

# REMOTE POINT-OF-GAZE ESTIMATION WITH SINGLE-POINT PERSONAL CALIBRATION BASED ON THE PUPIL BOUNDARY AND CORNEAL REFLECTIONS

*Elias D. Guestrin<sup>1,2</sup> and Moshe Eizenman<sup>1,2,3</sup>*

<sup>1</sup> Department of Electrical and Computer Engineering, <sup>2</sup> Institute of Biomaterials and Biomedical Engineering, <sup>3</sup> Department of Ophthalmology and Vision Sciences, University of Toronto, Toronto, Ontario, Canada. E-mail: elias.guestrin@utoronto.ca

## ABSTRACT

This paper describes a new method for remote, non-contact point-of-gaze (PoG) estimation that tolerates head movements and requires a simple personal calibration procedure in which the subject has to fixate a single calibration point. This method uses the pupil boundary and at least two corneal reflections (virtual images of light sources) that are extracted from eye images captured by at least two video cameras. Experimental results obtained with a prototype system exhibited RMS PoG estimation errors of approx.  $0.4\text{-}0.6^\circ$  of visual angle. Such accuracy is comparable to that of the best commercially available systems, which use multiple-point personal calibration procedures, and significantly better than that of any other systems that use a single-point personal calibration procedure and have been previously described in the literature.

**Index Terms**— Point-of-gaze, pupil boundary, corneal reflections, single-point personal calibration, remote gaze estimation

## 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 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, analysis of reading and driving behavior, pilot training, ergonomics, marketing and advertising research, human-computer interfaces and assistive devices for motor-disabled persons.

Accurate PoG estimation requires the determination of a set of subject-specific parameters through a personal calibration procedure in which subjects are required to fixate one or more specific points at specific times. Minimizing personal calibration requirements is critical in applications requiring minimal subject cooperation such as those with infants and mentally challenged individuals. By using at least two video cameras and at least two near-infrared light sources (near-infrared is invisible to the human eye), it is

possible to estimate the PoG in the presence of head movements and without any contact with the user (remote operation) after completing a personal calibration procedure in which the subject is required to fixate only a single point [5]-[8]. These single-point calibration gaze estimation methodologies are based on the center of the pupil and the corneal reflections (Fig. 1 - Inset) extracted from the images captured by the cameras. The corneal reflections (first Purkinje images) are virtual images of the light sources that are created by the front surface of the cornea, which acts as a convex mirror.

Gaze estimation methodologies that are based on the pupil center inherently assume that the center of the pupil extracted from the images captured by the cameras corresponds to the image of the pupil center. However, in general, due to the refraction at the cornea (primarily its front surface) followed by the perspective projection at the camera, the center of the pupil image is not coincident with the image of the pupil center [9]-[10]. The discrepancy between these two points increases with the pupil diameter and with the angle between the optic axis of the eye and the optic axis of the camera [7], [11]. The methodology presented in this paper attempts to overcome this problem by using the pupil boundary instead of the pupil center.

The next section presents the mathematical model used to estimate the PoG from the pupil boundary and multiple corneal reflections extracted from images captured by at least two synchronized cameras. Section 3 describes a prototype system, Section 4 shows experimental results, and Section 5 presents the conclusions of this work.

## 2. MATHEMATICAL MODEL

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).

The first stage in the estimation of the PoG consists of determining the 3-D position of the center of curvature of the cornea,  $\mathbf{c}$ , from the corneal reflections extracted from the

