User-Calibration-Free Remote Gaze Estimation System

Dmitri Model
University of Toronto
dmitri.model@utoronto.ca

Moshe Eizenman
University of Toronto
eizenm@ecf.utoronto.ca

Abstract

Gaze estimation systems use calibration procedures that require active subject participation to estimate the point-of-gaze accurately. In these procedures, subjects are required to fixate on a specific point or points in space at specific time instances. This paper describes a gaze estimation system that does not use calibration procedures that require active user participation. The system estimates the optical axes of both eyes using images from a stereo pair of video cameras without a personal calibration procedure. To estimate the point-of-gaze, which lies along the visual axis, the angles between the optical and visual axes are estimated by a novel automatic procedure that minimizes the distance between the intersections of the visual axes of the left and right eyes with the surface of a display while subjects look naturally at the display (e.g., watching a video clip). Experiments with four subjects demonstrate that the RMS error of this point-of-gaze estimation system is 1.3º.

Keywords: Point-of-gaze, remote gaze estimation, calibration free, minimal subject cooperation.

1 Introduction

The point-of-gaze (PoG) is the point within the visual field that is imaged on the highest acuity region of the retina that is known as the fovea. Systems that estimate the PoG are used in a large variety of applications [Duchowski 2002] such as studies of emotional and cognitive processes [Eizenman et al. 2006],[Rayner 1998], driver behavior [Harbluk et al. 2007], marketing and advertising [Loshe 1997], pilot training [Wetzel et al. 1997], ergonomics [Goldberg and Kotval 1999] and human-computer interfaces [Hutchinson et al. 1989].

Remote Eye-Gaze Estimation (REGT) systems use calibration procedures to estimate subject-specific parameters that are needed for the accurate calculation of the PoG. In these procedures, subjects are required to fixate specific points at specific time instances. However, these calibration routines cannot be reliably performed in some studies, such as studies with young children or mentally challenged people, as well as studies for which covert gaze monitoring is required. To enable such studies it is crucial to estimate human PoG without calibration procedures that require active user participation.

ETRA 2010, Austin, TX, March 22 – 24, 2010.
© 2010 ACM 978-1-60558-994-7/10/0003 $10.00

This paper presents such a system. The system uses images from a stereo pair of video cameras to estimate the center of curvature of the cornea and the optical axis of each eye without any personal calibration [Guestrin and Eizenman 2006; Guestrin and Eizenman 2008; Shih et al. 2000; Shih and Liu 2004]. Then, the angles between the optical and visual axes (human gaze is directed along the visual axis) are estimated using a novel AAE (Automatic Angles Estimation) algorithm that rely on the assumption that the visual axes of left and right eyes intersect on the surface of the display [Eizenman et al. 2009; Model et al. 2009]. The estimation of subject-specific angles between the visual and optical axes is performed while subject looks naturally at the display.

This paper is organized as follows. A methodology for the estimation of the optical axis of the eye that does not require a priori knowledge of subject-specific eye parameters is described in the next section. The AAE algorithm and numerical simulations are presented in Section 3. Experimental results and conclusions are presented in Sections 4 and 5, respectively.

2 Estimation of the Optical Axis

Figure 1 presents a general model of gaze estimation systems with any number of cameras and any number of light sources. 3-D points are indicated in a bold font and are represented as 3-D column vectors in a right-handed Cartesian world coordinate system. Light sources are modeled as point sources, video cameras are modeled as pinhole cameras and the front surface of the cornea is modeled as a spherical section. In this model, the line connecting the center of curvature of the cornea, $\mathbf{c}$, and the pupil center, $\mathbf{p}$, defines the optical axis of the eye. The line connecting the center of curvature of the cornea, $\mathbf{c}$, and the center of the fovea defines the visual axis. The PoG is at the intersection of the visual axis with the display.

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 ξ_1, ξ_2 and ξ_3 satisfy $\xi_1 \times \xi_2 \bullet \xi_3 = 0$, this condition can be formalized as

