

Remote Point-of-Gaze Estimation Requiring a Single-Point Calibration for Applications with Infants

Elias Daniel Guestrin^{*}
University of Toronto

Moshe Eizenman[†]
University of Toronto

Abstract

This paper describes a method for remote, non-contact point-of-gaze estimation that tolerates free head movements and requires a simple calibration procedure in which the subject has to fixate only on a single point. This method uses the centers of the pupil and at least two corneal reflections that are estimated from eye images captured by at least two cameras. Experimental results obtained with three adult subjects exhibited RMS point-of-gaze estimation errors ranging from 7 to 12 mm (equivalent to about $0.6 - 1^\circ$ of visual angle) for head movements in a volume of about 1 dm^3 . Preliminary results with two infants demonstrated the ability of a system that requires a single-point calibration procedure to estimate infants' point-of-gaze. The ability to record infants' visual scanning behavior can be used for the study of visual development, the determination of attention allocation and the assessment of visual function in preverbal infants.

Keywords: Point-of-gaze, remote gaze estimation, single-point calibration, minimal subject cooperation, infants' gaze.

1 Introduction

The point-of-gaze is the point within the visual field that is imaged on the highest acuity region of the retina known as the fovea. Systems that estimate the point-of-gaze are used in a large variety of applications [Eizenman et al. 2003; Duchowski 2002; Harbluk et al. 2007; Frey et al. 1990] such as studies of mood disorders, reading behavior, driver behavior, pilot training, ergonomics, marketing and advertising research, human-computer interfaces and assistive devices for motor-disabled persons.

Most modern approaches to remote, non-contact point-of-gaze estimation are based on the analysis of eye features extracted from video images. The most common features are the centers of the pupil and one or more corneal reflections (Fig. 1). The corneal reflections (first Purkinje images) are virtual images of infrared light sources that illuminate the eye, and are created by the front surface of the cornea, which acts as a convex mirror.

Typically, point-of-gaze estimation systems have to be calibrated for each subject by having the subject fixate on multiple points in the scene. A multiple-point calibration procedure, however, can be an obstacle in applications requiring minimal subject



Figure 1: Image of the face of a 6-month old infant. The pupil centers are indicated by crosses and the corneal reflections are enclosed by squares.

cooperation such as applications with infants. As shown in [Shih and Liu 2004; Guestrin and Eizenman 2006], if at least two cameras and at least two light sources are used, it is possible to estimate the point-of-gaze in the presence of head movements after completing a simpler calibration procedure in which the subject is required to fixate on a single point. A single-point calibration procedure could be performed with infants by presenting a bright flashing stimulus on a dark uniform background to attract their attention.

A system that can estimate infants' point-of-gaze could be used, for example, as part of an automatic system for the objective assessment of visual acuity. For preverbal subjects, two of the most common techniques to assess visual acuity are the Optokinetic Nystagmus (OKN) technique and the Forced Preferential Looking (FPL) technique. The OKN technique is based on the fact that an awake subject whose visual field is filled with moving stripes cannot avoid showing an eye movement pattern that includes slow smooth eye movements in the direction of the motion that are followed by fast corrective saccades in the opposite direction (see Fig. 9). The smallest stripe width that elicits these eye movements is correlated with the infant's visual acuity [Gorman et al. 1957]. The FPL technique is based on the fact that infants prefer to look at patterned surfaces rather than homogeneous surfaces. With the FPL technique, infants are presented with a pair of stimuli: black and white stripes of a given width and a gray field of matched space-average luminance. A human observer that monitors the infant's visual scanning behavior determines if the two stimuli are scanned differentially. Visual acuity is correlated with the smallest stripe width that is scanned differentially [Teller et al. 1974]. By replacing the human observer with a point-of-gaze estimation system, visual acuity can be determined from the analysis of the infant's visual scanning patterns.

^{*}e-mail: elias.guestrin@utoronto.ca

[†]e-mail: eizennm@ecf.utoronto.ca

Copyright © 2008 by the Association for Computing Machinery, Inc.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions Dept, ACM Inc., fax +1 (212) 869-0481 or e-mail permissions@acm.org.

ETRA 2008, Savannah, Georgia, March 26–28, 2008.