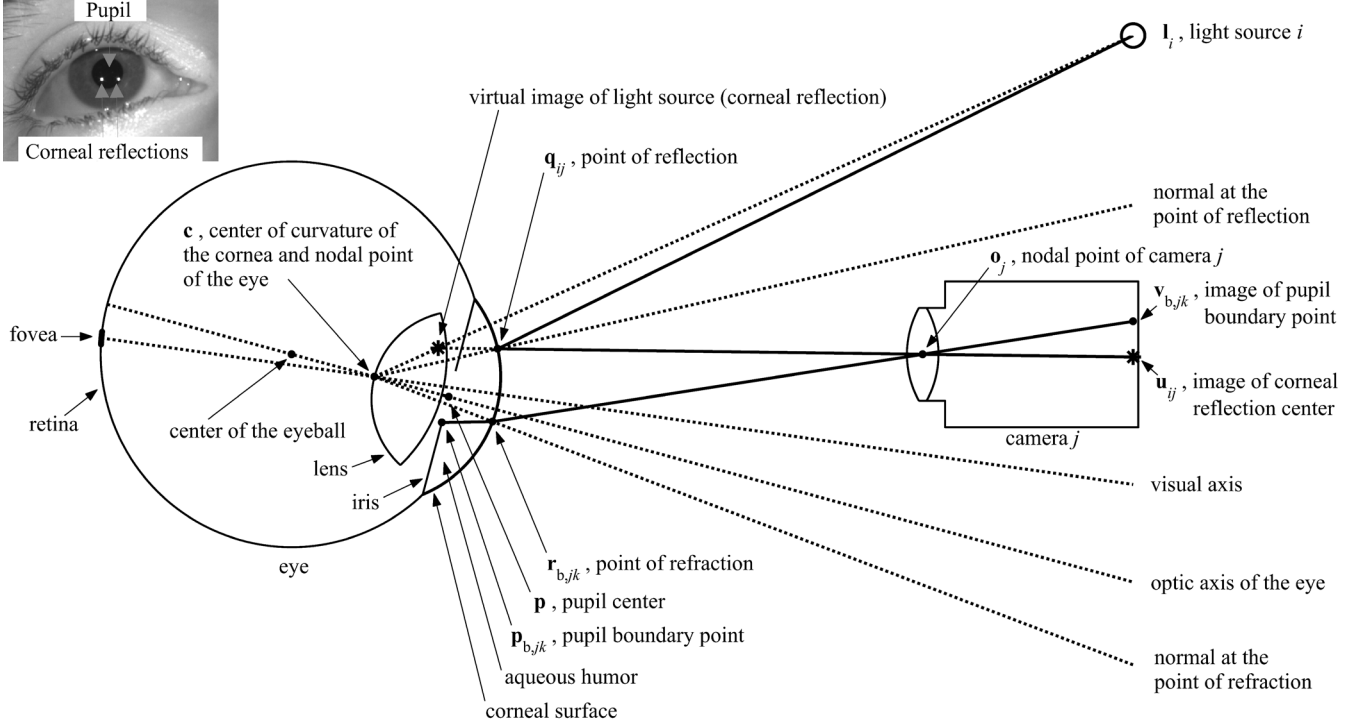


Fig. 1: Ray-tracing diagram (not to scale in order to be able to show all the elements of interest), showing schematic representations of an eye, a camera and a light source. Inset: eye image indicating the pupil and two corneal reflections.

images captured by the video cameras. Towards this end, consider 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  $\mathbf{u}_{ij}$  (the image of the corneal reflection produced by light source  $i$  in the image plane of camera  $j$ ).

The condition that the ray coming from the point of reflection  $\mathbf{q}_{ij}$  and passing through the nodal point of camera  $j$ ,  $\mathbf{o}_j$ , intersects the camera image plane at point  $\mathbf{u}_{ij}$ , can be expressed in parametric form as

$$\mathbf{q}_{ij} = \mathbf{o}_j + k_{q,ij}(\mathbf{o}_j - \mathbf{u}_{ij}) \quad \text{for some } k_{q,ij}. \quad (1)$$

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 (the angles of incidence and reflection are measured with respect to the normal at the point of reflection). 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 the line defined by  $\mathbf{c}$  and  $\mathbf{q}_{ij}$  must bisect the angle defined by  $\mathbf{l}_i$ ,  $\mathbf{q}_{ij}$ , and  $\mathbf{o}_j$ . Noting that given two vectors  $\mathbf{a}$  and  $\mathbf{b}$  such that  $\|\mathbf{a}\| = \|\mathbf{b}\|$  and  $\mathbf{a} \neq -\mathbf{b}$ , vector  $\mathbf{a} + \mathbf{b}$  bisects the angle formed by vectors  $\mathbf{a}$  and  $\mathbf{b}$ , the preceding condition can be expressed as

$$\mathbf{c} = \mathbf{q}_{ij} - R \frac{\frac{\mathbf{l}_i - \mathbf{q}_{ij}}{\|\mathbf{l}_i - \mathbf{q}_{ij}\|} + \frac{\mathbf{o}_j - \mathbf{q}_{ij}}{\|\mathbf{o}_j - \mathbf{q}_{ij}\|}}{\left\| \frac{\mathbf{l}_i - \mathbf{q}_{ij}}{\|\mathbf{l}_i - \mathbf{q}_{ij}\|} + \frac{\mathbf{o}_j - \mathbf{q}_{ij}}{\|\mathbf{o}_j - \mathbf{q}_{ij}\|} \right\|}, \quad (2)$$