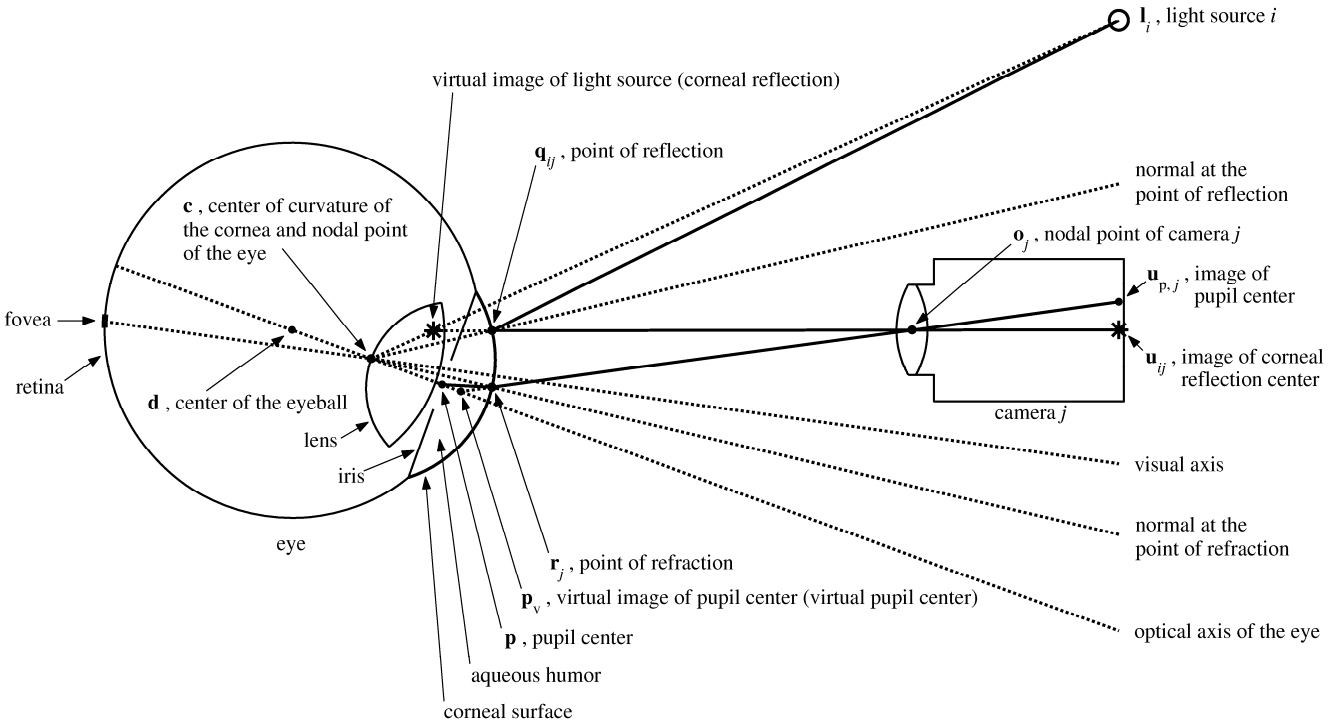


Figure 1 *Ray-tracing diagram (not to scale in order to be able to show all the elements of interest), showing schematic representations of the eye, a camera and a light source.*

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

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 $\boldsymbol{\eta}_j$ is a vector in the direction of the line of intersection of the planes, then

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

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

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

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

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

Next, consider an imaginary ray that originates 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 $\mathbf{u}_{\mathrm{p},j}$. This refraction results in the formation of a virtual image of the pupil center (virtual pupil center), $\mathbf{p}_{\mathrm{v},j}$, located on the extension of the refracted ray, i.e.,

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

In strict terms, the spatial location of $\mathbf{p}_{\mathrm{v},j}$ 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 $\mathbf{p}_{\mathrm{v},j}$ will be slightly different for each of the two cameras. Despite this, an approximate virtual image of the pupil center, $\mathbf{p}_\mathrm{v}$, 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.,

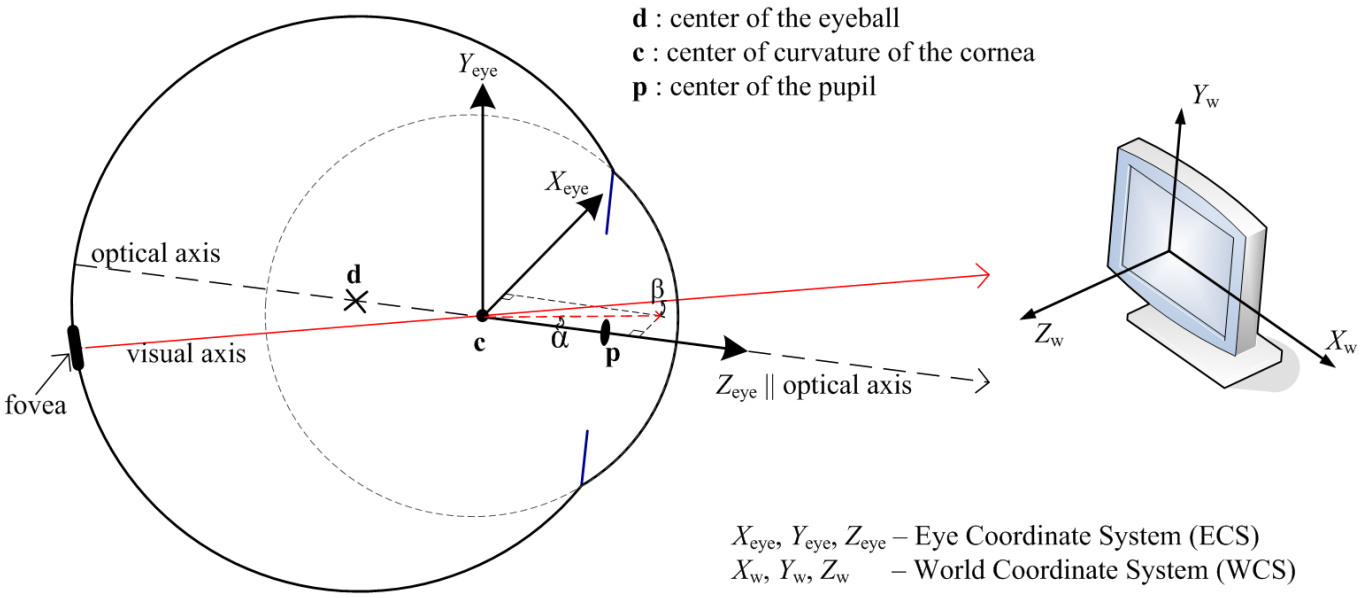


Figure 2 *Simplified schematics of the eye. The optical axis of the eye connects the center of the pupil with the center of curvature of the cornea. Gaze is directed along the visual axis, which connects the center of the region of highest acuity of the retina (fovea) with the center of curvature of the cornea.*