© 2008 ACM 978-1-59593-982-1/08/0003 \$5.00

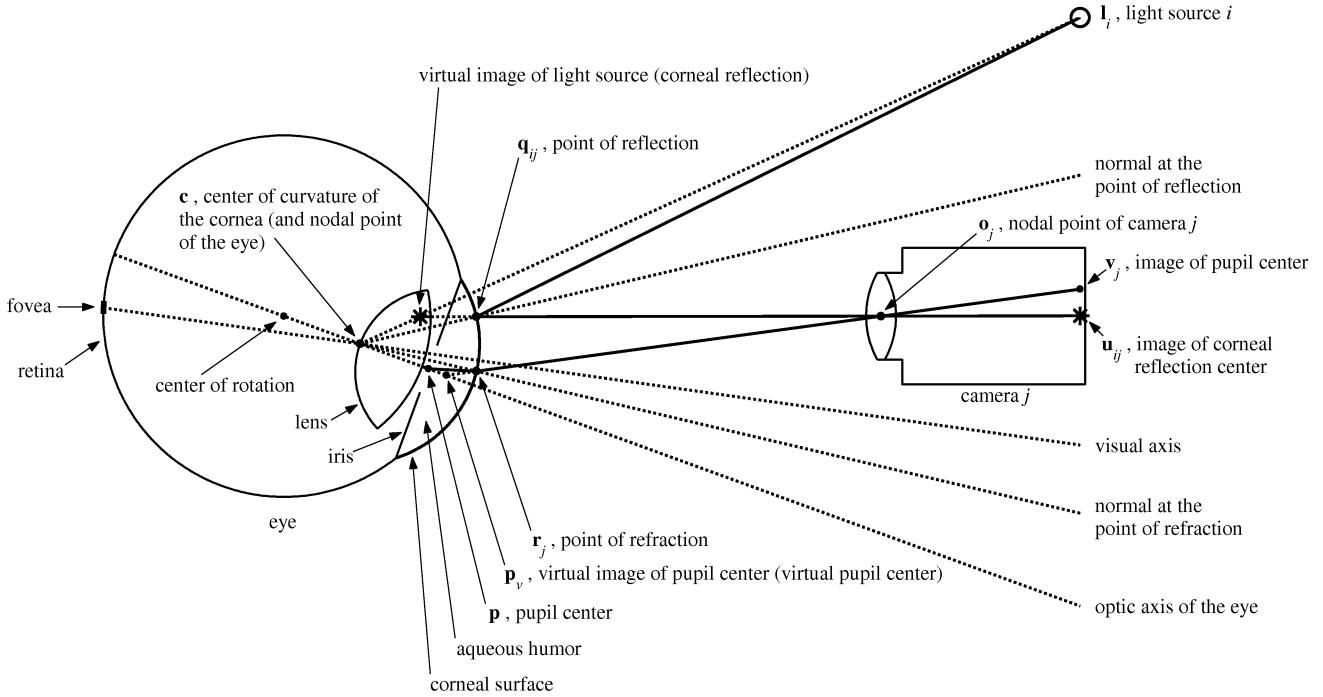


Figure 2: Ray-tracing diagram (not to scale in order to be able to show all the elements of interest), showing schematic representations of the eye, a camera and a light source.

In this paper, a point-of-gaze estimation system developed for applications with infants is described. The next section presents the mathematical model of a novel method for point-of-gaze estimation that requires a single-point calibration procedure. Section 3 describes the set-up of a prototype system with two cameras and multiple light sources. Section 4 presents an evaluation of the proposed method through numerical simulations. Section 5 shows experimental results with three adult subjects and a 7-month old infant (one of the two infants tested). The results with the adult subjects provide a measurement of the expected point-of-gaze estimation accuracy, while the results with the infant demonstrate that the proposed method can be used in applications with infants. Finally, Section 6 summarizes the conclusions.

2 Mathematical Model

This section presents a mathematical model for remote point-of-gaze estimation using the coordinates of the centers of the pupil and corneal reflections that are estimated from images captured by two video cameras (the extension to more cameras is trivial). Under the assumptions that the light sources are modeled as point sources, the video cameras are modeled as pinhole cameras and the corneal surface is modeled as a spherical section, Fig. 2 presents a ray-tracing diagram, where all points are represented as 3-D column vectors (bold font) in a right-handed Cartesian world coordinate system (WCS).

First, 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 (a.k.a. camera center, center of projection) of camera j , \mathbf{o}_j , and intersects the camera image plane at a point \mathbf{u}_{ij} . According to the law of reflection, the incident ray,

the reflected ray and the normal at the point of reflection are coplanar. Since any line going through the center of curvature of the cornea, \mathbf{c} , is normal to the spherical corneal surface, vector $(\mathbf{q}_{ij} - \mathbf{c})$ is normal to the corneal surface at the point of reflection \mathbf{q}_{ij} . It then follows that points \mathbf{l}_i , \mathbf{q}_{ij} , \mathbf{o}_j , \mathbf{u}_{ij} , and \mathbf{c} are coplanar. In other words, the center of curvature of the cornea, \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} . Noting that three coplanar vectors \mathbf{a}_1 , \mathbf{a}_2 and \mathbf{a}_3 satisfy $\mathbf{a}_1 \times \mathbf{a}_2 \cdot \mathbf{a}_3 = 0$, this condition can be formalized as

$$\underbrace{(\mathbf{l}_i - \mathbf{o}_j) \times (\mathbf{u}_{ij} - \mathbf{o}_j)}_{\text{normal to the plane defined by } \mathbf{l}_i, \mathbf{o}_j \text{ and } \mathbf{u}_{ij}} \bullet (\mathbf{c} - \mathbf{o}_j) = 0. \quad (1)$$

Notice that (1) shows that, for each camera j , all the planes defined by \mathbf{o}_j , \mathbf{l}_i and \mathbf{u}_{ij} contain the line defined by points \mathbf{c} and \mathbf{o}_j . If the light sources, \mathbf{l}_i , are positioned such that at least two of those planes are not coincident, the planes intersect at the line defined by \mathbf{c} and \mathbf{o}_j . If \mathbf{b}_j is a vector in the direction of the line of intersection of the planes, then

$$\mathbf{c} = \mathbf{o}_j + k_{c,j} \mathbf{b}_j \text{ for some } k_{c,j} \quad (2)$$

In particular, if two light sources are considered ($i = 1, 2$),

$$\mathbf{b}_j = \frac{[(\mathbf{l}_1 - \mathbf{o}_j) \times (\mathbf{u}_{1j} - \mathbf{o}_j)] \times [(\mathbf{l}_2 - \mathbf{o}_j) \times (\mathbf{u}_{2j} - \mathbf{o}_j)]}{\|[(\mathbf{l}_1 - \mathbf{o}_j) \times (\mathbf{u}_{1j} - \mathbf{o}_j)] \times [(\mathbf{l}_2 - \mathbf{o}_j) \times (\mathbf{u}_{2j} - \mathbf{o}_j)]\|} \quad (3)$$

where $[(\mathbf{l}_1 - \mathbf{o}_j) \times (\mathbf{u}_{1j} - \mathbf{o}_j)]$ is the normal to the plane defined by \mathbf{o}_j , \mathbf{l}_1 and \mathbf{u}_{1j} , and $[(\mathbf{l}_2 - \mathbf{o}_j) \times (\mathbf{u}_{2j} - \mathbf{o}_j)]$ is the normal to the plane defined by \mathbf{o}_j , \mathbf{l}_2 and \mathbf{u}_{2j} .

