Proceedings of the 29th Annual International
Conference of the IEEE EMBS
Cité Internationale, Lyon, France
August 23-26, 2007.

Remote Point-of-Gaze Estimation with Free Head Movements Requiring a Single-Point Calibration

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

Abstract—**This paper describes a method for remote, non-contact point-of-gaze (POG) 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 (virtual images of light sources) that are estimated from eye images captured by at least two cameras. Experimental results obtained with a prototype system that tolerates head movements in a volume of about 1 dm^3, exhibited RMS POG estimation errors of approximately 0.6-1° of visual angle. This system can enable applications with infants that, otherwise, would not be possible with existing POG estimation methods, which typically require multiple-point calibration procedures.**

I. Introduction

THE point-of-gaze (POG) 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 POG are used in a large variety of applications [1]-[4] such as studies of mood disorders, reading behavior and 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 POG 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 – Inset). 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, POG 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 with infants, such as the study of the development of the visual and oculomotor systems, and the assessment of visual function in preverbal infants. As shown in [5], [6], if at least two cameras and at least two light sources are used, it is possible to estimate the POG 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 could be performed even with babies by presenting a bright flashing stimulus on a dark uniform background to attract their attention.

The next Section presents a mathematical model for POG estimation with single-point calibration that is less sensitive to noise than the model from [5], [6]. Section III describes the set-up of a prototype system with two cameras and multiple light sources. Section IV shows experimental results in the presence of head movements. Finally, Section V summarizes the conclusions.

II. Mathematical Model

This section presents a mathematical model for remote POG 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. 1 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 $\mathbf{a}_1 \times \mathbf{a}_2 \bullet \mathbf{a}_3 = 0$, this condition can be formalized as

Manuscript received April 16, 2007. This work was supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC), and in part by scholarships from the Ontario Graduate Scholarship program (OGS). *Asterisk indicates corresponding author.*

*E. D. Guestrin is a 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 (phone: 1-416-978-2255; fax: 1-416-978-4317; 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).

