A Novel Approach to 3-D Gaze Tracking Using Stereo Cameras

Sheng-Wen Shih and Jin Liu

Sheng-Wen Shih is with the Department of Computer Science and Information Engineering, National Chi Nan University, Nantou 545, Taiwan (e-mail: swshih@ncnu.edu.tw).

Jin Liu is with the T-Systems, T-Systems International GmbH, BERKOM , 10589 Berlin, Germany (e-mail: jin.liu@t-systems.com).

Abstract

A novel approach to 3-D gaze tracking using 3-D computer vision techniques is proposed in this paper. This method employs multiple cameras and multiple point light sources to estimate the optical axis of user's eye without using any user-dependent parameters. Thus, it renders the inconvenient system calibration process which may produce possible calibration errors unnecessary. A real-time 3-D gaze tracking system has been developed which can provide 30 gaze measurements per second. Moreover, a simple and accurate calibration method is proposed to calibrate the gaze tracking system. Before using the system, each user only has to stare at a target point for a few (2–3) seconds so that the constant angle between the 3-D line of sight and the optical axis can be estimated. The test results of six subjects showed that the gaze tracking system is very promising achieving an average estimation error of under 1 degree.

Keywords

Gaze tracking, Eye tracking, Line of Sight, Human-Computer Interface.

I. Introduction

When one stares at a 3-D point, his/her oculomotor mechanism controls his/her eyes such that the image of the point appears at the fovea region. The 3-D LoS (Line of Sight) of the eye is a straight line that passes through the 3D point, the fovea spot and the optical center of the eye. 3-D gaze tracking is to determine and track the 3-D LoS of a person from the appearance of his eyes and has numerous applications in the area of human computer interaction. Hutchinson et al. [1] presented an eye-gaze-response interface (called Erica) as a prosthetic device. Because the gaze input was used to select several large menu boxes, the gaze tracking accuracy requirement is not high; therefore, the calibration process is very simple and can be accomplished within several seconds. Frey et al. developed an text entry system from Erica [2]. Jacob [3] described how to process the gaze tracking results to extract stable fixation points so that people can use eye movement to point and select objects on the screen. Zhai et al. [4] proposed combining both the gaze input and the mouse pointer to control the cursor position efficiently. Sharkey and Murray [5] reported the feasibility of incorporating gaze tracking results to increase resolution for visual telepresence. Pastoor et al. [6] and Redert et al. [7] showed that, by incorporating a gaze tracker, 3-D rendering can be accomplished in a more realistic way. Note that since the width of the fovea is approximately equivalent to a view-field of one degree [3], the angular accuracy of most of the existing gaze tracking methods is also about one degree.

In this paper, a novel method for estimating the 3-D LoS using 3-D computer vision techniques is proposed. The gaze tracking theory in this work developed is from the simplified eye model proposed by Le Grand [8]. It is shown that by using just two cameras and two point light sources located at known positions, the optical axis (OA) of the eye can be determined by solving linear equations. The OA of a human eye is defined to be the 3-D line connecting the center of the pupil and the curvature center of the cornea. If the geometric relation between the cameras and the screen has been calibrated, the OA of the eye, once determined, can then be transformed into the screen coordinate system to obtain the fixation point on the screen. In this way, users no longer have to participate in intensive user-dependent

calibration process as required by other methods. As a result, a sophisticated calibration method can be adopted to calibrate accurately the tracking system to provide a nearly ready-to-use gaze tracker for users who do not wear glasses. The only user-dependent parameter that has to be calibrated by each user is the angle between the OA and the LoS. This has to be determined only once for each user, and it usually takes only 2–3 seconds.

A. Related Works

Existing gaze tracking techniques can be classified into the following three categories.

1. 2-D techniques: Those gaze tracking techniques which cannot provide full information of the 3-D LoS are classified as 2-D techniques. In 2-D gaze tracking systems, the 3-D position of the eye is usually unknown and only the relative orientation of user's eye with respect to user's head is measured. In general, 2-D techniques require users to hold their head very still. Furthermore, if highly accurate gaze tracking results are expected, some auxiliary apparatus, such as head rest, chin rest or bite bar, have to be used in conjunction with the 2-D gaze tracking techniques. 2-D gaze tracking techniques include the electro-oculographic potential (EOG) technique [9], the pupil tracking technique, the artificial neural network technique [10], the pupil and corneal reflection tracking technique [1], [2], [11], the dual Purkinje image tracking technique [12], and the sclera coil (or search coil) technique [13]. Any 2-D gaze tracking system can be extended to a 3-D tracking system, if the absolute 3-D position of the eye (it could be the center of the eyeball, the rotation center of the eyeball, or any other fixed point on the head) can be determined in the system.

2. Model-based 3-D techniques: As implied by the name, these techniques use a 3-D model of some facial feature points to provide 3-D information of the LoS from a monocular image. The gaze tracking system implemented by Stiefelhagen et al. tracks six facial feature points for estimating the head pose and the gaze direction [14]. Collet et al. have also developed a similar gaze tracking system [15]. They proposed to improve the robustness of their facial feature tracking subsystem by introducing a recognition module to verify the tracking results. In order to track the facial features, the view-field of the tracking camera has to be made sufficiently large to cover the entire head of a user. The advantage of this approach is that users are allowed to move their head freely in a specified area, but using a camera with a large field of view yields a lower gaze tracking accuracy. To overcome this drawback, one can use another camera to acquire a high definition image of user's eye.

3. 3-D techniques: Pastoor et al. have implemented a gaze tracking system which consists of a head tracker for tracking the 3-D eye locations and a pan-tilt-zoom camera for tracking the gaze direction using corneal reflection and pupil images [6]. Sugioka et al. have also developed a gaze tracking system with a pan-tilt-zoom camera for gaze tracking [16]. An individual calibration process is required in both algorithms. The main difference between the gaze tracking systems proposed in [6] and [16] is that Sugioka et al. utilized directly an ultrasonic distance measurement device to estimate the distance between the pan-tilt-zoom

camera and the eye. However, the direction of the LoS is still estimated by the 2-D techniques using pupil and corneal reflection as done in [1], [2], [11].

The gaze tracking method proposed in this paper is different from the existing methods in the following ways.

1. A pair of calibrated stereo cameras are used to track the gaze information of one eye while other methods use one camera for tracking one eye.
2. The positions of the LEDs should be calibrated beforehand.
3. Both the position and the orientation of the OA of user's eye can be explicitly determined by solving linear equations. Thus, the gaze tracking results can be used not only for inferring fixation points on the screen, but also for inferring 3-D fixation points in the real world.
4. Because the proposed method employs 3-D computer vision technique, users neither have to keep their head still, nor have to spend a lot of time calibrating the gaze tracking system.

B. Organization of This Paper

The proposed gaze tracking method developed in this paper from the eye model is described in Section II. This method uses the first Purkinje images of point light sources and multiple cameras to determine the curvature center of the cornea (which will be referred to as the cornea center hereafter for convenience) as detailed in Section III. Section IV discusses how the estimated 3-D position of the eye and the observed centers of the pupil images are used to compute the gaze direction. Overview of the gaze tracking system is introduced in Section V. A simple and accurate calibration method for calibrating the gaze tracking system is proposed in Section VI. Experimental results are shown in Section VII. Conclusions are given in Section VIII.

II. Simplified Structure of a Human Eye

The simplified eye model adopted in this work (see Fig. 1) is proposed by Le Grand [8]. In Le Grand's simplified eye model, the cornea and the aqueous humor, which is the clear liquid between the cornea and the lens, are assumed to have the same refraction index. The shape of the cornea is assumed to be a portion of an imaginary sphere whose radius is 8mm. The sclera is a white and opaque membrane that surrounds the cornea and covers the other portion of the eye ball. The surface of the sclera is relatively coarser than that of the cornea, and its curvature is smaller than that of the cornea. The iris is an opaque tissue with a hole, namely, the pupil, on its center. Different species of mankind may have iris of different colors. The diameter of the pupil varies from 2mm to 8mm in response to different ambient light levels and some psychological factors. The most complicated structure in the simplified eye model is the lens. The shape of the lens changes when staring at objects of different distances. The unknown relationship between the lens shape and the object distance leads to difficulties in predicting the reflection and refraction of the light which passes through the lens. In order to avoid more complicated optical

properties introduced by the lens, we will use the reflection light of the first surface of the cornea and the refracted image of the pupil to infer the LoS of the eye. Accordingly, only the reflection and refraction of the cornea have to be considered. The proposed method has made the following assumptions.

1. The cornea surface has a constant curvature $\frac{1}{\rho}$ (i.e., the cornea has a spherical surface with radius ρ).
2. The cornea and the aqueous humor can be modeled as a convex spherical lens having uniform refraction index.

III. Computing Cornea Center from the First Purkinje Image

When a beam of light enters the cornea surface of a human eye, part of the light is reflected by the surface and part of the light is refracted after traveling through the surface. The light which is reflected by the exterior cornea surface is called the first Purkinje image (see Fig. 2(a)). Since the first Purkinje image is a result of specular reflection, observers at different directions will observe different images. Figure 2(b) shows the first Purkinje image of an infrared light emitting diode (LED). LEDs are commonly used in place of point light sources in gaze tracking systems since they are inexpensive.

