General Theory of Remote Gaze Estimation Using the Pupil Center and Corneal Reflections

Elias Daniel Guestrin* , *Student Member, IEEE*, and Moshe Eizenman

***Abstract*—This paper presents a general theory for the remote estimation of the point-of-gaze (POG) from the coordinates of the centers of the pupil and corneal reflections. Corneal reflections are produced by light sources that illuminate the eye and the centers of the pupil and corneal reflections are estimated in video images from one or more cameras. The general theory covers the full range of possible system configurations. Using one camera and one light source, the POG can be estimated only if the head is completely stationary. Using one camera and multiple light sources, the POG can be estimated with free head movements, following the completion of a multiple-point calibration procedure. When multiple cameras and multiple light sources are used, the POG can be estimated following a simple one-point calibration procedure. Experimental and simulation results suggest that the main sources of gaze estimation errors are the discrepancy between the shape of real corneas and the spherical corneal shape assumed in the general theory, and the noise in the estimation of the centers of the pupil and corneal reflections. A detailed example of a system that uses the general theory to estimate the POG on a computer screen is presented.**

***Index Terms*—Model, point of regard, pupil center and corneal reflection(s), remote gaze estimation, system configurations, video based gaze estimation.**

I. Introduction

THE point-of-gaze (POG) is the point in space that is imaged on the center of the highest acuity region of the retina (fovea) of each eye. Systems that estimate the POG are primarily used in the analysis of visual scanning patterns and in human-machine interfaces. Since visual scanning patterns closely follow shifts in attentional focus, they provide insight into human cognitive processes [1]. As such, analysis of visual scanning patterns is used in the quantification of mood disorders [2], studies of perception, attention and learning disorders [3], [4], driving research and safety [5]–[7], pilot training [8], and ergonomics [9]. In the area of human-machine interfaces, the POG can be used as an input modality in multimodal human-computer interfaces [10] and assistive devices such as gaze-controlled interfaces to allow nonverbal motor-disabled persons to communicate and control the environment [11], [12].

Manuscript received March 28, 2005; revised November 6, 2005. This work was supported in part by a grant from the Canadian Institute of Health Research (CIHR), and in part by scholarships from the University of Toronto, Toronto, ON, Canada, and the Universidad Tecnológica Nacional, Argentina. *Asterisk indicates corresponding author.*

*E. D. Guestrin is with the Department of Electrical and Computer Engineering, and the Institute of Biomaterials and Biomedical Engineering, University of Toronto, 164 College Street, Toronto, ON M5S 3G9, Canada (e-mail: elias.guestrin@utoronto.ca).

M. Eizenman is with the Departments of Electrical and Computer Engineering and Ophthalmology, and the Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, ON M5S 3G9, Canada (e-mail: eizenm@ecf.utoronto.ca).

Digital Object Identifier 10.1109/TBME.2005.863952

There are two main classes of gaze estimation systems: head-mounted systems and head-free or remote systems [13]. In head-mounted systems, gaze direction is measured relative to the head. In order to calculate the POG in space, the three-dimensional (3-D) head pose (position and orientation) has to be estimated. Various types of transducers can be used to measure head pose, of which the most common is the magnetic position transducer [14]. Another approach involves the use of a head-mounted camera that is used to record the scene in front of the subject. Visual cues extracted from images obtained by the scene camera are used to determine the head pose relative to the observed scene [15].

Even though head-mounted gaze estimation systems are preferred for applications that require large and fast head movements, they cannot be used in applications that require continuous gaze monitoring over long periods of time (e.g., aids for motor-disabled persons, monitoring driver's behavior) or in applications that involve infants. For these applications, remote gaze estimation systems are preferred.

Most modern approaches to remote gaze estimation are based on the analysis of eye features and, sometimes, head features extracted from video images. One approach consists of tracking facial features to estimate the 3-D head pose and thus derive the positions of the center of rotation of the eyes [16], [17]. By combining this information with the estimated positions of the iris or pupil centers, the POG can be calculated. Another approach uses the perspective projection of the iris-sclera boundary (limbus) to estimate the position and orientation of the eye in space in order to calculate the POG [18], [19]. The most common approach to remote POG estimation uses the estimates of the centers of the pupil and one or more corneal reflections [7], [11], [13], [20]–[23]. The corneal reflections (first Purkinje images, glints) are virtual images of light sources (usually infrared) that illuminate the eye and are created by the front surface of the cornea, which acts as a convex mirror. The pupil center and corneal reflection(s) have been used in gaze estimation systems for over 40 years but a general theory that applies to all possible system configurations and explains the performance, limitations and potential of such systems, is not available.

The following section presents a general mathematical model for remote gaze estimation systems that utilize the estimates of the centers of the pupil and one or more corneal reflections extracted from video images. The general model covers the full range of possible system configurations, from the simplest, that includes one video camera and one light source, to the most complex, that include multiple cameras and multiple light sources. System configurations are described in order of increasing complexity while highlighting the benefits of the added complexity. Section III describes the details of a specific system implementation that can be used to estimate the POG

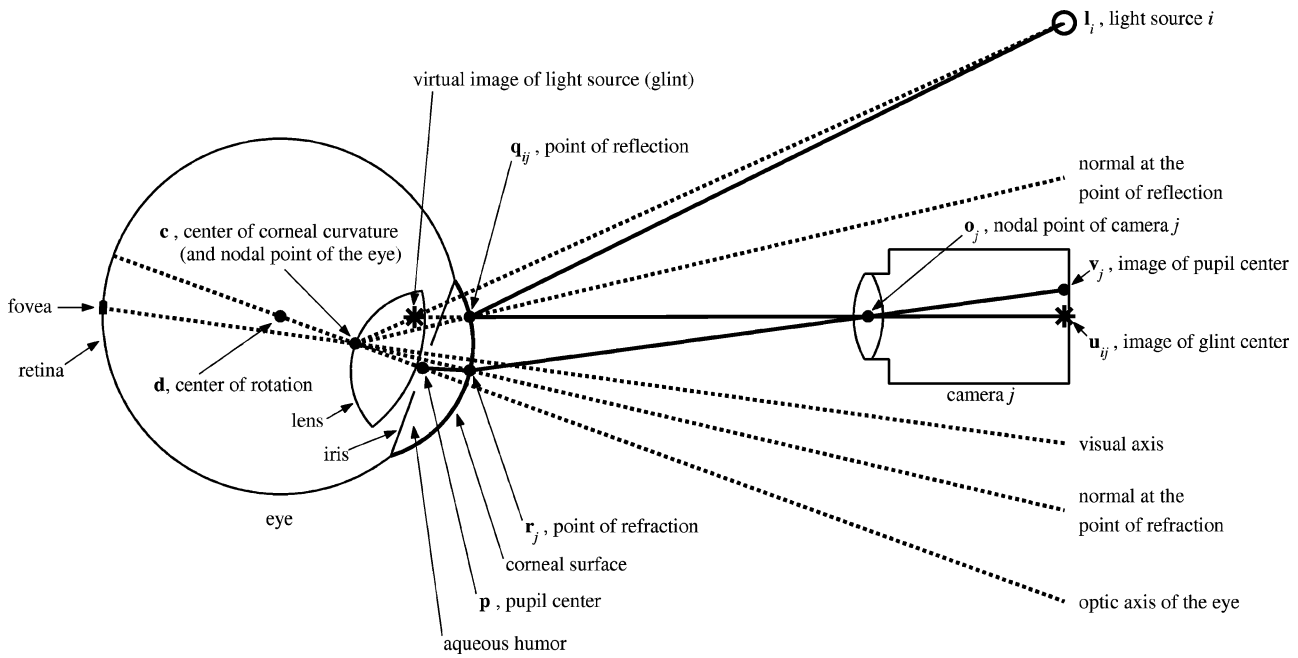

Fig. 1. Ray-tracing diagram (not to scale in order to be able to show all the elements of interest), showing schematic representations of the eye, a camera, and a light source.

while allowing for free head movements. Section IV provides a brief summary of the conclusions of this work.

II. Mathematical Model

This section presents a general model for video-based remote POG estimation using the coordinates of the centers of the pupil and one or more corneal reflections (glints) estimated from images captured by one or more video cameras. The POG is formally defined as the intersection of the visual axes of both eyes with the 3-D scene. The visual axis is the line connecting the center of the fovea with the nodal point[1] of the eye's optics (Fig. 1). Since in the human eye the visual axis deviates from the optic axis [13], the development that follows is divided into two parts. The first part considers the problem of reconstructing the optic axis of the eye from the centers of pupil and glint(s) in the images of the eye. The second part deals with the reconstruction of the visual axis from the optic axis, and the estimation of the POG.