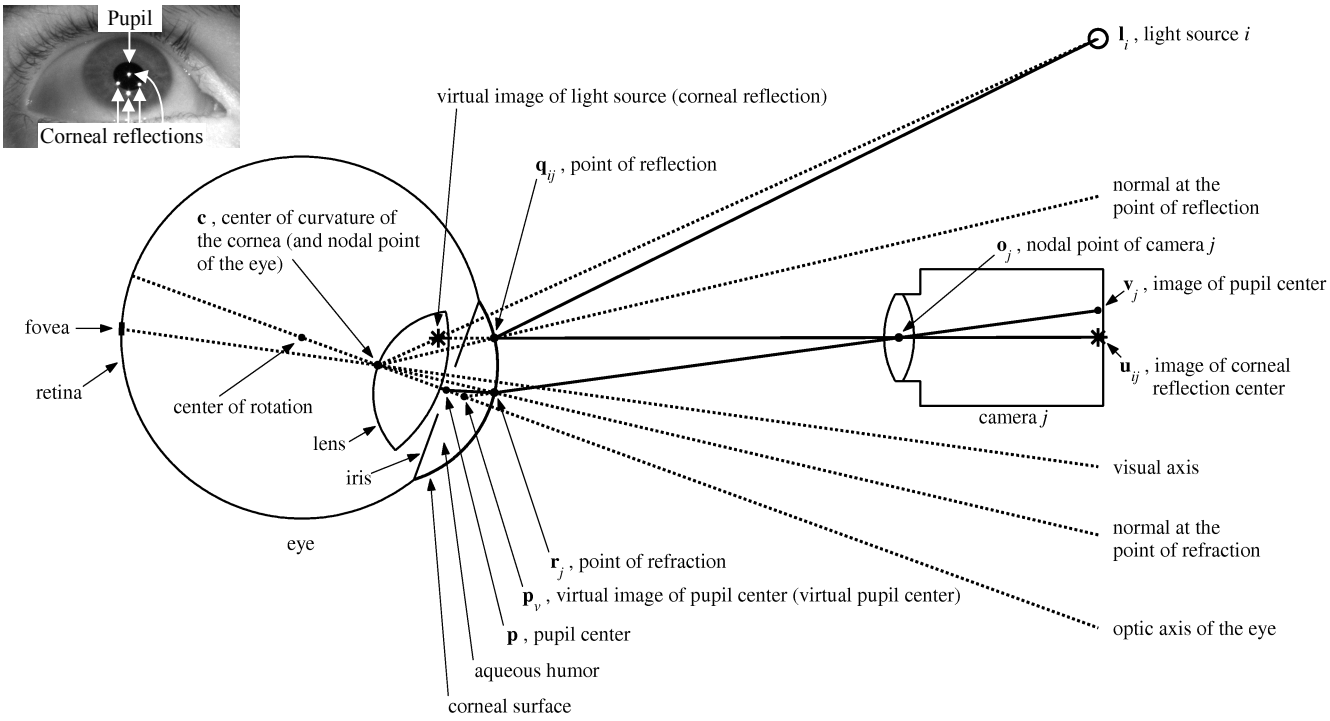

Fig. 1. Ray-tracing diagram (not to scale in order to be able to show all the elements of interest), showing schematic representations of the eye, a camera and a light source. Inset: close-up eye image indicating the pupil and corneal reflections.

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

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

$$\mathbf{c} = \mathbf{o}_j + k_{c,j}\,\mathbf{b}_j \;\text{ for some } k_{c,j} \,. \tag{2}$$

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

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

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 reflection centers, $\mathbf{u}_{ij}$, 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

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

This expression for $\mathbf{c}$ was found to be somewhat less sensitive to noise in the estimates of the image coordinates of the corneal reflections than the expression presented in [6].

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) 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 (virtual pupil center), $\mathbf{p}_v$, located on the extension of the refracted ray, i.e.,

$$\mathbf{p}_v = \mathbf{o}_j + k_{p,j}\underbrace{(\mathbf{o}_j - \mathbf{v}_j)}_{\mathbf{h}_j} \;\text{for some } k_{p,j} \,. \tag{5}$$

Having two cameras, the position of the virtual pupil center, $\mathbf{p}_v$, can be approximately found as the intersection of the lines given by (5), j = 1, 2, i.e.,

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

Since $\mathbf{c}$ is on the optic axis of the eye and assuming that $\mathbf{p}_v$ is also on the optic axis (Fig. 1), (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 strict terms, due to spherical aberration, the actual location of the virtual image of the pupil center is slightly different for each camera, and, for at least one of the cameras, it is not on the optic axis of the eye. Therefore, in general, $\mathbf{p}_v$ calculated from (6) may not be exactly on the optic axis of the eye. However, the POG estimation error due to the assumption that $\mathbf{p}_v$ is a unique point on the optic axis is relatively small and the resulting solution is significantly less sensitive to noise in the estimated coordinates of the images of the pupil center, $\mathbf{v}_j$, than the exact solution described in [5], [6].

The POG, $\mathbf{g}$, 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 5º [7]. Since the nodal point is within 1 mm of the center of curvature of the cornea [7], $\mathbf{c}$, it can be assumed to be coincident with $\mathbf{c}$ (Fig. 1). 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 $\mathbf{c}$ and $\mathbf{p}_v$ by solving

$$\frac{\mathbf{p}_v-\mathbf{c}}{\|\mathbf{p}_v-\mathbf{c}\|} = \begin{bmatrix}\cos\varphi_{eye}\sin\theta_{eye}\\ \sin\varphi_{eye}\\ -\cos\varphi_{eye}\cos\theta_{eye}\end{bmatrix}, \quad (7)$$

and the visual axis can be described in parametric form as

$$\mathbf{g} = \mathbf{c} + k_g\begin{bmatrix}\cos(\varphi_{eye}+\beta_{eye})\sin(\theta_{eye}+\alpha_{eye})\\ \sin(\varphi_{eye}+\beta_{eye})\\ -\cos(\varphi_{eye}+\beta_{eye})\cos(\theta_{eye}+\alpha_{eye})\end{bmatrix} \quad (8)$$

for all k_g. Since the scene plane is at $Z = 0$, the POG is given by (8) for a value of k_g such that the Z-component of $\mathbf{g}$, g_Z, equals 0, that is,

$$k_g = \frac{c_Z}{\cos(\varphi_{eye}+\beta_{eye})\cos(\theta_{eye}+\alpha_{eye})}\,. \quad (9)$$

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)–(7), α_{eye} and β_{eye} are found from (8) and (9).

In order to estimate the POG 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 [6]. 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.

III. System Set-Up

A prototype system to estimate the POG 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. 2).

