NON-CONTACT CALIBRATION-FREE EYE-TRACKER

Elias D. Guestrin[a,b] and Moshe Eizenman[a,b,c]
[a] *Department of Electrical & Computer Engineering,* [b] *Institute of Biomaterials & Biomedical Engineering,* [c] *Department of Ophthalmology and Vision Sciences, University of Toronto, Toronto, Ontario, Canada.* E-mail: elias.guestrin@utoronto.ca

INTRODUCTION

Commercially available eye-trackers need to be calibrated for each subject by following a personal calibration procedure in which the subject is required to fixate multiple target points. This paper describes a system that can measure eye movements accurately without personal calibration, in the presence of head movements and without contact with the subject. While eliminating the need for personal calibration can be beneficial in general, it is critical for applications in which subjects cannot follow instructions (e.g., infants, mentally challenged individuals, non-human subjects), when subjects cannot maintain an accurate and stable fixation (e.g., Age-related Macular Degeneration), when subjects cannot fixate (e.g., oscillopsia) or when subjects are not willing to cooperate.

An example of an application that can benefit from this technology is the objective assessment of visual function in preverbal infants. One of the most common techniques to assess visual acuity in preverbal infants is the Optokinetic Nystagmus (OKN) technique. This technique is based on the fact that an awake subject whose visual field is filled with moving stripes cannot avoid showing the OKN eye movement pattern as long as the subject can resolve the stripes. The OKN pattern includes slow smooth tracking eye movements in the direction of the motion that are followed by fast corrective saccades in the opposite direction. The smallest stripe width that elicits these eye movements is correlated with the infant's visual acuity [1].

PRINCIPLES OF OPERATION

The measurement of eye movements with this system is based on the centers of the pupil and corneal reflections extracted from images captured by a pair of synchronized video cameras. 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.

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, Figure 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).

To derive the mathematical model used to measure eye movements, consider first a ray that originates 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