Let O denote the optical center of a camera and Q denote the position of a point light source. Assume that the camera parameters and the position of point light source relative to the camera coordinate system, denoted by $\overrightarrow{OQ}$, have been accurately calibrated (the calibration process will be discussed in Section VI). Given the 2-D point of the first Purkinje image on the CCD sensor plane, denoted by P_{img}, and the calibrated optical center of the camera, O, the 3-D vector $\overrightarrow{OP_{img}}$ (see Fig. 3) can be computed by back projecting P_{img}. For clarity of deriving a nonlinear constraint and a linear constraint of the location of cornea center, an auxiliary 2-D coordinate system is defined as follows:

- the origin is defined by O,
- the x-axis is defined by ray $\overrightarrow{OQ}$, and
- the y-axis is selected such that the x-y plane contains the vector $\overrightarrow{OP_{img}}$.

A. Nonlinear Constraint for Computing the Cornea Center Using One Camera and One Point Light Source

By the law of reflection in the basic geometric optics that the incident angle equals the reflectional angle, we have $\measuredangle MP_0Q = \measuredangle MP_0O = \frac{\pi-\alpha-\beta}{2}$, which yields $\measuredangle P_0MQ = \frac{\pi+\alpha-\beta}{2}$ (see Fig. 3). Note that the angle, α, can be easily determined since both $\overrightarrow{OP_{img}}$ and $\overrightarrow{OQ}$ are known.

Let the 2-D coordinates of point P_0 be $(x_0,\ y_0)$ where $y_0 = x_0 \tan(\alpha)$ and x_0 is an unknown parameter. Since $\overrightarrow{P_0P_1} = \left[\rho\cos\left(\frac{\pi+\alpha-\beta}{2}\right),\ \rho\sin\left(\frac{\pi+\alpha-\beta}{2}\right)\right]^t$, the 2-D coordinates of the cornea center, i.e., P_1, can be easily derived as follows:

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} x_0 - \rho\sin\left(\frac{\alpha-\beta}{2}\right) \\ x_0\tan(\alpha) + \rho\cos\left(\frac{\alpha-\beta}{2}\right) \end{pmatrix} \tag{1}$$

If the cornea radius, ρ, is known, equation (1) defines a 2-D curve with one independent parameter x_0. This 2-D curve will pass through the cornea center with a specific value of x_0. In the case that ρ is

unknown, equation (1) will become a planar equation determined by two independent parameters, x_0 and ρ. This linear planar constraint will be discussed in detail in the next subsection.

Equation (1) indicates that the 3-D location of the eye cannot be uniquely determined by using only one point light source and one camera. The following proposition states the necessary condition for computing the cornea center with one camera.

Proposition 1: If the cornea radius, ρ, is known, then we will need at least one camera and at least two point light sources with known positions to compute the cornea center.

Proof: With just one camera and a point light source, the solution to equation (1) is a 2-D curve passing through the cornea center. If there are two point light sources, then we will have two 2-D curves derived from the first Purkinje images of both point light sources passing the cornea center. ■

B. Linear Constraint for Computing the Cornea Center Using One Camera and Multiple Point Light Sources

According to the law of reflection, the location of the point light source, the beams of incident and the reflected light, and the normal vector of the reflectional surface are all coplanar. That is, the normal vector $\overrightarrow{P_0M}$ is on the plane QP_0O. Since the cornea surface is spherical and the normal vector of a spherical surface always passes the spherical center, it follows that the cornea center, P_1, is also on the plane QP_0O.

With the location of the point light source and its first Purkinje image recorded by a camera, the linear constraint for computing the location of the cornea center can be formulated by the following plane equation in the camera coordinate system:

$$s \cdot k_c = 0, \tag{2}$$

where the normal vector s is given by the cross-product of $\overrightarrow{OP_{img}}$ and $\overrightarrow{OQ}$, and $k_c = \overrightarrow{OP_1}$ is the vector representing the cornea center. If there are N point light sources, Q_i, $i =$, 1, 2, ..., N and N identified Purkinje images for all point light sources, we will have N plane equations with the form of equation (2). However, the solution to these plane equations is not unique. The following proposition summarizes this property.

Proposition 2: If the optical center of the camera, O, and a set of point light sources, Q_i's, are not collinear, then the solutions to the plane equations in the form of equation (2) is a 3-D line.

Proof: It is obvious that the coefficient matrix of the plane equations is not of full rank because all the planes contains both the optical center of the camera, O, and the cornea center, k_c. Furthermore, if there exist two point light sources, Q_i and Q_j, such that $\overrightarrow{OQ_i}$ and $\overrightarrow{OQ_j}$ are not collinear, then $s_i = \overrightarrow{OP_{img,i}} \times \overrightarrow{OQ_i}$ and $s_j = \overrightarrow{OP_{img,j}} \times \overrightarrow{OQ_j}$ are not collinear; hence, the rank of the coefficient matrix of the plane equations is two. ■

It follows from Proposition 2 that if the cornea radius, ρ, of a user is unknown, gaze tracking systems using one camera can only determine the 3-D position of the cornea center up to a scale factor with

respect to the camera coordinate system. The condition for obtaining a unique solution for the cornea center when the cornea radius is unknown is summarized in the following proposition.

Proposition 3: If the cornea radius, ρ, is unknown, at least two cameras and two point light sources are needed to recover the 3-D position of the cornea center. Furthermore, in order to obtain a unique solution for the 3-D cornea center, the point light sources and at least one of the optical centers of the cameras should not be collinear.

Proof: According to Proposition 2, a 3-D line which contains the optical center of a camera and the cornea center in the camera coordinate system can be determined by using a camera and at least two point light sources where the optical center of the camera and the point light sources are not collinear. To obtain a unique solution, a second camera is required to provide either a plane equation (point light sources and the optical center are collinear) or a line equation (point light sources and the optical center are not collinear). In general, the optical centers of two cameras will not coincide unless special optical devices are adopted. Therefore, the cornea center can be determined uniquely by using at least two cameras and at least two point light sources. ■

Proposition 3 states the sufficient condition for obtaining a unique solution of the cornea center. In order to find an optimal placement of the LEDs, the estimation error of the cornea center is analyzed (refer to Appendix-A) and the result shows that the LEDs should be placed such that for each of the cameras, there exist two LEDs locating at the x- and y-axes of the camera, respectively.

IV. Orientation of the 3-D Line of Sight

A. Virtual Image of the 3-D Pupil

The pupil of a human eye is located inside a convex lens composed of the cornea and the aqueous humor. Due to the refraction of the convex lens, it is the virtual image of the pupil, not the pupil itself, that is observed from a camera. For clarity of deriving the relationship between the location of a real object point inside the convex lens and the location of the corresponding virtual image point observed from the outside of the convex lens, an auxiliary 3-D coordinate system is defined as follows:

- the origin is defined by the cornea center,
- the x-axis is defined to be parallel to $-k_c$ (refer to the definition following equation (2)), and
- the y-axis is defined such that the x-y plane contains the OA of the eye (see Fig. 4).

Suppose that the angle between the pupil normal and the x-axis is ϕ, the radius of the pupil is r_p, and the distance between the cornea center and the pupil center is d. The 3-D coordinates of the circular edges of the pupil are given by

$$\begin{pmatrix} p_x \\ p_y \\ p_z \end{pmatrix} = \begin{pmatrix} d\cos(\phi) - r_p\cos(\theta)\sin(\phi) \\ d\sin(\phi) + r_p\cos(\theta)\cos(\phi) \\ r_p\sin(\theta) \end{pmatrix}, \tag{3}$$

where $\theta \in [0,\ 2\pi]$. The coordinates of the virtual pupil image can be computed using basic geometric optics [17], and the derived coordinates of the virtual pupil image are given by

$$P_{vp} = \begin{pmatrix} p'_x \\ p'_y \\ p'_z \end{pmatrix} = \begin{pmatrix} \rho - \frac{n'l}{R\cdot l+n} \\ m \cdot p_y \\ m \cdot p_z \end{pmatrix}, \tag{4}$$

where $l = \rho - p_x$, $R = \left|\frac{n'-n}{\rho}\right|$, $m = \frac{n}{R\cdot l+n}$, n is the refraction index of the cornea, and $n' \approx 1$ is the refraction index of the air.

It can be shown that the virtual image of the pupil is still a planar object, and the center of the virtual pupil can be determined by setting r_p to zero in equation (4):

$$P'_c = \lim_{r_p \to 0} P_{vp} = \begin{pmatrix} \rho - \frac{n'\,(\rho - d\,\cos(\phi))}{n + \frac{(n-n')\,(\rho - d\,\cos(\phi))}{\rho}} \\ \frac{d\,n\,\sin(\phi)}{n + \frac{(n-n')\,(\rho - d\,\cos(\phi))}{\rho}} \\ 0 \end{pmatrix} \tag{5}$$

Likewise, when r_p is set to zero in equation (3), we obtain the center of the real pupil

$$P_c = [d\cos(\phi)\ d\sin(\phi)\ 0]^t, \tag{6}$$

which is on the x-y plane of the auxiliary 3-D coordinate system. Since the center of the virtual pupil image shown in equation (5) is also on the x-y plane, it follows that P_c (i.e., the OA of the eye), P'_c and the x-axis (i.e., $-k_c$) of the auxiliary 3-D coordinate system are coplanar.