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

Since **c** is on the optical axis of the eye and assuming that $\mathbf{p}_v$ is also on the optical axis (Fig. 1), (3)–(6) provide a closed-form solution for the reconstruction of the optical axis of the eye in 3-D space without the knowledge of any subject-specific eye parameter. In particular, the direction of the optical axis of the eye is given by the unit vector

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

The PoG, **g**, is defined as the intersection of the *visual* axis, rather than the *optical* axis, with the scene. The visual axis might deviate from the optical axis by as much as 5° [Young and Sheena 1975]. Therefore, if the PoG is estimated from the intersection of the *optical* axis with the scene, the estimation might suffer from large, subject dependent errors. As an example, for the experiments described in Section 4 the RMS errors in PoG estimation by the intersection of the *optical* axis with the computer monitor ranges from 1° to 4.5° (2.5° on average for all four subjects). If the angle between the *optical* and *visual* axes is estimated with a one-point user-calibration procedure [Guestrin and Eizenman 2008], during which the subject is required to fixate a known point in space for several seconds, the PoG estimation error is reduced to 0.8° RMS. In the next section, we present a novel AAE algorithm, in which the angles between the *optical* and *visual* axes are estimated automatically.

3 The AAE Algorithm

3.1 Theory

The AAE algorithm is based on the assumption that at each time instant the visual axes of both eyes intersect on the observation surface (e.g., display). The unknown angles between the *optical* and *visual* axes can be estimated by minimizing the distance between the intersections of the left and right visual axes with that observation surface.

Two coordinate systems are used to describe the relation between the *optical* and *visual* axes of the eye. The first is a stationary right-handed Cartesian World Coordinate System (WCS) with the origin at the center of the display, the X_w-axis in the horizontal direction, the Y_w-axis in the vertical direction and the Z_w-axis perpendicular to the display (see Figure 2). The second is a non-stationary right-handed Cartesian Eye Coordinate System (ECS), which is attached to the eye, with the origin at the center of curvature of the cornea, **c**, the Z_{eye} axis that coincides with the optical axis of the eye and X_{eye} and Y_{eye} axes that, in the primary gaze position, are in the horizontal and vertical directions, respectively. The X_{eye}-Y_{eye} plane rotates according to Listing's law [Helmholtz 1924] around the Z_{eye} axis for different gaze directions.

In the ECS, the unknown 3-D angle between the optical and the visual axes of the eye can be expressed by the horizontal[1], α, and vertical[2], β, components of this angle (see Figure 2). The unit vector in the direction of the visual axis with respect to the ECS, $\mathbf{v}_{ECS}$, is then expressed as

[1] The angle between the projection of the visual axis on the X_{eye}-Z_{eye} plane and the Z_{eye} axis. It is equal to 90° if the visual axis is in the $-X_{eye}$ direction.

[2] The angle between the visual axis and its projection on the X_{eye}-Z_{eye} plane. It is equal to 90° if the visual axis is in the $+Y_{eye}$ direction.

$$\mathbf{v}_{\mathrm{ECS}}(\alpha,\beta)=\begin{bmatrix}-\sin(\alpha)\cos(\beta)\\ \sin(\beta)\\ \cos(\alpha)\cos(\beta)\end{bmatrix}. \quad (8)$$

The unit vector in the direction of the visual axis with respect to the WCS, $\mathbf{v}$, can be expressed as

$$\mathbf{v}(\alpha,\beta)=\mathbf{R}\,\mathbf{v}_{\mathrm{ECS}}(\alpha,\beta) \quad (9)$$

where **R** is the rotation matrix from the ECS to the WCS (independent of α and β), which can be calculated from the orientation of the optical axis of the eye and Listing's law [Helmholtz 1924].

Because the visual axis goes through the center of curvature of the cornea, **c**, and the PoG is defined by the intersection of the visual axis with the display (Z_w = 0), the PoG in the WCS is given by

$$\mathbf{g}(\alpha,\beta)=\mathbf{c}+k(\alpha,\beta)\,\mathbf{v}(\alpha,\beta)=\mathbf{c}+k(\alpha,\beta)\,\mathbf{R}\,\mathbf{v}_{\mathrm{ECS}}(\alpha,\beta) \quad (10)$$

where $k(\alpha,\beta)$ is a line parameter defined by the intersection of the visual axis with the surface of the display:

$$k(\alpha,\beta)=-\frac{\mathbf{c}\bullet\mathbf{n}}{\mathbf{v}(\alpha,\beta)\bullet\mathbf{n}} \quad (11)$$

where $\mathbf{n}=[0\ 0\ 1]^T$ is the normal to the display surface and "^{T}" denotes transpose.

The estimation of α^{L}, β^{L}, α^{R} and β^{R} is based on the assumption that at each time instant the visual axes of both eyes intersect on the surface of the display (superscripts "L" and "R" are used to denote parameters of the left and right eyes, respectively). The unknown angles α^{L}, β^{L}, α^{R} and β^{R} can be estimated by minimizing the distance between the intersections of the left and right visual axes with that surface (left and right PoGs).