where  $R$  is the radius of curvature of the cornea (subject-specific and unknown a priori). Notice that the case “ $\mathbf{a} = -\mathbf{b}$ ” would happen only if the line connecting the light source  $\mathbf{l}_i$  and the nodal point of the camera  $\mathbf{o}_j$  were tangent to the corneal surface at  $\mathbf{q}_{ij}$ , which does not happen in practice.

Substituting (1) into (2),  $\mathbf{c}$  is expressed as a function of  $k_{q,ij}$  and  $R$ :  $\mathbf{c}_{ij}(k_{q,ij}, R)$ . Ideally, all the  $\mathbf{c}_{ij}(k_{q,ij}, R)$  should be equal to each other. However, in general, due to noise in the coordinates of the corneal reflections extracted from the images captured by the cameras, and because real corneas are not exactly spherical, the condition that all the  $\mathbf{c}_{ij}(k_{q,ij}, R)$  are equal to each other does not hold. To be able to solve the problem in such situation, it is restated as finding the values of  $R$  and  $k_{q,ij}$  that minimize the sum of the squared distances between all possible pairs of  $\mathbf{c}_{ij}(k_{q,ij}, R)$ . This minimization problem can be expressed in a generalized form as

$$\min_{k_{q,ij}, R} \sum_{i_a, j_a, i_b, j_b} w_{i_a j_a} w_{i_b j_b} \left\| \mathbf{c}_{i_a j_a}(k_{q, i_a j_a}, R) - \mathbf{c}_{i_b j_b}(k_{q, i_b j_b}, R) \right\|^2, \quad (3)$$

where the  $w_{ij}$  are weight factors. The use of binary weight

factors to select a subset of all available corneal reflection images is discussed in Section 3. Having obtained the optimum values of  $R$  and  $k_{q,ij}$ ,  $\mathbf{c}$  is finally obtained as the mean of all the  $\mathbf{c}_{ij}(k_{q,ij}, R)$ :

$$\mathbf{c} = \text{mean}_{i,j}(w_{ij} \mathbf{c}_{ij}(k_{q,ij}, R)) \cdot \quad (4)$$

It should be noted that although this method for the calculation of the 3-D coordinates of the center of curvature of the cornea is more complex than those described in [5]-[8], it is more robust to noise in the coordinates of the corneal reflections extracted from the eye images captured by the cameras and to corneal asphericity (deviation of the shape of real corneas from the ideal spherical shape assumed in the mathematical model) [11].

Having determined the 3-D coordinates of the center of curvature of the cornea,  $\mathbf{c}$ , and the radius of curvature of the cornea,  $R$ , the next stage in the estimation of the PoG consists of determining the orientation of the optic axis of the eye from the boundary of the pupil extracted from the eye images captured by the video cameras.

The orientation of the optic axis of the eye has 2 degrees of freedom and can thus be described by a unit vector,  $\boldsymbol{\omega}$ , that is a function of 2 parameters such as its pan and tilt angles with respect to the WCS. To provide a formal expression for  $\boldsymbol{\omega}$ , 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  $\theta_{\text{eye}}$  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  $\varphi_{\text{eye}}$  be the tilt angle (angle between the optic axis of the eye and the  $XZ$ -plane of the WCS), with  $\theta_{\text{eye}} > 0$  for rotations to the right, and  $\varphi_{\text{eye}} > 0$  for rotations upwards. With these definitions,  $\boldsymbol{\omega}$  is expressed as

$$\boldsymbol{\omega} \equiv \boldsymbol{\omega}(\theta_{\text{eye}}, \varphi_{\text{eye}}) = \begin{bmatrix} \sin \theta_{\text{eye}} \cos \varphi_{\text{eye}} \\ \sin \varphi_{\text{eye}} \\ -\cos \theta_{\text{eye}} \cos \varphi_{\text{eye}} \end{bmatrix}. \quad (5)$$