Suppose that the edge points of the 2-D pupil image are found to be p_i, $i = 1, 2, ..., N$, on the image plane of a camera. Because the center of the 3-D virtual pupil image is invisible, its 2-D image location, denoted by p'_c, can only be inferred by using the projected 2-D image points of the pupil edge, p_i's. Estimation of p'_c using p_i is difficult because of the nonlinearity introduced by the perspective projection. However, since the radius of the virtual pupil image is very small compared with the distance between the eye and the gaze tracking cameras, affine projection can be used to describe accurately the relationship between the 3-D virtual pupil image and its projected 2-D pupil image. The image location of the virtual pupil image center, p'_c, can be determined by $\hat{p}'_c = \frac{1}{N}\sum_{i=1}^{N} p_i$ owing to the linearity of affine projection,. When $\hat{p}'_c$ is projected back into the 3-D space, the resulting 3-D vector $\widetilde{P}'_c$ will be parallel to ${}^C\mathbf{T}_A\, P'_c$ (see Fig. 4), where ${}^C\mathbf{T}_A$ is the transformation matrix from the 3-D auxiliary coordinate system to the camera coordinate system. The image location of the virtual pupil image center will be used to estimate the orientation of the LoS in the following subsections.

B. Linear Constraint for the Direction of the Optical Axis of the Eye

Because each camera can record only one pupil image, to determine the 3-D position of P_c without knowing the values of n and d, one will have to use at least two cameras. The following proposition summarizes the sufficient condition for the unique solution of the OA direction.

Proposition 4: Suppose that a gaze tracking system equipped with at least two cameras has been calibrated so that the transformation matrices between any two coordinate systems of the cameras are known. If the cornea center, k_c, is also known (which can be estimated by using the method described in Section III), the 3-D coordinates of k_c with respect to the ith camera coordinate system, denoted by k_{ci}, can be computed with the calibrated transformation matrices. Let the estimated direction of the center of the virtual pupil image in the ith camera coordinate system be denoted by $\widetilde{P}'_{ci}$. Because the OA of the eye is coplanar with $\widetilde{P}'_{ci}$ and k_{ci}, if the 3-D planes defined by $\left(\widetilde{P}'_{ci}, k_{ci}\right)$'s do not coincide, then the OA of the eye can be uniquely determined.

Proof: Let the unknown OA direction of the eye measured in the coordinate system of the first camera be denoted by P_{c1}. If $\widetilde{P}'_{ci}$ and k_{ci} are not collinear, then a plane equation can be defined as follows:

$$\left(k_{ci} \times \widetilde{P}'_{ci}\right)^t \mathbf{R}_{i,1}\, P_{c1} = 0, \tag{7}$$

where $\mathbf{R}_{i,1}$ is the relative orientation matrix between camera coordinate systems i and 1. It is obvious that if the 3-D planes defined by $\left(\widetilde{P}'_{ci}, k_{ci}\right)$'s do not coincide, then the OA of the eye can be uniquely determined. In particular, if a gaze tracking system is composed of two cameras, then the OA of user's eye can not be uniquely determined when the user is staring at a point on the line connecting the optical centers of the two cameras. In this case, $\widetilde{P}'_{c1}$, $\widetilde{P}'_{c2}$, k_{c1}, and k_{c2} are coplanar. ■

Error analysis of the orientation estimation algorithm can be accomplished by using the technique described in the appendices. However, since the analysis results involve parameters of the eye and the cameras, the results are so complicated that no rule was found from the derived results; hence, they are not included in this paper. Nevertheless, an intuitive rule can be concluded from Proposition 4 that the distance between the images of the cornea center and the pupil center should be made as large as possible (compared with the image noise) in order to increase the signal-to-noise ratio. Therefore, the cameras and the LEDs should be carefully arranged to maximize the signal-to-noise ratio.

V. Overview of the Gaze Tracking System

To verify the gaze tracking theory proposed in the preceding sections, a gaze tracking system has been developed. It is composed of a pair of stereo cameras and three IR LEDs, namely D_1, D_2 and D_3, as shown in Fig. 5. The IR LEDs are used as point light sources for locating the cornea center. The number and the placement of the LEDs used in the system are not fixed as described in the paper, but can be varied. The only restriction is to ensure that the total number of the first Purkinje images of the LEDs observed by the two stereo cameras is always enough to determine uniquely the cornea center (that means, at least two reflexes). In order to filter out un-wanted visible light, long-pass optical filters were mounted in front of the lenses of the stereo cameras. The stereo cameras are synchronized to acquire stereo images simultaneously.

VI. Calibration of the Gaze Tracking System

Calibration of the gaze tracking system involves five major steps.

1. Calibrating the stereo cameras so that they can provide accurate 3-D measurements. In this work, the camera parameters were first estimated by using a linear method [18] and then refined by using the well-known Levenberg-Marquardt optimization algorithm with the nonlinear camera model proposed in [19]. For convenience, the coordinate system of the stereo cameras will be referred to as the CCS (Camera Coordinate System) hereafter.
2. Estimating the positions of the IR LEDs in order to estimate the cornea center.
3. Resolving the geometric properties of the monitors.
4. Determining the relationship between the stereo cameras and the screen so that once a 3-D LoS is determined, it can be transformed into the coordinate system of the screen to compute the intersection point of the LoS.

Note that although the above-mentioned calibration steps appear to be complicated, these four steps need to be performed only once and they are user-independent. On the other hand, the last step of the gaze system calibration is user-dependent.

5. Determining the angle between the LoS and the OA of user's eye. Since this step needs to be done by users, it is designed to be very simple and can be accomplished within a few seconds.

The last four calibration steps of the gaze tracking system will be detailed in the following subsections.

A. Estimation of the Positions of the LEDs

After the stereo cameras have been calibrated, the easiest way to determine the 3-D coordinates of the IR LEDs is to use the calibrated stereo cameras themselves to measure the 3-D position of the IR LEDs. Since the IR LEDs are not located in the view-field of the stereo cameras, they cannot be observed directly by the stereo cameras. By introducing a planar mirror with at least three fiducial marks attached on the mirror surface into the measurement system, the stereo cameras can observe the fiducial marks as well as the IR LEDs located beside them. The 3-D position of the fiducial marks can be used to reconstruct the reflection surface of the mirror. Thus, the actual 3-D positions of the LEDs can be easily determined. The error analysis result derived in Appendix-B shows that the distance between the LED and the mirror should be minimized in order to obtain accurately the 3-D position of the LED.

B. Estimation of the Geometric Properties of the Monitor

Monitor calibration comprises the following two main tasks:

1. Calibration of the position, the orientation and the shape parameters of the monitor: Monitors are usually frabricated with a flat or spherical screen surface whose parameters has to be determined in order to compute the intersection point of the screen surface and the LoS. One can use another pair of calibrated stereo cameras (referred to as the auxiliary stereo cameras), which have a large field of view, to measure

the parameters of the screen surface. A reference coordinate system of the screen (referred to as the Screen Coordinate System (SCS)) is defined after the screen is calibrated. Three fiducial marks were attached to the monitor and their 3-D coordinates in SCS can also be determined using the auxiliary stereo cameras, which will also be used to compute the coordinate transformation matrices from other coordinate systems to SCS (refer to Section VI-C).

2. Calibration of the 2-D geometry parameters: In general, the 2-D window coordinates (in pixels) and the corresponding 3-D coordinates of a point on the screen surface can be related by a linear transformation. Given a set of 2-D-to-3-D correspondences of points on the screen surface, parameters of the linear transformation can be determined. Thus, we can transform the 3-D coordinates of the fixation point on the surface into 2-D window coordinates and vice versa.

C. Estimation of the Coordinate Transformation Matrix from the Coordinate System of the Gaze-Tracking Stereo Cameras to the Screen Coordinate System

Because the geometry of the monitor and the parameters of the gaze-tracking stereo cameras are calibrated independently, the coordinate transformation matrix between the CCS and SCS has to be calibrated with another procedure. In this work, a 3-D digitizer (i.e., Microscribe, Immersion Corporation) is used to calculate ${}^{S}T_{C}$, the coordinate transformation matrix from the CCS to the SCS. A calibration plate with some calibration points is mounted in front of the stereo cameras so that the 3-D coordinates of the calibration points can be measured using the stereo cameras. Then, the 3-D coordinates of the calibration points are measured by the MicroSribe. Using the 3-D correspondences, we can compute a transformation matrix from the CCS to the MicroSribe Coordinate System (MCS), ${}^{M}\mathbf{T}_{C}$, by using the closed-form solution proposed by Arun et al. [20]. Likewise, the 3-D digitizer is employed to measure the 3-D coordinates of the three fiducial marks stuck on the monitor. Since the 3-D coordinates of the fiducial marks with respect to the SCS are also known, the transformation matrix from the MCS to the SCS, denoted by ${}^{S}\mathbf{T}_{M}$, can be determined by using the method proposed in [20]. The unknown transformation matrix can be determined by ${}^{S}T_{C} = {}^{S}T_{M}\,{}^{M}T_{C}$.

D. Estimation of the Angle between the 3-D LoS and the Optical Axis of User's Eye

After the gaze tracking system has been calibrated, it can be used to measure the OA of a user's eye. However, since the OA of a user's eye is usually different from his/her LoS [21], unless the relationship between them is known, the user's fixation point can not be determined. In this work, it is assumed that the angle between the 3-D LoS and the OA is constant. Since the distance between the cornea center and the optical center of the eye is negligible compared with that between the eye and the monitor, it is further assumed that the optical center and the cornea center of the eye are coincident. To formulate the relation between the two 3-D vectors of the LoS and the OA, an eye coordinate system (ECS) is defined as follows.

1. The origin of the ECS is aligned with the measured curvature center of the cornea, i.e., ${}^{C}k_c$.

2. Define the direction of the z-axis of the ECS by the estimated OA direction of the eye, which is denoted by the unit vector d_O.