The objective function to be minimized is then

$$F(\alpha^{\mathrm{L}},\beta^{\mathrm{L}},\alpha^{\mathrm{R}},\beta^{\mathrm{R}})=\sum_i\left\|\mathbf{g}_i^{\mathrm{L}}(\alpha^{\mathrm{L}},\beta^{\mathrm{L}})-\mathbf{g}_i^{\mathrm{R}}(\alpha^{\mathrm{R}},\beta^{\mathrm{R}})\right\|_2^2 \quad (12)$$

where the subscript *i* identifies the *i*-th gaze sample.

The above objective function is non-linear, and thus a numerical optimization procedure is required to solve for the unknown angles α^{L}, β^{L}, α^{R} and β^{R}. However, since the deviations of the unknown angles α^{L}, β^{L}, α^{R} and β^{R} from the expected "average" values α_0^{L}, β_0^{L}, α_0^{R} and β_0^{R} are relatively small, a linear approximation of (10) can be obtained by using the first three terms of its Taylor's series expansion:

$$\mathbf{g}(\alpha,\beta)\approx\mathbf{g}(\alpha_0,\beta_0)+\left.\frac{\partial\mathbf{g}(\alpha,\beta)}{\partial\alpha}\right|_{\alpha_0,\beta_0}(\alpha-\alpha_0)+\left.\frac{\partial\mathbf{g}(\alpha,\beta)}{\partial\beta}\right|_{\alpha_0,\beta_0}(\beta-\beta_0) \quad (13)$$

Let

$$\mathbf{g}_0=\mathbf{g}(\alpha_0,\beta_0)=\mathbf{c}+k_0\mathbf{v}_0 \quad (14)$$

where $\mathbf{v}_0=\mathbf{v}(\alpha_0,\beta_0)$ and $k_0=k(\alpha_0,\beta_0)$.

Then, using (10),

$$\mathbf{a}=\left.\frac{\partial\mathbf{g}(\alpha,\beta)}{\partial\alpha}\right|_{\alpha_0,\beta_0}=k_\alpha\mathbf{v}_0+k_0\mathbf{v}_\alpha \quad (15)$$

$$\mathbf{b}=\left.\frac{\partial\mathbf{g}(\alpha,\beta)}{\partial\beta}\right|_{\alpha_0,\beta_0}=k_\beta\mathbf{v}_0+k_0\mathbf{v}_\beta \quad (16)$$

where, from (8)-(9),

$$\mathbf{v}_\alpha=\left.\frac{\partial\mathbf{v}(\alpha,\beta)}{\partial\alpha}\right|_{\alpha_0,\beta_0}=\mathbf{R}\begin{bmatrix}-\cos(\alpha_0)\cos(\beta_0)\\ 0\\ -\sin(\alpha_0)\cos(\beta_0)\end{bmatrix} \quad (17)$$

$$\mathbf{v}_\beta=\left.\frac{\partial\mathbf{v}(\alpha,\beta)}{\partial\beta}\right|_{\alpha_0,\beta_0}=\mathbf{R}\begin{bmatrix}\sin(\alpha_0)\sin(\beta_0)\\ \cos(\beta_0)\\ -\cos(\alpha_0)\sin(\beta_0)\end{bmatrix} \quad (18)$$

and, from (11),

$$k_\alpha=\left.\frac{\partial k(\alpha,\beta)}{\partial\alpha}\right|_{\alpha_0,\beta_0}=-k_0\frac{\mathbf{v}_\alpha\bullet\mathbf{n}}{\mathbf{v}_0\bullet\mathbf{n}} \quad (19)$$

$$k_\beta=\left.\frac{\partial k(\alpha,\beta)}{\partial\beta}\right|_{\alpha_0,\beta_0}=-k_0\frac{\mathbf{v}_\beta\bullet\mathbf{n}}{\mathbf{v}_0\bullet\mathbf{n}}\ . \quad (20)$$

Using the above linear approximation, the sum of the squared distances between the left and right PoGs in the objective function (12) can be expressed as

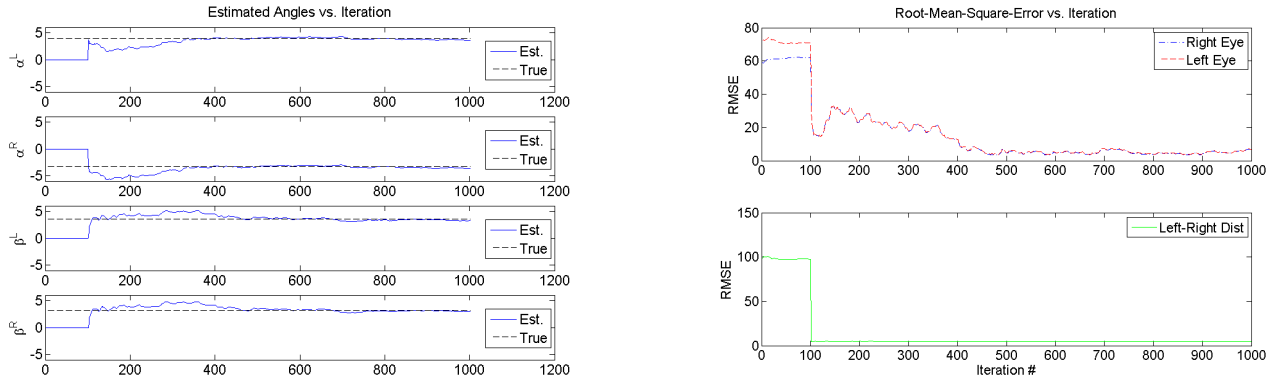
$$F(\alpha^{\mathrm{L}},\beta^{\mathrm{L}},\alpha^{\mathrm{R}},\beta^{\mathrm{R}})=\sum_i\left\|\mathbf{M}_i\,\mathbf{x}+\mathbf{y}_i\right\|_2^2 \quad (21)$$