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 of about ±5 cm laterally, ±4 cm vertically, and ±5 cm backwards/forward, before the eye features are no longer in the field of view of the cameras or are out of focus.

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 head position and POG on the screen, and in the presence of eyelid and eyelash interferences.

In order to be able to estimate the POG, 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 II and with the origin of the WCS at the center of the screen), are measured directly using rulers and calipers. The intrinsic camera

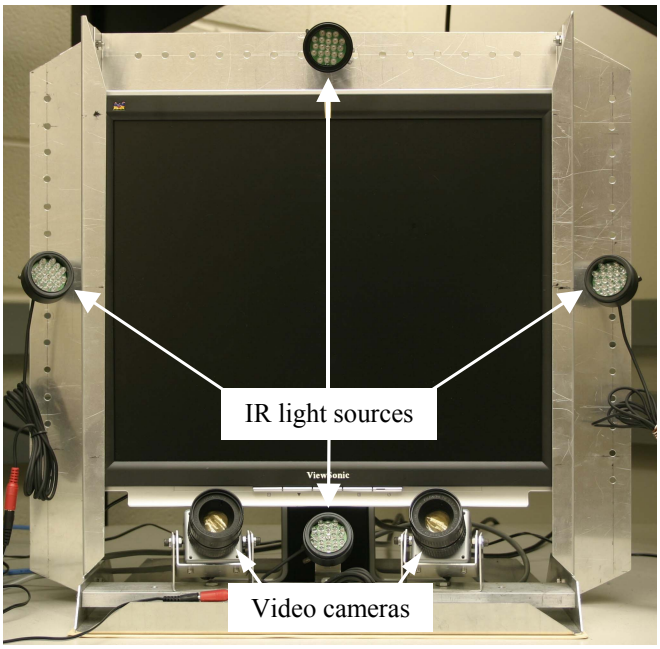

Fig. 2. System set-up.

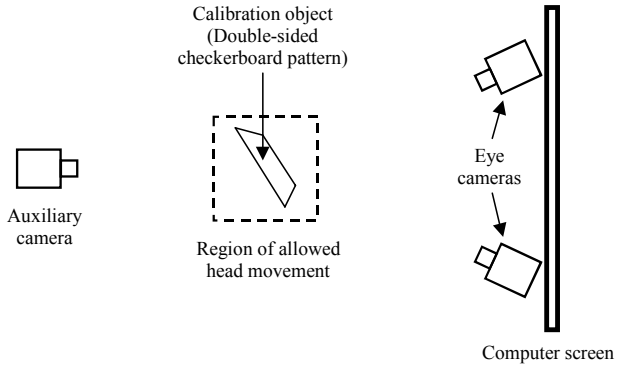

Fig. 3. Top-view schematic representation of the camera calibration set-up.

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. 2), and therefore cannot observe the monitor, an auxiliary camera that views both the computer screen and the region of allowed head movements (Fig. 3) is used together with a double-sided planar checkerboard pattern. The calibration is based on a camera calibration toolbox for MATLAB® [8] 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.

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 [5].

IV. Experimental Results

A preliminary evaluation of the performance of the prototype system was carried out through experiments with 3 adult subjects without eyeglasses or contact lenses. In these experiments, 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 (eye for which the POG was estimated) 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-by-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.

The POG was estimated using the two light sources that produced corneal reflections closest to the pupil center, in order to reduce the effect of the deviation of the shape of real corneas from the ideal spherical shape assumed in the model of Section II (corneal asphericity [6]). 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 [7]. By using the two corneal reflections that are closest to the pupil center in the eye images, it is therefore possible to reduce the POG estimation bias due to corneal asphericity.

Fig. 4 shows the experimental results for all 3 subjects and all 5 head positions, where the large crosses (+) indicate the target fixation points. The POG 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 POG estimates varied among the 3 subjects (△, ○, □), which can be attributed primarily to different degrees of corneal asphericity. For Subject 1 (△) and Subject 2 (○) most of the POG estimates are clustered together fairly close to the corresponding target fixation points, exhibiting relatively small bias. For Subject 3