3. Let the x-axis of the ECS be parallel to the horizontal plane. For example, in the implemented gaze tracking system, the LEDs D_1 and D_2 (see Fig. 5) are arranged horizontally. Therefore, the x-axis of the ECS in our gaze tracking system is determined by the difference vector of the 3-D positions of LEDs D_1 and D_2.

4. The direction of the y-axis of the ECS is determined by computing the cross-product of the z- and x-axes.

Note that by defining the x-axis of the ECS to be parallel to the horizontal plane, it is equivalent to asking users not to roll their head while using the gaze tracking system (yawing and pitching movements do not affect the gaze tracking results). This is not only a constraint of our gaze tracking system, but also a common drawback of all the gaze tracking systems that do not compute the torsional motion of the eye ball. Should a user roll his/her head while using the gaze tracking system, the gaze tracking error will be proportional to the angle between the LoS and the OA and the net rolling angle of user's eye (i.e., the rolling angle of user's head minus the rolling angle of user's eye).

Since the cornea center, ${}^{C}k_c$, the OA direction, d_O, and the 3-D positions of the LEDs measured with repsect to the CCS are all known, the transformation matrix, ${}^{E}\mathbf{T}_C$, varying with the position and the orientation of user's eye, from the CCS to the ECS can be easily computed. In order to estimate the constant LoS vector with respect to the ECS, the user is asked to stare at the point (x_d, y_d) on the screen for a while and the corresponding gaze information is recorded. Since the monitor has been calibrated, the 3-D coordinates of this calibration point in the SCS can be computed and then transformed into the CCS. Let the 3-D coordinates of the calibration point in the CCS be denoted by ${}^{C}F$. Each time when a new measurement, say measurement i, of the optical axis of the eye is obtained, a new transformation matrix ${}^{E}\mathbf{T}_C(i)$ is computed and then ${}^{C}F$ is transformed into the ECS and is denoted by ${}^{E}F_i$. Define the direction of ${}^{E}F_i$ as ${}^{E}D_i = \frac{{}^{E}F_i}{\|{}^{E}F_i\|}$. In order to minimize the effects of measurement noise and spontaneous eye movements, the constant LoS vector in the ECS, denoted by ${}^{E}L$, is estimated by using the following equation:

$$ {}^{E}L = \begin{pmatrix} \text{median}\left({}^{E}D_{i,x}\right) \\ \text{median}\left({}^{E}D_{i,y}\right) \\ \sqrt{1 - \text{median}\left({}^{E}D_{i,x}\right)^2 - \text{median}\left({}^{E}D_{i,y}\right)^2}, \end{pmatrix} \tag{8} $$

where ${}^{E}D_{i,x}$ and ${}^{E}D_{i,y}$ are the x- and y-components of the unit vector ${}^{E}D_i$, respectively. The 3-D LoS represented in the SCS can now be determined by the following equation:

$$ \tilde{p}_L = {}^{S}\mathbf{T}_C \, {}^{C}\tilde{k}_c + \lambda \, {}^{S}\mathbf{T}_C \, {}^{E}\mathbf{T}_C^{-1} \, {}^{E}\tilde{L}, \tag{9} $$

where ${}^{C}\tilde{k}_c = \left[{}^{C}k_c{}^{t} \; 1 \right]^{t}$ and ${}^{E}\tilde{L} = \left[{}^{E}L^{t} \; 0 \right]^{t}$ are the homogeneous coordinate representation of the 3-D point ${}^{C}k_c$ and the 3-D direction ${}^{E}L$, respectively.

VII. Experiments

In the gaze tracking system, two B/W cameras (Mintron OS-60D) are used to acquire stereo image sequences of user's eye. The output video signal format is RS-170. A dual channel frame grabber (Luetron PicPort-Stereo) is used to digitize and to transfer the stereo images to the main memory of a Pentium III/500MHz PC with 128-megabyte memory. The frame grabber is configured to store the even field and odd field of the input interlaced images at two contiguous memory blocks, respectively. Only one field (effective image size is 240×640 pixels) is used for computing the gaze information. The display resolution of the PC is adjusted to 768×1024 pixels. The monitor (Proview PK-786) of the PC has a 17" CRT with a 16" viewable area. The calibrated pixel spacings of the monitor in both horizontal and vertical directions are approximately 0.29 millimeters. The distance between user's eye and the monitor is approximately 450 millimeters. Three IR LEDs are used as point light sources to shine on one of the user's eyes. The LEDs were placed according to the guideline proposed in Appendix-A. Then, the locations of the IR LEDs were adjusted such that wherever the user is looking at on the monitor, the stereo cameras can detect enough number of the first Purkinje images of the LEDs for determining the cornea center. The gaze tracking system is developed under the Microsoft Windows 2000 environment.

Before performing the gaze tracking experiments, the gaze tracking system was calibrated by using the calibration method described in Section VI. The 3-D estimation error of the calibrated stereo cameras is approximatly 0.05 millimeters when the 2-D feature location error is 0.5 pixels. The setup for estimating the positions of the LEDs are shown in Fig. 7 (refer to Section VI-A). The stereo images acquired are shown in Fig. 8. The estimation error of the 3-D positions of the LEDs, i.e., $\sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2}$, can be evaluated by using the equations derived in Appendix-B, which is approximately 1.5 millimeters (where the distance between any two of the fiducial marks, w, is 15 millimeters). The amount of the position estimation error of the LEDs are only 0.4% of the actual distance between the LEDs and the cameras. In order to test the influence of 0.4% estimation error of the LEDs, a computer simulation was conducted. By using the calibrated parameters of our gaze tracker and the parameters of the Le Grand's eye model, a simulation system was developed. The position of each point light source is perturbed by a 3-D random vector having its length set at a certain percentage of the distance between the LED and the optical center of the left camera. Six hundred random trials were performed and the root-mean-square error (RMSE) of the orientation error versus the percentage of position errors of the LEDs is plotted in Fig. 9(a). According to the computer simulation results, 0.4% of the position calibration errors of the LEDs causes an orientation estimation error of the LoS of approximately 0.1° which is negligible. To test the robustness of the proposed algorithm against the image noise, we further performed another computer simulation and the results are shown in Figs. 9(b)–(c). Each datum shown in Figs. 9(b)–(c) was obtained by computing

the RMSE of 600 random trials while the percentage of position errors of the LEDs were set at 0.4%. The computer simulation results shows that the cornea center estimation algorithm is rather robust. The error analysis results described in Appendix-A shows that if the LEDs are properly placed, the image location estimation error of the cornea center is approximately of the same order as the image noise. Therefore, when the image noise is increased to two pixels, the estimated image location of the cornea center is also about two pixels. Since the 3-D estimation error of the calibrated stereo cameras is 0.05mm when the image noise is 0.5 pixels, if the image noise is increased to two pixels ($= 4 \times 0.5$ pixles), the 3-D estimation error will become 0.2 mm ($= 4 \times 0.05$ mm). Simulation results show that the position estimation error is 0.32mm (see Fig. 9(c)) which is slightly greater than 0.2 mm because the LEDs are not placed at the optimal positions. In short, the position estimation error of the cornea center can be obtained accurately provided that the cameras are accurately calibrated and that the LEDs are properly placed. On the other hand, the simulation results also show that the orientation estimation error of the LoS using the proposed algorithm is somewhat sensitive to image noise. This is because the orientation of the LoS is determined by using the 2-D image difference vectors of the pupil center and the cornea center, if the lengths of the difference vectors are short compared with the image noise, then the orientation estimation error will be very sensitive to image noise. In the configuration of our gaze tracker, the length of the difference vector is only about 50 pixels. Two pixels of the images noise amounts to the signal-to-noise ratio of only about 25. This is the common weakness of all the gaze tracking techniques using cornea reflection and pupil image.

Figure 10 shows four pairs of typical stereo images containing the extracted features (such as the pupil center and the first Purkinje images of the IR LEDs) for computing the gaze information. In the implemented gaze tracking system, the pupil center is computed to sub-pixel accuracy, whereas the glare points are only computed to pixel-level accuracy (i.e., with ± 0.5 pixels of quantization error). A lot of work has been done to optimize the run-time performance of the gaze tracking algorithm. The system currently processes 30 pairs of stereo images per second and uses only 3% of the CPU time on the Pentium III/500MHz PC. Therefore, if the RS-170 stereo cameras are replaced by some high-speed cameras, then the sampling rate of the gaze data can be improved accordingly. Every time a 2-D fixation point is successfully computed, an event is posted onto the message queue of the operation system. If the image area of the pupil is smaller than a specified threshold value, then the event is marked with a "blink." Note that since the image sampling rate is only 30Hz, when the eye undergoes a saccade motion, the acquired images will be corrupted due to motion blur effect. The motion blur problem can be diminished either by adopting high-speed cameras or by activating the electronic shutter of the cameras.

In order to test the accuracy of the gaze tracking system, six subjects who do not wear glasses were asked to participate in the tests. Before using the system for gaze tracking, the angle between LoS and OA has to be determined for each subject individually by using the method described in Section VI-D. Each subject was asked to read out a sequence of randomly generated English letters (updated once per

TABLE I

Root-mean-square error (RMSE) of the gaze tracking error of the six subjects.