where $\mathbf{M}_i=\begin{bmatrix}\mathbf{a}_i^{\mathrm{L}} & \mathbf{b}_i^{\mathrm{L}} & -\mathbf{a}_i^{\mathrm{R}} & -\mathbf{b}_i^{\mathrm{R}}\end{bmatrix}$ is a 3x4 matrix, $\mathbf{y}_i=\mathbf{g}_{0,i}^{\mathrm{L}}-\mathbf{g}_{0,i}^{\mathrm{R}}$ is a 3x1 vector and $\mathbf{x}=\left[(\alpha^{\mathrm{L}}-\alpha_0^{\mathrm{L}})\ \ (\beta^{\mathrm{L}}-\beta_0^{\mathrm{L}})\ \ (\alpha^{\mathrm{R}}-\alpha_0^{\mathrm{R}})\ \ (\beta^{\mathrm{R}}-\beta_0^{\mathrm{R}})\right]^T$ is a 4x1 vector of unknown angles. The subscript "*i*" is used to explicitly indicate the correspondence to the specific time instance "*i*" or *i*-th gaze sample.

The solution to (21) can be obtained in a closed form using least squares as

$$\mathbf{x}_{\mathrm{opt}}=-\left(\mathbf{M}^T\mathbf{M}\right)^{-1}\mathbf{M}^T\mathbf{y} \quad (22)$$

where the optimization over several time instances is achieved by stacking the matrices on top of each other:



(a) (b)

Figure 3 *Estimation results with a 40 cm x 30 cm observation surface (e.g., 20" monitor). Noise STD in the optical axis was set to 0.1° and in the center of curvature of the cornea to 0.5 mm. (a) Estimated subject-specific angles; (b) Root-Mean-Square Error (RMSE) of PoG estimation (top) and Left PoG - Right PoG distance (bottom).*

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_1 \\ \mathbf{M}_2 \\ \vdots \\ \mathbf{M}_N \end{bmatrix} \quad ; \quad \mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_N \end{bmatrix} . \tag{23}$$

Finally, the estimates of the subject-specific angles are given by

$$\begin{bmatrix} \hat{\alpha}^{\mathrm{L}} & \hat{\beta}^{\mathrm{L}} & \hat{\alpha}^{\mathrm{R}} & \hat{\beta}^{\mathrm{R}} \end{bmatrix}^T = \begin{bmatrix} \alpha_0^{\mathrm{L}} & \beta_0^{\mathrm{L}} & \alpha_0^{\mathrm{R}} & \beta_0^{\mathrm{R}} \end{bmatrix}^T + \mathbf{x}_{\mathrm{opt}} . \tag{24}$$

Since the objective function (21) is a linear approximation of the objective function (12), several iterations of (14)-(24) might be needed to converge to the true minimum of the objective function (12). In the first iteration, α_0^{L}, β_0^{L}, α_0^{R} and β_0^{R} are set to zero. In subsequent iterations, α_0^{L}, β_0^{L}, α_0^{R} and β_0^{R} are set to the values of $\hat{\alpha}_0^{\mathrm{L}}, \hat{\beta}_0^{\mathrm{L}}, \hat{\alpha}_0^{\mathrm{R}}$ and $\hat{\beta}_0^{\mathrm{R}}$ from the preceding iteration.

The above methodology to estimate the angle between the *optical* and *visual* axes is suitable for "on-line" estimation as a new matrix $\mathbf{M}_i$ is added to $\mathbf{M}$ and a new vector $\mathbf{y}_i$ is added to $\mathbf{y}$ for each new estimate of the centers of curvature of the corneas and optical axes.

3.2 Numerical Simulations

The performance of the AAE algorithm was evaluated as a function of noise in the estimates of the direction of the optical axis and the coordinates of the center of curvature of the cornea of each eye. In the simulations, the angles between the optical and visual axes were randomly drawn from a uniform distribution with a range of (-5, 0) for α^{R}, (0, 5) for α^{L} and (-5, 5) for β^{R} and β^{L}. The PoGs were randomly drawn from a uniform distribution over the observation surface and the optical axes of the two eyes were calculated. Noise in the estimates of the center of curvature of the cornea and the optical axis was simulated by adding independent zero-mean white Gaussian processes to the coordinates of the centers of curvature of the cornea (X,Y and Z) and to the horizontal and vertical components of the direction of the optical axis. The AAE algorithm was implemented in MATLAB®. One thousand PoGs were used to estimate the angles between the optical and visual axes for each set of eye parameters ($\alpha^{\mathrm{R}}, \alpha^{\mathrm{L}}, \beta^{\mathrm{R}}$ and β^{L}). The initial value for the angle between the optical and visual axes was set to zero (the initial value did not affect the final results as long as it was within ±20° of the actual value). The first update of the algorithm was done after 100 PoGs to prevent large fluctuations during start-up.

