# REMOTE POINT-OF-GAZE ESTIMATION FOR STUDIES OF SELECTIVE ATTENTION AND MOOD DISORDERS

Elias Guestrin[a,b] and Moshe Eizenman[a,b,c]
[a] *Department of Electrical and Computer Engineering,* [b] *Institute of Biomaterials and Biomedical Engineering,* [c] *Department of Ophthalmology, University of Toronto, 164 College St., Toronto, Ontario, Canada M5S 3G9. E-mail: elias.guestrin@utoronto.ca*

## INTRODUCTION

The point-of-gaze is the point in space that is imaged on the center of the highest acuity region of the retina (fovea) of each eye. Estimation of the point-of-gaze is required for the determination of visual scanning patterns. Since visual scanning patterns closely follow shifts in attentional focus, they provide insight into human cognitive processes. As such, the analysis of visual scanning patterns is used in studies of mood, attention, perception and learning disorders. In the past, our group performed a clinical study on visual selective attention and mood disorders [1]. A head-mounted point-of-gaze estimation system [2] was used to record visual scanning patterns while subjects viewed a sequence of slides, each containing images with different themes. The study showed that depressed subjects spent significantly more time than normal control subjects looking at images with dysphoric themes. This observation can provide an objective measure of depression.

Head-mounted systems, however, become uncomfortable when used for long periods of time. In order to improve subjects' comfort, a system that estimates the point-of-gaze on a computer screen remotely was developed. The next section describes the principles of operation of the novel system. The third section shows experimental results that illustrate the system performance while the forth section summarizes the conclusions.

## PRINCIPLES OF OPERATION

The novel remote point-of-gaze estimation system utilizes two near-IR (850 nm) light sources that are symmetrically positioned at the sides of a 19" computer monitor, and a video camera (640 x 480 pixels, 1/3" CCD with a 35 mm lens) that is centered under the screen (Fig. 1). Since the front corneal surface resembles a convex mirror, each light source results in the formation of a virtual image (first Purkinje image) that appears as a bright spot called corneal reflection or glint (Fig. 2 - Insert). By using the centers of the pupil and corneal reflections that are estimated

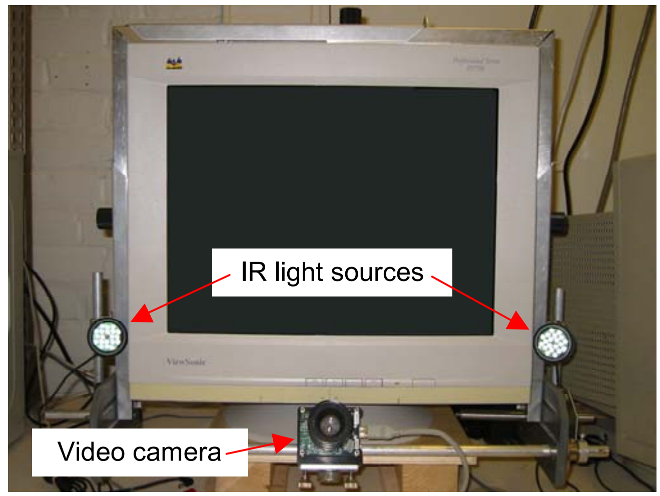


Figure 1: System setup.

in each eye image and the novel mathematical model described below, the point-of-gaze is calculated.

The point-of-gaze is formally defined as the intersection of the visual axes of both eyes with the 3D scene. The visual axis is the line connecting the center of the fovea with the nodal point[1] of the eye's optics (Fig. 2). Since in the human eye the visual axis deviates from the optic axis [3], the estimation of the point-of-gaze can be divided into two stages: (i) the reconstruction of the optic axis of the eye from the centers of the pupil and corneal reflections in the images of the eye; (ii) the reconstruction of the visual axis from the optic axis, and the estimation of the point-of-gaze.

Under the assumptions that the light sources are modeled as point sources and the video camera is modeled as a pinhole camera, Fig. 2 presents a ray-tracing diagram of the system and the eye, where all points are represented as 3D column vectors (bold font) in a right-handed Cartesian world coordinate system (WCS). In order to derive the mathematical

[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.