Having two cameras, the position of the center of curvature of the cornea, \mathbf{c} , can be found as the intersection of the lines given by (2)–(3), $j = 1, 2$. Since, in practice, the estimated coordinates of the images of the corneal reflections, \mathbf{u}_{ij} , are corrupted by noise, those lines may not intersect. Therefore, \mathbf{c} is found as the midpoint of the shortest segment defined by a point belonging to each of those lines. It can be shown that, in such case, \mathbf{c} is given by

$$\begin{aligned}\mathbf{c} = & \frac{1}{2} [\mathbf{b}_1 \quad \mathbf{b}_2] \begin{bmatrix} \mathbf{b}_1 \bullet \mathbf{b}_1 & -\mathbf{b}_1 \bullet \mathbf{b}_2 \\ -\mathbf{b}_1 \bullet \mathbf{b}_2 & \mathbf{b}_2 \bullet \mathbf{b}_2 \end{bmatrix}^{-1} \begin{bmatrix} -\mathbf{b}_1 \bullet (\mathbf{o}_1 - \mathbf{o}_2) \\ \mathbf{b}_2 \bullet (\mathbf{o}_1 - \mathbf{o}_2) \end{bmatrix} \\ & + \frac{1}{2} (\mathbf{o}_1 + \mathbf{o}_2).\end{aligned}\quad (4)$$

Next, consider an imaginary ray that originates at the pupil center, \mathbf{p} , travels through the aqueous humor and cornea (effective index of refraction ≈ 1.3375 [Corbett et al. 1999]) and refracts at a point \mathbf{r}_j on the corneal surface as it travels into the air (index of refraction ≈ 1), such that the refracted ray passes through the nodal point of camera j , \mathbf{o}_j , and intersects the camera image plane at a point \mathbf{v}_j . This refraction results in the formation of a virtual image of the pupil center, \mathbf{p}_{vj} , located on the extension of the refracted ray, i.e.,

$$\mathbf{p}_{vj} = \mathbf{o}_j + k_{pj} \underbrace{(\mathbf{o}_j - \mathbf{v}_j)}_{\mathbf{h}_j} \text{ for some } k_{pj}. \quad (5)$$

In strict terms, the spatial location of \mathbf{p}_{vj} depends on the position of the nodal point of the camera, \mathbf{o}_j , relative to the eye. Therefore, in general, the spatial location of \mathbf{p}_{vj} will be slightly different for each of the two cameras. Despite this, an approximate virtual image of the pupil center, \mathbf{p}_v , can be found as the midpoint of the shortest segment defined by a point belonging to each of the lines given by (5), $j = 1, 2$, i.e.,

$$\begin{aligned}\mathbf{p}_v = & \frac{1}{2} [\mathbf{h}_1 \quad \mathbf{h}_2] \begin{bmatrix} \mathbf{h}_1 \bullet \mathbf{h}_1 & -\mathbf{h}_1 \bullet \mathbf{h}_2 \\ -\mathbf{h}_1 \bullet \mathbf{h}_2 & \mathbf{h}_2 \bullet \mathbf{h}_2 \end{bmatrix}^{-1} \begin{bmatrix} -\mathbf{h}_1 \bullet (\mathbf{o}_1 - \mathbf{o}_2) \\ \mathbf{h}_2 \bullet (\mathbf{o}_1 - \mathbf{o}_2) \end{bmatrix} \\ & + \frac{1}{2} (\mathbf{o}_1 + \mathbf{o}_2).\end{aligned}\quad (6)$$

The optic axis of the eye can be defined as the line going through the center of curvature of the cornea, \mathbf{c} , and the pupil center, \mathbf{p} (Fig. 2). Assuming that \mathbf{p}_v is also on the optic axis (this assumption is discussed in Section 4), (3)–(6) provide a closed-form solution for the reconstruction of the optic axis of the eye in 3-D space without the knowledge of any eye parameter. In particular, the direction of the optic axis of the eye is given by the unit vector

$$\boldsymbol{\omega} = \frac{\mathbf{p}_v - \mathbf{c}}{\|\mathbf{p}_v - \mathbf{c}\|}. \quad (7)$$

It should be noted that the method to reconstruct the optic axis proposed in this paper removes the constraint of the method from [Shih and Liu 2004; Guestrin and Eizenman 2006] by which the optic axis of the eye cannot be reconstructed when the optic axis of the eye intersects the line connecting the nodal points of the cameras. Also, the proposed method can have better performance in the presence of noise in the estimated coordinates of the pupil center in the eye images than the method from [Shih and Liu (2004); Guestrin and Eizenman (2006)] (Section 4).

The point-of-gaze, \mathbf{g} , is defined as the intersection of the visual axis, rather than the optic axis, with the scene. The visual axis is the line defined by the nodal point of the eye and the center of the fovea, and deviates from the optic axis by as much as 5° [Young and Sheena 1975]. Since the nodal point is within 1 mm of the center of curvature of the cornea [Young and Sheena 1975], \mathbf{c} , it can be assumed to be coincident with \mathbf{c} (Fig. 2). If the orientation of the optic axis is described by the pan (horizontal) angle θ_{eye} and the tilt (vertical) angle φ_{eye} , and the horizontal and vertical angles from the optic axis to the visual axis are, respectively, α_{eye} and β_{eye} , then the orientation of the visual axis is given by the pan angle ($\theta_{eye} + \alpha_{eye}$) and the tilt angle ($\varphi_{eye} + \beta_{eye}$).

In order to formalize the relation between the two axes, suppose that the scene is a vertical plane (e.g., a computer screen) and that the WCS has its XY -plane coincident with the scene plane, with the X -axis horizontal, the positive Y -axis vertical pointing up and the positive Z -axis coming out of the scene plane. Furthermore, let $\theta_{eye} = \varphi_{eye} = 0$ when the optic axis is normal to the scene plane (parallel to the Z -axis of the WCS), $\theta_{eye} > 0$ for rotations to the right, and $\varphi_{eye} > 0$ for rotations upwards. Then, θ_{eye} and φ_{eye} can be obtained from the unit vector in the direction of the optic axis of the eye, $\boldsymbol{\omega}$, by solving

$$\boldsymbol{\omega} = \begin{bmatrix} \cos \varphi_{eye} \sin \theta_{eye} \\ \sin \varphi_{eye} \\ -\cos \varphi_{eye} \cos \theta_{eye} \end{bmatrix}, \quad (8)$$

and the visual axis can be described in parametric form as

$$\mathbf{g} = \mathbf{c} + k_g \begin{bmatrix} \cos(\varphi_{eye} + \beta_{eye}) \sin(\theta_{eye} + \alpha_{eye}) \\ \sin(\varphi_{eye} + \beta_{eye}) \\ -\cos(\varphi_{eye} + \beta_{eye}) \cos(\theta_{eye} + \alpha_{eye}) \end{bmatrix} \quad (9)$$

for all k_g . Since the scene plane is at $Z = 0$, the point-of-gaze is given by (9) for a value of k_g such that the Z -component of \mathbf{g} , g_z , equals 0, that is,

$$k_g = \frac{c_z}{\cos(\varphi_{eye} + \beta_{eye}) \cos(\theta_{eye} + \alpha_{eye})}. \quad (10)$$

The values of α_{eye} and β_{eye} are subject-specific and can be determined through a calibration procedure in which the subject is required to fixate on a single point. By having the subject fixate on a known point \mathbf{g} , after calculating \mathbf{c} , θ_{eye} and φ_{eye} from (3)–(8), α_{eye} and β_{eye} are found from (9) and (10).