	RMSE(ϵ_x)	RMSE(ϵ_y)
Subject 1	20.2 pixels (0.67°)	32.6 pixels (1.09°)
Subject 2	31.0 pixels (1.03°)	19.9 pixels (0.72°)
Subject 3	15.4 pixels (0.50°)	11.6 pixels (0.40°)
Subject 4	18.4 pixels (0.59°)	26.6 pixels (0.85°)
Subject 5	14.7 pixels (0.48°)	13.8 pixels (0.46°)
Subject 6	18.1 pixels (0.54°)	20.9 pixels (0.60°)

second to attract subject's attention) showing at one of the 16 pre-specified grid positions for 10 seconds per position. To record the gaze information, the subject double clicks on the target point and the system sounds a drum roll for two seconds. Moreover, every time the randomly generated letter is about to be updated, a tick is sounded. Through different sounds, the subject is informed what the system is going to do and can get prepared. Therefore, the acquired data can reflect the accuracy of the gaze tracking system. If a subject feels tired, he can leave the seat and have a break. Since the angle between the OA and the LoS has been calibrated, the subject can immediately resume the tests on the next fixation point whenever he/she is back. The experimental results are shown in Fig. 11, where the 16 target points are marked with crosses and the ellipses denote the standard deviation of the computed 2-D fixation points. Since the target point used for estimating the angle between the LoS and the OA of each user's eye is also located at one of the 16 grid positions (at the second row and the second column), the estimated fixation points shown in Fig. 11 match well the target point.

Some statistics of the overall gaze tracking error are listed in Table I, where ϵ_x and ϵ_y stand for the horizontal and vertical measurement error (i.e., the distance between the estimated fixation point and the target point measured in the 2-D window coordinate system), respectively. The experimental results show that the accuracy of the estimated fixation points are user-dependent due to some subtle differences among the eyes of different subjects. For example, since the cornea area of subject 4 is smaller than that of the others, when he looked to the left part of the monitor, one of the first Purkinje images of the LEDs appeared at the junction between the cornea and the sclera of his eye leading the estimated fixation points to wrong positions outside the screen. However, this is not a flaw of the proposed gaze tracking algorithm and can be easily resolved by adjusting the positions of the LEDs. Generally speaking, the estimated fixation points at the border of the monitor tend to be less accurate than those at the central portion of the monitor. Two causes were identified that degraded the accuracy of gaze tracking results when the subjects stared at the border area of the monitor.

1. Since the fovea region of the retina is not an infinitely small spot, the constant orientation difference

model for the LoS and the OA of user's eye is merely an approximate model. The angle between the LoS and the optical axis may vary with the location of the fixation point which yields gaze tracking error.
2. The cornea surface of a person is so smooth that the central portion of the cornea can be approximately modeled as a spherical surface. However, the spherical model may not be suitable for modeling the border area of the cornea. When the subjects stared at the border area of the monitor, the observed glare points moved toward the border of the cornea. Thus, the gaze tracking accuracy may be degraded if the curvature of the cornea of the subject varies greatly.
Moreover, since the stereo cameras are adjusted to focus at a distance of about 70 millimeters, this close-range focusing arrangement has a drawback of short field depth. As the subject moved away or toward the camera, the captured stereo images became blurred. The out-of-focus effect is amplified in the vertical direction as the vertical image resolution is only a half of the horizontal one. The overall accuracy of the gaze tracking system is better than 1 degree and is satisfactory for many applications.

VIII. Conclusions

A novel method for estimating the 3-D line of sight of a user has been proposed in this paper. The proposed method employs stereo cameras to determine the cornea center and the gaze direction using linear solutions. Furthermore, a simple and accurate method for calibrating the gaze tracking system is proposed in this paper. The contributions of this work are summarized in the following.
1. A linear solution to computing the position and the orientation of the LoS is proposed in this paper. Sufficient conditions for obtaining a unique solution of the 3-D LoS are derived.
2. A simple calibration method is proposed for estimating the parameters of the gaze tracking system.
3. Most of the calibration process is user-independent. It takes only a few seconds (normally, 2–3 seconds) for a new user to complete the user-dependent parameter estimation process.
4. Error analysis results of the proposed method are provided which can be used as a guideline to design a new gaze tracking system.
The gaze tracking algorithm is implemented on a Pentium III/500MHz computer with the Microsoft Windows 2000 operation system. The system can process 30 pairs of stereo images per second to provide 3-D information about user's LoS while consuming only 3% of the CPU time. The accuracy of the implemented gaze tracking system is about one degree, comparable with the view-field error of the fovea of a human[3]. For people having different intra eye parameters different from those used in this eye model (e. g. the cornea surface can not be modeled with a constant curvature), a local calibration procedure [3] can be employed to improve the tracking accuracy. Moreover, if the cameras are replaced by progressive-scan ones such that the full resolution of the images could be used, the system accuracy will be improved further. The proposed gaze tracking algorithm can be used as a prosthetic device. With the gaze tracking algorithm of such high accuracy, an efficient interface can be developed for paralyzed patients. Furthermore, for some applications, e.g., 3-D rendering, that require only the positions of the

eyes to be known, the cornea center estimation algorithm can be used to provide very robust estimation results. Thus, 3-D objects can be rendered according to the eye positions to achieve more realistic effects (such as motion parallax).

Acknowledgement

This work is supported in part by the National Science Council of Taiwan under Grant NSC 87-2218-E-260-003.

Appendices

A. Error Analysis of the Cornea Center Position Estimation Algorithm Using Two Point Light Sources

For the ease of illustration, proposition 2 is derived using 3-D vector representation. However, to obtain concise error analysis results, we will use the projective geometry representation [22] in the following. Also, since only two LEDs are needed in order to determine the cornea center, the error analysis focuses on how to place these two LEDs in order to obtain the most accurate estimation results.

Let the projective image coordinates of the two LEDs be denoted by p_1 and p_2, respectively. Since the LEDs have to be placed outside the view-filed of the camera, it is reasonable to assumed that the LEDs are placed on the x-y plane of the camera coordinate system (so that they will not obstruct the view of the camera), i.e., $p_1 = [\cos(\theta_1)\ \sin(\theta_1)\ 0]^t$ and $p_2 = [\cos(\theta_2)\ \sin(\theta_2)\ 0]^t$. Let q_1 and q_2 denote the projective image coordinates of the first Purkinje images of the LEDs. It can be shown that the projective coordinates of the cornea center can be computed as follows:

$$[c_1\ c_2\ c_3]^t = (p_1 \times q_1) \times (p_2 \times q_2)\,. \tag{10}$$

The inhomogeneous coordinates of the image of the cornea center are given by

$$[c_x\ c_y]^t = \left[\frac{c_1}{c_3}\ \frac{c_2}{c_3}\right]^t. \tag{11}$$

By computing the first-order Taylor series expansion of the above equation, we have

$$[\hat{c}_x\ \hat{c}_y]^t \approx [c_x\ c_y]^t + \mathbf{J}\delta P, \tag{12}$$

where $\delta P = [\delta q_1^t\ \delta q_2^t]^t$ is the image noise vector and $\mathbf{J}$ is the Jacobian matrix of equation (11) with respect to δP. Suppose that all components of the image noise vector have identical independent Gaussian distributions having zero mean and variance σ^2. Then, the approximate covariance matrix of $[\hat{c}_x\ \hat{c}_y]^t$ can be derived using Mathematica:

$$\sigma^2\left(\mathbf{J}\mathbf{J}^t\right) = \frac{\sigma^2\begin{pmatrix} \frac{2+\cos(2\,\theta_1)+\cos(2\,\theta_2)}{2} & \cos(\theta_1)\,\sin(\theta_1)+\cos(\theta_2)\,\sin(\theta_2) \\ \cos(\theta_1)\,\sin(\theta_1)+\cos(\theta_2)\,\sin(\theta_2) & \frac{2-\cos(2\,\theta_1)-\cos(2\,\theta_2)}{2} \end{pmatrix}}{\sin(\theta_1-\theta_2)^2}. \tag{13}$$

According to the above equation, the variance of the estimated image location of the cornea center is minimized when $\theta_1 = \theta_2 + \frac{\pi}{2}$. Let $L_1 = A^{-1}\, p_1$ and $L_2 = A^{-1}\, p_2$ denote the vectors obtained by back-projecting p_1 and p_2 into the Euclidian space, respectively, where A is the calibration matrix containing the intrinsic camera parameters. Substituting $\theta_1 = \theta_2 + \frac{\pi}{2}$ into the scalar product of L_1 and L_2, we have

$$L_1^t\, L_2 = \frac{\delta_v^2 - \delta_u^2}{f^2} \cos(\theta_2) \sin(\theta_2), \tag{14}$$

where δ_u, δ_v, and f are the horizontal and vertical pixel spacings, and the effective focal length of the camera, respectively. If $\theta_2 = 0$, i.e., the second LED is placed along the x-axis of the CCS, then equation (14) becomes zero even when $\delta_u \neq \delta_v$. This implies that the first LED is located at the y-axis of the CCS. Therefore, in order to estimate accurately the cornea center location, one can place the two LEDs at the x- and y-axes of the camera coordinate system, respectively. Otherwise, the position estimation error may be amplified. In particular, when the LEDs and the optical center of the camera are collinear, i.e., $\theta_1 = \theta_2$, the variance of the image error will approach infinity and thus should be avoided. The derived image error of the cornea center can be used to infer the 3-D position estimation error of the cornea center by using the error analysis result proposed by Blostein and Huang [23].

