31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

An Automatic Calibration Procedure for Remote Eye-Gaze Tracking Systems

Dmitri Model, Elias D. Guestrin, *Member, IEEE*, and Moshe Eizenman

***Abstract*— Remote gaze estimation systems use calibration procedures to estimate subject-specific parameters that are needed for the calculation of the point-of-gaze. In these procedures, subjects are required to fixate on a specific point or points at specific time instances. Advanced remote gaze estimation systems can estimate the optical axis of the eye without any personal calibration procedure, but use a single calibration point to estimate the angle between the optical axis and the visual axis (line-of-sight). This paper presents a novel automatic calibration procedure that does not require active user participation. To estimate the angles between the optical and visual axes of each eye, this procedure 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). Simulation results demonstrate that the performance of the algorithm improves as the range of viewing angles increases. For a subject sitting 75 cm in front of an 80 cm x 60 cm display (40" TV) the standard deviation of the error in the estimation of the angles between the optical and visual axes is 0.5°.**

***Index Terms*—Remote gaze estimation, calibration.**

I. Introduction

THE point-of-gaze (PoG) is the point within the visual field that is imaged on the highest acuity region of the retina that is known as the fovea. Systems that estimate the PoG are used in a large variety of applications [1] such as studies of emotional and cognitive processes [2], [3], driver behavior [4], marketing and advertising [5], pilot training [6], ergonomics [7] and human-computer interfaces [8].

All gaze estimation systems use calibration procedures to estimate subject-specific parameters that are needed for the calculation of the PoG. In these procedures, subjects are required to fixate on specific points at specific time instances. One of the most important goals in the field of gaze tracking technology is to estimate human PoG without calibration procedures that require active user participation.

Remote eye-gaze tracking (REGT) systems that use a stereo pair of video cameras can estimate the center of curvature of the cornea and the *optical* axis of the eye without any user calibration [9]-[11]. However, since human gaze is directed along the *visual* axis [12], [13], a one-point user-calibration procedure is still required for the estimation of the angle between the optical and visual axes. This paper describes a new methodology to estimate the 3-D angle between the optical and visual axes of the eye. This methodology does not require subjects to fixate on specific calibration points. Instead, it relies on the fact that when subjects look naturally at a display the visual axes of their eyes intersect on the surface of the display.

The next section presents the algorithm for the automatic estimation of the subject-specific angles between the optical and visual axes. Section III presents simulation results and the conclusions are presented in Section IV.

II. Automatic Calibration Methodology

A. Model

Fig. 1 presents a simplified schematic diagram of the eye. The line connecting the center of curvature of the cornea with the center of the pupil defines the *optical* axis. The line connecting the center of the fovea with the center of curvature of the cornea defines the *visual* axis or the line-of-sight. The average magnitude of the angle between the *optical* and *visual* axes is 5°. This angle has both horizontal (nasal) and vertical components, which exhibit considerable inter-personal variation [13].

To develop an algorithm that estimates the angle between the *optical* and *visual* axes, two coordinate systems are defined. The first coordinate system 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 Fig. 1). The second coordinate system 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, 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 [13], are in the horizontal and vertical directions, respectively. The X_{eye}-Y_{eye} plane rotates according to Listing's law [13] around the Z_{eye} axis for different gaze directions.

Manuscript received April 6, 2009. This work was supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC), and in part by scholarships from NSERC and the Vision Science Research Program (Toronto Western Research Institute, University Health Network, Toronto, ON, Canada).

D. Model is a Ph.D. student with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada (phone: 1-416-978-2255; fax: 1-416-978-4317; e-mail: dmitri.model@utoronto.ca).

E. D. Guestrin is a Ph.D. student with the Department of Electrical and Computer Engineering, and the Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, ON M5S 3G9, Canada (e-mail: elias.guestrin@utoronto.ca).

M. Eizenman is with the Department of Electrical and Computer Engineering, the Department of Ophthalmology and Vision Sciences, and the Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, ON M5S 3G9, Canada (e-mail: eizenm@ecf.utoronto.ca).