Assuming that (i) the optic axis of the eye contains both the center of curvature of the cornea,  $\mathbf{c}$ , and the pupil center,  $\mathbf{p}$ , (ii) the optic axis of the eye is perpendicular to the plane of the pupil, and (iii) the distance between  $\mathbf{c}$  and the plane of the pupil is represented by  $K$  (subject-specific and unknown a priori), it follows that  $\mathbf{p}$  is given by

$$\mathbf{p} \equiv \mathbf{p}(\theta_{\text{eye}}, \varphi_{\text{eye}}, K) = \mathbf{c} + K\boldsymbol{\omega}(\theta_{\text{eye}}, \varphi_{\text{eye}}), \quad (6)$$

and the pupil boundary points,  $\mathbf{p}_{b,jk}$ , have to satisfy

$$(\mathbf{p}_{b,jk} - \mathbf{c}) \bullet \boldsymbol{\omega}(\theta_{\text{eye}}, \varphi_{\text{eye}}) = K. \quad (7)$$

Let  $\mathbf{p}_{b,jk}$  be a point in the pupil boundary such that a ray

of light originating from it travels through the aqueous humor and cornea (effective index of refraction  $n_1 = 1.3375$  [12]) and refracts at point  $\mathbf{r}_{b,jk}$  on the front corneal surface as it travels into the air (index of refraction  $n_2 = 1$ ), such that the refracted ray goes through the nodal point of camera  $j$ ,  $\mathbf{o}_j$ , and intersects the camera image plane at point  $\mathbf{v}_{b,jk}$  (the  $k$ -th boundary point of the pupil image captured by camera  $j$ ). Given  $\mathbf{c}$ ,  $R$ ,  $\mathbf{o}_j$  and  $\mathbf{v}_{b,jk}$ , the point of refraction  $\mathbf{r}_{b,jk}$  is obtained as the intersection of the line defined by  $\mathbf{o}_j$  and  $\mathbf{v}_{b,jk}$  and the spherical surface with center  $\mathbf{c}$  and radius  $R$ :

$$\begin{aligned} \mathbf{r}_{b,jk} &= \mathbf{o}_j + k_{rb,jk} \boldsymbol{\rho}_{b,jk}, \\ k_{rb,jk} &= \gamma_j \bullet \boldsymbol{\rho}_{b,jk} - \sqrt{(\gamma_j \bullet \boldsymbol{\rho}_{b,jk})^2 - \|\gamma_j\|^2 + R^2}, \\ \gamma_j &= \mathbf{c} - \mathbf{o}_j \end{aligned} \quad (8)$$

where

$$\boldsymbol{\rho}_{b,jk} = \frac{\mathbf{o}_j - \mathbf{v}_{b,jk}}{\|\mathbf{o}_j - \mathbf{v}_{b,jk}\|} \quad (9)$$

is the unit vector in the direction of the refracted ray passing through the nodal point of the camera.

Having obtained  $\mathbf{r}_{b,jk}$ , the unit vector normal to the corneal surface at the point of refraction can be expressed as

$$\boldsymbol{\eta}_{b,jk} = \frac{\mathbf{r}_{b,jk} - \mathbf{c}}{\|\mathbf{r}_{b,jk} - \mathbf{c}\|} = \frac{\mathbf{r}_{b,jk} - \mathbf{c}}{R}. \quad (10)$$

It then follows [9] that the unit vector in the direction of the incident ray originating from the pupil boundary point is given by

$$\mathbf{u}_{b,jk} = \frac{n_2}{n_1} \left\{ \boldsymbol{\rho}_{b,jk} - \left[ \boldsymbol{\eta}_{b,jk} \bullet \boldsymbol{\rho}_{b,jk} + \sqrt{\left(\frac{n_1}{n_2}\right)^2 - 1 + (\boldsymbol{\eta}_{b,jk} \bullet \boldsymbol{\rho}_{b,jk})^2} \right] \boldsymbol{\eta}_{b,jk} \right\} \quad (11)$$

Having  $\mathbf{r}_{b,jk}$  and  $\mathbf{u}_{b,jk}$ ,  $\boldsymbol{\rho}_{b,jk}$  is given by

$$\boldsymbol{\rho}_{b,jk} = \mathbf{r}_{b,jk} + k_{pb,jk} \mathbf{u}_{b,jk} \quad \text{for some } k_{pb,jk}. \quad (12)$$