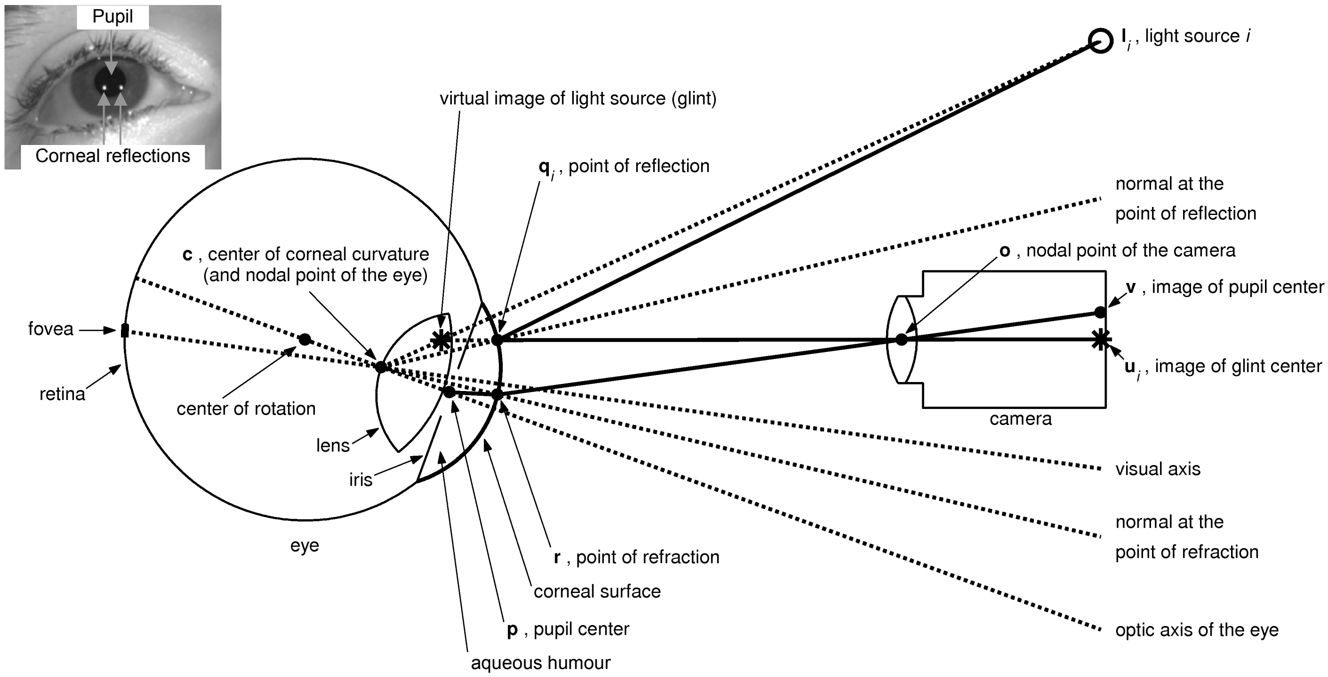


Figure 2: Ray-tracing diagram (not to scale in order to be able to show all the elements of interest), showing schematic representations of the eye, the camera and a light source. Insert: close-up eye image.

model, first consider a ray that comes from light source $i$, $\mathbf{l}_i$, and reflects at a point $\mathbf{q}_i$ on the corneal surface such that the reflected ray passes through the nodal point of the camera (a.k.a. camera center, center of projection), $\mathbf{o}$, and intersects the camera image plane at a point $\mathbf{u}_i$. The condition that the ray coming from the point of reflection $\mathbf{q}_i$ and passing through the nodal point of the camera, $\mathbf{o}$, intersects the camera image plane at point $\mathbf{u}_i$, can be written in parametric form as

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

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

$$\|\mathbf{q}_i - \mathbf{c}\| = R, \tag{2}$$

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

The law of reflection states two conditions: (i) the incident ray, the reflected ray and the normal at the point of reflection are coplanar; and (ii) the angles of incidence and reflection are equal. Since vector ($\mathbf{q}_i$ – $\mathbf{c}$) is normal to the spherical surface at the point of reflection $\mathbf{q}_i$, condition (i) implies that points $\mathbf{l}_i$, $\mathbf{q}_i$, $\mathbf{c}$ and $\mathbf{o}$ 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 (i) can be formalized as

