, which corresponds to around 0.77° angular accuracy. Also, it shows that our proposed technique can handle head movements very well.

When the user moves his head away from the camera, the eye in the image will become smaller. Due to the increased pixel measurement error caused by the lower image resolution, the gaze accuracy of the eye gaze tracker will decrease as the user moves away from the camera.

In another experiment, the effect of the distance to the camera on the gaze accuracy of our system is analyzed. The second user was asked to perform the gaze calibration when he was sitting around 360 mm to the camera. After the calibration, the user was positioned at five different locations, which have different distances to the camera as listed in Table V. At each location, the user was asked to follow the moving objects that will display 12

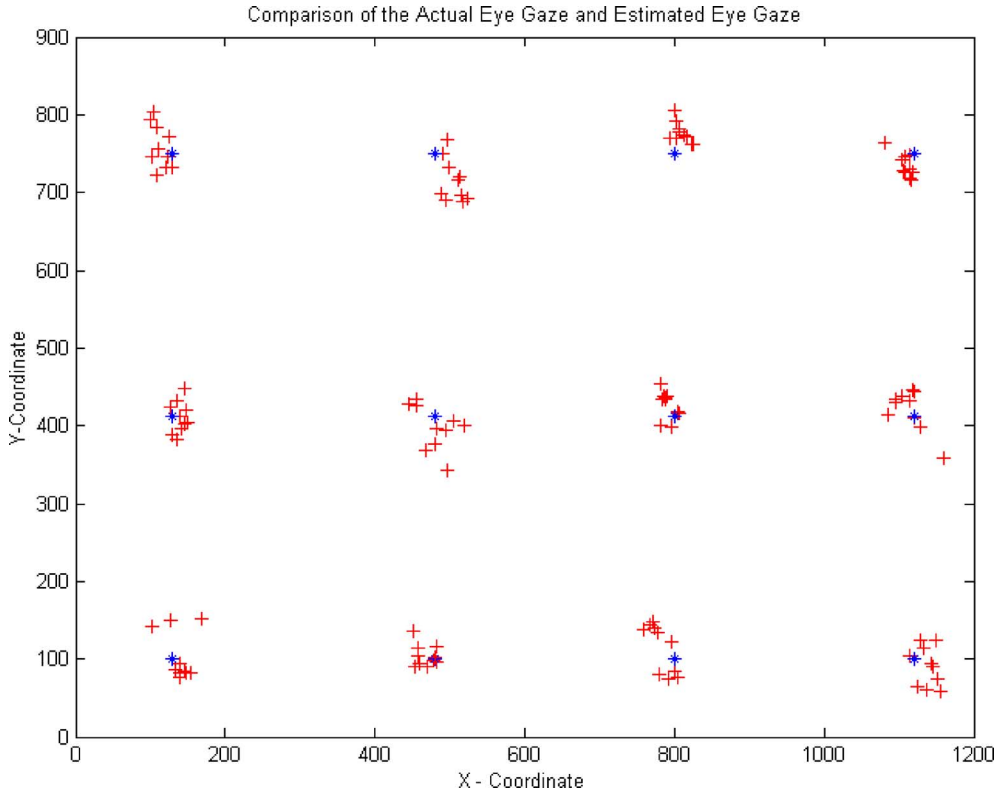


Fig. 13. Plot of the estimated gaze points and the true gaze points, where, "+" represents the estimated gaze point and "∗" represents the actual gaze point.

TABLE V
GAZE ESTIMATION ACCURACY UNDER DIFFERENT EYE IMAGE RESOLUTIONS

Distance to the camera (mm)	Horizontal accuracy (λ, σ)	Vertical accuracy (λ, σ)
300.26	$0.52° \pm 0.19°$	$0.61° \pm 0.22°$
340.26	$0.68° \pm 0.24°$	$0.83° \pm 0.27°$
400.05	$1.31° \pm 0.32°$	$1.41° \pm 0.36°$
462.23	$1.54° \pm 0.52°$	$1.90° \pm 0.60°$
552.51	$1.73° \pm 0.65°$	$2.34° \pm 0.77°$

predefined positions across the screen. Table V lists the gaze estimation accuracy (mean λ and standard deviation σ) at these five different locations, which shows that as the user moves away from the camera, the gaze resolution will decrease. However, within this space volume allowed for the head movement, approximately $200 \times 200 \times 300$ mm (width $\times$ height $\times$ depth) at 450 mm from the camera, the average horizontal angular accuracy is around 1.15° and the average vertical angular accuracy is around 1.40°, which is acceptable for most Human Computer Interaction applications. Also, this space volume allowed for the head movement is large enough for a user to sit comfortably in front of the camera and communicate with the computer naturally.