Substituting (12) into (7), solving for  $k_{pb,jk}$  and substituting back into (12), it follows that

$$\boldsymbol{\rho}_{b,jk}(\theta_{\text{eye}}, \varphi_{\text{eye}}, K) = \mathbf{r}_{b,jk} + \frac{K - (\mathbf{r}_{b,jk} - \mathbf{c}) \bullet \boldsymbol{\omega}(\theta_{\text{eye}}, \varphi_{\text{eye}})}{\mathbf{u}_{b,jk} \bullet \boldsymbol{\omega}(\theta_{\text{eye}}, \varphi_{\text{eye}})} \mathbf{u}_{b,jk}. \quad (13)$$

Assuming that the pupil is circular with radius  $r_{\text{pupil}}$  (unknown and time-varying), all the  $\mathbf{p}_{b,jk}$  also have to satisfy

$$\|\mathbf{p}_{b,jk}(\theta_{\text{eye}}, \varphi_{\text{eye}}, K) - \mathbf{p}(\theta_{\text{eye}}, \varphi_{\text{eye}}, K)\| = r_{\text{pupil}}. \quad (14)$$

The problem of reconstructing the optic axis of the eye consists then of finding  $\theta_{\text{eye}}$ ,  $\varphi_{\text{eye}}$ ,  $K$  and  $r_{\text{pupil}}$  such that (14) is satisfied for all  $j$  and all  $k$ . Since, in practice, the coordinates of  $\mathbf{v}_{b,jk}$ , and therefore the coordinates of  $\mathbf{r}_{b,jk}$  given by (8)-(9) and the components of  $\mathbf{u}_{b,jk}$  given by (9)-(11), are corrupted by noise, the condition expressed by (14) for all  $j$  and all  $k$  is turned into the following minimization problem:

$$\min_{\theta_{\text{eye}}, \varphi_{\text{eye}}, K, r_{\text{pupil}}} \sum_j \sum_k \left( \left\| \mathbf{p}_{b,jk}(\theta_{\text{eye}}, \varphi_{\text{eye}}, K) - \mathbf{p}(\theta_{\text{eye}}, \varphi_{\text{eye}}, K) \right\| - r_{\text{pupil}} \right)^2, \quad (15)$$

where the optimization over  $r_{\text{pupil}}$  can be avoided by simply substituting

$$r_{\text{pupil}} = \frac{1}{N_{\text{boundary points (total of all cameras)}}} \sum_j \sum_k \left\| \mathbf{p}_{b,jk}(\theta_{\text{eye}}, \varphi_{\text{eye}}, K) - \mathbf{p}(\theta_{\text{eye}}, \varphi_{\text{eye}}, K) \right\| \quad (16)$$

Note that, theoretically, three pupil boundary points are enough to reconstruct the pupil (three non-collinear points define a circle). However, also due noise in the coordinates of the  $\mathbf{v}_{b,jk}$ 's, many more pupil boundary points need to be used in practice to obtain a stable solution.

The methodology described so far can be used to reconstruct the optic axis of the eye in 3-D space without a priori knowledge of the values of subject-specific eye parameters ( $R$  and  $K$ ). In other words, the optic axis of the eye can be reconstructed without the need for a personal calibration procedure, thereby allowing for the calibration-free measurement of relative eye movements.

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 exhibits a subject-specific deviation from the optic axis of as much as  $5^\circ$  [13]. Since the nodal point is within 1 mm of the center of curvature of the cornea [13],  $\mathbf{c}$ , it can be assumed to be coincident with  $\mathbf{c}$  (Fig. 1).

The last stage in the estimation of the PoG, then, requires the reconstruction of the visual axis in 3-D space. To accomplish this, the relation between the visual and optic axes has to be modeled. A general approach to model such relation relies on the definition of an Eye Coordinate System (ECS) that is rigidly attached to the eye. This ECS is a right-handed 3-D Cartesian coordinate system whose axes are represented by  $x_{\text{eye}}$ ,  $y_{\text{eye}}$ ,  $z_{\text{eye}}$ . The origin of the ECS is defined at the center of the eyeball and the  $z_{\text{eye}}$ -axis is coincident with the optic axis of the eye, pointing forward. The  $x_{\text{eye}}$ -axis and  $y_{\text{eye}}$ -axis are defined such that, when the eye is in the primary position, the  $x_{\text{eye}}$ -axis is horizontal and the  $y_{\text{eye}}$ -axis points up. The primary position can be defined as the position of the eyes in the skull when the head is in a natural erect position and the visual axis is horizontal and

perpendicular to the line connecting the centers of rotation of both eyes [14].