Figure **3** shows an example of one simulation. The parameters for the simulations were as follows: 1) Standard deviation (STD) of the noise in the components of the direction of the optical axis 0.1°, 2) STD of the noise in the coordinates of the center of curvature of the cornea 0.5mm, 3) observation surface 40cm x 30cm (20" monitor) 4) the center of curvature of the cornea of the subject's right eye, $\mathbf{c}^{\mathrm{R}}$, at $[3\ 0\ 75]^T$ (i.e., approximately 75 cm from the display's surface), 5) the center of curvature of the cornea of the subject's left eye, $\mathbf{c}^{\mathrm{L}}$, at $[-3\ 0\ 75]^T$ (i.e., inter-pupillary distance = 6 cm). As can be seen from

Figure **3**(a), after approximately 500 PoGs the solution converges to within ±0.5° of the true values of $\alpha^{\mathrm{R}}, \alpha^{\mathrm{L}}, \beta^{\mathrm{R}}$ and β^{L}.

Figure **3**(b) shows the root-mean-square (RMS) error in the PoG estimation (top) and the value of the objective function (bottom). The objective function was effectively minimized after the first update (i.e., after 100 iterations) while the RMS error continues to decline until iteration 500. This is due to the fact that even though after 100 iterations the distance between the points-of-gaze of the left and right eyes was minimized, PoGs of the both eyes had similar biases relative to the actual PoGs. From iteration 100 to 500 the PoGs of the two eyes drifted simultaneously towards their true values.

As shown in [Model and Eizenman 2009], the performance of the algorithm for the estimation of the angles between the optical and visual axes depends on the range of the viewing angles between the visual axes of both eyes and the vectors normal to the observation surface. Therefore, the simulations used four different observation surfaces: a) a plane that provides viewing angles in the range of ±14.9° horizontally and ±11.3° vertically (40 cm x 30 cm observation surface that is similar in size to a

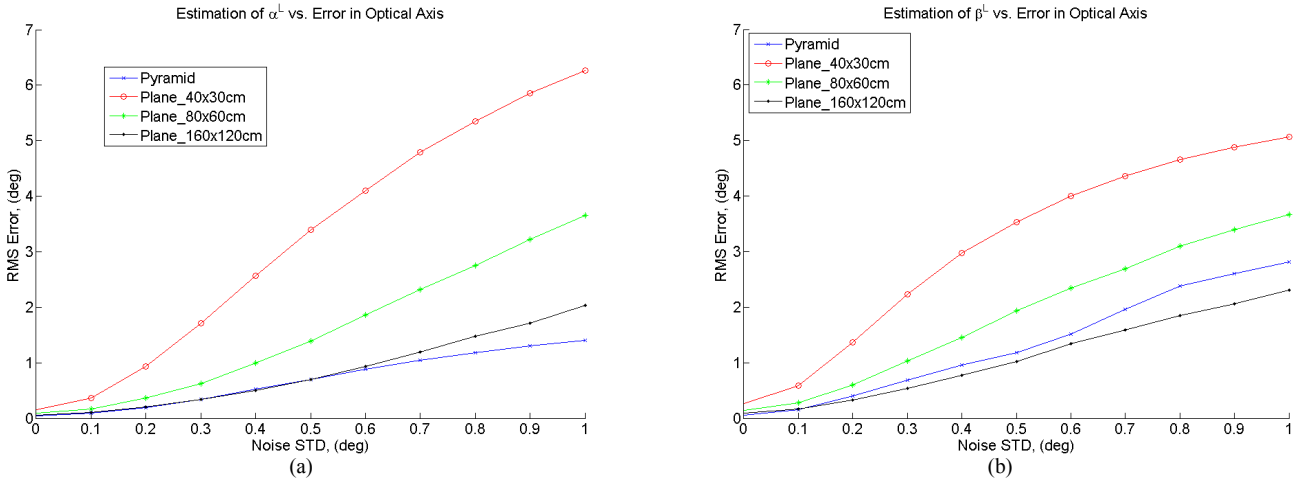


Figure 4 *RMS errors for the angle between the optical and visual axes, for four different surfaces, as a function of the noise in the estimation of the components of the direction of the optical axis. 27 head positions (all the combinations of X_{head} = -10 cm, 0 cm, 10 cm; Y_{head} = -10 cm, 0 cm, 10 cm; Z_{head} = -10 cm, 0 cm, 10 cm). (a) α^L ; (b) β^L .*

20" monitor), b) a plane that provides viewing angles in the range of ±28º horizontally and ±21.8º degrees vertically (80 cm x 60 cm observation surface that is similar in size to a 40" monitor), c) a plane that provide viewing angles in the range of ±46.8º horizontally and ±38.7º (160 cm x 120 cm) and d) a pyramid (30 cm x 30 cm base, 15 cm height, 45º angle between the base and each of the facets).

For these simulations, the coordinates of the center of curvature of the cornea of the right eye, $\mathbf{c}^R$, were set at 27 different positions (all the combinations of X= -7 cm, 3 cm, 13 cm; Y= -10 cm, 0 cm, 10 cm; Z= 85 cm, 75 cm, 65 cm, e.g., $\mathbf{c}^R = [-7\ 0\ 75]^T$, $[3\ 0\ 75]^T$, $[13\ 0\ 75]^T$, *etc*). The simulations were repeated for different noise levels in the components of the direction of the optical axis. For all the simulations, the noise level in the coordinates of the center of the curvature of the cornea was set to 1 mm. For each eye position and noise level in the optical axis the simulations were repeated 100 times. Figure 4 shows the aggregate (all head positions) RMS error in the estimations of α and β as a function of the noise level in the components of the direction of the optical axis.