Under the assumptions that the light sources are modeled as point sources and the video cameras are modeled as pinhole cameras, Fig. 1 presents a ray-tracing diagram of the system and the eye, where all points are represented as 3-D column vectors (bold font) in a right-handed Cartesian world coordinate system (WCS). Consider a ray that comes from light source i, $\mathbf{l}_i$, and reflects at a point $\mathbf{q}_{ij}$ on the corneal surface such that the reflected ray passes through the nodal point[2] of camera j, $\mathbf{o}_j$, and intersects the camera image plane at a point $\mathbf{u}_{ij}$. The condition that the ray coming from the point of reflection $\mathbf{q}_{ij}$ and passing through the nodal point of camera j, $\mathbf{o}_j$, intersects the camera image plane at point $\mathbf{u}_{ij}$, can be expressed in parametric form as

$$\mathbf{q}_{ij} = \mathbf{o}_j + k_{q,ij}(\mathbf{o}_j - \mathbf{u}_{ij}) \text{ for some } k_{q,ij} \tag{1}$$

whereas, if the corneal surface is modeled as a convex spherical mirror of radius R, the condition that $\mathbf{q}_{ij}$ is on the corneal surface can be written as

$$\|\mathbf{q}_{ij} - \mathbf{c}\| = R \tag{2}$$

where $\mathbf{c}$ is the center of corneal curvature.

The law of reflection states two conditions: 1) the incident ray, the reflected ray and the normal at the point of reflection are coplanar; 2) the angles of incidence and reflection are equal. Since vector $(\mathbf{q}_{ij} - \mathbf{c})$ is normal to the spherical surface at the point of reflection $\mathbf{q}_{ij}$, condition 1) implies that points $\mathbf{l}_i$, $\mathbf{q}_{ij}$, $\mathbf{c}$ and $\mathbf{o}_j$ are coplanar. Noting that three coplanar vectors $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$ satisfy $\mathbf{a}_1 \times \mathbf{a}_2 \bullet \mathbf{a}_3 = 0$, condition 1) can be formalized as

$$(\mathbf{l}_i - \mathbf{o}_j) \times (\mathbf{q}_{ij} - \mathbf{o}_j) \bullet (\mathbf{c} - \mathbf{o}_j) = 0. \tag{3}$$

Since the angle θ between two vectors $\mathbf{a}$ and $\mathbf{b}$ can be obtained from $\mathbf{a} \bullet \mathbf{b} = \|\mathbf{a}\|\|\mathbf{b}\| \cos\theta$, condition 2) can be expressed as

$$(\mathbf{l}_i - \mathbf{q}_{ij}) \bullet (\mathbf{q}_{ij} - \mathbf{c}) \cdot \|\mathbf{o}_j - \mathbf{q}_{ij}\| = (\mathbf{o}_j - \mathbf{q}_{ij}) \bullet (\mathbf{q}_{ij} - \mathbf{c}) \cdot \|\mathbf{l}_i - \mathbf{q}_{ij}\|. \tag{4}$$

Next, consider a ray that comes from the pupil center, $\mathbf{p}$, and refracts at point $\mathbf{r}_j$ on the corneal surface such that the refracted

[1]The nodal point of an optical system is the point on the optic axis where all lines that join object points with their respective image points intersect.

[2]The nodal point of a camera is also known as center of projection, camera center, and, sometimes, lens center.

ray passes through the nodal point of camera j, $\mathbf{o}_j$, and intersects the camera image plane at a point $\mathbf{v}_j$. The condition that the ray coming from the point of refraction $\mathbf{r}_j$ and passing through the nodal point of camera j, $\mathbf{o}_j$, intersects the camera image plane at point $\mathbf{v}_j$, can be expressed in parametric form as

$$\mathbf{r}_j = \mathbf{o}_j + k_{r,j}(\mathbf{o}_j - \mathbf{v}_j) \text{ for some } k_{r,j} \tag{5}$$

whereas the condition that $\mathbf{r}_j$ is on the corneal surface can be written as

$$||\mathbf{r}_j - \mathbf{c}|| = R. \tag{6}$$

The law of refraction states two conditions: 1) the incident ray, the refracted ray and the normal at the point of refraction are coplanar; 2) the angle of incidence, θ_1, and the angle of refraction, θ_2, satisfy Snell's law (i.e., $n_1 \sin\theta_1 = n_2 \sin\theta_2$, where n_1 and n_2 are the indices of refraction of mediums 1 and 2). Since vector $(\mathbf{r}_j - \mathbf{c})$ is normal to the spherical surface at the point of refraction $\mathbf{r}_j$, condition 1) implies that points $\mathbf{p}$, $\mathbf{r}_j$, $\mathbf{c}$ and $\mathbf{o}_j$ are coplanar, which can be formalized as

$$(\mathbf{r}_j - \mathbf{o}_j) \times (\mathbf{c} - \mathbf{o}_j) \bullet (\mathbf{p} - \mathbf{o}_j) = 0. \tag{7}$$

Since the sine of the angle θ between two vectors $\mathbf{a}$ and $\mathbf{b}$ can be obtained from $||\mathbf{a} \times \mathbf{b}|| = ||\mathbf{a}||||\mathbf{b}|| \sin\theta$, condition 2) can be expressed as

$$n_1 \cdot ||(\mathbf{r}_j - \mathbf{c}) \times (\mathbf{p} - \mathbf{r}_j)|| \cdot ||\mathbf{o}_j - \mathbf{r}_j|| = n_2 \cdot ||(\mathbf{r}_j - \mathbf{c}) \times (\mathbf{o}_j - \mathbf{r}_j)|| \cdot ||\mathbf{p} - \mathbf{r}_j|| \tag{8}$$

where n_1 is the effective index of refraction of the aqueous humor and cornea combined and n_2 is the index of refraction of air ($\approx$1). In this model, the refraction at the aqueous humor-cornea interface is neglected since the difference in their indices of refraction is small relative to that of the cornea-air interface. Only the refraction at the cornea-air interface is taken into account and the aqueous humor and cornea are considered as a homogenous medium.

Finally, considering the distance K between the pupil center and the center of corneal curvature leads to

$$||\mathbf{p} - \mathbf{c}|| = K. \tag{9}$$

Since the optic axis of the eye passes through the pupil center $(\mathbf{p})$ and the center of corneal curvature $(\mathbf{c})$, if the above system of equations is solved for $\mathbf{c}$ and $\mathbf{p}$, the optic axis of the eye in space can be reconstructed as the line defined by these two points. Notice that in order to solve the above system of equations, the eye parameters R, K and n_1 have to be known. These eye parameters are subject-specific and are not easily measured directly. Therefore, in general, they are obtained through a calibration procedure that is performed for each subject (an example is provided in Section III). The typical values of these eye parameters are given in Appendix A.

Since the POG is defined as the intersection of the visual axis rather than the optic axis with the scene, the relation between these two axes has to be modeled. The visual axis is the line defined by the nodal point of the eye[3] and the center of the fovea (i.e., the highest acuity region of the retina corresponding to 0.6 to 1° of visual angle), and deviates from the optic axis [13] (Fig. 1). In a typical adult human eye, the fovea falls about 4–5° temporally and about 1.5° below the point of intersection of the optic axis and the retina [24].

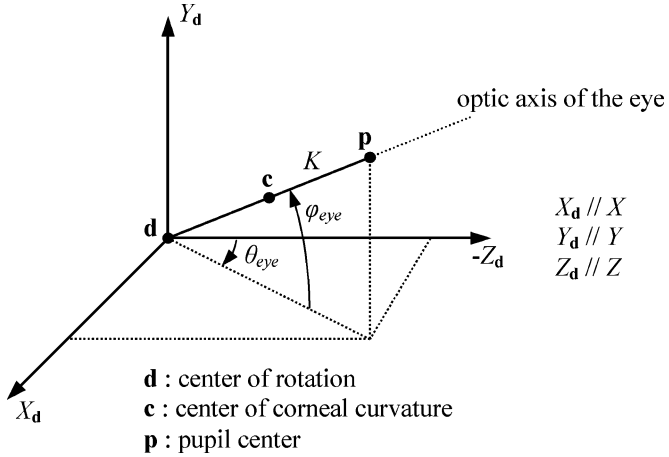

Fig. 2. Orientation of the optic axis of the eye.