Let the direction of the visual axis with respect to the ECS be described by the subject-specific unit vector  $\mathbf{v}_{\text{ECS}}$ . Then, the unit vector in the direction of the visual axis with respect to the WCS is given by

$$\mathbf{v} = \mathbf{R}_{\text{eye}}(\theta_{\text{eye}}, \varphi_{\text{eye}}, \kappa_{\text{eye}}) \mathbf{v}_{\text{ECS}}, \quad (17)$$

where  $\mathbf{R}_{\text{eye}}(\theta_{\text{eye}}, \varphi_{\text{eye}}, \kappa_{\text{eye}})$  is the rotation matrix of the ECS with respect to the WCS,  $\theta_{\text{eye}}$  and  $\varphi_{\text{eye}}$  are the pan and tilt angles defined previously, and  $\kappa_{\text{eye}}$  is the torsion angle about the optic axis of the eye (the complete analytical expression can be found in [11]). For the remainder of this paper it will be assumed that  $\kappa_{\text{eye}} = 0$  (refer to [11], [15] for a discussion about the torsion of the eye). With this assumption, it can be shown that the rotation matrix  $\mathbf{R}_{\text{eye}}$  is given by

$$\mathbf{R}_{\text{eye}} = \begin{bmatrix} -\cos \theta_{\text{eye}} & -\sin \theta_{\text{eye}} \sin \varphi_{\text{eye}} & \sin \theta_{\text{eye}} \cos \varphi_{\text{eye}} \\ 0 & \cos \varphi_{\text{eye}} & \sin \varphi_{\text{eye}} \\ -\sin \theta_{\text{eye}} & \cos \theta_{\text{eye}} \sin \varphi_{\text{eye}} & -\cos \theta_{\text{eye}} \cos \varphi_{\text{eye}} \end{bmatrix} \quad (18)$$

Since the visual axis is assumed to go through  $\mathbf{c}$ , the entire visual axis in 3-D space is given by

$$\mathbf{g} = \mathbf{c} + k_g \mathbf{v}, \quad k_g \in \mathbb{R}, \quad (19)$$

and the PoG is given by (19) with a value of  $k_g$  such that the visual axis intersects the scene. In the particular case of a planar scene at  $Z = 0$ , that value of  $k_g$  is given by

$$k_g = -\frac{c_z}{v_z}. \quad (20)$$

To be able to calculate the PoG with the above methodology, the subject-specific deviation of the visual axis from the optic axis of the eye modeled by the unit vector  $\mathbf{v}_{\text{ECS}}$  needs to be known a priori. The unit vector  $\mathbf{v}_{\text{ECS}}$  can be determined through a single-point personal calibration procedure, which, conceptually, can be described as follows. The center of curvature of the cornea,  $\mathbf{c}$ , and the orientation of the optic axis of the eye with respect to the WCS described by its pan and tilt angles,  $\theta_{\text{eye}}$  and  $\varphi_{\text{eye}}$ , are determined while the subject fixates the calibration point  $\mathbf{g}$ . Having  $\mathbf{c}$  and  $\mathbf{g}$ , the unit vector in the direction of the visual axis with respect to the WCS,  $\mathbf{v}$ , is found as

$$\mathbf{v} = \frac{\mathbf{g} - \mathbf{c}}{\|\mathbf{g} - \mathbf{c}\|}. \quad (21)$$

Finally,  $\mathbf{v}_{\text{ECS}}$  can be readily obtained from (17)-(18) as

$$\mathbf{v}_{\text{ECS}} = \mathbf{R}_{\text{eye}}^{-1} \mathbf{v} = \mathbf{R}_{\text{eye}}^T \mathbf{v}. \quad (22)$$

To estimate the PoG with the preceding mathematical model, it is necessary to know the world coordinates of the positions of the light sources ( $\mathbf{l}_i$ ), the nodal points of the cameras ( $\mathbf{o}_j$ ), and the pupil boundary points ( $\mathbf{v}_{b,jk}$ ) and corneal reflections ( $\mathbf{u}_{ij}$ ) in the eye images. Since the coordinates of the pupil boundary points 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 [11], [16]. 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 estimate the PoG, 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.