In order to estimate the point-of-gaze with the above system of equations, the world coordinates of the positions of the light sources (\mathbf{l}_i), the nodal points of the cameras (\mathbf{o}_j), and the centers of the pupil (\mathbf{v}_j) and corneal reflections (\mathbf{u}_{ij}) in the eye images, must be known. Since the centers of the pupil and corneal reflections that are estimated in each eye image are measured in pixels in an image coordinate system, they have to be transformed into the WCS [Guestrin and Eizenman 2006]. This transformation requires all intrinsic and extrinsic camera parameters (notice that the position of \mathbf{o}_j is one of the extrinsic camera parameters). The positions of the light sources and the camera parameters are obtained as described in the next section.

3 System Set-Up

A prototype system to estimate the point-of-gaze on a computer screen was implemented. This system uses two synchronized monochrome CCD cameras (Scorpion SCOR-14SOM, Point Grey Research, Vancouver, BC, Canada) with 35 mm lenses and four infrared light sources (850 nm) attached to a 19" LCD monitor by a custom aluminum frame (Fig. 3).

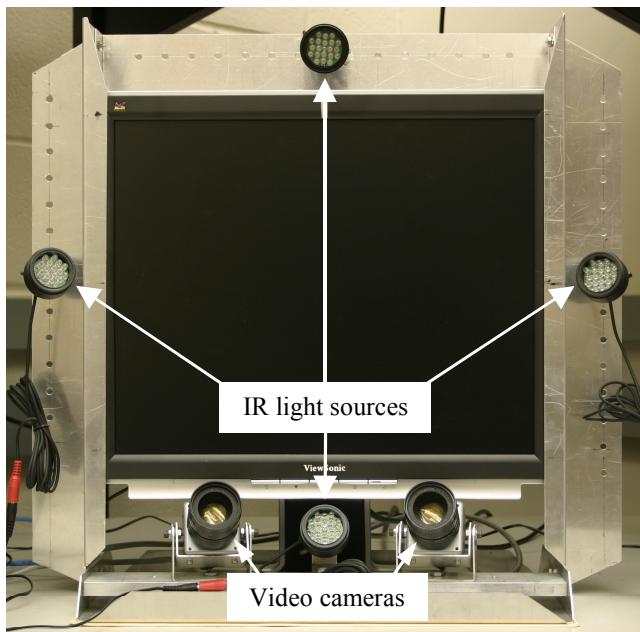


Figure 3: System set-up.

The two cameras were set at a resolution of 1280 pixels by 960 pixels and oriented such that their optic axes intersect at a distance of approximately 65 cm from the screen (typical viewing distance). In these conditions, the prototype system can tolerate moderate head movements before the eye features are no longer in the field of view of the cameras or are out of focus. The range of lateral head movements is about 10 cm when the point-of-gaze is estimated for a specific eye, about 16 cm when the point-of-gaze is estimated for either eye and about 4 cm when the point-of-gaze is estimated for both eyes simultaneously (binocular mode). The range of vertical head movements is about 8 cm and the range of backwards/forward head movements is about 10 cm.

The use of more than the minimum of two light sources to illuminate the eye helps to improve the robustness of the system

by increasing the likelihood that at least two corneal reflections are available regardless of the head position and the point-of-gaze on the screen, and in the presence of eyelid and eyelash interferences. Moreover, the use of more than two light sources helps to reduce the effect of the deviation of the shape of real corneas from the ideal spherical shape assumed in the model of Section 2 (corneal asphericity [Guestrin and Eizenman 2006]). Typically, the radius of curvature of the front corneal surface increases towards the boundary with the sclera and only the central part is approximately spherical [Young and Sheena 1975]. Therefore, by using the two light sources that produce corneal reflections that are closest to the pupil center in the eye images, it is possible to reduce the point-of-gaze estimation bias due to corneal asphericity.

In order to be able to estimate the point-of-gaze, the positions of the light sources and the intrinsic and extrinsic camera parameters must be known accurately. The positions of the light sources, \mathbf{l}_i , with respect to the WCS, which is attached to the LCD monitor (as explained in Section 2 and with the origin of the WCS at the center of the screen), are measured directly using rulers and calipers. The intrinsic camera parameters and the position and orientation of the cameras with respect to the WCS (extrinsic camera parameters) are determined through the camera calibration procedure outlined in the next paragraph.

Since the cameras that capture images of the eye (eye cameras) are positioned under the screen (Fig. 3), and therefore cannot observe the monitor, an auxiliary camera that views both the computer screen and the region of allowed head movements (Fig. 4) is used together with a double-sided planar checkerboard pattern. The calibration is based on a camera calibration toolbox for MATLAB® [Bouguet] and the entire calibration procedure can be summarized as follows: (1) Images of the calibration pattern, at several different orientations within the region of allowed head movement, are captured simultaneously by the two eye cameras and the auxiliary camera. (2) By using the corners of the checkerboard pattern from all the different views, the intrinsic parameters of the 3 cameras, and the relative position and orientation of the two eye cameras with respect to the auxiliary camera are calculated. (3) A checkerboard pattern is then displayed on the screen, and the position and orientation of the auxiliary camera with respect to the WCS is calculated. (4) Using this information, the position and orientation of the eye cameras with respect to the WCS are determined.