B. Error Analysis of the LED Position Estimation Method Proposed in Section VI-A Using Three Fiducial Marks

Let the true 3-D position vector of the three equally-spaced fiducial marks and the virtual image of the LED be denoted by P_1, P_2, P_3, and Q, respectively. The position of the real LED can be computed by using the following equation:

$$Q' = Q - 2\left((Q - \bar{P})^t n\right)\, n, \tag{15}$$

where $\bar{P}$ and n are the centroid and the normal vector of the triangle defined by these three fiducial marks, respectively. By computing the Taylor series expansion of (15), we have

$$Q'\left(\hat{P}_1, \hat{P}_2, \hat{P}_3, \hat{Q}\right) = Q'\left(P_1, P_2, P_3, Q\right) + \mathbf{J}\,\delta P + O(2), \tag{16}$$

where $\hat{P}_1 = P_1 + \delta P_1$, $\hat{P}_2 = P_2 + \delta P_2$, $\hat{P}_3 = P_3 + \delta P_3$, $\hat{Q} = Q + \delta Q$ and $\mathbf{J}$ is the Jacobian matrix of (15) with respect to $\delta P = [\delta P_1^t\ \delta P_2^t\ \delta P_3^t\ \delta Q^t]^t$.

The goal of the error analysis is to find the variance of the estimation error of Q'. Assume that P_1, P_2, P_3, and Q are represented in a coordinate system defined by using the three fiducial marks. In particular, let the coordinate system be defined such that $P_1 = [0, 0, 0]^t$, $P_2 = [w, 0, 0]^t$ and $P_3 = \left[\frac{w}{2}, \frac{\sqrt{3}w}{2}, 0\right]^t$, where w is the distance between any two of the three fiducial marks. If the estimation error of Q' is small and the estimation noise of the 12 components of δP are identical independent Gaussian distributions with zero mean and variance σ^2, then the diagonal components of the covariance matrix of Q' ($\mathrm{cov}\left(\hat{Q}'\right) \approx \sigma^2 \mathbf{J}\mathbf{J}^t$) can be computed by using Mathematica:

$$\sigma_x^2 \approx \sigma^2 \left(1 + \frac{8\, z^2}{w^2}\right), \tag{17}$$

$$\sigma_y^2 \approx \sigma^2 \left(1 + \frac{8\,z^2}{w^2}\right), \tag{18}$$

and

$$\sigma_z^2 \approx \sigma^2 \frac{15\,w^2 - 8\,w\,\left(3\,x + \sqrt{3}\,y\right) + 24\,\left(x^2 + y^2\right)}{3\,w^2} \tag{19}$$

where $[x\ y\ z]^t = Q$. Although the above three values are dependent on the orientation of the coordinate system, it can be shown that the sum of the above three variances is independent of the coordinate system. Threrefore, we can use $\sigma_x^2 + \sigma_y^2 + \sigma_z^2$ as an estimate of the position estimation error of an LED regardless of the reference coordinate system used. Equations (17)–(19) show that the estimation error can be minimized if w is maximized and the distance between the LED and the mirror is minimized.

References

[1] T. E. Hutchinson, J. K. P. White, W. M. Martin, K. C. Reichert, and L. A. Frey, "Human-computer interaction using eye-gaze input," IEEE Transactions on Systems, Man and Cybernetics, vol. 19, no. 6, pp. 1527–1534, 1989.

[2] L. A. Frey, J. K. P. White, and T. E. Hutchinson, "Eye-gaze word processing," IEEE Transactions on Systems, Man, and Cybernetics, vol. 20, no. 4, pp. 944–950, 1990.

[3] R. J. K. Jacob, "The use of eye movements in human-computer interaction techniques: What you look at is what you get," ACM Transactions on Information Systems, vol. 9, no. 3, pp. 152–169, 1991.

[4] S. Zhai, C. H. Morimoto, and S. Ihde, "Manual and gaze input cascaded (magic) pointing," in Proceedings of the ACM SIGCHI—Human Factors in Computing Systems Conference, pp. 246–253, 1999.

[5] P. M. Sharkey and D. W. Murray, "Feasibility of using eye tracking to increase resolution for visual tele-presence," in Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, vol. 2, pp. 1078–1083, 1997.

[6] S. Pastoor, J. Liu, and S. Renault, "An experimental multimedia system allowing 3-d visualization and eye-controlled interaction without user-worn devices," IEEE Transactions on Multimedia, vol. 1, no. 1, pp. 41–52, 1999.

[7] A. Redert, E. Hendriks, and J. Biemond, "3-d scene reconstruction with viewpoint adaptation on stereo displays," IEEE Transactions on Circuits and Systems for Video Technology, vol. 10, no. 4, pp. 550–562, 2000.

[8] Y. L. Grand, Light, Color and Vision. Wiley, 1957.

[9] J. Gips, P. Olivieri, and J. Tecce, "Direct control of the computer through electrodes placed around the eyes," in Proceedings of the Fifth International Conference on Human-Computer Interaction, Elsevier, Orlando, Florida, pp. 630–635, 1993.

[10] S. Baluja and D. Pomerleau, "Non-intrusive gaze tracking using artificial neural networks," Tech. Rep. CMU-CS-94-102, School of Computer Science, Carnegie Mellon Univesity, Pittsburgh PA, USA, 1994.

[11] J. K. P. White, T. E. Hutchinson, and J. M. Carley, "Spatially dynamic calibration of an eye-tracking system," IEEE Transactions on Systems, Man and Cybernetics, vol. 23, no. 4, pp. 1162–1168, 1993.

[12] T. Cornweet and H. Crane, "Accurate two-dimensional eye tracker using first and fourth purkinje images," Journal of Optical Society America, vol. 63, no. 8, pp. 921–928, 1973.

[13] L. Bour, "Dmi-search scleral coil," Tech. Rep. H2-214, Dept. of Neurology, Clinical Neurophysiology, Academic Medical Centre, AZUA, Meibergdreef 9, 1105AZ Amsterdam, Netherlands, 1997.

[14] R. Stiefelhagen, J. Yang, and A. Waibel, "A model-based gaze tracking system," in Proceedings of the IEEE International Joint Symposia on Intelligence and Systems, pp. 304–310, 1996.

[15] C. Collet, A. Finkel, and R. Gherbi, "Capre: a gaze tracking system in man-machine interaction," in Proceedings of the IEEE International Conference on Intelligent Engineering Systems, pp. 577–581, 1997.

[16] A. Sugioka, Y. Ebisawa, and M. Ohtani, "Noncontact video-based detection method allowing large head displacements," in Proceedings of the International Conference on Engineering in Medicine and Biology Society, vol. 2, pp. 526–528, 1996.

[17] M. Freeman, Optics. London, Boston, Butterworths, 1999.

[18] O. D. Faugeras and G. Toscani, "The calibration problem for stereo," in Conference on Computer Vision and Pattern Recognition, pp. 15–20, 1986.

[19] R. Y. Tsai, "A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses," IEEE Transactions on Robotics and Automation, vol. 3, no. 4, pp. 323–344, 1987.

[20] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-square fitting of two 3-d point sets," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 9, no. 5, pp. 698–700, 1987.

[21] R. A. Applegate, L. N. Thibos, A. Bradley, S. Marcos, A. Roorda, T. O. Salmon, and D. A. Atchison, "Reference axis selection: Subcommittee report of the osa working group to establish standards for measurement and reporting of optical aberrations of the eye," Journal of Refractive Surgery, vol. 16, pp. 656–658, 2000.

[22] R. Hartley and A. Zisserman, Multiple View Geometry in Computer Vision. Cambridge University Press, 2000.

[23] S. D. Blostein and T. S. Huang, "Error analysis in stereo determination of 3-d point positions," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 9, no. 6, pp. 752–765, 1987.

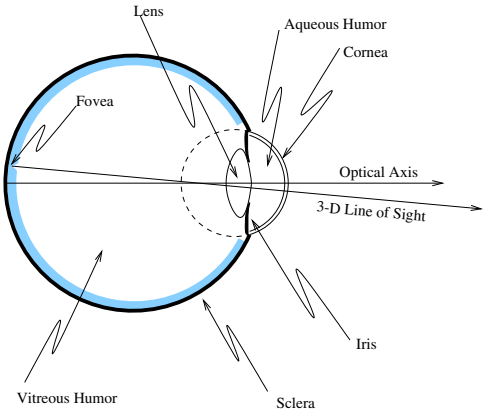

Fig. 1. Structure of Le Grand's simplified eye.

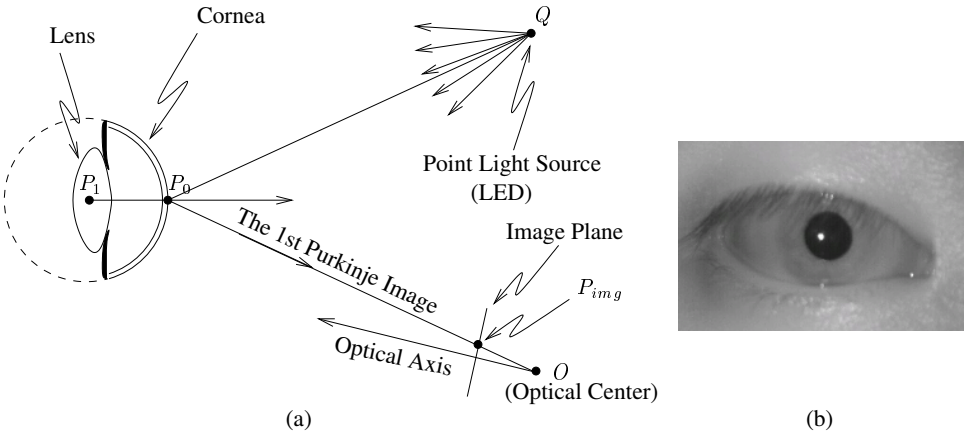

Fig. 2. (a). A schematic diagram illustrating the geometric relationships between the point light source, the cornea, the first Purkinje image, and the camera. (b). The bright spot inside the pupil is the recorded first Purkinje image (a virtual image) of an infrared LED.