As can be seen from Figure 4, the performance of the AAE algorithm improves when the range of angles between the subject's gaze vectors, as he/she looks at the displays, and vectors normal to the observation surfaces increases. For a plane that provides viewing angles in the range of ±14.9º horizontally and ±11.3º vertically (40 cm x 30 cm observation surface), a noise level of 0.4 degrees in the components of the direction of the optical axis results in a RMS error of 2.6º in the estimation of α^L . For the same noise level in the optical axis, the RMS error in the estimation of α^L is reduced to only 0.5º when a plane that provides viewing angles in the range of ±46.8º horizontally and ±38.7º vertically (160 cm x 120 cm) is used. This is similar to the performance of the algorithm with the pyramidal observation surface.

4 Experiments

A two-camera REGT system [Guestrin and Eizemnan 2007] was used for experiments with four subjects. The STD of the noise in the estimates of the components of the direction of the optical axis in this system is 0.4° and the STD of the noise in the estimates of the coordinates of the center of curvature of the cornea is 1.0 mm. Subjects were seated at a distance of approximately 75cm from a computer monitor (plane Z=0) and head movements were not restrained. Given the limited tracking range of the system (approximately ±15º horizontally and vertically), only a relatively small observation planes (<20" monitor) or more complicated pyramid observation surfaces could be used for the experiments. Because the expected RMS errors with a 20" observation surface ($\alpha > 2.5$º and $\beta > 3$º, see Figure 4) are larger than the expected errors when α^R , α^L , β^R and β^L are set a-priori to -2.5º, 2.5º, 0, 0, respectively, there was no point in evaluating the performance of the algorithm with such an observation surface. Therefore, the experiments were performed with a pyramid observation surface (30 cm x 30 cm base, 15 cm height, and pinnacle towards the viewer). The base of the pyramid was mounted on the plane Z=9 (cm) and the pinnacle was located at $[0,-1.5,24]^T$.

During the experiment, each subject was asked to look at the pyramid and thousand estimates of the center of curvature of the cornea and the direction of the optical axis were obtained for each eye. The subject-specific angles between the optical and visual axes were estimated, off-line, using the AAE algorithm. Next, the pyramid was removed and the subject was asked to complete a standard one-point calibration procedure [Guestrin and Eizemnan 2007]. The values of the estimated subject-specific angles between the optical and visual axes with the AAE procedure and a one-point calibration procedure are shown in Table I.

TABLE I

HORIZONTAL (α) AMD VERTICAL (β) COMPONENTS OF THE ANGLE BETWEEN THE OPTICAL AND VISUAL AXES

Subjects:	Subject-specific angles, (°)							
	AAE				1-point calib			
	α^L	α^R	β^L	β^R	α^L	α^R	β^L	β^R
1	1.8	-2.6	-0.1	-0.7	0.9	-2.9	-0.2	-1.0
2	1.1	-1.0	-2.8	-0.7	0.9	-1.3	-3.1	-1.5
3	0.3	-0.7	2.7	0.3	-0.1	-0.3	2.1	0.4
4	-0.2	-3.1	0.6	0.1	0.1	-2.3	0.5	0.7

'AAE' – α and β estimated with the AAE procedure. '1-point calib' – α and β estimated with a one-point calibration procedure.

Since the estimates of the subject-specific angles between the optical and visual axes that are obtained by the one-point calibration procedure minimize the RMS error between the estimated PoG and the actual PoG, the values obtained by the one-point calibration procedure will serve as a reference ("gold standard") for the calculations of the errors of the AAE procedure. The estimation errors of the AAE procedure are -0.08±0.59 [°] in α (left and right combined) and -0.19±0.43 [°] in β (left and right combined). Given the noise characteristics of the gaze estimation system, these estimation errors are in the range predicted by the numerical simulations (Figure 4).

Following the estimation of the angles between the optical and visual axes, the subjects looked at a grid of nine points (3x3, 8.5° apart) displayed on a computer monitor. Fifty PoGs were collected at each point. The RMS errors in the estimation of the PoG for all four subjects are summarized in Table II.

TABLE II

POG ESTIMATION ERROR

Subjects:	RMS error, (°)			
	UCF-REGT		1-point calib REGT	
	Left Eye	Right Eye	Left Eye	Right Eye
1	1.2	0.8	0.6	0.5
2	1.3	1.4	1.0	0.7
3	1.4	1.0	0.8	0.8
4	1.6	1.7	0.9	1.1

'UCF-REGT' – PoG is estimated with the user-calibration-free REGT . '1-point calib REGT' – PoG is estimated with the REGT system that uses one-point user-calibration procedure.

For the User-Calibration-Free REGT (UCF-REGT) system described in this paper the average RMS error in PoG estimation for all four subjects is 1.3°. For the REGT system that uses the one-point calibration procedure the average RMS error is 0.8°.