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

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),

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

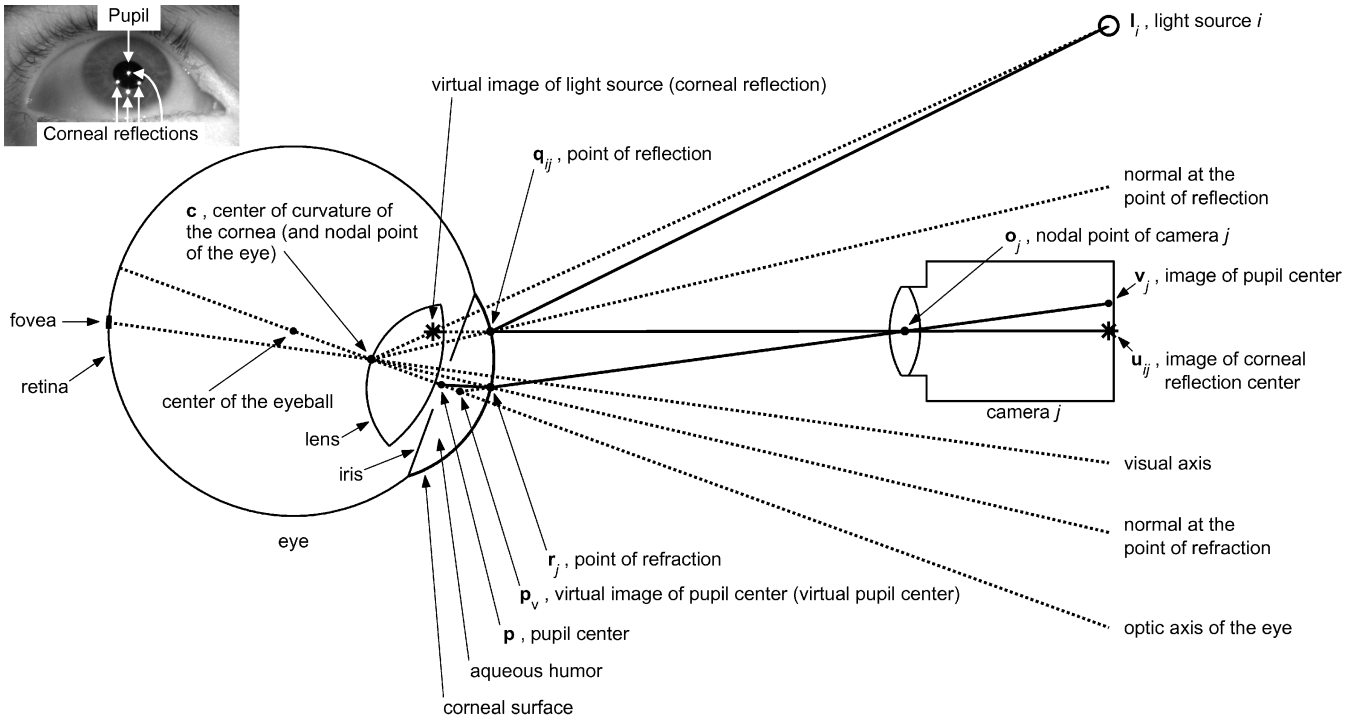

Figure 1: Ray-tracing diagram (not to scale in order to be able to show all the elements of interest), showing schematic representations of an eye, a camera and a light source. Inset: close-up eye image indicating the pupil and corneal reflections.

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 centers 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

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

Next, consider an imaginary ray that originates from the pupil center, $\mathbf{p}$, travels through the aqueous humor and cornea (effective index of refraction ≈ 1.3375 [2]) 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, $\mathbf{p}_{v,j}$, located on the extension of the refracted ray, i.e.,

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

In strict terms, the spatial location of $\mathbf{p}_{v,j}$ 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 $\mathbf{p}_{v,j}$ 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.,

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

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, $\mathbf{p}$ (Figure 1). Assuming that $\mathbf{p}_v$ is also on the optic axis, (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

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

In order to provide useful measurements of eye movements, the direction of the optic axis of the eye has to be expressed as angular components such as the pan (horizontal) angle and the tilt (vertical) angle. To obtain a formal expression for the calculation of the pan and tilt angles, suppose that the WCS has its *X*-axis horizontal, its positive *Y*-axis vertical pointing up and its positive *Z*-axis pointing towards the subject. Then, let θ_{eye} be the pan angle (angle between the projection of the optic axis of the eye on the *XZ*-plane and the –*Z*-axis of the WCS) and φ_{eye} be the tilt angle (angle between the optic axis of the eye and the *XZ*-plane of the WCS), with $\theta_{eye} > 0$ for rotations to the right, and $\varphi_{eye} > 0$ for rotations upwards. With these definitions, 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

$$\boldsymbol{\omega} = \begin{bmatrix} \cos\varphi_{eye} \sin\theta_{eye} \\ \sin\varphi_{eye} \\ -\cos\varphi_{eye} \cos\theta_{eye} \end{bmatrix} . \tag{8}$$

To be able to reconstruct the optic axis of the eye in 3-D space using the preceding set 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 coordinates of the centers of the pupil and corneal reflections that are estimated in each eye image are initially measured in pixels in an image coordinate system, they have to be transformed into the WCS [3], [4]. This transformation requires all intrinsic and extrinsic camera parameters (note that the position of $\mathbf{o}_j$ is one of the extrinsic camera parameters). In summary, to be able to use the preceding methodology to reconstruct the optic axis of the eye, the position of the light sources and the intrinsic and extrinsic camera parameters must be measured / estimated accurately. An example of how to do so is provided briefly in the next section.

SYSTEM SETUP

The prototype system used for the experiment described in the next section is shown in Figure 2. This system uses two synchronized monochrome CCD cameras (Scorpion SCOR-14SOM, Point Grey Research, Richmond, BC, Canada) with 35 mm lenses and four infrared light sources (850 nm) attached to a 19" LCD monitor by a custom aluminum frame.

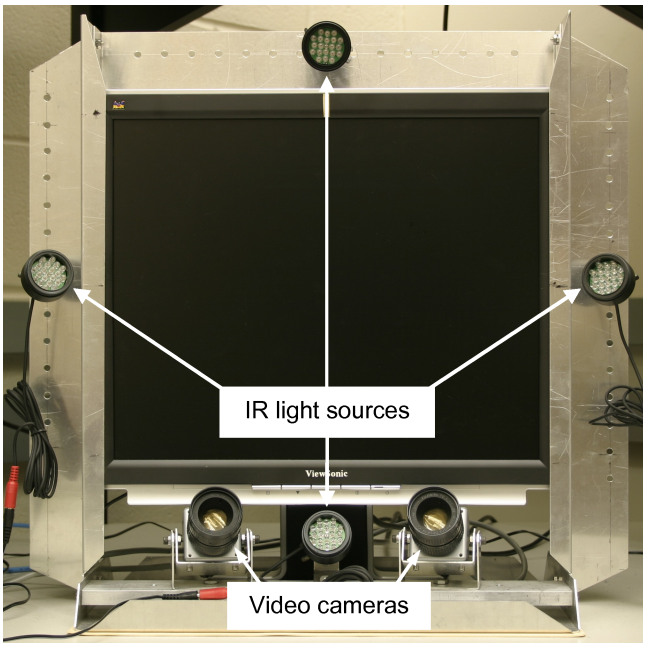

Figure 2: System setup.