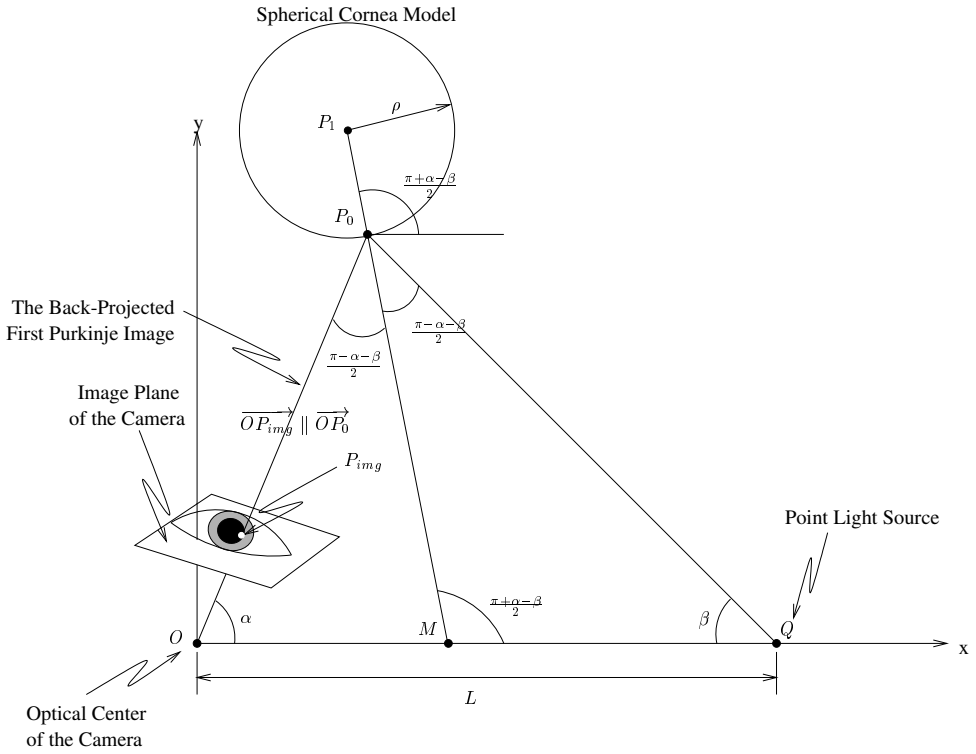

Fig. 3. The relation between the first Purkinje image, P_{img}, the location of the point light source, Q, and the location of the cornea center.

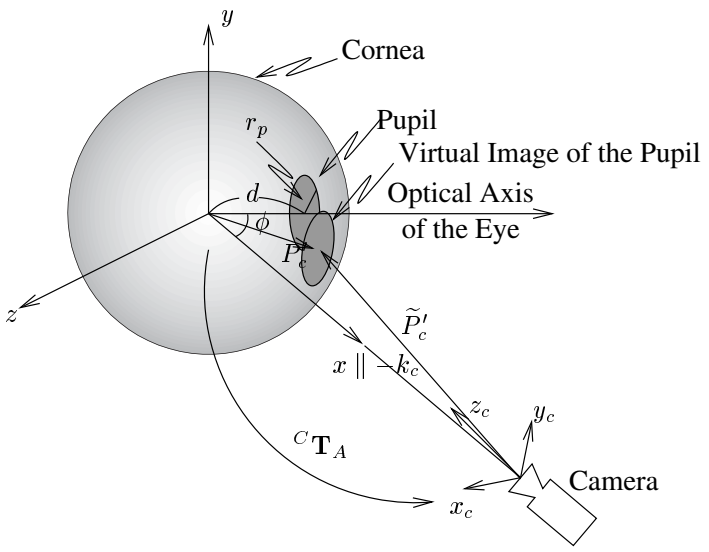

Fig. 4. The auxiliary 3-D coordinate system defined at the cornea center.

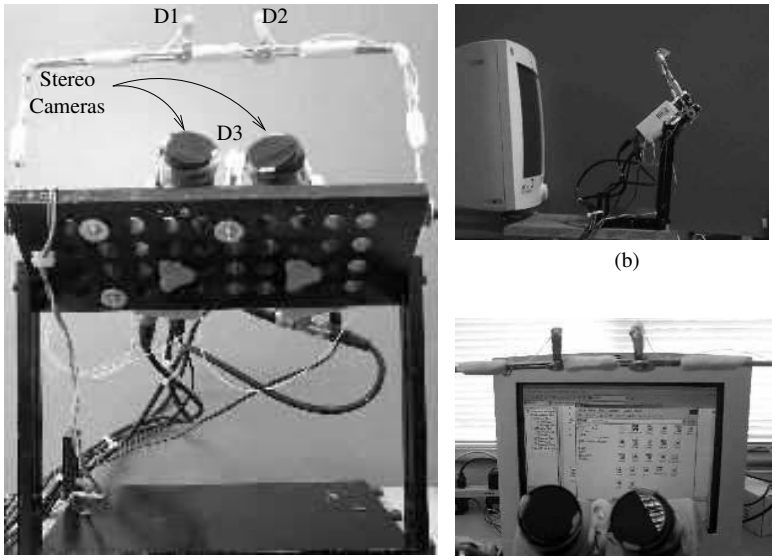

Fig. 5. (a). Front view of the gaze tracking system. (b). Side view of the gaze tracking system. (c). Users can see the whole screen through the rectangular window in between the IR LEDs and the stereo cameras without being obstructed.

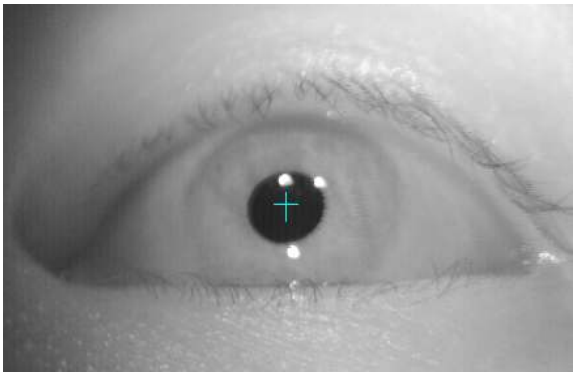

Fig. 6. The image feature points are located at the cornea area of the eye.

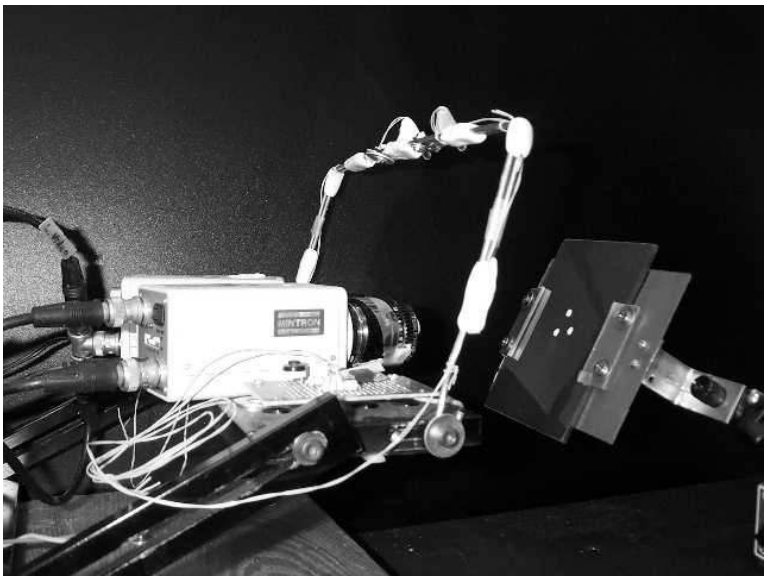

Fig. 7. The 3-D coordinates of the LEDs measured in the coordinate system of the stereo cameras can be determined using a first-surface planar mirror with three fiducial marks.

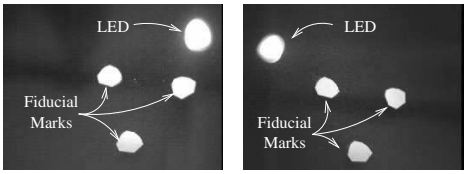

Fig. 8. A pair of stereo images of an LED and the three fiducials captured by the stereo cameras.