In order to formalize the relation between the visual and optic axes, suppose that the scene is a vertical plane (e.g., a projection screen or computer monitor) and that the WCS is a right-handed 3-D Cartesian coordinate system whose XY-plane is coincident with the scene plane, with the X-axis horizontal, the Y-axis vertical and the positive Z-axis coming out of the scene plane. Then, the orientation of the optic axis of the eye can be described by the pan (horizontal) angle θ_{eye} and the tilt (vertical) angle φ_{eye} defined in Fig. 2, where the WCS is translated to the center of rotation of the eye, $\mathbf{d}$. As it can be derived from this figure, the angles θ_{eye} and φ_{eye} can be obtained from $\mathbf{c}$ and $\mathbf{p}$ by solving the following equation:

$$\frac{\mathbf{p} - \mathbf{c}}{||\mathbf{p} - \mathbf{c}||} = \begin{bmatrix} \cos\varphi_{\text{eye}} \sin\theta_{\text{eye}} \\ \sin\varphi_{\text{eye}} \\ -\cos\varphi_{\text{eye}} \cos\theta_{\text{eye}} \end{bmatrix}. \tag{10}$$

If the horizontal and vertical angles between the visual and optic axes are given by α_{eye} and β_{eye}, respectively, the orientation of the visual axis can be expressed by the pan angle $(\theta_{\text{eye}} + \alpha_{\text{eye}})$ and the tilt angle $(\varphi_{\text{eye}} + \beta_{\text{eye}})$, where all angles are signed. In particular, $\alpha_{\text{eye}} < 0$ for the right eye while $\alpha_{\text{eye}} > 0$ for the left eye, and $\beta_{\text{eye}} > 0$. The eye parameters α_{eye} and β_{eye} are subject-specific and are usually estimated through a calibration procedure that is performed for each subject. The typical values of these two eye parameters are included in Appendix A.

To completely define the visual axis in space, in addition to its orientation, a point through which it passes is required. The visual axis and the optic axis intersect at the nodal point of the eye. Since the nodal point remains within 1 mm of the center of corneal curvature for different degrees of eye accommodation [13], for the sake of simplicity, the nodal point is assumed to be coincident with the center of corneal curvature $(\mathbf{c})$.

[3]Actually, the eye has two nodal points, 0.3 mm apart. For the sake of simplicity, a single nodal point is considered.

From the above discussion, it follows that the visual axis can be then described in parametric form as

$$\mathbf{g} = \mathbf{c} + k_g \begin{bmatrix} \cos(\varphi_{\text{eye}} + \beta_{\text{eye}}) \sin(\theta_{\text{eye}} + \alpha_{\text{eye}}) \\ \sin(\varphi_{\text{eye}} + \beta_{\text{eye}}) \\ -\cos(\varphi_{\text{eye}} + \beta_{\text{eye}}) \cos(\theta_{\text{eye}} + \alpha_{\text{eye}}) \end{bmatrix} \quad (11)$$

for all k_g. Since it was assumed that the scene plane is at $Z = 0$, the POG is given by (11) for a value of k_g such that the Z-component of $\mathbf{g}$, g_Z, equals 0, that is

$$k_g = \frac{c_Z}{\cos(\varphi_{\text{eye}} + \beta_{\text{eye}}) \cos(\theta_{\text{eye}} + \alpha_{\text{eye}})}. \quad (12)$$

In the remainder of this Section, it is assumed that the world coordinates of the positions of the light sources $(\mathbf{l}_i)$, the nodal point(s) of the camera(s) $(\mathbf{o}_j)$ and the centers of the pupil $(\mathbf{v}_j)$ and glints $(\mathbf{u}_{ij})$ in the eye images, are known. Since the centers of the pupil and glints that are estimated in each eye image are measured in pixels in an image coordinate system (ICS), they have to be transformed into world coordinates (Appendix B). In order to transform from image coordinates into world coordinates, all camera parameters, including the position of the nodal point $(\mathbf{o}_j)$, must be known. Typically, the camera parameters are estimated through a camera calibration procedure [25], whereas the positions of the light sources are measured directly. In general, the system structure is fixed and, hence, these system parameters are measured/estimated only once during system set up.

The above development shows that 1) the reconstruction of the optic axis of the eye as the line defined by the center of corneal curvature $(\mathbf{c})$ and the pupil center $(\mathbf{p})$, using (1)–(9), depends on the system configuration (i.e., number of cameras and light sources) and 2) once the optic axis of the eye is obtained, the reconstruction of the visual axis and the estimation of the POG, using (10)–(12), are independent of the system configuration. For this reason, the following subsections concentrate on the reconstruction of the optic axis of the eye for different system configurations, which are presented in order of increasing complexity. The purpose of increasing system complexity is to relax constraints on subject's head movements and simplify the calibration procedure.

A. One Camera and One Light Source

The simplest system configuration consists of a single camera and a single light source. In this case, if the eye parameters R, K and n_1 are known, the system of equations (1)–(9) with $i = 1$ and $j = 1$, is equivalent to 13 scalar equations with 14 scalar unknowns. This means that the problem cannot be solved unless another constraint such as

$$\|\mathbf{o}_1 - \mathbf{c}\| = known \quad (13)$$

is introduced. This constraint can be satisfied if the head is fixed relative to the system or if the distance between the eye and the camera is estimated somehow (e.g., magnetic head tracker, ultrasonic transducer, auto-focus system, etc.).

In general, gaze estimation systems that use one corneal reflection and one light source do not solve the above system of equations but rather use the vector from the pupil center to the corneal reflection in the eye image to compute the gaze direction relative to the camera axis [13], and either assume that the head movements are negligible or have means to estimate the position of the eye in space (e.g., combination of a moving camera or moving mirrors that track the eye and an auto-focus system or an ultrasonic transducer) [7], [11], [21]–[23]. However, the above system of equations demonstrates the limitations of the single camera-single light source configuration when the head is not completely stationary. The next subsection presents the simplest configuration that allows for the estimation of the POG from the centers of pupil and glints, without any constraints on head movements and without using any additional device to estimate the position of the eye in space.

B. One Camera and Multiple Light Sources

The use of multiple light sources allows for the solution of the system of equations (1)–(9) with $i = 1, \ldots, N$ and $j = 1$, if the eye parameters (R, K and n_1) are obtained through a calibration procedure. In this case, it is advantageous to substitute (1) into (3) to obtain

$$\underbrace{(\mathbf{l}_i - \mathbf{o}_j) \times (\mathbf{u}_{ij} - \mathbf{o}_j)}_{\substack{\text{normal to the plane defined by} \\ \mathbf{l}_i, \mathbf{o}_j \text{ and } \mathbf{u}_{ij}}} \bullet (\mathbf{c} - \mathbf{o}_j) = 0. \quad (14)$$

This equation means that the center of corneal curvature, $\mathbf{c}$, belongs to each plane defined by the nodal point of camera j, $\mathbf{o}_j$, light source i, $\mathbf{l}_i$, and its corresponding image point $\mathbf{u}_{ij}$. Moreover, for each camera j, all those planes intersect at the line defined by points $\mathbf{c}$ and $\mathbf{o}_j$. Since in this case there is only one camera, the subscript that identifies the camera can be dropped for simplicity of notation and, by noting that $\mathbf{a} \bullet \mathbf{b} = \mathbf{a}^T \mathbf{b}$, (14), $i = 1, \ldots, N$ can be written in matrix form as

$$\underbrace{\begin{bmatrix} [(\mathbf{l}_1 - \mathbf{o}) \times (\mathbf{u}_1 - \mathbf{o})]^T \\ [(\mathbf{l}_2 - \mathbf{o}) \times (\mathbf{u}_2 - \mathbf{o})]^T \\ \vdots \\ [(\mathbf{l}_N - \mathbf{o}) \times (\mathbf{u}_N - \mathbf{o})]^T \end{bmatrix}}_{\mathbf{M}} (\mathbf{c} - \mathbf{o}) = \mathbf{0}. \quad (15)$$

From the interpretation of (14) it follows that matrix $\mathbf{M}$ has, at most, rank 2. If $\mathbf{M}$ has rank 2, the solution to (15) is given by an equation of the form

$$\mathbf{c} - \mathbf{o} = k_{c,b} \mathbf{b}_{\text{norm}} \quad (16)$$

which defines vector $(\mathbf{c} - \mathbf{o})$ up to a scale factor. From this reasoning, it follows that (1), (2), (4), $i = 1, \ldots, N$, $j = 1$, and (16) are equivalent to $(5N + 3)$ scalar equations with $(4N + 4)$ scalar unknowns. In particular, when $N = 2$, $\mathbf{b}_{\text{norm}}$ is a unit vector in the direction of the line of intersection of the planes whose normals are given by $[(\mathbf{l}_1 - \mathbf{o}) \times (\mathbf{u}_1 - \mathbf{o})]$ and $[(\mathbf{l}_2 - \mathbf{o}) \times (\mathbf{u}_2 - \mathbf{o})]$, thus

$$\mathbf{b}_{\text{norm}} = \frac{\mathbf{b}}{\|\mathbf{b}\|}$$
$$\mathbf{b} = [(\mathbf{l}_1 - \mathbf{o}) \times (\mathbf{u}_1 - \mathbf{o})] \times [(\mathbf{l}_2 - \mathbf{o}) \times (\mathbf{u}_2 - \mathbf{o})] \quad (17)$$

and (1), (2), (4), $i = 1, 2$, and (16) are equivalent to 13 scalar equations with 12 scalar unknowns.