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

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

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

Next, consider a ray that comes from the pupil center, $\mathbf{p}$, and refracts at point $\mathbf{r}$ on the corneal surface such that the refracted ray passes through the nodal point of the camera, $\mathbf{o}$, and intersects the camera image plane at a point $\mathbf{v}$. The condition that the ray coming from the point of refraction $\mathbf{r}$ and passing through the nodal point of the camera, $\mathbf{o}$, intersects the camera image plane at point $\mathbf{v}$, can be expressed in parametric form as

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

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

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

The law of refraction states two conditions: (i) the incident ray, the refracted ray and the normal at the point of refraction are coplanar; and (ii) the angle of incidence, $\theta_1$, and the angle of refraction, $\theta_2$, 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 ($\mathbf{r}$ – $\mathbf{c}$) is normal to the spherical surface at the point of refraction **r**, condition (i) implies that points **p**, **r**, **c** and **o** are coplanar, which can be formalized as

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

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

$$\begin{aligned} &n_1\cdot\|(\mathbf{r}-\mathbf{c})\times(\mathbf{p}-\mathbf{r})\|\cdot\|\mathbf{o}-\mathbf{r}\| \\ &= n_2\cdot\|(\mathbf{r}-\mathbf{c})\times(\mathbf{o}-\mathbf{r})\|\cdot\|\mathbf{p}-\mathbf{r}\| \ , \end{aligned} \tag{8}$$

where $n_1$ is the effective index of refraction of the aqueous humour and cornea combined and $n_2$ is the index of refraction of air (≈ 1). In this model, the refraction at the aqueous humour-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 humour and cornea are considered as a homogenous medium.

Finally, 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 (**p**) and the center of corneal curvature (**c**), if the above system of equations is solved for **c** and **p**, the optic axis of the eye in space can be reconstructed as the line defined by these two points.

The point-of-gaze, **g**, is defined as the intersection of the visual axis with the scene. The visual axis is the line defined by the nodal point of the eye[2] and the center of the fovea (i.e. the highest acuity region of the retina corresponding to 0.6 to 1º of visual angle) [3]. Since the visual axis deviates from the optic axis (Fig. 2), the relation between these two axes has to be modeled. The visual and optic axes 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 [3], for the sake of simplicity, the nodal point is assumed to be coincident with the center of corneal curvature, **c**.

In a typical adult human eye, the intersection of the visual axis with the retina (center of the fovea) falls about 4-5º temporally and about 1.5º below the point of intersection of the optic axis and the retina [4]. In order to formalize the relation between the two axes, suppose that the scene is a vertical plane (i.e., the computer screen) and that the WCS has its *XY*-plane 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 $\theta_{eye}$ and the tilt (vertical) angle $\varphi_{eye}$, where $\theta_{eye} = \varphi_{eye} = 0$ when the optic axis is normal to the scene plane (parallel to the *Z*-axis of the WCS) and the pan angle is positive for rotations to the right while the tilt angle is positive for rotations upwards. The angles $\theta_{eye}$ and $\varphi_{eye}$ can be obtained from **c** and **p** by solving the following equation:

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

If the horizontal and vertical angles between the visual and optic axes are given by $\alpha_{eye}$ and $\beta_{eye}$, respectively, the orientation of the visual axis can be expressed by the pan angle ($\theta_{eye} + \alpha_{eye}$) and the tilt angle ($\varphi_{eye} + \beta_{eye}$), where all angles are signed ($\alpha_{eye} < 0$ for the right eye, $\alpha_{eye} > 0$ for the left eye, and $\beta_{eye} > 0$).

Following the above discussion, the visual axis can be then described in parametric form as

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

for all $k_g$. Since it was assumed that the scene plane is at $Z = 0$, the point-of-gaze is given by (11) for a value of $k_g$ such that the *Z*-component of **g**, $g_Z$, equals 0, i.e.,

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

In the above system of equations, the coordinates of the centers of the pupil (**v**) and corneal reflections ($\mathbf{u}_i$) in the eye images are measured in the WCS. Since the centers of the pupil and glints that are estimated in each eye image are measured in pixels in an image coordinate system, they have to be transformed into the WCS [5]. This transformation requires all intrinsic and extrinsic camera parameters, including the position of the nodal point (**o**).