In order to test the accuracy of the proposed gaze estimation algorithm on the other five subjects, an experiment that contains five 1-min sessions is designed. During the experiment, after the gaze mapping function is calibrated for each user, a marker will display at 12 fixed locations across the screen randomly, and each user is asked to gaze at the marker when it appears at each location. At each session, each user is required to position his head at a different position purposely within the allowed head movement space, approximately $200 \times 200 \times 300$ mm (width$\times$ height $\times$ depth) at 450 mm from the camera. Table VI summarizes the computed average gaze estimation accuracy for the five subjects. Specifically, the average angular gaze accuracy in the horizontal direction and vertical direction is 1.17° and 1.38° respectively for all these five users, allowing natural head movement.

VI. COMPARISON OF BOTH TECHNIQUES

A. System Setup Complexity and Measurement Accuracy

In this section, the differences between the proposed two gaze tracking techniques are discussed. The 2-D mapping-based gaze

TABLE VI
AVERAGE GAZE ESTIMATION ACCURACY FOR FIVE SUBJECTS

Subject	Horizontal accuracy	Vertical accuracy
1	1.12°	1.32°
2	1.16°	1.35°
3	1.17°	1.38°
4	1.20°	1.41°
5	1.22°	1.48°

estimation method does not require knowledge of the 3-D direction of the eye gaze to determine the gaze point on an object; instead, it estimates the gaze point on the object from a gaze-mapping function directly by inputting a set of features extracted from the eye image. The gaze-mapping function is usually obtained through a calibration procedure repeated for each person.

The calibrated gaze mapping function is very sensitive to head motion; consequently, a complicated head-motion compensation model is proposed to eliminate the effect of head motion on the gaze-mapping function. Thus, the 2-D mapping-based method can work under natural head movement. Since the 2-D mapping-based method is proposed to estimate the gaze points on a specific object, a new gaze-mapping function calibration must be performed each time when a different object is presented.

In contrast, the 3-D gaze estimation technique estimates the 3-D direction of the eyeball's visual axis directly, and determines the gaze by intersecting the visual axis with the object in the scene. Thus, it can be used to estimate the gaze point on any object in the scene without the use of tedious gaze-mapping function calibration. Furthermore, since this method is not constrained by head position, the complicated head-motion compensation model can be avoided. But the 3-D technique needs an accurate stereo camera system, and the accuracy of the 3-D gaze estimation technique is affected by the accuracy of the stereo camera system significantly.

In terms of accuracy, the experiments indicate that the 2-D mapping-based gaze estimation technique is more accurate than the 3-D gaze tracking technique. For example, for a user who is sitting approximately 340 mm from the camera, the 2-D mapping-based gaze estimation technique can achieve 0.68° accuracy in the horizontal direction and 0.83° accuracy in the vertical direction; on the other hand, the direct 3-D gaze estimation technique only achieves 1.14° accuracy in the horizontal direction and 1.58° accuracy in the vertical direction. The main source of errors for the 3-D method results from the calibration errors, both with the estimated camera parameters and with the estimated IR light position. Therefore, we can improve the calibration accuracy to improve the 3-D gaze estimation accuracy.

Both gaze tracking techniques proposed in the paper are implemented using C++ on a PC with a Xeon (TM) 2.80 GHz CPU and a 1.00 GB RAM. The image resolution of the cameras is 640×480 pixels, and the built gaze tracking systems can run at approximately 25 fps comfortably.

VII. CONCLUSION

In this paper, based on exploiting eye's anatomy, two novel techniques are proposed to improve the existing gaze tracking techniques. First, a simple direct method is proposed to estimate the 3-D optic axis of a user without using any user-dependent eyeball parameters. The method is more feasible to work on different individuals without tedious calibration. Second, a novel 2-D mapping-based gaze estimation technique is proposed to allow free head movement and minimize the number of personal calibration procedures. Therefore, the eye gaze can be estimated with high accuracy under natural head movement, with the personal calibration being minimized simultaneously. By the novel techniques proposed in this paper, two common drawbacks of the existing eye gaze trackers can be eliminated or minimized nicely so that the eye gaze of a user can be estimated under natural head movement, with minimum personal calibration.