5 Discussion and Conclusions

A user-calibration-free REGT system was presented. The system uses the assumption that at each time instant both eyes look at the same point in space to eliminate the need for a user-calibration procedure that requires subjects to fixate on known targets at specific time intervals to estimate the angles between the optical and visual axes. Experiments with a REGT system that estimates the components of the direction of the optical axis with an accuracy of 0.4° [Guestrin and Eizemnan 2007] showed that the AAE algorithm can estimate the angles between the optical and visual axes with an RMS error of 0.5°. These results are consistent with the numerical analysis (see Section 3.2) and were achieved with an observation surface that included four planes. Based on the numerical analysis (see Figure 4), an eye-tracking system that estimates the components of the direction of the optical axis with an accuracy of 0.1° can use a 30 cm x 40 cm observation plane (e.g., 20" computer monitor) to achieve a similar RMS error. Increasing the observation plane to 60 cm x 80 cm will allow eye trackers that can estimate the components of the direction of the optical axis with an accuracy of 0.2° to achieve similar RMS errors. As the accuracy of eye-tracking systems will improve, it will become feasible to use the AAE algorithm presented in this paper to estimate accurately the angle between the optical and visual axis using a single observation plane. The use of a single observation surface will improve considerably the utility of the UCF-REGT system.

Experiments with four subjects demonstrated that when a two-camera REGT system [Guestrin and Eizemnan 2007] is used with a single point calibration procedure, the RMS error is 0.8°. With the UCF-REGT system, the RMS error of the PoG estimation is increased to 1.3° (see Table II). User-calibration-free gaze estimation systems that estimate the PoG from the intersection of the *optical* axis of one of the eyes with the display (e.g., [Shih et al. 2000]) or from the midpoint of the intersections of the *optical* axes of both eyes with the display (e.g., [Nagamatsu et al. 2009]) can exhibit large, subject dependent, gaze estimation errors. For the experiments described in Section 4 (see Table II), calculations of the PoG by the intersection of the *optical* axis of one eye with the computer monitor results in RMS errors ranging from 1° to 4.5° (average 2.5°). By using the midpoint, the RMS errors for the four subjects are reduced to a range between 1.5° to 3.2° (average 2.0°). With the UCF-REGT system, the RMS error is in the range of 0.8° to 1.7° (1.3° on average for all four subjects). The UCF-REGT system reduces the variability of PoG estimation errors between subjects and improves significantly the accuracy of "calibration-free" REGT systems.

References

DUCHOWSKI, A. T. (2002). "A breadth-first survey of eye-tracking applications." *Behavior Research Methods Instruments & Computers* 34(4): 455-470.

EIZENMAN, M., MODEL, D. and GUESTRIN, E. D. (2009). Covert Monitoring of the Point-of-Gaze. *IEEE TIC-STH*, Toronto, ON, Canada, 551-556.

EIZENMAN, M., SAPIR-PICHHADZE, R., WESTALL, C. A., WONG, A. M., LEE, H. and MORAD, Y. (2006). "Eye-movement responses to disparity vergence stimuli with artificial monocular scotomas." *Current Eye Research* 31(6): 471-480.

GOLDBERG, J. H. and KOTVAL, X. P. (1999). "Computer interface evaluation using eye movements: methods and constructs." *International Journal of Industrial Ergonomics* 24(6): 631-645.

GUESTRIN, E. D. and EIZEMNAN, M. (2007). Remote point-of-gaze estimation with free head movements requiring a

single-point calibration. *Proc. of Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 4556-4560.

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.

GUESTRIN, E. D. and EIZENMAN, M. (2008). Remote point-of-gaze estimation requiring a single-point calibration for applications with infants. Proc. of the 2008 Symposium on Eye Tracking Research & Applications. Savannah, GA, USA, ACM**:** 267-274.

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 & Prevention* 39(2): 372-379.

HELMHOLTZ, H. (1924). Helmholtz's treatise on physiological optics. Translated from the 3rd German ed. Edited by J. P. C. Southall Rochester, NY, Optical Society of America.

HUTCHINSON, T. E., WHITE, K. P., JR., MARTIN, W. N., REICHERT, K. C. and FREY, L. A. (1989). "Human-computer interaction using eye-gaze input." *Systems, Man and Cybernetics, IEEE Transactions on* 19(6): 1527-1534.

LOSHE, G. (1997). "Consumer eye movement patterns of Yellow Pages advertising." *Journal of Advertising* 26(1): 61-73.

MODEL, D. and EIZENMAN, M. (2010). "An Automatic Personal Calibration Procedure for Advanced Gaze Estimation Systems." *IEEE Transactions on Biomedical Engineering,* Accepted.

MODEL, D., GUESTRIN, E. D. and EIZENMAN, M. (2009). An Automatic Calibration Procedure for Remote Eye-Gaze Tracking Systems. *31st Annual International Conference of the IEEE EMBS*, Minneapolis, MN, USA, 4751-4754.

NAGAMATSU, T., KAMAHARA, J. and TANAKA, N. (2009). Calibration-free gaze tracking using a binocular 3D eye model. *Proceedings of the 27th International Conference on Human factors in Computing Systems*, Boston, MA, USA, ACM

RAYNER, K. (1998). "Eye movements in reading and information processing: 20 years of research." *Psychological Bulletin* 124(3): 372-422.

SHIH, S.-W., WU, Y.-T. and LIU, J. (2000). A calibration-free gaze tracking technique. *In Proceedings of 15th Int. Conference on Pattern Recognition* 201-204.

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, Part B: Cybernetics* 34(1): 234-245.

WETZEL, P. A., KRUEGER-ANDERSON, G., POPRIK, C. and BASCOM, P. (1997). An eye tracking system for analysis of pilots' scan paths, United States Air Force Armstrong Laboratory.

YOUNG, L. R. and SHEENA, D. (1975). "Survey of eye-movement recording methods." *Behavior Research Methods & Instrumentation* 7(5): 397-429.