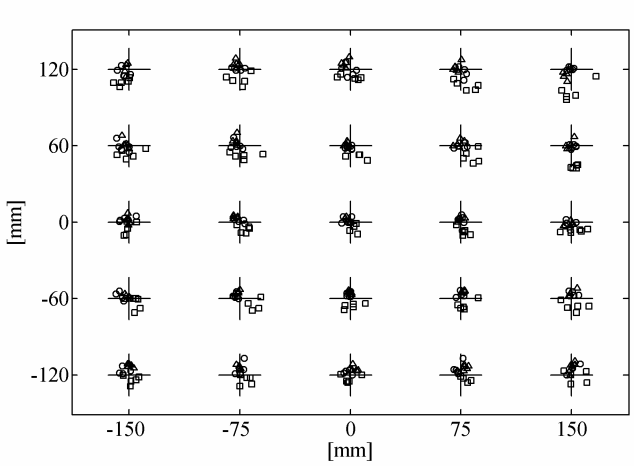

Fig. 4. Experimental POG estimates for all subjects, obtained using the average image coordinates of the eye features for each fixation point and each head position (1 estimate per fixation point and head position).

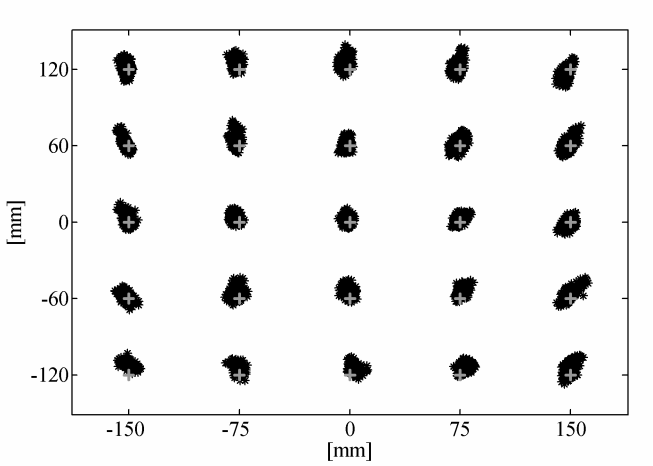

Fig. 5. Experimental POG estimates for Subject 1 and all 5 head positions, obtained using the individual estimates of the image coordinates of the eye features (50 estimates per fixation point and head position).

(□), the POG estimates exhibit the largest bias due to corneal asphericity.

Fig. 5 shows the POG 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 POG 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 POG estimates from their respective target fixation points, although relatively small, is primarily due to corneal asphericity.

Table I summarizes the RMS POG estimation errors, for each subject and all 5 head positions, when the POG was calculated using the individual estimates of the image coordinates of the eye features (50 estimates per fixation point and head position). The RMS POG 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. 4 and 5 suggest that it is possible to distinguish unambiguously more than 25 points on the screen of a 19" monitor.

TABLE I
EXPERIMENTAL POINT-OF-GAZE ESTIMATION ERRORS

	RMS error	Equivalent to
Subject 1	6.55 mm	~ 0.6º
Subject 2	6.68 mm	~ 0.6º
Subject 3	11.24 mm	~ 1º

V. CONCLUSIONS

A remote, non-contact POG estimation method that tolerates head movements and requires a single-point calibration procedure was presented. Experimental results obtained with a prototype system that tolerates head movements in a volume of about 1 dm^3, exhibited RMS POG estimation errors of less than 12 mm (equivalent to about 1° of visual angle). The main sources of POG 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.

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 and an accurate system calibration. In addition, a calibration procedure that relies on fixation on a single point is, in general, less robust than calibration procedures that use multiple points. However, despite this, the method described in this paper can enable applications with infants that, otherwise, would not be possible with existing methods that require multiple-point calibration procedures.

REFERENCES

[1] A. T. Duchowski, "A breadth-first survey of eye-tracking applications," *Behav. Res. Meth., Instrum., Comput.*, vol. 34, no. 4, pp. 455-470, Nov. 2002.

[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] J. L. Harbluk, Y. I. Noy, P. L. Trbovich, and M. Eizenman "An on-road assessment of cognitive distraction: impacts on drivers' visual behaviour and braking performance," *Accid. Anal. Prev.*, vol. 39, no. 2, pp. 372-379, Mar. 2007.

[4] 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.

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

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

[7] 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.

[8] J.-Y. Bouguet. (2006, Nov.) Camera Calibration Toolbox for MATLAB. California Institute of Technology, Department of Electrical Engineering, Pasadena, CA. [Online]. Available: http://www.vision.caltech.edu/bouguetj/calib_doc