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 a specific eye is tracked, about 16 cm when either eye is tracked and about 4 cm when both eyes are tracked 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 where the subject is looking at 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 the previous section (corneal asphericity [3], [4]). 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 [5]. 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 optic axis reconstruction bias due to corneal asphericity.

As indicated in the previous section, to be able to reconstruct the optic axis of the eye, the position of the light sources and the intrinsic and extrinsic camera parameters must be measured / estimated accurately. The positions of the light sources with respect to the WCS, which is attached to the LCD monitor (the *XY*-plane of the WCS is coincident with the plane of the screen and the origin of the WCS is 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 [4], [6].

EXPERIMENT

To demonstrate the ability of the prototype system to measure eye movements without personal calibration, a proof-of-concept experiment with a 7-month old infant was carried out. In this experiment, the infant sat on her mother's lap with her face at about 65 cm from the screen. Eye movements were measured for both eyes simultaneously.

The visual stimulus consisted of a full-screen black and white vertical grating (square wave, 50 % duty cycle, full contrast) with a spatial frequency of about 0.7 cycles/degree that moved from left to right at a speed of about 2.2 degrees/second. A recording of the horizontal eye movements elicited by this stimulus is shown in Figure 3. This recording exhibits four distinct segments of the slow phase of optokinetic nystagmus (slow smooth eye movements that track the motion of the grating). 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 could resolve the stripes of the moving grating pattern.

CONCLUSIONS

To the best of our knowledge, this was the first time that accurate measurements of eye movements were obtained remotely without having the subject complete a personal calibration procedure. The ability to measure eye movements accurately without personal calibration can enable many applications, such as the objective assessment of visual function in preverbal infants, that were previously very difficult or impossible.

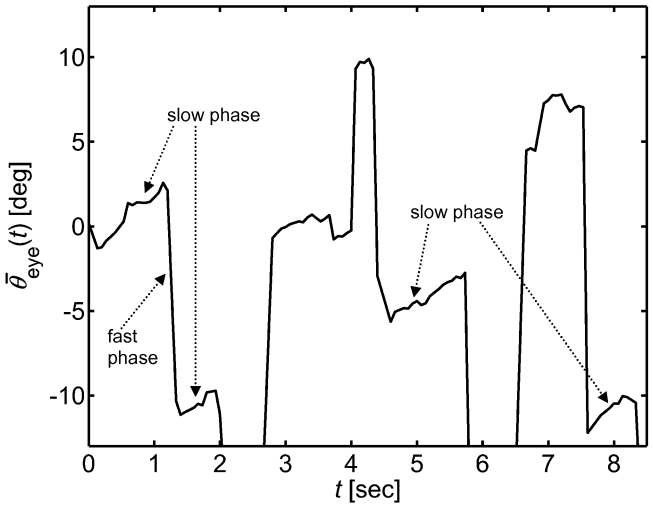

Figure 3: Optokinetic nystagmus in a 7-month old infant. The horizontal eye movements (mean of the right and left eyes) is shown. The slow phase eye movements track the motion of the grating and the fast phase eye movements correspond to corrective saccades in the opposite direction.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the funding from the Ontario Graduate Fellowship (OGS) program to Elias Guestrin and from the Natural Sciences and Engineering Research Council of Canada (NSERC) to Moshe Eizenman. The authors would also like to express their gratitude to the parents of the infant that participated in the experiment.

REFERENCES

[1] J. J. Gorman, D. G. Cogan, and S. S. Gellis, "An apparatus for grading the visual acuity on the basis of optokinetic nystagmus," *Pediatrics*, vol. 19, pp. 1088-1092, 1957.

[2] M. C. Corbett, E. S. Rosen, and D. P. S. O'Brart, *Corneal Topography: Principles and Applications*. London, UK: BMJ Books, 1999, p. 6.

[3] 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.

[4] E. D. Guestrin, "Remote, non-contact gaze estimation with minimal subject cooperation," PhD Thesis, Department of Electrical & Computer Engineering, and Institute of Biomaterials & Biomedical Engineering, University of Toronto, Toronto, ON, Canada, 2010. (http://hdl.handle.net/1807/24349)

[5] 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.

[6] E. D. Guestrin and M. Eizenman, "Remote point-of-gaze estimation requiring a single-point calibration for applications with infants," in *Proc. ETRA 2008*, Mar. 2008, pp.267-274.