In order to be able to estimate the point-of-gaze 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, **o**) are measured / estimated only once during system set up. The subject-specific eye

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

parameters ($R$, $K$, $n_1$, $\alpha_{eye}$ and $\beta_{eye}$) are obtained through a calibration procedure that is performed once for each subject. In the calibration procedure, the subject fixates on a set of points that are presented sequentially on the screen. Using the coordinates of the centers of pupil and corneal reflections corresponding to the fixation points, 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.

Knowing the system and subject-specific eye parameters, the point-of-gaze is estimated as follows. First, the centers of the pupil and corneal reflections are estimated in the image of the eye and their image coordinates are transformed into world coordinates (**v** and $\mathbf{u}_i$, $i$ = 1,2). Then, equations (1)-(4), $i$ = 1,2, are solved for the center of corneal curvature, **c**. Knowing **c**, (5) and (6) are used to compute the point of refraction, **r**. Knowing **c** and **r**, (7)-(9) are used to compute the pupil center, **p**. Knowing **c** and **p**, (10) is used to compute the orientation of the optic axis of the eye in space. Finally, (11) and (12) are used to estimate the point-of-gaze on the screen. A more detailed discussion on the solution of the above system of equations is provided in [5], where it is also shown that one camera and two light sources is the simplest system configuration that allows for the estimation of the point-of-gaze from the centers of the pupil and corneal reflections while allowing for free head movements.

## SYSTEM PERFORMANCE

This specific system can estimate the point-of-gaze in real time at 30 Hz while allowing for moderate head movements of about ±3 cm laterally, ±2 cm vertically, and ±4 cm backwards/forward, before the eye features are no longer in the field of view of the camera or are out of focus (a new version of the system that will allow for head movements within a cubic foot of space is under development). Experimental results obtained with four subjects yielded RMS point-of-gaze estimation errors that, for all experimental conditions, were less than 10 mm (about 0.9º of visual angle when the eye is at a typical viewing distance of 65 cm from the computer screen). This implies that it is possible to distinguish unambiguously about 18 x 14 points on the screen of a 19” monitor. Fig. 3 shows a subject’s visual scanning pattern on a slide containing four different images, similar to the slides used in the clinical study on visual selective attention and mood disorders [1], [2]. The circles represent point-of-gaze estimates while the lines joining consecutive estimates represent saccadic eye movements. As it can be seen from this figure, this system allows for the determination of not only which image is the focus of attention but also which detail within the image is being examined.

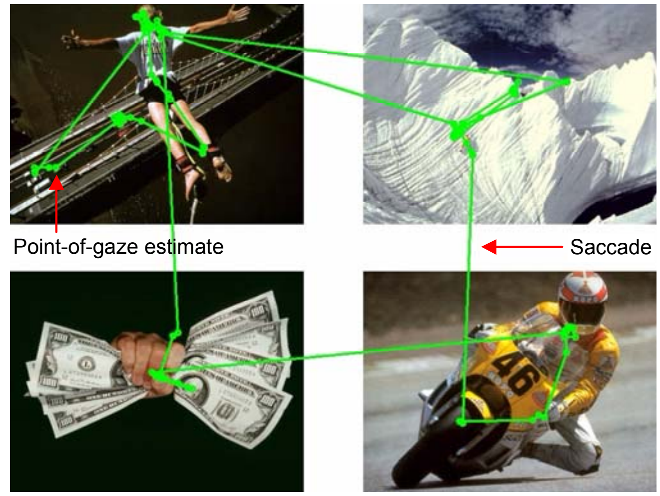

Figure 3: Example of a visual scanning pattern.

## CONCLUSIONS

This paper presented a novel remote point-of-gaze estimation system based on the centers of the pupil and corneal reflections. Using a novel mathematical model, the point-of-gaze can be estimated in the presence of head movements after having each subject complete a relatively simple calibration procedure. The system performance is sufficient for the analysis of characteristics of visual scanning patterns (e.g., viewing time and viewing frequency on each image) that can be used as indicators of attentional bias for the assessment of mood disorders.

## REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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, June 2006.