In the special case that $\mathbf{M}$ in (15) has rank 1 [$\mathbf{b} = \mathbf{0}$ in (17)], which means that all normals given by $(\mathbf{l}_i - \mathbf{o}) \times (\mathbf{u}_i - \mathbf{o})$ are parallel, the effective number of scalar equations decreases to $(5N + 2)$. In the case that $N = 2$, it results in the equivalent to 12 scalar equations with 12 scalar unknowns.

Consequently, if multiple light sources are used, there are enough equations to solve for the center of corneal curvature $\mathbf{c}$. Knowing $\mathbf{c}$, (5) and (6) are used to compute the point of refraction $\mathbf{r} = \mathbf{r}_1$ (4 scalar unknowns and 4 scalar equations). Knowing $\mathbf{c}$ and $\mathbf{r}$, (7)–(9) are used to compute $\mathbf{p}$ (3 scalar unknowns and 3 scalar equations). Knowing $\mathbf{c}$ and $\mathbf{p}$, the optic axis of the eye can be reconstructed as the line defined by these two points. Notice that the eye parameters R, K and n_1 must be known in order to reconstruct the optic axis of the eye and thus be able to estimate the POG.

The above discussion shows that one camera and two light sources is the simplest configuration that allows for the reconstruction of the optic axis of the eye from the centers of pupil and glints while allowing for free head movements. The above analysis also shows that knowing $\mathbf{c}$ (the center of corneal curvature), the calculation of $\mathbf{p}$ (the pupil center) is independent of the number of light sources (7 scalar equations and 7 scalar unknowns regardless of the number of light sources). In the next subsection, system configurations that allow for the reconstruction of the optic axis of the eye without the need for a subject-specific calibration procedure are discussed.

C. Multiple Cameras and Multiple Light Sources

When multiple cameras and multiple light sources are used, it is possible to discard all equations that contain the eye parameters R, K, and n_1, while still being able to reconstruct the optic axis of the eye by using the remaining equations. In order to keep the notation simple, the discussion that follows is carried out for two cameras, noting that the extension to more cameras is trivial. When two cameras and multiple light sources are used, (14), $i = 1, \ldots, N$ and $j = 1, 2$, can be written in matrix form as

$$\underbrace{\begin{bmatrix} [(\mathbf{l}_1 - \mathbf{o}_1) \times (\mathbf{u}_{11} - \mathbf{o}_1)]^T \\ [(\mathbf{l}_1 - \mathbf{o}_2) \times (\mathbf{u}_{12} - \mathbf{o}_2)]^T \\ \vdots \\ [(\mathbf{l}_N - \mathbf{o}_1) \times (\mathbf{u}_{N1} - \mathbf{o}_1)]^T \\ [(\mathbf{l}_N - \mathbf{o}_2) \times (\mathbf{u}_{N2} - \mathbf{o}_2)]^T \end{bmatrix}}_{\mathbf{M}_2} \mathbf{c} = \underbrace{\begin{bmatrix} (\mathbf{l}_1 - \mathbf{o}_1) \times (\mathbf{u}_{11} - \mathbf{o}_1) \bullet \mathbf{o}_1 \\ (\mathbf{l}_1 - \mathbf{o}_2) \times (\mathbf{u}_{12} - \mathbf{o}_2) \bullet \mathbf{o}_2 \\ \vdots \\ (\mathbf{l}_N - \mathbf{o}_1) \times (\mathbf{u}_{N1} - \mathbf{o}_1) \bullet \mathbf{o}_1 \\ (\mathbf{l}_N - \mathbf{o}_2) \times (\mathbf{u}_{N2} - \mathbf{o}_2) \bullet \mathbf{o}_2 \end{bmatrix}}_{\mathbf{h}} \tag{18}$$

after applying the distributive property for the dot product, rearranging terms and noting that $\mathbf{a} \bullet \mathbf{b} = \mathbf{a}^T\mathbf{b}$. If $\mathbf{M}_2$ has rank 3, $\mathbf{c}$ can be obtained from (18) by using the left pseudoinverse of $\mathbf{M}_2$ as

$$\mathbf{c} = \left(\mathbf{M}_2^T\mathbf{M}_2\right)^{-1}\mathbf{M}_2^T\mathbf{h}. \tag{19}$$

If only 3 linearly independent rows of $\mathbf{M}_2$ and the corresponding rows of $\mathbf{h}$ are considered in (18), then (19) reduces to $\mathbf{c} = \mathbf{M}_2^{-1}\mathbf{h}$.

Note that when (5) and (7) are combined, they correspond to the physical condition that for each camera (refer also to Fig. 1), the pupil center $(\mathbf{p})$, the point of refraction $(\mathbf{r}_j)$, the nodal point of the camera $(\mathbf{o}_j)$, the image of the pupil center $(\mathbf{v}_j)$, and the center of corneal curvature $(\mathbf{c})$ are coplanar. For this system configuration with two cameras, it is convenient to represent this physical condition as

$$\begin{aligned} (\mathbf{o}_1 - \mathbf{v}_1) \times (\mathbf{c} - \mathbf{o}_1) \bullet (\mathbf{p} - \mathbf{c}) &= 0 \\ (\mathbf{o}_2 - \mathbf{v}_2) \times (\mathbf{c} - \mathbf{o}_2) \bullet (\mathbf{p} - \mathbf{c}) &= 0. \end{aligned} \tag{20}$$

Since the optic axis of the eye is defined by points $\mathbf{c}$ and $\mathbf{p}$, these equations mean that the optic axis of the eye belongs to the plane defined by points $\mathbf{c}$, $\mathbf{o}_1$ and $\mathbf{v}_1$ (normal given by $[(\mathbf{o}_1 - \mathbf{v}_1) \times (\mathbf{c} - \mathbf{o}_1)]$) and to the plane defined by points $\mathbf{c}$, $\mathbf{o}_2$ and $\mathbf{v}_2$ (normal given by $[(\mathbf{o}_2 - \mathbf{v}_2) \times (\mathbf{c} - \mathbf{o}_2)]$). Therefore, the optic axis of the eye is the line of the intersection of those two planes and its direction is given by

$$\begin{aligned} \mathbf{s}_{\text{norm}} &= \frac{\mathbf{s}}{\|\mathbf{s}\|} \\ \mathbf{s} &= [(\mathbf{o}_1 - \mathbf{v}_1) \times (\mathbf{c} - \mathbf{o}_1)] \times [(\mathbf{o}_2 - \mathbf{v}_2) \times (\mathbf{c} - \mathbf{o}_2)]. \end{aligned} \tag{21}$$

If $\mathbf{s} \neq \mathbf{0}$, the solution to (20) can be expressed as

$$\mathbf{p} - \mathbf{c} = k_{pc}\mathbf{s}_{\text{norm}} \tag{22}$$

which defines vector $(\mathbf{p} - \mathbf{c})$ up to a scale factor (notice that $|k_{pc}| = K$, the distance between the pupil center and the center of corneal curvature). In this way, knowing $\mathbf{c}$ and the direction of vector $(\mathbf{p} - \mathbf{c})$, i.e., the direction of vector $\mathbf{s}$, the optic axis of the eye can be reconstructed without actually knowing the eye parameters R, K, and n_1.