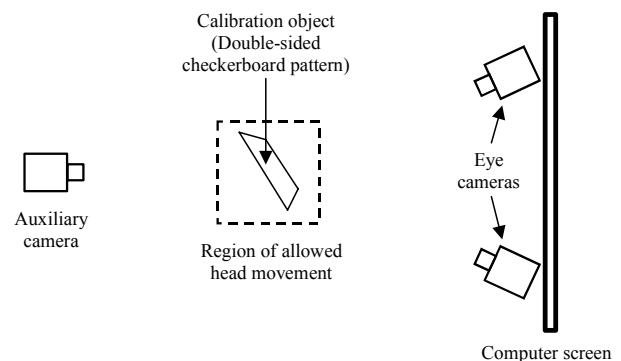


Figure 4: Top-view schematic representation of the camera calibration set-up.

Since all the system components are fixed with respect to each other, the system calibration procedure needs to be performed

only once during system set-up and is simpler than the one described in [Shih and Liu 2004].

Before describing the experimental results, an evaluation of the point-of-gaze estimation bias resulting from model assumptions and the sensitivity of the proposed method to noise in the coordinates of the estimated eye features (centers of pupil and corneal reflections) will be presented.

4 Evaluation of Model Assumptions and Noise Sensitivity

In strict terms, the reconstruction of the optic axis of the eye using \mathbf{p}_v (equations (5)–(7)) is only approximate as it relies on two assumptions. The first assumption is, similarly to [Shih and Liu 2004; Guestrin and Eizenman 2006], that \mathbf{v}_j is the image of the pupil center in the camera image plane. However, in practice, \mathbf{v}_j is estimated as the center of the pupil image in the camera image plane. As pointed out in [Ohno et al. 2002; Ohno and Mukawa 2004; Villanueva and Cabeza 2007], in general, due to the refraction at the front corneal surface followed by the perspective projection at the camera, the image of the pupil center does not coincide with the center of the pupil image. The second assumption is that \mathbf{p}_v calculated from (6) is on the optic axis of the eye. However, this is not exact even if \mathbf{v}_j is the image of the pupil center. These two assumptions result in point-of-gaze estimation bias.

The point-of-gaze estimation bias resulting from the above two assumptions, as well as the sensitivity of the proposed method to noise in the image coordinates of the centers of the pupil and corneal reflections, were evaluated through numerical simulations. The numerical simulations corresponded to the experimental system described in the previous section and a right eye with spherical cornea whose parameter values were taken from [Guestrin and Eizenman 2006] (i.e., distance between the center of rotation of the eye and the center curvature of the cornea: 5.3 mm; radius of curvature of the cornea: 7.8 mm; distance between the pupil plane and the center of curvature of the cornea: 4.2 mm; effective index of refraction of the aqueous humor and cornea combined: 1.3375; angular deviation of the visual axis from the optic axis: $\alpha_{eye} = -5^\circ$ and $\beta_{eye} = 1.5^\circ$). Since the bias depends on the pupil size, the pupil diameter was varied from 2 to 8 mm in steps of 1 mm. For each value of the pupil diameter, the center of rotation of the simulated eye was set at 27 ($3 \times 3 \times 3$ grid) positions covering the entire field of view / depth of field of the system (region of allowed head movements when tracking a specific eye). For each eye position, the simulated eye fixated on 25 (5×5 grid) points that spanned the entire computer screen, including the center and the four corners. Similarly to the experimental system, the point-of-gaze was estimated using the two light sources that produced corneal reflections that were closest to the pupil center in the eye images.

For comparison, the simulations were also carried out determining the direction of the optic axis of the eye using the method described previously by Shih and Liu [2004] and Guestrin and Eizenman [2006]. More specifically, in that case, after the 3-D position of the center of curvature of the cornea, \mathbf{c} , was found using (3)–(4), the direction of the optic axis was determined as

$$\omega = \frac{[(\mathbf{o}_1 - \mathbf{v}_1) \times (\mathbf{c} - \mathbf{o}_1)] \times [(\mathbf{o}_2 - \mathbf{v}_2) \times (\mathbf{c} - \mathbf{o}_2)]}{\|[(\mathbf{o}_1 - \mathbf{v}_1) \times (\mathbf{c} - \mathbf{o}_1)] \times [(\mathbf{o}_2 - \mathbf{v}_2) \times (\mathbf{c} - \mathbf{o}_2)]\|}, \quad (11)$$

and then the point-of-gaze was calculated using (8)–(10). For simplicity, the determination of the direction of the optic axis using (11) will be hereafter referred to as “SL/GE method”, while the determination of the direction of the optic axis of the eye as proposed in this paper, i.e., using (5)–(7), will be hereafter referred to as “GE method”.

4.1 Point-of-gaze estimation bias

To quantify the point-of-gaze estimation bias due to the above assumptions, the following performance measures were calculated for each pupil diameter: the root-mean-square value of the bias (RMS bias) and the bias for the fixation point at the center of the screen, which is usually used for the calibration of α_{eye} and β_{eye} (Bias SC).

The results of these simulations are summarized in Table 1, showing the ranges observed as the pupil diameter varied from 2 to 8 mm. The simulations showed that, as expected, the bias increases with the pupil diameter (conversely, as the pupil diameter gets smaller, the center of the pupil image approaches the image of the pupil center and the bias decreases). The bias for a target calibration point at the center of the screen (Bias SC) is relatively small and suggests that the calibrated values of α_{eye} and β_{eye} would differ from their actual values by no more than 0.04° for the SL/GE method and 0.06° for the GE method.

The RMS bias exhibited by the SL/GE method, which relies only on the first of the above assumptions (i.e., that the image of the pupil center falls at the center of the pupil image), is small and can be neglected in practice. The RMS bias exhibited by the GE method, which relies on both of the above assumptions, is larger. Since the only difference between the GE method and the SL/GE method is the second assumption (i.e., that \mathbf{p}_v is on the optic axis of the eye), these simulation results suggest that the second assumption is the main cause of the bias exhibited by the GE method. In fact, if the point-of-gaze is estimated using the actual image of the pupil center, thereby eliminating the bias due to the first assumption, the RMS bias exhibited by the GE method is 1.87 mm (as it can be seen from Table 1 it accounts for most of the bias in the GE method).