REFERENCES

[1] R. J. K. Jacob, "The use of eye movements in human computer interaction techniques: What you look at is what you get," *ACM Trans. Inf. Syst.*, vol. 9, no. 3, pp. 152–169, 1991.

[2] S. Zhai, C. H. Morimoto, and S. Ihde, "Manual and gaze input cascaded (magic) pointing," in *Proc. ACM SIGHCI-Human Factors Comput. Syst. Conf.*, 1999, pp. 246–253.

[3] Z. Zhu and Q. Ji, "Eye and gaze tracking for interactive graphic display," *Mach. Vis. Applicat.*, vol. 15, no. 3, pp. 139–148, 2004.

[4] S. P. Liversedge and J. M. Findlay, "Saccadic eye movements and cognition," *Trends in Cognitive Science*, vol. 4, no. 1, pp. 6–14, 2000.

[5] M. F. Mason, B. M. Hood, and C. N. Macrae, "Look into my eyes: Gaze direction and person memory," *Memory*, vol. 12, pp. 637–643, 2004.

[6] S. Milekic, "The more you look the more you get: Intention-based interface using gaze-tracking," in *Trant, J.(Des.) Museums and the Web 2002: Selected Papers from an Int. Conf., Archives and Museum Informatics*, 2002, pp. 1–27.

[7] K. Hyoki, M. Shigeta, N. Tsuno, Y. Kawamuro, and T. Kinoshita, "Quantitative electro-oculography and electroencephalography as indexes of alertness," *Electroencephalogr. Clinical Neurophysiol.*, vol. 106, pp. 213–219, 1998.

[8] K. H. Tan, D. Kriegman, and H. Ahuja, "Appearance based eye gaze estimation," in *Proce. IEEE Workshop Applications of Computer Vision*, 2002, pp. 191–195.

[9] J. Zhu and J. Yang, "Subpixel eye gaze tracking," in *Proc. IEEE Inr. Conf. Automatic Face and Gesture Recognition*, Washington, D.C., 2002, pp. 131–136.

[10] C. H. Morimoto and M. Mimica, "Eye gaze tracking techniques for interactive applications," *Comput. Vi. Image Understand., Special Issue on Eye Detection and Tracking*, vol. 98, no. 1, pp. 4–24, 2005.

[11] D. Beymer and M. Flickner, "Eye gaze tracking using an active stereo head," in *Proc. Int. Conf. Computer Vision and Pattern Recognition*, 2003, pp. 451–458.

[12] S. W. Shih and J. Liu, "A novel approach to 3-D gaze tracking using stereo cameras," *IEEE Trans. Syst. Man Cybern. B*, vol. 34, no. 1, pp. 234–245, Feb. 2004.

[13] J. Wang, E. Sung, and R. Venkateswarlu, "Eye gaze estimation from a single image of one eye," in *Proc. Int. Conf. Computer Vision*, 2003, pp. 136–143.

[14] T. Ohno, N. Mukawa, and A. Yoshikawa, "Freegaze: A gaze tracking system for everyday gaze interaction," in *Proc. Symp. ETRA 2002*, 2002, pp. 125–132.

[15] C. H. Morimoto, A. Amir, and M. Flickner, "Detecting eye position and gaze from a single camera and 2 light sources," in *Proc. Int. Conf. Pattern Recognition*, 2002, pp. 314–317.

[16] Y. Matsumoto, T. Ogasawara, and A. Zelinsky, "Behavior recognition based on head pose and gaze direction measurement," in *Proc. 2000 IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2000, pp. 2127–2132.

[17] E. D. Guestrin and M. Eizenman, "General theory of remote gaze estimation using the pupil center and corneal reflections," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 6, pp. 1124–1133, Jun. 2006.