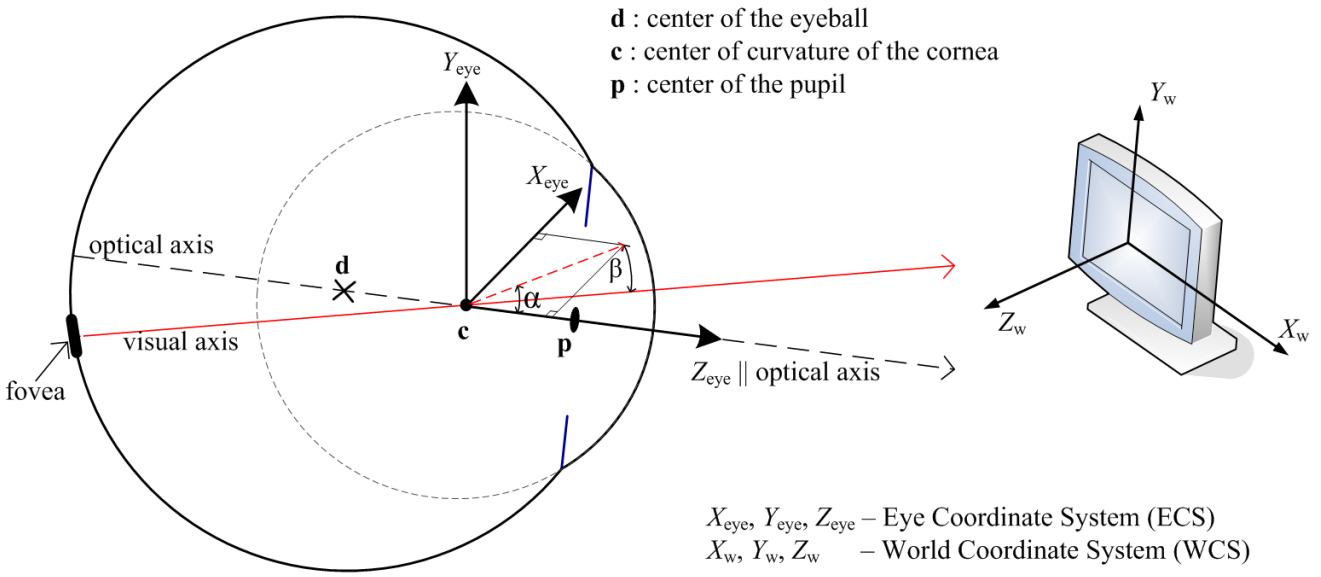


Fig. 1. Simplified schematics of the eye. The *optical* axis of the eye connects the center of the pupil with the center of curvature of the cornea. Gaze is directed along the *visual* axis, which connects the center of the region of highest acuity of the retina (fovea) with the center of curvature of the cornea.

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 Fig. 1). The unit vector in the direction of the *visual* axis with respect to the ECS, $\mathbf{v}_{ECS}$, is then expressed as

$$\mathbf{v}_{ECS}(\alpha,\beta)=\begin{bmatrix}-\sin(\alpha)\cos(\beta)\\ \sin(\beta)\\ \cos(\alpha)\cos(\beta)\end{bmatrix}. \quad (1)$$

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

$$\mathbf{v}(\alpha,\beta)=\mathbf{R}\,\mathbf{v}_{ECS}(\alpha,\beta) \quad (2)$$

where $\mathbf{R}$ is the rotation matrix from the ECS to the WCS (independent of α and β), which can be readily calculated from the orientation of the optical axis of the eye and Listing's law [13].

Because the visual axis goes through the center of curvature of the cornea, $\mathbf{c}$, 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

$$\mathbf{g}(\alpha,\beta)=\mathbf{c}+k(\alpha,\beta)\,\mathbf{v}(\alpha,\beta)=\mathbf{c}+k(\alpha,\beta)\,\mathbf{R}\,\mathbf{v}_{ECS}(\alpha,\beta) \quad (3)$$

with

$$k(\alpha,\beta)=-\frac{\mathbf{c}\bullet\mathbf{n}}{\mathbf{v}(\alpha,\beta)\bullet\mathbf{n}} \quad (4)$$

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

[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.

Note that since REGT systems that use a stereo pair of video cameras can estimate the center of curvature of the cornea and the *optical* axis of the eye without any user calibration [9]-[11], both $\mathbf{c}$ and $\mathbf{R}$ are known.

B. Automatic Calibration Algorithm

Throughout this subsection, the superscripts "L" and "R" are used to denote parameters of the left and right eyes, respectively.