	SL/GE [mm]	GE [mm]
RMS bias	0.019 – 0.27	1.92 – 2.90
Bias SC	0 – 0.39	0.10 – 0.58

Table 1: Point-of-gaze estimation bias.

Fig. 5 shows the worst-case simulation results (largest pupil diameter, i.e., 8 mm) for the GE method, where the crosses (+) indicate the actual points-of-gaze (in this case, the original 25 fixation points and 56 additional intermediate points were considered), the circles indicate the estimates of the point-of-gaze and the thin dotted line indicates the screen boundary. As can be seen in this figure, the largest bias occurs when the eye fixates on the top right region of the screen (largest angular deviation between the optic axis of the eye and the optic axis of the cameras for these simulations). Nevertheless, the region of the screen where the bias is relatively large is rather limited and the RMS bias is under 3 mm.

4.2 Noise sensitivity

To evaluate the sensitivity to noise, for each eye position and each fixation point, 100 independent realizations of zero-mean

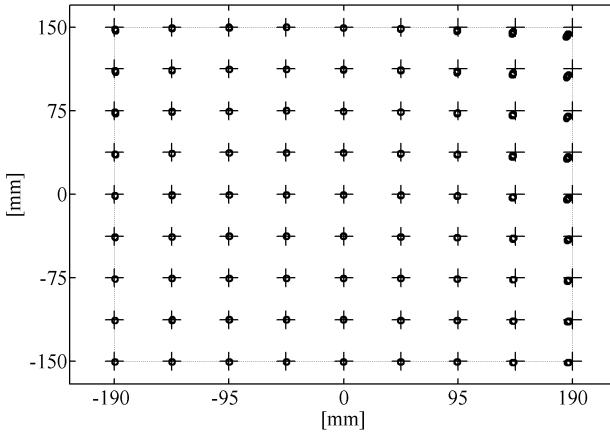


Figure 5: Point-of-gaze estimation bias for a simulated eye when the pupil diameter is 8 mm (+: actual point-of-gaze; o: estimated point-of-gaze; dotted line: screen boundary).

Gaussian noise with a standard deviation of 0.1 pixel were added to each coordinate of the simulated images of the corneal reflections and center of the pupil image before estimating the point-of-gaze. To quantify the dispersion of the point-of-gaze estimates due to noise by itself, for each pupil diameter, the RMS distance between the noisy point-of-gaze estimates and the corresponding noise-free point-of-gaze estimates was determined for each fixation point (RMS Dispersion for ea. fix. pt.). The maximum value of these RMS dispersions (MAX(RMS Dispersion for ea. fix. pt.)) was then determined for each pupil diameter. The range of the MAX(RMS Dispersion for ea. fix. pt.) observed for the range of pupil diameters considered is summarized in Table 2. The results from this table show that the maximum dispersion of the noisy point-of-gaze estimates around the corresponding noise-free estimates exhibited by the GE method is about 38 % the maximum dispersion exhibited by the SL/GE method.

To quantify the total point-of-gaze estimation error due to bias and dispersion combined, for each pupil diameter, the RMS point-of-gaze estimation error for each fixation point (RMS Error for ea. fix. pt.) was calculated. The maximum value of these RMS errors (MAX(RMS Error for ea. fix. pt.)) was then determined for each pupil diameter. The range of the MAX(RMS Error for ea. fix. pt.) observed for the range of pupil diameters considered is shown in Table 2. As it can be seen, the MAX(RMS Error for ea. fix. pt.) for the GE method is 60-70 % the MAX(RMS Error for ea. fix. pt.) for the SL/GE method. The total RMS point-of-gaze estimation error for all fixation points (RMSE) was also calculated for each pupil diameter and included in Table 2. These results show that the RMSE is similar for both methods, with the RMSE for the GE method being slightly smaller.

	SL/GE [mm]	GE [mm]
MAX(RMS Dispersion for ea. fix. pt.)	13.46 – 13.76	5.07 – 5.28
MAX(RMS Error for ea. fix. pt.)	13.46 – 13.76	7.83 – 9.94
RMSE	4.95 – 5.06	4.34 – 4.76

Table 2: Noise sensitivity.

In summary, even though the GE method exhibits a larger bias, it is less sensitive to noise in the estimated coordinates of the eye features than the SL/GE method. Overall, the GE method proposed in this paper exhibits a better performance.

5 Experimental Results

A preliminary evaluation of the performance of the prototype system was carried out through experiments with three adult subjects without eyeglasses or contact lenses, and two infants (6 and 7 months of age). The results obtained with the adult subjects provide a measurement of the expected point-of-gaze estimation accuracy.

5.1 Evaluation of the expected point-of-gaze estimation accuracy with adult subjects

In these experiments, the point-of-gaze was estimated for a single eye (right eye) and the head of each subject was placed at the center and at 4 positions at the boundaries of the region of allowed head movements. For the central position, the right eye was at 65 cm from the screen and at the center of the field of view of both eye cameras. The 4 boundary positions corresponded to lateral and forward/backwards head movements. For each head position, each subject was asked to fixate on 25 points (5 x 5 rectangular grid) on the computer screen and 50 estimates (≈ 3.3 seconds @ 15 estimates/second) of the image coordinates of the centers of the pupil and corneal reflections were obtained for each fixation point. The subject-specific angular deviation of the visual axis from the optic axis, α_{eye} and β_{eye} , was determined when the head was at the central position and the subject fixated on the center of the screen.

Fig. 6 shows the experimental results for all 3 subjects and all 5 head positions, where the large crosses (+) indicate the target fixation points. The point-of-gaze estimates shown in this figure were obtained using the average of the estimated image coordinates of the centers of the pupil and corneal reflections for each head position and fixation point. Due to the averaging, the effect of noise in the estimates of the image coordinates of the eye features is marginal. It can be observed that the accuracy of the point-of-gaze estimates varied among the 3 subjects (\triangle , \circ , \square), which can be attributed primarily to different degrees of corneal asphericity. For Subject 1 (\triangle) and Subject 2 (\circ) most of the point-of-gaze estimates are clustered together fairly close to the corresponding target fixation points, exhibiting relatively small bias. For Subject 3 (\square), the point-of-gaze estimates exhibit the largest bias due to corneal asphericity.