Equation (22) is only valid when $\mathbf{s} \neq \mathbf{0}$. The following discussion considers the conditions that result in $\mathbf{s} = \mathbf{0}$. To have $\mathbf{s} = \mathbf{0}$, it is sufficient that $(\mathbf{o}_1 - \mathbf{v}_1) \times (\mathbf{c} - \mathbf{o}_1) = \mathbf{0}$ or $(\mathbf{o}_2 - \mathbf{v}_2) \times (\mathbf{c} - \mathbf{o}_2) = \mathbf{0}$, or that $[(\mathbf{o}_1 - \mathbf{v}_1) \times (\mathbf{c} - \mathbf{o}_1)]$ and $[(\mathbf{o}_2 - \mathbf{v}_2) \times (\mathbf{c} - \mathbf{o}_2)]$ are parallel. The condition that $(\mathbf{o}_j - \mathbf{v}_j) \times (\mathbf{c} - \mathbf{o}_j) = \mathbf{0}$ implies that the center of corneal curvature, $\mathbf{c}$, the nodal point of the camera, $\mathbf{o}_j$, and the image of the pupil center, $\mathbf{v}_j$, are collinear. Since any line passing through point $\mathbf{c}$ is normal to the spherical corneal surface, the line defined by points $\mathbf{o}_j$ and $\mathbf{v}_j$ is normal to the corneal surface. Since points $\mathbf{o}_j$ and $\mathbf{v}_j$ are on the refracted ray coming from the pupil center, $\mathbf{p}$, the refracted ray is normal to the corneal surface (i.e., there is no refraction) and, therefore, point $\mathbf{p}$ is collinear with points $\mathbf{o}_j$, $\mathbf{v}_j$ and $\mathbf{c}$. Since the optic axis of the eye is defined by points $\mathbf{c}$ and $\mathbf{p}$, it implies that the optic axis goes through the nodal point of the camera, $\mathbf{o}_j$. In summary, if the optic axis

of the eye passes through the nodal point of camera j, $\mathbf{o}_j$, then $(\mathbf{o}_j - \mathbf{v}_j) \times (\mathbf{c} - \mathbf{o}_j) = \mathbf{0}$ and, hence, $\mathbf{s} = \mathbf{0}$.

The condition that $[(\mathbf{o}_1 - \mathbf{v}_1) \times (\mathbf{c} - \mathbf{o}_1)]$ and $[(\mathbf{o}_2 - \mathbf{v}_2) \times (\mathbf{c} - \mathbf{o}_2)]$ are parallel implies that points $\mathbf{c}$, $\mathbf{o}_1$, $\mathbf{v}_1$, $\mathbf{o}_2$ and $\mathbf{v}_2$ are all on a single plane. Equation (20) implies that the optic axis of the eye, defined by the center of corneal curvature, $\mathbf{c}$, and the pupil center, $\mathbf{p}$, is also in that plane. Consequently, this situation occurs when the optic axis of the eye is coplanar with the line defined by the nodal points of the cameras, $\mathbf{o}_1$ and $\mathbf{o}_2$.

Since the condition that the optic axis of the eye passes through the nodal point of a camera is a particular case of the condition that the optic axis of the eye is coplanar with the line connecting the nodal points of the cameras, and since in practice the condition that the optic axis of the eye is parallel to the line connecting the nodal points of the cameras is unrealistic, the above discussion can be summarized by saying that $\mathbf{s} \neq \mathbf{0}$ and, hence, (22) is valid as long as the optic axis of the eye does not intersect the line defined by the nodal points of the cameras.

From the discussion in this section, it follows that the simplest configuration that allows for the reconstruction of the optic axis of the eye, without knowledge of the values of the subject-specific eye parameters R, K, and n_1, consists of two cameras and two light sources [26]. In order to reconstruct the visual axis of the eye and thus be able to estimate the POG, the eye parameters α_{eye} and β_{eye} still need to be estimated. These two parameters, which are subject-specific, can be estimated through a simple calibration procedure in which the subject is required to fixate on a single point. A single point calibration could be performed even with infants by presenting a flashing object to attract their attention.

Section III provides a detailed description of a specific POG estimation system.

III. Experimental Results

This section presents a specific system implementation that is used to estimate the POG on a computer screen. The system utilizes two near-infrared (850 nm) light sources that are symmetrically positioned at the sides of a 19-in computer monitor, and a video camera (640 × 480 pixels, 1/3-in charge-coupled device with a 35-mm lens) that is centered under the screen. A typical image from the video camera for a subject sitting at a distance of 65 cm from the monitor (typical viewing distance), with his head approximately at the center of the region of allowed head movement, is shown in Fig. 3. This specific system can tolerate only moderate head movements of about ±3 cm laterally, ±2 cm vertically, and ±4 cm backward/forward, before the eye features are no longer in the field of view of the camera or are out of focus.

To estimate the POG on the screen, a set of system and subject-specific eye parameters has to be measured/estimated. Since the system components are fixed relative to the computer monitor, the system parameters (the position of the two light sources, $\mathbf{l}_1$ and $\mathbf{l}_2$, and the extrinsic and intrinsic camera parameters, which include the nodal point of the camera, $\mathbf{o} = \mathbf{o}_1$) are measured/estimated only once during system set up. The subject-specific eye parameters (R, K, n_1, α_{eye}, and β_{eye}) are obtained through a calibration procedure that is performed once for each subject. In the calibration procedure, the subject fixates on 9 evenly distributed points that are presented sequentially on the screen. For each fixation point, 100 estimates of the image coordinates of the centers of pupil and glints are obtained and the average coordinates of these features are computed. Using the average coordinates of the centers of pupil and glints, the eye parameters are optimized to minimize the sum of the square errors between the points on the screen and the estimated points-of-gaze [27]. The initial guess for the optimization corresponds to the typical values of the eye parameters given in Appendix A. During the calibration procedure the head is positioned at the center of the region of allowed head movements (central position).

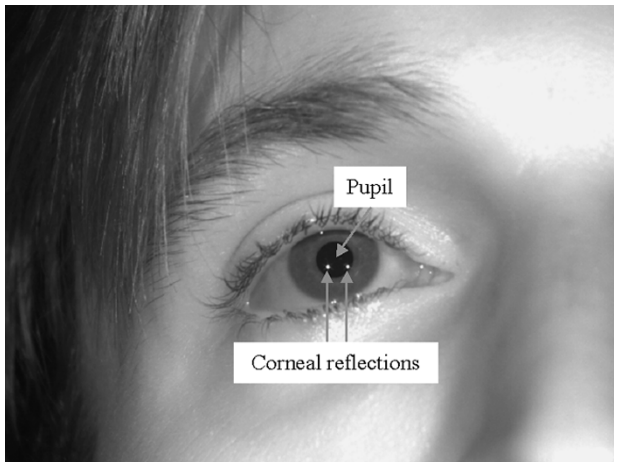

Fig. 3. Sample eye image showing the pupil and the two corneal reflections (glints).

To estimate the POG, the coordinates of the centers of pupil and glints are first estimated in each image captured by the video camera [28]. These image coordinates are then transformed into world coordinates (Appendix B) as $\mathbf{v} = \mathbf{v}_1$ for the pupil center, and $\mathbf{u}_1 = \mathbf{u}_{11}$ and $\mathbf{u}_2 = \mathbf{u}_{21}$ for the glints. Next, the center of corneal curvature, $\mathbf{c}$, is calculated from (1), (2), (4), $i = 1, 2$, $j = 1$ and (16), (17). Knowing $\mathbf{c}$, (5) and (6) are used to compute the point of refraction $\mathbf{r} = \mathbf{r}_1$. Knowing $\mathbf{c}$ and $\mathbf{r}$, (7)–(9) are used to compute $\mathbf{p}$. Knowing $\mathbf{c}$ and $\mathbf{p}$, the optic axis of the eye in space is reconstructed as the line defined by these two points. Finally, using (10)–(12), the visual axis of the eye is obtained and the POG on the screen is estimated.

A preliminary evaluation of the performance of this POG estimation system was carried out through experiments with 4 subjects. In these experiments, the head of each subject was placed in the central position and 4 positions at the edges of the region of allowed head movements. The 4 edge positions correspond to lateral and backward/forward head displacements. For each head position, the subject was asked to fixate on 9 points on the computer screen and 100 estimates (≈3.3 s at 30 estimates/s) of the POG were obtained for each fixation point. The resulting root-mean-square (RMS) errors in the estimation of the POG for the central position and the edge positions are summarized in Table I (RMS error). Table I also shows the RMS errors when the POG was estimated for the average of the coordinates of the centers of pupil and glints (ACPG-RMS error). A typical example of POG estimation errors for the central head position is shown in Fig. 4. The ACPG-RMS errors (Table I) correspond to the deviation of the white crosses from the centers of the dotted circles in Fig. 4 and are the result of bias in the estimation of the POG. The dispersion of the asterisks around the white crosses is

TABLE I
EXPERIMENTAL RMS POINT-OF-GAZE ESTIMATION ERRORS

Subject	Head position(s)	ACPG-RMS [a] error (mm)	RMS error (mm)
M. E.	Central	2.93	5.15
	Edges	4.25	6.12
J. K.	Central	4.88	6.41
	Edges	5.33	6.73
R. H.	Central	4.93	7.01
	Edges	5.74	7.58
E. G.	Central	4.68	6.25
	Edges	8.12	9.04

[a]RMS error when the point-of-gaze was estimated for the average coordinates of the centers of pupil and glints.