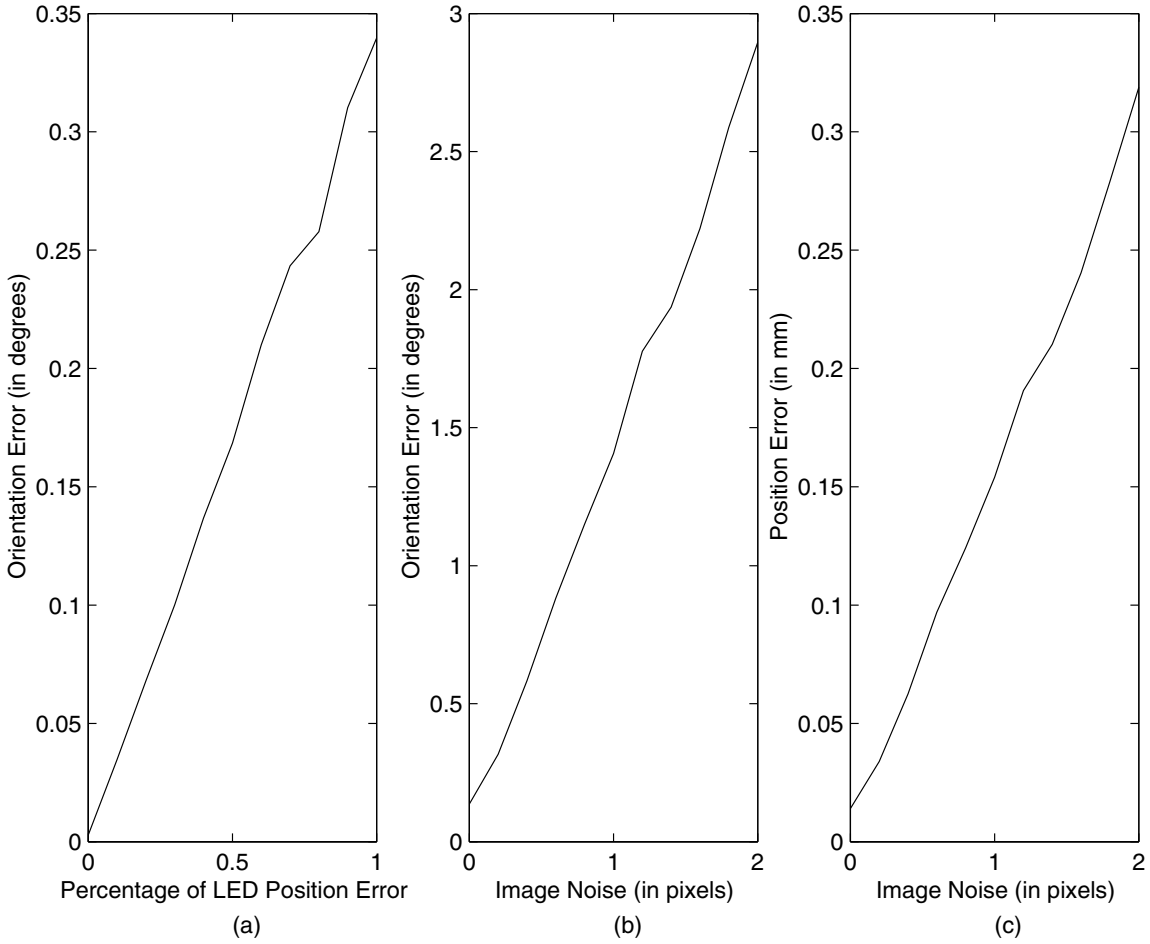

Fig. 9. Computer simulation results: (a). Influence of the position estimation errors of the LEDs on the orientation estimation error of the LoS. (b). Orientation estimation error of the LoS versus the standard deviation of the image noise. (c). Position estimation error of the cornea center versus the standard deviation of image noise.

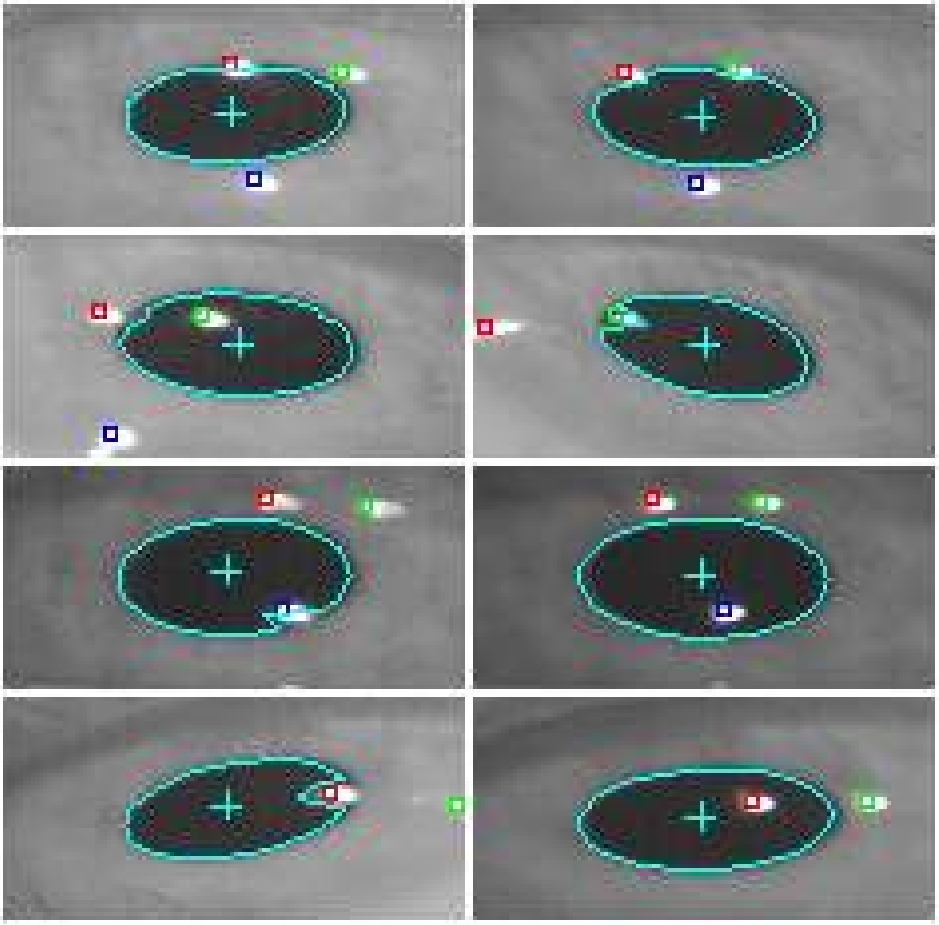

Fig. 10. Features extracted from stereo images for computing the gaze information. Images shown on the left and right columns are clipped from the left and the right stereo images, respectively.

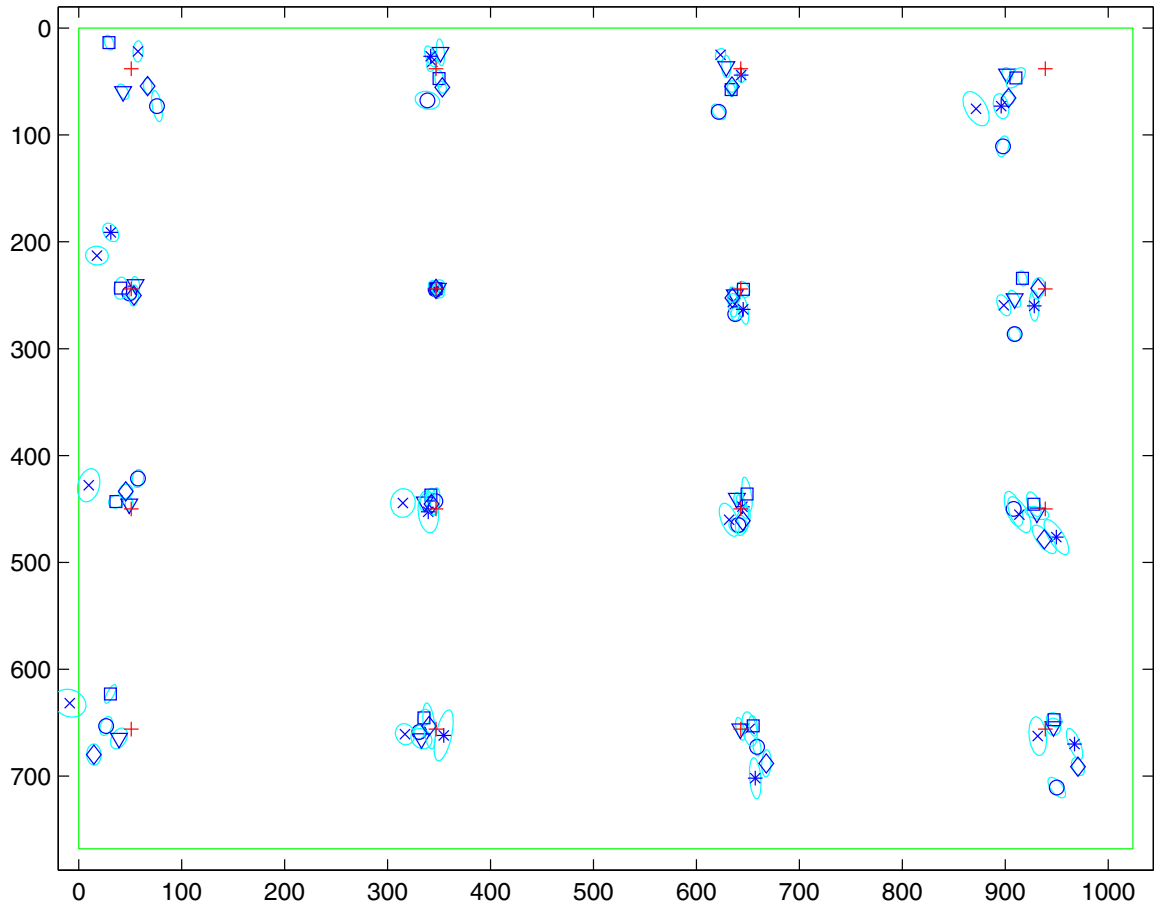

Fig. 11. Test results of six subjects where '$\circ$,' '$\times$,' '$\triangledown$,' '$*$,' '$\square$,' and '$\Diamond$' mark the sample means of the estimated fixation points of the six subjects, respectively.