### 3. PROTOTYPE SYSTEM

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, Richmond, BC, Canada) with 35 mm lenses (aperture set to  $f/5.6$ ) and four near-infrared light sources (each light source has 20 infrared LEDs with nominal wavelength of 850 nm) attached to a 19" LCD monitor (5:4 aspect ratio) by a custom aluminum frame (Fig. 2).

The cameras are oriented such that their optic axes intersect at a distance of approximately 65 cm from the screen (typical viewing distance). Both cameras operate at a frame rate of 20 Hz with a resolution of 1392 pixels (maximum horizontal resolution) by 900 pixels (highest vertical resolution that allows for 20 Hz operation). In these conditions, the range of lateral head movements is about 11 cm when the PoG is estimated for a specific eye, about 17 cm when the PoG is estimated for either eye and about 5 cm when the PoG is estimated for both eyes simultaneously (binocular mode). The range of vertical head movements is about 7 cm and the range of backwards/forward head movements is about 10 cm.

The use of more than the minimum of two light sources can help to improve the system robustness by increasing the likelihood that at least two corneal reflections are available for each camera regardless of the head position and the PoG on the screen, and in the presence of eyelid and eyelash interferences. Moreover, the use of more than two light sources can help to reduce the effect of corneal asphericity. 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 [13]. By using the two light sources that produce corneal reflections that are closest to the pupil center in the eye images, it is then possible to reduce the PoG estimation bias due to corneal asphericity [11]. For the results shown in this paper, the light sources are selected in such a way that the same pair of light

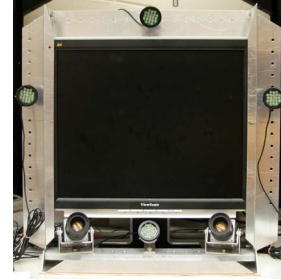


Fig. 2: System setup with two cameras and four light sources.

sources is used with both cameras. Formally, the selection of the light sources for each gaze estimate is carried out as follows. Let  $DPCCR_{ij}$  be the distance between the center of the pupil and the corneal reflection corresponding to light source  $i$  in the image from camera  $j$ . Then, after finding  $M_i = \max(DPCCR_{i1}, DPCCR_{i2})$  for each of the light sources, the two light sources corresponding to the two smallest  $M_i$  are selected. After selecting the two light sources and their corresponding corneal reflection images, there are two options. One of them is to use both corneal reflection images from both cameras, for a total of 4 corneal reflection images (4CR variant). The other option is to discard the corneal reflection image with the largest  $DPCCR_{ij}$  out of those 4 corneal reflection images (3CR variant). While the 4CR variant exhibits a lower sensitivity to noise in the coordinates of the corneal reflections extracted from the eye images, the 3CR variant exhibits a lower sensitivity to corneal asphericity [11]. Note that this selection of corneal reflection images corresponds to setting  $w_{ij} = 1$  in (3)-(4) for the selected ones and  $w_{ij} = 0$  for the rest.

As indicated in the previous section, 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 (attached to the LCD monitor, with the  $XY$ -plane coincident with the plane of the screen and the origin 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 [7], [11], which is based on [16].

### 4. EXPERIMENTAL RESULTS

Proof-of-concept experiments were carried out with three adults without eyeglasses or contact lenses in binocular mode. The head of each subject was placed at 65 cm (nominal distance), 60 cm (minimum distance) and 70 cm (maximum distance) from the screen. For each head position, each subject completed a number of trials (2 for Subjects 1 and 2, and 7 for Subject 3). In each trial, the subject was asked to sequentially fixate 25 target points (arranged in a 5-by-5 rectangular grid) on the computer

