A Calibration Method for Eye-Gaze Estimation Systems Based on 3D Geometrical Optics

Hwaran Lee, Nadeem Iqbal, Wonil Chang, and Soo-Young Lee

***Abstract*— A new calibration method is presented for robust eye-gaze estimation systems which are utilized to understand the human mind. Although current eye-gaze systems with one camera and a few infrared light sources have been developed to allow users' head motion, they still require one to know the relative positions between the camera and light sources with very high accuracy. The developed calibration method utilizes a three-dimensional geometrical relationship between the light source positions and corresponding camera images reflected on a mirror at several positions and rotations. The best estimates of the light source positions are obtained from noisy measurements by minimizing a cost function, which ensures the integrity of the camera and light sources. The developed calibration method makes it possible to convert many camera-and-display devices into robust eye-gaze estimation systems.**

***Index Terms*— Eye-gaze estimation system, sensor system calibration, model-based eye-gaze estimation.**

I. Introduction

EYES are windows of the brain. The brain processes sensory data by putting attention to specific components for efficiency. This is especially true to human visual system, which handles huge information in real-time by selectively processing only parts of the whole visual scene [1]. Therefore, the tracking of eye-gaze provides clues on the current status of brain information processing for both normal and atypical. Especially, the eye-gaze estimation requires only non-invasive process and has been widely used for both children and adults for both clinical and research purposes [2]–[7].

Eye-gaze tracking technologies are categorized into two approaches, i.e., interpolation (or regression) and model-based (or geometric) methods [8]. The interpolation methods obtain a mapping function from 2-dimensional (2D) eye-image features (or raw images) to determine the gaze position [9]. Although the interpolation methods are quite simple, they have difficulty to allow head movement and therefore are not suitable for many real-world applications. The model-based approaches utilize the 3-dimensional (3D) geometric relationship among the eye(s), camera, and infrared light sources [10]–[16]. The optical axis is estimated as the line between the cornea and the pupil centers, from which the visual axis is obtainable in connection with user-dependent offset angles. The model-based approaches can allow relative motion between the device and the user's eye and have potential advantages over the interpolation approaches.

However, in order to estimate the visual axis accurately, the model-based approaches require accurate values for a few system and user parameters. The former includes camera focal length and relative positions among the camera and light emitting diodes (or display screen), while the latter includes offset angles between the optical and visual axes. The visual axis is very sensitive to these parameters, and it is still challenging to obtain accurate eye-gaze estimation from consumer or custom-made devices with one camera and a small display. Therefore, a stereo camera [17] or a large display screen [18] had been used. In this paper, a novel method is presented to calibrate the system parameters for the eye-gaze estimation systems based on mobile consumer devices.

The remainder of this paper is organized as follows: Section II presents the eye-gaze system hardware and algorithm, and Section III describes the proposed calibration method. Experimental setup and results are presented in Section V.

II. Geometrical Properties of the System

Figure 1(a) shows an eye-gaze estimation system containing a single camera and multiple light sources. It may be a consumer device (i.e., a smartphone) or custom-made device (i.e., personal computer or touchpad with an additional camera). The L light sources denoted by $\mathbf{s}_i$ $(i = 1, \ldots, L)$ are located on the edges of the display screen, or some display pixels serve as the light sources. When eyes are located in front of this system, the camera captures glints of these lights at cornea surface as well as the pupil. One can estimate the cornea center from the geometrical relationship among the glint camera images and light source positions. Also, the pupil center is estimated from the elliptic pupil images. The eye-gaze optical axis is obtainable by connecting the cornea and pupil centers. The visual axis may be estimated with slight rotations from the optical axis, of which parameters are user dependent and estimated from binocular images.

Figure 1(b) shows the geometric relation among the light source $\mathbf{s}$, camera center $\mathbf{o}$, and cornea center $\mathbf{c}$. In this paper a symbol with lower-case bold fonts denotes a physical position, and also a 3D vector from the camera center $\mathbf{o}$. A spherical cornea is assumed with its center $\mathbf{c}$ and radius r. Let's consider the i-th light source $\mathbf{s}_i$, the corresponding glint image $\mathbf{u}_i$ on

Manuscript received February 17, 2013; revised April 6, 2013; accepted April 23, 2013. Date of publication June 12, 2013; date of current version July 30, 2013. This work was supported in part by the Basic Science Research Program of the National Research Foundation of Korea, funded by the Ministry of Education, Science and Technology under Grants 2011-0029816 and 2012-0008681. The associate editor coordinating the review of this paper and approving it for publication was Prof. Aime Lay-Ekuakille.

The authors are with the Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea (e-mail: hwaran.lee@gmail.com; nadeemi77@kaist.ac.kr; chang.wonil@gmail.com; sylee@kaist.ac.kr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSEN.2013.2268247