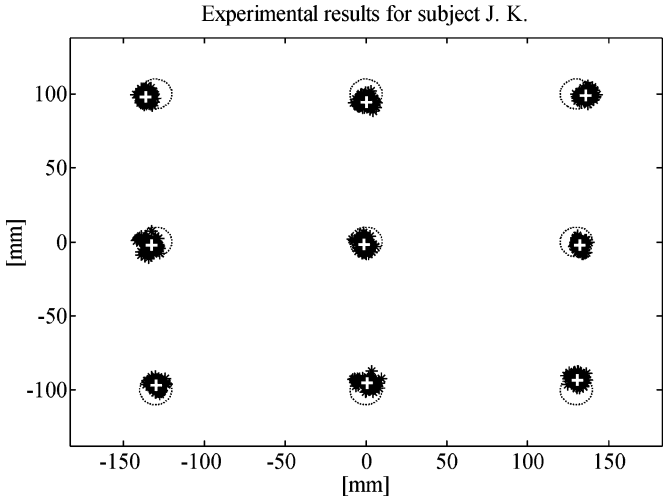

Fig. 4. Experimental POG estimation results for subject J. K. The centers of the dotted circles (10 mm radius) indicate the intended fixation points, the asterisks represent the estimates of the POG, and the white crosses represent the estimates of the POG for the average coordinates of the centers of pupil and glints.

caused by noise in the estimates of the image coordinates of the centers of pupil and glints. The RMS errors shown in the last column of Table I correspond to the combined effects of bias and dispersion of the POG estimates. It can be also observed that the RMS errors for the edge head positions are larger than the RMS errors for the central head position.

In order to understand the observed errors, the effects of differences between the shape of real human corneas and the ideal spherical corneal shape assumed in the model of Section II (corneal asphericity), as well as the effects of noise in the estimation of the centers of pupil and glints in the eye images, were studied through numerical simulations. The effects of corneal asphericity were studied using an ellipsoidal corneal model [29]. In this model, the corneal surface is modeled as a section of an ellipsoid that has one of its axes coincident with the optic axis of the eye and whose cross-sections perpendicular to the optic axis are circular. This ellipsoidal corneal model can be completely characterized by the distance between the apex of the cornea and the center of rotation of the eye (13.1 mm—see Appendix A), the radius of curvature at the apex of the cornea (7.8 mm—see Appendix A) and the radius of curvature of the cornea at the boundary with the sclera (at 6 mm from the optic axis, R_6). Using this ellipsoidal corneal model, the image coordinates of the centers of pupil and glints were computed for the same fixation points and the same head positions that

TABLE II
SIMULATION RMS POINT-OF-GAZE ESTIMATION ERRORS

R_6 (mm)	Head position(s)	NFD-RMS [a] error (mm)	RMS error (mm)
7.8 (spherical)	Central	0	3.53
	Edges	0	3.57
8	Central	0.22	3.61
	Edges	0.26	3.68
9	Central	1.27	3.76
	Edges	1.33	3.79
10	Central	2.35	4.14
	Edges	2.45	4.17
11	Central	3.41	4.67
	Edges	3.55	4.87
12	Central	4.46	5.49
	Edges	4.65	5.70
13	Central	5.49	6.37
	Edges	5.74	6.60

[a]Simulated noise-free image coordinates of the centers of pupil and glints.

were used in the experiments. As in the experiments, the central head position was used to calibrate the eye parameters. The resulting RMS POG estimation errors for different degrees of corneal asphericity (different values of the radius of curvature at the cornea-sclera boundary, R_6) are summarized in Table II (NFD-RMS error). These POG estimation errors are only due to the difference between the ellipsoidal corneal model used to calculate the image coordinates of the centers of pupil and glints and the spherical corneal model (Section II) used to estimate the POG. It is clear from Table II that the POG estimation errors increase with the degree of corneal asphericity. In particular, corneal asphericity results in sensitivity to head displacements, making the RMS error for the edge head positions larger than the RMS error for the central head position. Furthermore, the sensitivity to head displacements also increases with the degree of corneal asphericity. If the cornea were truly spherical, head displacements would have no effect on the POG estimation error.

In order to simulate the effect of noise in the estimates of the centers of pupil and glints in the video images, each coordinate of the centers of pupil and glints obtained with the ellipsoidal corneal model was contaminated with 100 independent realizations of an additive zero-mean Gaussian process with a standard deviation of 0.1 pixel. The properties of the noise were similar to those observed in the system using a stationary artificial eye. As for the noise-free simulations described above, the POG was estimated for the central head position and for the 4 edge head positions. The RMS errors of the POG estimates for different degrees of corneal asphericity are summarized in the last column of Table II (RMS error). Fig. 5 shows simulation results for the central head position and for corneal asphericity $(\mathrm{R}_6 = 11\,\mathrm{mm})$ that produces an error pattern that is similar to that of the example in Fig. 4. The bias of the white crosses from the centers of the dotted circles is only due to corneal asphericity. The dispersion of the asterisks around the white crosses is caused by the simulated noise in the estimation of the image coordinates of the centers of pupil and glints. The combined effects of estimation bias and dispersion result in the RMS errors shown in the last column of Table II. It can be observed that the errors obtained through simulations (Fig. 5) are consistent with the experimental errors (Fig. 4). This example clearly demonstrates

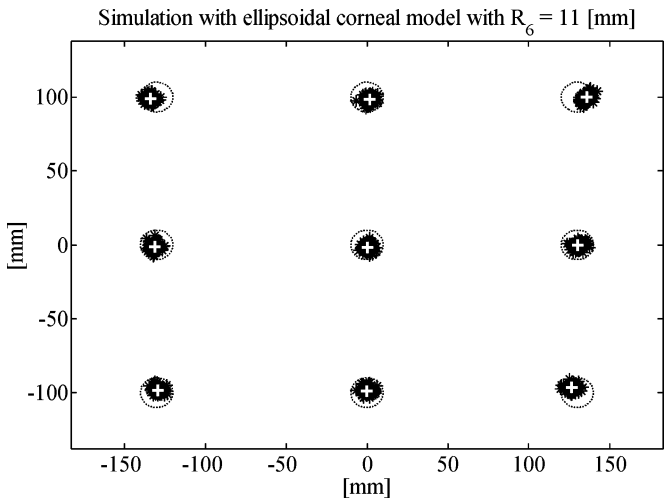

Fig. 5. Simulation results for the ellipsoidal corneal model with $\mathrm{R}_6 = 11\,\mathrm{mm}$. The centers of the dotted circles (10 mm radius) indicate the actual fixation points. The white crosses represent the estimates of the POG for the noise-free data. The asterisks represent the estimates of the POG when zero-mean Gaussian noise with a standard deviation of 0.1 pixel was added to the coordinates of the centers of pupil and glints.

the effects of corneal asphericity and noise in the estimation of the centers of pupil and glints in the video images of the eye on the accuracy of the POG estimation.

Comparison between the RMS errors of the POG estimates for the averaged experimental data (Table I, ACPG-RMS error) and the RMS errors for the noise-free simulated data (Table II, NFD-RMS error) shows that for the central head position, the range of RMS errors for the experimental data is within the range of RMS errors for the simulated data. Since the RMS errors for the averaged experimental data are only marginally affected by noise in the estimates of the centers of pupil and glints, the results in Table II suggest that the RMS errors for the averaged experimental data are mainly due to corneal asphericity.

The differences between the RMS errors for the edge head positions and the RMS errors for the central head position are larger for the experimental data than for the simulated data. This can be attributed to increased bias and noise in the estimation of the coordinates of the centers of the pupil and glints due to nonuniform illumination of the eye. When the head moves with respect to the system, the angle and the intensity of the illumination at the eye changes, resulting in changes in image brightness and contrast. Also, the size and the shape of the glints change for different head positions. The changes in illumination, size and shape translate into biases in the estimation of the centers of pupil and glints and can result in POG estimation errors. Notice that for the system described in this section, when the eye is at a distance of 65 cm from the screen, the pupil center moves about 0.122 pixel per mm shift in the POG, while the glints move about 0.064 pixel/mm. This means that even relatively small biases in the estimation of the centers of pupil and glints can result in relatively large errors in the estimation of the POG.