The proposed automatic calibration algorithm for the estimation of α^L, β^L, α^R and β^R is based on the fact that at each time instant the visual axes of both eyes intersect on the surface of the display. 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

$$F(\alpha^L,\beta^L,\alpha^R,\beta^R)=\sum_i\left\|\mathbf{g}_i^L(\alpha^L,\beta^L)-\mathbf{g}_i^R(\alpha^R,\beta^R)\right\|_2^2 \quad (5)$$

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 (3) can be obtained by using the first three terms of its Taylor's series expansion:

$$\mathbf{g}(\alpha,\beta)\approx\mathbf{g}(\alpha_0,\beta_0)+\left.\frac{\partial\mathbf{g}(\alpha,\beta)}{\partial\alpha}\right|_{\alpha_0,\beta_0}(\alpha-\alpha_0)+\left.\frac{\partial\mathbf{g}(\alpha,\beta)}{\partial\beta}\right|_{\alpha_0,\beta_0}(\beta-\beta_0) \quad (6)$$

Let

$$\mathbf{g}_0 = \mathbf{g}(\alpha_0, \beta_0) = \mathbf{c} + k_0 \mathbf{v}_0 \tag{7}$$

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

Then, using (3),

$$\mathbf{a} = \left.\frac{\partial \mathbf{g}(\alpha,\beta)}{\partial \alpha}\right|_{\alpha_0,\beta_0} = k_\alpha \mathbf{v}_0 + k_0 \mathbf{v}_\alpha \tag{8}$$

$$\mathbf{b} = \left.\frac{\partial \mathbf{g}(\alpha,\beta)}{\partial \beta}\right|_{\alpha_0,\beta_0} = k_\beta \mathbf{v}_0 + k_0 \mathbf{v}_\beta \tag{9}$$

where, from (1)-(2),

$$\mathbf{v}_\alpha = \left.\frac{\partial \mathbf{v}(\alpha,\beta)}{\partial \alpha}\right|_{\alpha_0,\beta_0} = \mathbf{R}\begin{bmatrix} -\cos(\alpha_0)\cos(\beta_0) \\ 0 \\ -\sin(\alpha_0)\cos(\beta_0) \end{bmatrix} \tag{10}$$

$$\mathbf{v}_\beta = \left.\frac{\partial \mathbf{v}(\alpha,\beta)}{\partial \beta}\right|_{\alpha_0,\beta_0} = \mathbf{R}\begin{bmatrix} \sin(\alpha_0)\sin(\beta_0) \\ \cos(\beta_0) \\ -\cos(\alpha_0)\sin(\beta_0) \end{bmatrix} \tag{11}$$

and, from (4),

$$k_\alpha = \left.\frac{\partial k(\alpha,\beta)}{\partial \alpha}\right|_{\alpha_0,\beta_0} = -k_0 \frac{\mathbf{v}_\alpha \bullet \mathbf{n}}{\mathbf{v}_0 \bullet \mathbf{n}} \tag{12}$$

$$k_\beta = \left.\frac{\partial k(\alpha,\beta)}{\partial \beta}\right|_{\alpha_0,\beta_0} = -k_0 \frac{\mathbf{v}_\beta \bullet \mathbf{n}}{\mathbf{v}_0 \bullet \mathbf{n}} \,. \tag{13}$$

Using the above linear approximations, the sum of the squared distances between the left and right PoGs in the objective function (5) can be expressed as

$$F(\alpha^{\mathrm{L}}, \beta^{\mathrm{L}}, \alpha^{\mathrm{R}}, \beta^{\mathrm{R}}) = \sum_i \left\| \mathbf{M}_i \, \mathbf{x} + \mathbf{y}_i \right\|_2^2 \tag{14}$$

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 $\mathbf{x} = \left[(\alpha^{\mathrm{L}} - \alpha_0^{\mathrm{L}}) \quad (\beta^{\mathrm{L}} - \beta_0^{\mathrm{L}}) \quad (\alpha^{\mathrm{R}} - \alpha_0^{\mathrm{R}}) \quad (\beta^{\mathrm{R}} - \beta_0^{\mathrm{R}}) \right]^T$ 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 (14) can be obtained in closed-form using least squares as

$$\mathbf{x}_{\mathrm{opt}} = -(\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T \mathbf{y} \tag{15}$$

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

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_1 \\ \mathbf{M}_2 \\ \vdots \\ \mathbf{M}_N \end{bmatrix} \quad ; \quad \mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_N \end{bmatrix} . \tag{16}$$

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

$$\begin{bmatrix} \hat{\alpha}^{\mathrm{L}} & \hat{\beta}^{\mathrm{L}} & \hat{\alpha}^{\mathrm{R}} & \hat{\beta}^{\mathrm{R}} \end{bmatrix}^T = \begin{bmatrix} \alpha_0^{\mathrm{L}} & \beta_0^{\mathrm{L}} & \alpha_0^{\mathrm{R}} & \beta_0^{\mathrm{R}} \end{bmatrix}^T + \mathbf{x}_{\mathrm{opt}} \,. \tag{17}$$

Since the objective function (14) is a linear approximation of the objective function (5), several iterations of (7)-(17) might be needed to converge to the true minimum of the objective function (5). 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 $\hat{\alpha}^{\mathrm{L}}$, $\hat{\beta}^{\mathrm{L}}$, $\hat{\alpha}^{\mathrm{R}}$ and $\hat{\beta}^{\mathrm{R}}$ 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 $\mathbf{M}$ for each new estimate of the centers of curvature of the corneas and optical axes that are provided by the REGT.