[18] Y. Ebisawa, M. Ohtani, and 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 *Proc. 18th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 1996, pp. 523–525.
[19] C. H. Morimoto, D. Koons, A. Amir, and M. Flickner, "Frame-rate pupil detector and gaze tracker," in *IEEE ICCV'99 Frame-Rate Workshop*, 1999, pp. 1–6.
[20] A. T. Duchowski, *Eye Tracking Methodology: Theory and Practice*. New York: Springer Verlag, 2002.
[21] R. J. K. Jacob and K. S. Karn, "Eye tracking in human-computer interaction and usability research: Ready to deliver the promises," in *The Mind's Eyes: Cognitive and Applied Aspects of Eye Movements*, J. Hyona, R. Radach, and H. deubel, Eds. Oxford, U.K.: Elsevier Science, 2003.
[22] D. H. Yoo and M. J. Chung, "A novel nonintrusive eye gaze estimation using cross-ratio under large head motion," *Comput. Vis. Image Understand., Special Issue on Eye Detection and Tracking*, vol. 98, no. 1, pp. 25–51, 2005.
[23] LC Technologies, Inc., McLean, VA [Online]. Available: http://www.eyegaze.com
[24] T. E. Hutchinson, K. P. White Jr., K. C. Reichert, and L. A. Frey, "Human-computer interaction using eye-gaze input," *IEEE Trans. Syst., Man, Cybern.*, vol. 19, no. 6, pp. 1527–1533, Nov. 1989.
[25] R. J. K. Jacob, *Eye-Movement-Based Human-Computer Interaction Techniques: Towards Non-Command Interfaces*. Norwood, NJ: Ablex, 1993, vol. 4, pp. 151–190.
[26] C. H. Morimoto, D. Koons, A. Amir, and M. Flickner, "Pupil detection and tracking using multiple light sources," *Image Vis. Comput.*, vol. 18, pp. 331–336, 2000.
[27] Applied Science Laboratories [Online]. Available: http://www.a-s-l.com
[28] SensoMotoric [Online]. Available: http://www.smi.de
[29] C. W. Oyster, *The Human Eye: Structure and Function*. Sunderland, MA: Sinauer Associates, Inc., 1999.
[30] M. Katz, *Introduction to Geometrical Optics*. Singapore: World Scientific, 2002.
[31] The Reflection and the Ray Model of Light [Online]. Available: http://www.glenbrook.k12.il.us/gbssci/phys/Class/refln/u13l4a.html
[32] Z. Zhu and Q. Ji, "Robust real-time eye detection and tracking under variable lighting conditions and various face orientations," *Comput. Vis. Image Understand., Special Issue on Eye Detection and Tracking*, vol. 38, no. 1, pp. 124–154, 2005.
[33] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.

Zhiwei Zhu received the B.S. degree in computer science from the University of Science and Technology, Beijing, China, in 2000, the M.S. degree in computer science from University of Nevada, Reno, in 2002, and the Ph.D. degree in electrical and computer engineering from Rensselaer Polytechnic Institute, Troy, NY, in December 2005.

He is currently a Member of Technical Staff in the Computer Vision Laboratory, Sarnoff Corporation in Princeton, NJ. His interests are in computer vision, pattern recognition, image processing and human-computer interaction.

Qiang Ji (SM'03) received the Ph.D. degree in electrical engineering from University of Washington, Seattle, in 1998.

He is currently a tenured Associate Professor at the Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute (RPI), Troy, NY. Previously, he held teaching and research positions at University of Nevada, Carnegie Mellon University, and Air Force Research Laboratory. His research interests are in computer vision, probabilistic reasoning with graphical models, and decision making under uncertainty. Dr. Ji has published over 100 articles in peer-reviewed conferences and journals in these areas.

Dr. Ji has organized and served as PC members for many international conferences and workshops. He is on the editorial boards of image and vision computing journal, and is an associate editor for Pattern Recognition Letters journal and for IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART A: SYSTEMS AND HUMANS. He was a lead Guest Editor for a Special Issue on eye detection and tracking for the *Journal of Computer Vision and Image Understanding*, 2005. He has received several awards including the Research Excellence Award from RPI's School of Engineering, 2006; the Best Paper Award from the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, 2004; the Best Paper Award from IEEE Computer Vision and Pattern Recognition Workshop on Face Recognition Grand Challenge Experiments, 2005; and the Honda Initiation Award, 1998.