In the experiments described above, the RMS error of the POG estimation was less than 10 mm for all experimental conditions. This is equivalent to about 0.9° of visual angle when the eye is at a distance of 65 cm from the computer screen.

IV. Conclusion

This paper presented a general theory for remote POG estimation systems that use the coordinates of the centers of the pupil and corneal reflections (glints) estimated from video images. It was shown that as system complexity (i.e., the number of light sources and the number of cameras) increases, the number of subject-specific parameters that have to be estimated through calibration can be reduced and the constraints on head movements can be relaxed.

For a system configuration that consists of one camera and one light source, the POG cannot be determined from the coordinates of the centers of the pupil and the glint, unless the head is stationary or the head position is estimated by some other means.

The simplest configuration that allows for the estimation of the POG, while allowing for free head movements, consists of one camera and two light sources. To estimate the POG with this system configuration, five eye parameters (R, K, n_1, α_{eye} and β_{eye}) have to be estimated through a subject-specific calibration procedure that requires the subject to fixate on multiple points.

A specific system implementation that uses one camera and two light sources to estimate the POG on a computer screen was described in detail. It was shown that the main sources of errors in the estimation of the POG are associated with 1) corneal asphericity (deviation of the shape of real corneas from the ideal spherical cornea assumed in the model); and 2) noise in the estimation of the centers of pupil and glints in the eye images. Experimental results obtained with four subjects showed that by using the general theory, the POG on the computer screen can be estimated with an RMS error of less than 0.9° of visual angle.

If at least two cameras and at least two light sources are used, it is possible to reconstruct the optic axis of the eye without a subject-specific calibration procedure (i.e., calibration-free system). In order to reconstruct the visual axis of the eye and thus be able to estimate the POG, the angular deviation between the optic axis and the visual axis (α_{eye} and β_{eye}) still needs to be known. The angular deviation between the optic and visual axes can be estimated through a simple calibration procedure in which the subject is required to fixate on a single point. A single point calibration can be performed even with infants by presenting a flashing object to attract their attention.

A system with two cameras and multiple light sources is under development. Preliminary simulations for a system with two cameras and two light sources (under similar conditions to those described in Section III) yielded RMS POG estimation errors that were less than 7.75 mm (about 0.68° of visual angle). These preliminary results suggest that it is feasible to implement a POG estimation system that requires only a single-point calibration and has an accuracy of 1° of visual angle.

Appendix A
Typical Values of the Eye Parameters

Table III summarizes the typical values of the eye parameters R, K, n_1, α_{eye}, and β_{eye} found in the literature. The distance between the center of corneal curvature and the center of rotation of the eye was inferred from [13] to be $D = 5.3\,\mathrm{mm}$. From

TABLE III
TYPICAL VALUES OF THE EYE PARAMETERS

Parameter	Description	Typical value	References
R	Radius of corneal curvature	7.8 mm	[13], [30]
K	Distance between the center of the pupil and the center of corneal curvature	4.2 mm	[30]
n_1	Effective index of refraction of the cornea and the aqueous humor combined (a.k.a. Standard Keratometric Index, SKI)	1.3375	[31]
α_{eye}	Horizontal angle between visual and optic axes of the eye	–5° for the right eye, 5° for the left eye	[24], [30]
β_{eye}	Vertical angle between visual and optic axes of the eye	1.5°	[24], [30]

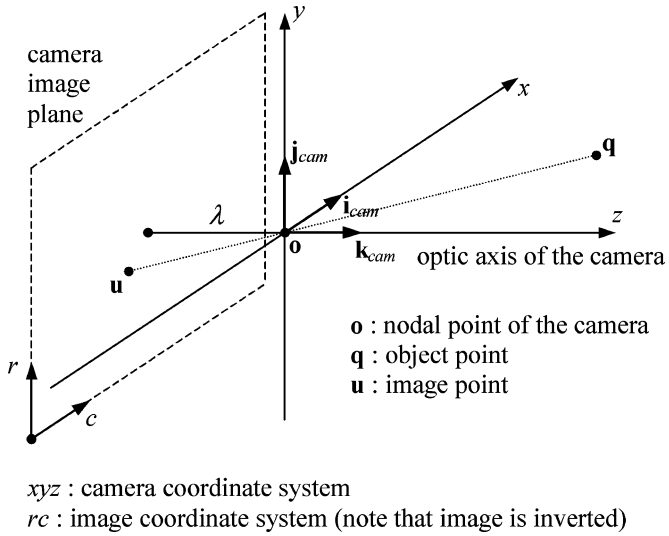

Fig. 6. Perspective projection and relation between the CCS and the ICS.

this value and the value of R from Table III, it follows that the distance between the apex of the cornea and the centre of rotation of the eye is $D + R = 13.1$ mm.

APPENDIX B
TRANSFORMATION FROM IMAGE COORDINATES INTO WORLD COORDINATES

In order to derive the transformation from the ICS to the WCSs, it is convenient to define a camera coordinate system (CCS), as shown in Fig. 6, which also illustrates the perspective projection (i.e., projection of 3-D world points onto the camera image plane). The CCS xyz is a 3-D right-handed Cartesian coordinate system where the x-axis is parallel to the rows of the image sensor, the y-axis is parallel to the columns of the image sensor, the z-axis is perpendicular to the plane of the image sensor, coincident with the optic axis of the camera, and the origin is coincident with the nodal point of the camera $(\mathbf{o})$. All points and vectors (bold font) in this figure are measured with respect to the WCS.

The ICS rc is a two-dimensional Cartesian coordinate system where the coordinates are measured in pixels. Formally, for a point with row coordinate r and column coordinate c in pixels in the ICS, the corresponding coordinates $[x\ y\ z]^T$ in units of length in the CCS are obtained as

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} pixel\,pitch_c(c - c_{\text{center}}) \\ pixel\,pitch_r(r - r_{\text{center}}) \\ -\lambda \end{bmatrix} \tag{23}$$

where c_{center} and r_{center} are, respectively, the column coordinate and row coordinate in pixels of the intersection of the optic axis and the image sensor (the principal point), while $pixel\,pitch_c$ and $pixel\,pitch_r$ are the distance in units of length between adjacent pixels across the columns and across the rows, respectively. The parameter λ is the distance between the nodal point and the image plane. For a pinhole camera, λ equals the focal length. For a camera with a lens that has to be focused, λ is related to the focal length by the Gaussian lens formula

$$\frac{1}{distance\,object\,-\,lens} + \frac{1}{\lambda} = \frac{1}{focal\,length}. \tag{24}$$

The parameters λ, c_{center}, r_{center}, $pixel\,pitch_c$, and $pixel\,pitch_r$ are the intrinsic camera parameters. Notice that, from the above definition of the CCS, all points in the image plane have $z = -\lambda$.

If $\mathbf{i}_{\text{cam}}$, $\mathbf{j}_{\text{cam}}$, and $\mathbf{k}_{\text{cam}}$ are the unit vectors in the direction of the axes of the CCS, measured in world coordinates, a point represented by $[x_u\ y_u\ z_u]^T$ in the CCS is transformed into $\mathbf{u} = [X_u\ Y_u\ Z_u]^T$ in the WCS as

$$\mathbf{u} = \underbrace{\begin{bmatrix} X_u \\ Y_u \\ Z_u \end{bmatrix}}_{\text{WCS}} = \underbrace{[\mathbf{i}_{\text{cam}} \quad \mathbf{j}_{\text{cam}} \quad \mathbf{k}_{\text{cam}}]}_{\mathbf{R}} \underbrace{\begin{bmatrix} x_u \\ y_u \\ z_u \end{bmatrix}}_{\text{CCS}} + \mathbf{o}. \tag{25}$$

The position of the nodal point of the camera $\mathbf{o}$ (the translation of the origin of the CCS with respect to the WCS) and the rotation matrix $\mathbf{R}$ (the rotation of the CCS with respect to the WCS) constitute the extrinsic camera parameters. Typically, the intrinsic and extrinsic camera parameters are obtained through a camera calibration procedure [25].