Fig. 7 shows the point-of-gaze estimates calculated for the individual estimates of the image coordinates of the eye features corresponding to Subject 1 for all 5 head positions. As before, the crosses (+) indicate the target fixation points. The dispersion of the point-of-gaze estimates (*) is due to noise in the estimates of the image coordinates of the centers of the pupil and corneal reflections. As explained above, the bias of the clusters of point-of-gaze estimates from their respective target fixation points, although relatively small, is primarily due to corneal asphericity.

Table 3 summarizes the RMS point-of-gaze estimation errors, for each subject and all 5 head positions, when the point-of-gaze was calculated using the individual estimates of the image coordinates of the eye features (50 estimates per fixation point and head position). The RMS point-of-gaze estimation errors are less than 12 mm (equivalent to about 1° of visual angle at a distance of 65

cm from the screen). The results from Figs. 6 and 7 suggest that it is possible to distinguish unambiguously more than 25 points on the screen of a 19" monitor.

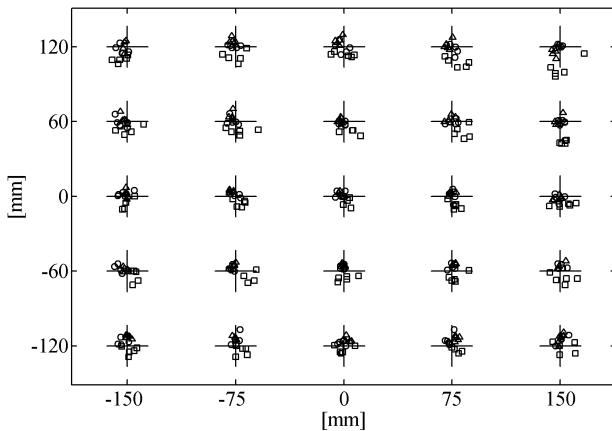


Figure 6: Experimental POG estimates for all subjects, obtained using the average image coordinates of the eye features for each fixation point and each head position (1 estimate per fixation point and head position).

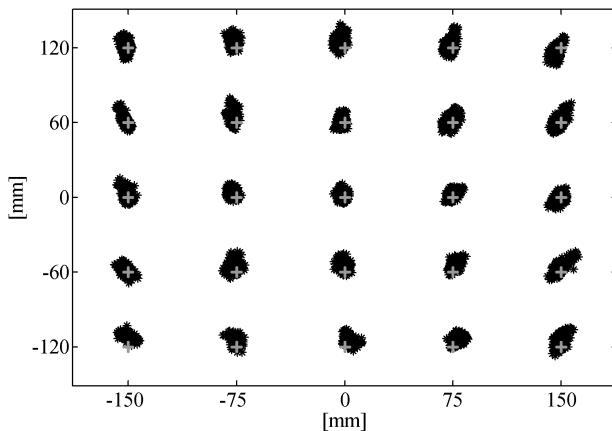


Figure 7: Experimental POG estimates for Subject 1 and all 5 head positions, obtained using the individual estimates of the image coordinates of the eye features (50 estimates per fixation point and head position).

	RMS error [mm]	Equivalent to
Subject 1	6.55	$\sim 0.6^\circ$
Subject 2	6.68	$\sim 0.6^\circ$
Subject 3	11.24	$\sim 1^\circ$

Table 3: Experimental point-of-gaze estimation errors.

5.2 Demonstration of system performance with infants

In these experiments, the infants sat on their parent's lap with their face at about 65 cm from the screen. The point-of-gaze was

estimated for both eyes simultaneously. In order to calibrate for the subject-specific angular deviation of the visual axis from the optic axis, α_{eye} and β_{eye} , a looming stimulus was presented at the center of the screen for about 3 seconds. A 500 ms tone was produced at the onset of the stimulus. After the calibration procedure, two different experiments were carried out.

In the first experiment, the same looming stimulus was presented sequentially at the center of the screen and at several pre-selected random positions on the screen. The looming stimulus was presented at each position for 3 seconds and a 500 ms tone was produced at the onset of the stimulus at each new position. The recordings of the X and Y-coordinates (with respect to the WCS defined above) of the stimulus position and the point-of-gaze estimates as a function of time, corresponding to one of the infants, are shown in Fig. 8. Fig. 8 demonstrates that it is relatively easy to determine from the point-of-gaze estimates that the visual stimulus was seen by the infant at all seven positions.

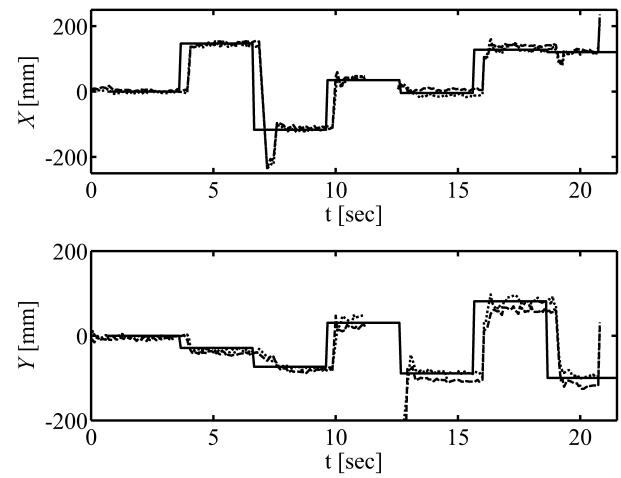


Figure 8: Recording of the stimulus position (solid line) and the estimates of the infant's point-of-gaze (dotted line: right eye, dashed line: left eye) as a function of time for the fixation experiment. The gap in the traces corresponds to a period without recording when the infant turned the head away.