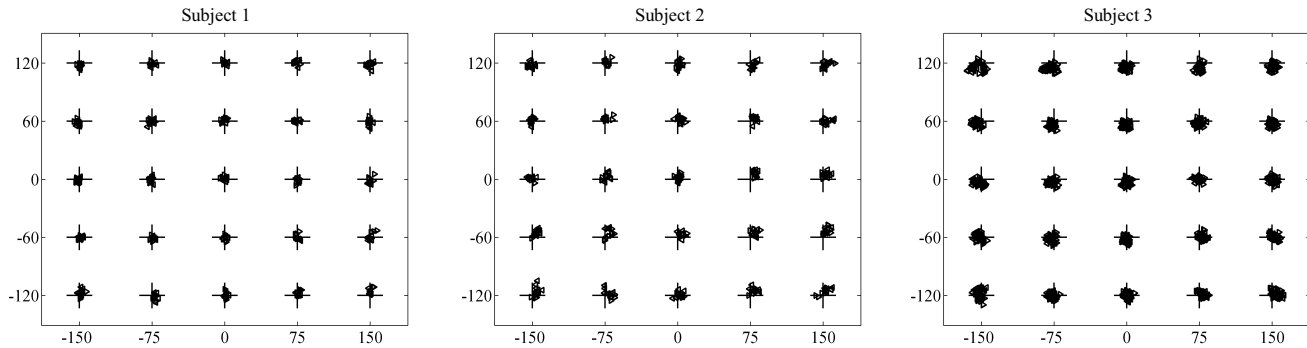


Fig. 3: Mean point-of-gaze estimates (+ : target fixation point, < : left eye, > : right eye). All coordinates are in mm.

screen. For each target point, 50 estimates (2.5 s) of the image coordinates of the centers of the corneal reflections and the parameters of the ellipse (center, major and minor axes, orientation of the major axis) fitted to the pupil-iris boundary points extracted from each eye image were recorded (8 points on the fitted ellipse were subsequently sampled at regular angular intervals). The subject-specific deviation of the visual axis from the optic axis, described by the unit vector  $\mathbf{v}_{\text{ECS}}$ , was determined when the head was at the nominal distance from the screen and the subject fixated a target point at the center of the screen. The PoG was estimated using the 3CR variant described in the previous section. The RMS PoG estimation error for the three subjects is provided in Table 1. Fig. 3 shows the mean PoG estimates obtained in each trial for both eyes.

Table 1: Experimental RMS point-of-gaze estimation error.

Subject	Eye	RMS Error
1	L	4.82 mm (0.40°)
	R	4.84 mm (0.40°)
2	L	7.26 mm (0.58°)
	R	6.62 mm (0.55°)
3	L	6.79 mm (0.56°)
	R	5.99 mm (0.49°)

## 5. CONCLUSIONS

The experimental results show that using the proposed methodology it is possible to estimate the PoG with an accuracy of 0.4-0.6° of visual angle after completing a single-point personal calibration procedure. Such accuracy is comparable to that of the best commercially available systems, which use multiple-point personal calibration procedures. Furthermore, to the best of our knowledge, the PoG estimation accuracy obtained with our prototype system is significantly better than that of any other systems that use a single-point personal calibration procedure and have been previously described in the literature (e.g., [5], [7], [8]).

## 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] P. Majaranta and K.-J. Rähkä, "Twenty years of eye typing: System and design issues," in *Proc. ETRA 2002*, Mar. 2002, pp. 15-22.
- [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] 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.
- [8] T. Nagamatsu, J. Kamahara, T. Iko, and N. Tanaka, "One-point calibration gaze tracking based on eyeball kinematics using stereo cameras," in *Proc. ETRA 2008*, Mar. 2008, pp. 95-98.
- [9] T. Ohno, N. Mukawa, and A. Yoshikawa, "FreeGaze: A gaze tracking system for everyday gaze interaction," in *Proc. ETRA 2002*, Mar. 2002, pp. 125-132.
- [10] A. Villanueva and R. Cabeza, "Models for gaze tracking systems," *EURASIP J. Image Video Process.*, vol. 2007, article ID 23570, 2007.
- [11] E. D. Guestrin, "Remote, non-contact gaze estimation with minimal subject cooperation," PhD Thesis, Univ. Toronto, Toronto, ON, Canada, 2010. (Available at <http://hdl.handle.net/1807/24349>)
- [12] M. C. Corbett, E. S. Rosen, and D. P. S. O'Brart, *Corneal Topography: Principles and Applications*. London, UK: BMJ Books, 1999, p. 6.
- [13] 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.
- [14] R. H. S. Carpenter, *Movements of the Eyes*. London, UK: Pion, 1977.
- [15] E. D. Guestrin and M. Eizenman, "Listing's and Donders' laws and the estimation of the point-of-gaze," in *Proc. ETRA 2010*, Mar. 2010, pp. 199-202.
- [16] J.-Y. Bouquet. (2011, Jan.) Camera Calibration Toolbox for MATLAB. California Institute of Technology, Department of Electrical Engineering, Pasadena, CA. [Online]. Available: [http://www.vision.caltech.edu/bouquet/calib\\_doc](http://www.vision.caltech.edu/bouquet/calib_doc)