REFERENCES

[1] N. Moray, "Designing for attention," in *Attention: Selection, Awareness, and Control*, A. Baddeley and L. Weiskrantz, Eds.. New York: Clarendon, 1993, pp. 111–134.
[2] M. Eizenman, L. H. Yu, L. Grupp, E. Eizenman, M. Ellenbogen, M. Gemar, and R. D. Levitan, "A naturalistic visual scanning approach to assess selective attention in major depressive disorder," *Psychiat. Res.*, vol. 118, no. 2, pp. 117–128, May 2003.
[3] C. Karatekin and R. F. Asarnow, "Exploratory eye movements to pictures in childhood-onset schizophrenia and attention-deficit/hyperactivity disorder (ADHD)," *J. Abnorm. Child Psychol.*, vol. 27, no. 1, pp. 35–49, Feb. 1999.
[4] M. De Luca, E. Di Pace, A. Judica, D. Spinelli, and P. Zoccolotti, "Eye movement patterns in linguistic and nonlinguistic tasks in developmental surface dyslexia," *Neuropsychologia*, vol. 37, no. 12, pp. 1407–1420, Nov. 1999.
[5] M. Eizenman, T. Jares, and A. Smiley, "A new methodology for the analysis of eye movements and visual scanning in drivers," presented at the 31st Annu. Conf. Erg. & Safety, Hall, QC, Canada, 1999.

[6] J. L. Harbluk, I. Y. Noy, and M. Eizenman, "The impact of cognitive distraction on driver visual and vehicle control," presented at the Transportation Research Board 81st Annual Meeting, Washington, DC, Jan. 2002.
[7] D. Cleveland, "Unobtrusive eyelid closure and visual point of regard measurement system," in *Proc. Tech. Conf. Ocular Measures of Driver Alertness, sponsored by The Federal Highway Administration—Office of Motor Carrier and Highway Safety and The National Highway Traffic Safety Administration—Office of Vehicle Safety Research*, Herndon, VA, 1999, pp. 57–74.
[8] P. A. Wetzel, G. Krueger-Anderson, C. Poprik, and P. Bascom, "An Eye Tracking System for Analysis of Pilots' Scan Paths," United States Air Force Armstrong Laboratory, Tech. Rep. AL/HR-TR-1996-0145, Apr. 1997.
[9] J. H. Goldberg and X. P. Kotval, "Computer interface evaluation using eye movements: methods and constructs," *Int. J. Ind. Erg.*, vol. 24, no. 6, pp. 631–645, Oct. 1999.
[10] R. Sharma, V. I. Pavlović, and T. S. Huang, "Toward multimodal human-computer interface," *Proc. IEEE*, vol. 86, no. 5, pp. 853–869, May 1998.
[11] T. E. Hutchinson, K. P. White, W. N. Martin, K. C. Reichert, and L. A. Frey, "Human-computer interaction using eye-gaze input," *IEEE Trans. Syst., Man, Cybern.*, vol. 19, no. 6, pp. 1527–1534, Nov./Dec. 1989.
[12] L. A. Frey, K. P. White, and T. E. Hutchinson, "Eye-gaze word processing," *IEEE Trans. Syst., Man, Cybern.*, vol. 20, no. 4, pp. 944–950, Jul./Aug. 1990.
[13] L. R. Young and D. Sheena, "Methods and designs—survey of eye movement recording methods," *Behav. Res. Meth. Instrum.*, vol. 7, no. 5, pp. 397–429, 1975.
[14] R. S. Allison, M. Eizenman, and B. S. K. Cheung, "Combined head and eye tracking system for dynamic testing of the vestibular system," *IEEE Trans. Biomed. Eng.*, vol. 43, no. 11, pp. 1073–1082, Nov. 1996.
[15] L. H. Yu and M. Eizenman, "A new methodology for determining point-of-gaze in head-mounted eye tracking systems," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 10, pp. 1765–1773, Oct. 2004.
[16] R. Newman, Y. Matsumoto, S. Rougeaux, and A. Zelinsky, "Real-time stereo tracking for head pose and gaze estimation," in *Proc. 4th IEEE Int. Conf. Automatic Face & Gesture Recognition*, 2000, pp. 122–128.
[17] Y. Matsumoto and A. Zelinsky, "An algorithm for real-time stereo vision implementation of head pose and gaze direction measurement," in *Proc. 4th IEEE Int. Conf. Automatic Face & Gesture Recognition*, 2000, pp. 499–504.
[18] J. G. Wang and E. Sung, "Gaze determination via images of irises," *Image Vis. Comput.*, vol. 19, no. 12, pp. 891–911, Oct. 2001.
[19] ——, "Study on eye gaze estimation," *IEEE Trans. Syst., Man, Cybern. B*, vol. 32, no. 3, pp. 332–350, Jun. 2002.
[20] A. Cowey, "The basis of a method of perimetry with monkeys," *Q. J. Exp. Psychol.*, vol. 15, pt. 2, pp. 81–90, 1963.
[21] J. Merchant, R. Morrissette, and J. L. Porterfield, "Remote measurement of eye direction allowing subject motion over one cubic foot of space," *IEEE Trans. Biomed. Eng.*, vol. BME-21, no. 4, pp. 309–317, Jul. 1974.
[22] A. Sugioka, Y. Ebisawa, and M. Ohtani, "Noncontact video-based eye-gaze detection method allowing large head displacements," in *Proc. 18th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 1996, vol. 2, pp. 526–528.
[23] Y. Ebisawa, M. Ohtani, A. Sugioka, and S. Esaki, "Single mirror tracking system for free-head video-based eye-gaze detection method," in *Proc. 19th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 1997, vol. 4, pp. 1448–1451.
[24] A. M. Slater and J. M. Findlay, "The measurement of fixation position in the newborn baby," *J. Exp. Child Psychol.*, vol. 14, pp. 349–364, 1972.
[25] E. Trucco and A. Verri, *Introductory Techniques for 3-D Computer Vision*. Upper Saddle River, NJ: Prentice-Hall, 1998, pp. 123–138.
[26] S.-W. Shih, Y.-T. Wu, and J. Liu, "A calibration-free gaze tracking technique," in *Proc. 15th Int. Conf. Pattern Recognition*, Sep. 2000, vol. 4, pp. 201–204.
[27] E. D. Guestrin, "A novel head-free point-of-gaze estimation system," M.A.Sc. thesis, Dept. Elect. Comput. Eng., Univ. Toronto, Toronto, ON, Canada, 2003.
[28] B. J. M. Lui, "A point-of-gaze estimation system for studies of visual attention," M.A.Sc. thesis, Dept. Elect. Comput. Eng., Univ. Toronto, Toronto, ON, Canada, 2003.
[29] A. I. Tew, "Simulation results for an innovative point-of-regard sensor using neural networks," *Neural Comp. Applicat.*, vol. 5, no. 4, pp. 230–237, 1997.
[30] A. G. Gale, "A note on the remote oculometer technique for recording eye movements," *Vis. Res.*, vol. 22, no. 1, pp. 201–202, 1982.
[31] M. C. Corbett, E. S. Rosen, and D. P. S. O'Brart, *Corneal Topography: Principles and Applications*. London, U.K.: BMJ Books, 1999, p. 6.

Elias Daniel Guestrin (S'98) was born in Paraná, Entre Ríos, Argentina, in 1974. He received the Electronics Engineer degree from the Universidad Tecnológica Nacional-Facultad Regional Paraná, in 2000. In 2003, he received the M.A.Sc. degree in electrical and computer engineering from the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada, for his work on the development of a novel head-free point-of-gaze estimation system. He is currently working toward the Ph.D. degree with the Department of Electrical and Computer Engineering, and the Institute of Biomaterials and Biomedical Engineering, University of Toronto, where he continues developing novel gaze estimation technology.

His research interests include gaze estimation, mathematical modeling of optical systems, and signal and image processing.

Moshe Eizenman was born in Tel-Aviv, Israel, in 1952. He received the B.A.Sc., M.A.Sc., and Ph.D. degrees in electrical engineering from the University of Toronto, Toronto, ON, Canada, in 1978, 1980, and 1984, respectively.

He joined the faculty of the University of Toronto in 1984, and he is currently an Associate Professor in the departments of Electrical and Computer Engineering, Ophthalmology, and at the Institute of Biomaterials and Biomedical Engineering. He is also a Research Associate at the Eye Research Institute of Canada and at the Hospital for Sick Children, Toronto. In cooperation with EL-MAR, Inc., he has developed advanced technologies for eye tracking and gaze estimation systems. These systems are used by universities and research institutes for medical, human-factors and driving research, and for pilot training. His research interests include detection and estimation of biological phenomena, eye tracking and gaze estimation systems, visual evoked potentials and the development of vision.