In the second experiment, a full-screen black and white vertical grating (square wave, 50 % duty cycle, full contrast) with a spatial frequency of 0.8 cycles/cm (equivalent to about 0.7 degrees/degree) was moving from left to right at a speed of 25 mm/second (equivalent to about 2.2 degrees/second). A recording of the horizontal eye movements elicited by this stimulus is shown in Fig. 9. This recording exhibits four clear segments of the slow phase (slow smooth eye movements that track the motion of the grating) of optokinetic nystagmus. One fast phase (corrective saccades in the opposite direction of the motion) is also indicated in this figure. The infant's eye movements demonstrate that she followed the moving grating pattern.

The results shown in Figs. 8 and 9 demonstrate that by monitoring the point-of-gaze it is relatively easy to determine if and when infants see the visual stimuli.

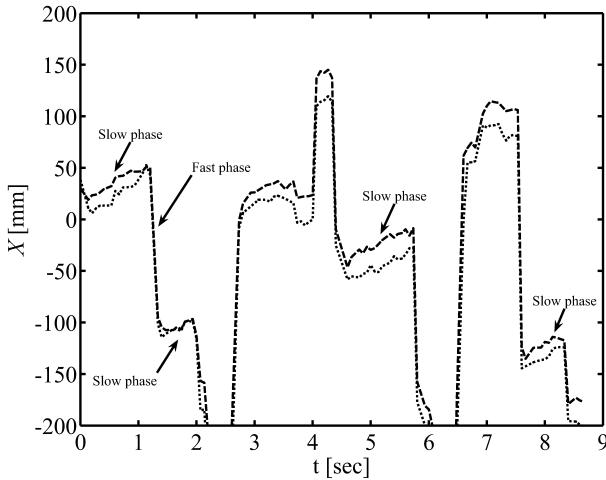


Figure 9: Optokinetic nystagmus in a 7-month old infant. The horizontal component of the infant's point-of-gaze estimates (dotted line: right eye, dashed line: left eye) is shown. The slow phase eye movements track the motion of the grating and the fast phase eye movements correspond to corrective saccades in the opposite direction.

6 Conclusions

A remote, non-contact point-of-gaze estimation system that requires a single-point calibration procedure and tolerates head movements in a volume of about 1 dm^3 was described. Experimental results with three adult subjects exhibited RMS point-of-gaze estimation errors of less than 12 mm (equivalent to about 1° of visual angle). The main sources of point-of-gaze estimation error are the deviation of the shape of real corneas from the spherical corneal shape assumed in the mathematical model, and the noise in the estimates of the centers of the pupil and corneal reflections in the eye images. Preliminary experimental results with two infants demonstrated the ability of the proposed system to monitor infants' point-of-gaze.

The simplification of the calibration procedure comes at the cost of system complexity: the need for at least two cameras and at least two light sources. In addition, a calibration procedure that relies on fixation on a single point can be less robust than calibration procedures that use multiple points. However, the method described in this paper can enable applications that rely on the analysis of visual scanning patterns of infants. These applications would be impossible or extremely difficult to carry out with point-of-gaze estimation methods that require multiple-point calibration procedures.

Acknowledgments

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Ontario Graduate Scholarship program (OGS).

References

- BOUGUET, J.-Y. Camera Calibration Toolbox for MATLAB. http://www.vision.caltech.edu/bouguetj/calib_doc/.
- CORBETT, M. C., ROSEN, E. S., and O'BRART, D. P. S. 1999. *Corneal Topography: Principles and Applications*. BMJ Books.
- DUCHOWSKI, A. T. 2002. A Breadth-First Survey of Eye-Tracking Applications, *Behavior Research Methods, Instruments, & Computers* 34, 4, 455-470.
- EIZENMAN, M., YU, L. H., GRUPP, L., EIZENMAN, E., ELLENBOGEN, M., GEMAR, M., and LEVITAN, R. D. 2003. A Naturalistic Visual Scanning Approach to Assess Selective Attention in Major Depressive Disorder, *Psychiatry Research* 118, 2, 117-128.
- FREY, L. A., WHITE, K. P., and HUTCHINSON, T. E. 1990. Eye-Gaze Word Processing, *IEEE Transactions on Systems, Man, and Cybernetics* 20, 4, 944-950.
- GORMAN, J. J., COGAN D. G., and GELLIS S. S. 1957. An Apparatus for Grading the Visual Acuity on the Basis of Optokinetic Nystagmus, *Pediatrics* 19, 1088-1092.
- GUESTRIN, E. D. and EIZENMAN, M. 2006. General Theory of Remote Gaze Estimation Using the Pupil Center and Corneal Reflections, *IEEE Transactions on Biomedical Engineering* 53, 6, 1124-1133.
- HARBLUK, J. L., NOY, Y. I., TRBOVICH, P. L., and EIZENMAN, M. 2007. An On-Road Assessment of Cognitive Distraction: Impacts on Drivers' Visual Behavior and Braking Performance, *Accident Analysis and Prevention* 39, 2, 372-379.
- OHNO, T., MUKAWA, N., and YOSHIKAWA, A. 2002. FreeGaze: A Gaze Tracking System for Everyday Gaze Interaction. In *Proceedings of the 2nd Symposium on Eye Tracking Research and Applications (ETRA 2002)*, 125-132.
- OHNO, T. and MUKAWA, N. 2004. A Free-Head, Simple Calibration, Gaze Tracking System that Enables Gaze-Based Interaction. In *Proceedings of the 3rd Symposium on Eye Tracking Research and Applications (ETRA 2004)*, 115-122.
- SHIH, S.-W. and LIU, J. 2004. A Novel Approach to 3-D Gaze Tracking Using Stereo Cameras, *IEEE Transactions on Systems, Man, and Cybernetics B* 34, 1, 234-245.
- TELLER, D. Y., MORSE R., BORTON R., and REGAL D. 1974. Visual Acuity for Vertical and Diagonal Gratings in Human Infants, *Vision Research* 14, 1433-1439.
- VILLANUEVA, A. and CABEZA, R. 2007. Models for Gaze Tracking Systems, *EURASIP Journal on Image and Video Processing* 2007, article ID 23570.
- YOUNG, L. R. and SHEENA, D. 1975. Methods and Designs - Survey of Eye Movement Recording Methods, *Behavior Research Methods & Instrumentation* 7, 5, 397-429.