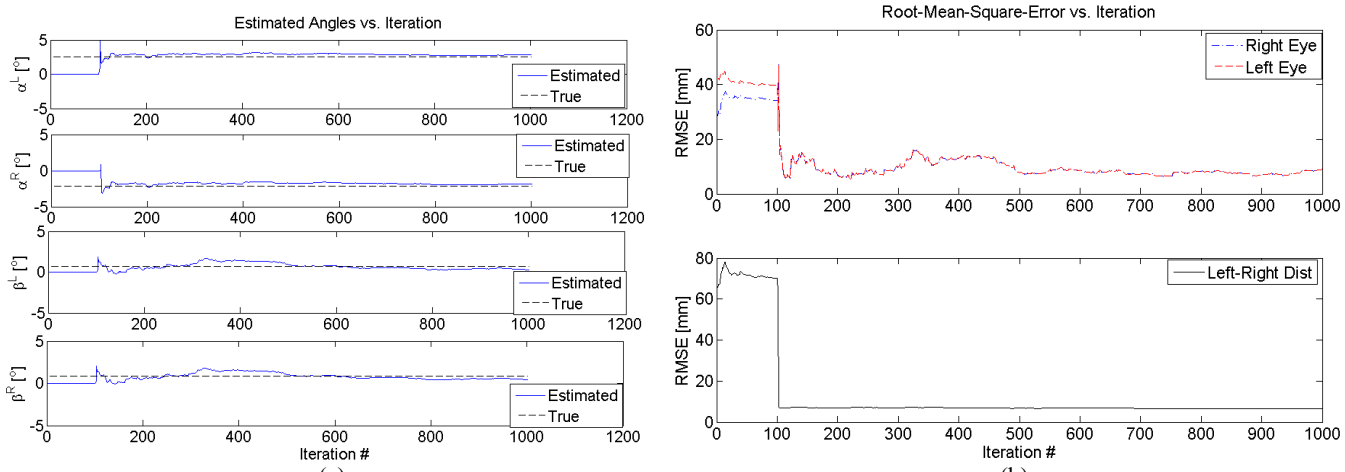
III. Numerical Simulations

The numerical simulations use system parameters (position of light sources and camera parameters) that are similar to the parameters of the two-camera REGT system described in [11]. The eye parameters are as follows: inter-pupillary distance = 60 mm; radius of curvature of the cornea = 7.8 mm; pupil radius = 4 mm; distance between the center of the pupil and the center of curvature of the cornea = 4.5 mm.

For the simulations, the position of the center of curvature of the cornea of the subject's right eye, $\mathbf{c}^{\mathrm{R}}$, was randomly selected from a uniform distribution in a 10 cm^3 volume around the point $[30\ 0\ 750]^T$ (i.e., 750 mm from the display surface) while the center of curvature of the cornea of the subject's left eye, $\mathbf{c}^{\mathrm{L}}$, was positioned at $\mathbf{c}^{\mathrm{R}} - [60\ 0\ 0]^T$ (i.e., 60 mm to the left of $\mathbf{c}^{\mathrm{R}}$). The subject-specific angles between the optical and visual axes were randomly drawn for each simulation from the uniform distribution in the 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 display surface, which was defined as the plane $Z_{\mathrm{w}} = 0$.

For each eye position and PoG, the pupil boundary and the corneal reflections were first reconstructed in the WCS and then projected onto the image planes of the two cameras of the REGT system. White Gaussian noise with a standard deviation of 0.1 pixels was then added to the coordinates of the projected features and the results were used to estimate the centers of curvature of the corneas and the optical axes of the two eyes. By following this procedure, the estimates of the positions of the centers of curvature of the corneas and the orientations of the optical axes of the two eyes exhibit similar noise and bias characteristics as the estimates of the remote two-camera system [11].

The proposed automatic calibration algorithm was implemented in MATLAB®. The reconstructed centers of curvature of the corneas and optical axes were used as inputs to the algorithm, which was used in an "on-line" mode, that is, at each iteration step a new PoG was added for the estimation of the angles between the optical and visual axes. The initial guess for each subject-specific angle was always zero. The first update of the algorithm was done



(a) (b)

Fig. 2. Estimation results with an 80 cm x 60 cm display. (a) Estimated subject-specific angles; (b) Root-Mean-Square Error (RMSE) of PoG estimation (top) and Left PoG - Right PoG distance (bottom).

after 100 PoGs were accumulated to prevent large fluctuations during start-up when only few PoGs are available.

Fig. 2 shows a typical simulation result where the subject looked at 1000 points on an 80 cm x 60 cm display (approximately the viewing area of a 40" monitor). As it can be seen from Fig. 2(a), the solution converged to the true values of α^L, β^L, α^R and β^R. The objective function was effectively minimized (Fig. 2 (b), bottom) after the first 100 iterations. However, the left and right PoGs continued to drift, simultaneously, around the true PoG locations until the true values of α^L, β^L, α^R and β^R were obtained (Fig. 2 (b), top).

Table I shows the means and standard deviations of the estimation errors for the horizontal, α, and, vertical, β, components of the angle between the *optical* and *visual* axes for three different display sizes (20", 30" and 40"). Each entry in the table is based on 100 simulations. Table I shows that as the range of viewing angles increases with the size of the display, the standard deviation of the error decreases. When the range of viewing angles is larger than ±28 degrees horizontally and ±22 degrees vertically, the standard deviation of the error in the estimation of the the subject-specific angles between the optical and visual axes is 0.5°.

TABLE I
ESTIMATION ERROR

Monitor Size	α^L, α^R	β^L, β^R
40 cm x 30 cm	-0.1 ± 0.61	0.1 ± 0.55
60 cm x 45 cm	0.0 ± 0.33	0.0 ± 0.45
80 cm x 60 cm	0.0 ± 0.32	0.0 ± 0.38

Mean and standard deviation of the error in the estimation of the subject-specific angles for different display sizes.

IV. CONCLUSIONS

This paper describes an automatic procedure for the estimation of the angle between the optical and visual axes while subjects look naturally at a display. The estimation errors are directly affected by the accuracy of the REGT system. For example, with a REGT system that generates "error-free" estimates of the centers of curvature of the corneas and optical axes, the error in the estimation of the angle between the optical and visual axis will be zero even with small displays. When the proposed algorithm is integrated with the state-of-the-art REGT system [11], the combined system can estimate human PoG without calibration procedures that require active user participation.

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," *Psychiatr. Res.*, vol. 118, no. 2, pp. 117-128, May 2003.

[3] K. Rayner, "Eye movements in reading and information processing: 20 years of research," *Psychol. Bull.*, vol. 124, no. 3, pp. 372-422, Nov. 1998.

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

[5] G. Loshe, "Consumer eye movement patterns of Yellow Pages advertising," *J. Advert.*, vol. 26, no. 1, pp. 61-73, 1997.

[6] P. A. Wetzel, G. Krueger-Anderson, C. Poprik, and P. Bascom, "An eye tracking system for analysis of pilots' scan paths," United States Air Force Armstrong Laboratory Tech. Rep. AL/HR-TR-1996-0145, Apr. 1997.

[7] J. H. Goldberg and X. P. Kotval, "Computer interface evaluation using eye movements: methods and constructs," *Int. J. Ind. Erg.*, vol. 24, no. 6, pp. 631-645, Oct. 1999.

[8] T. E. Hutchinson, K. P. White, W. N. Martin, K. C. Reichert, and L. A. Frey, "Human-computer interaction using eye-gaze input," *IEEE Trans. Syst., Man, Cybern.*, vol. 19, no. 6, pp. 1527-1534, Nov./Dec. 1989.

[9] 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.

[10] 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.

[11] E. D. Guestrin and M. Eizenman, "Remote point-of-gaze estimation requiring a single-point calibration for applications with infants," in *Proc. of the 2008 Eye Tracking Research & Applications Symposium (ETRA 2008), Savannah, GA, USA*, Mar. 2008, pp. 267-274.

[12] L. R. Young and D. Sheena, "Survey of eye-movement recording methods," *Behav. Res. Meth. Instrum.*, vol. 7, no. 5, pp. 397-429, 1975.

[13] R. H. S. Carpenter, *Movements of the eyes*. London, UK: Pion, 1977.