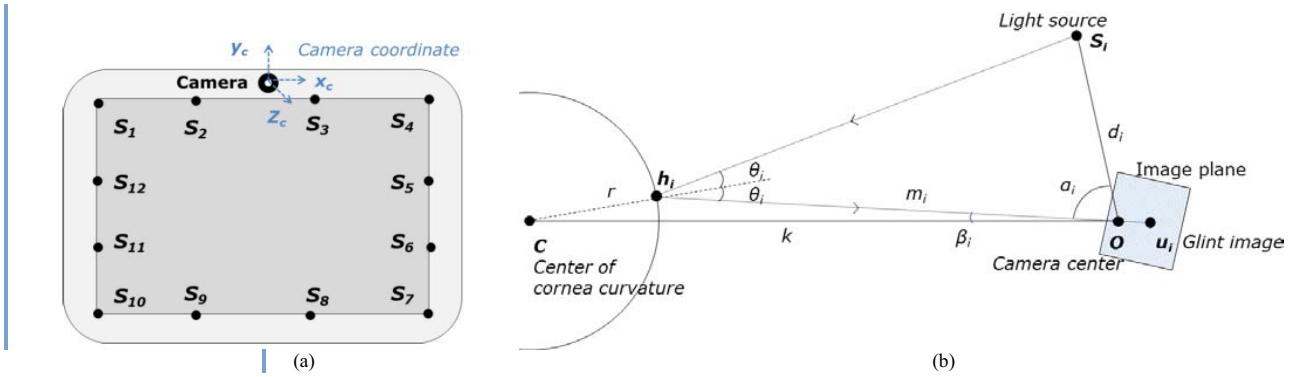


Fig. 1. (a) An eye-gaze estimation system with a single camera and multiple (infrared) light sources for glint generation. Some pixels of the display may serve as the light sources. (b) Geometrical relations among a light source, a cornea, and a camera. (The cornea and camera are shown to be much bigger than actual sizes.)

the camera imaging plane, and a reflection point $\mathbf{h}_i$ on the cornea surface. The subscript i represents the ith light source. In this system, $\mathbf{o}$, $\mathbf{s}_i$, $\mathbf{u_i}$, d_i and m_i are measured with some accuracy, but $\mathbf{c}$ and r need be estimated with high accuracy.

The cornea radius r is an important user parameter for the accurate estimation of the cornea center, and usually is estimated from the glint images. The model-base approaches utilize geometrical relation to find the cornea radius and also the cornea center.

Each light source defines a plane with three positions, $\mathbf{o}$, $\mathbf{s}_i$, and $\mathbf{u}_i$, and the cornea center $\mathbf{c}$ and the reflection point $\mathbf{h}_i$ are required to be located on the same plane. Also, the reflection law assures that the angle of incidence θ_i is equal to the angle of reflection. The cornea radius r is related to the distance $k = \|\mathbf{c}\|$ between the cornea center and the camera center as

$$\frac{k}{r} = \frac{\sin\theta_i}{\sin\beta_i} \tag{1}$$

$$\frac{r}{d_i} = \frac{\sin\beta_i}{\sin(\theta_i - \beta_i)} \frac{\sin(2\theta_i + \alpha_i)}{\sin 2\theta_i}, \tag{2}$$

where both the distance d_i between the light source and the camera center, and the angle α_i between the light source and the glint vector are obtainable by measurements or calibration process in advance.

Two planes defined by two light sources $\mathbf{s}_i$ and $\mathbf{s}_j$ form an intersect line denoted by a vector $\mathbf{w}_{ij}$, and both $\mathbf{o}$ and $\mathbf{c}$ should be located on this line. Since $\mathbf{c}$ and $\mathbf{w}_{ij}$ are on the same direction, the angle β_i between $-\mathbf{u}_i$ and $\mathbf{c}$ is uniquely determined from two light sources. Since r and k should be the same for all light sources, the optimum estimation of the cornea radius r is obtainable by minimizing a cost function defined as

$$f(r) = \frac{1}{2} \sum_{i=1}^{L} \sum_{j=1}^{i-1} \left(\frac{\sin\theta_i(r)}{\sin\beta_i} - \frac{\sin\theta_j(r)}{\sin\beta_j} \right)^2, \tag{3}$$

where θ_i (r) is defined as the inverse function of (2). It is important to note that only one variable r is used to minimize the cost function, and the simple gradient descent iteration is applicable as

$$r \leftarrow r - \eta \frac{\partial f(r)}{\partial r} \tag{4}$$

with a learning rate η.

At each iteration, new θ_i values are estimated by numerically solving the nonlinear equation (2) with new r value. Then, from (1), the value of k and the cornea center position $\mathbf{c}$ are obtained.

The robustness to unknown background illumination is a very critical requirement to many vision systems. To obtain this robustness one can utilize frame difference images. As shown in Figure 2, while the light sources are blinking at a high frame rate (60 Hz or higher), the system records images of user's eyes in synchronization with the blinking. Then, provided that the background illumination changed slowly, the frame difference images remove effects of the background illumination. Accurate position of the glint images is obtainable by a weighted average of the glint image intensities. For the higher accuracy the temporal average may also be applicable. The estimation of pupil center needs similar processing with two additional steps, i.e., the edge detection and elliptic curve fitting.

III. Noise Correction for Light Source Positions

Although the head motion is allowed in this model-based eye-gaze system, the eye-gaze angle is very sensitive to the system parameters such as the relative positions among the camera and light sources. Even a very small error on $\mathbf{s_i}$ causes a large error on the estimated cornea radius r, cornea center position $\mathbf{c}$, and the estimated gaze angle. In Figure 3 the eye-gaze angular error is plotted as a function of the light source position error for typical eye-gaze estimation systems Here, we assume that all light sources are on the display screen. As shown in Figure 3, it is required to have 0.1 mm light position accuracy for 1 degree accuracy in the eye-gaze optical axis.

The i-th light source position $\mathbf{s}_i$ needs be estimated from manufacturer's specification sheet or measured by a ruler, while the image position $\mathbf{u}_i$ is estimated from the camera images. Due to recent high-resolution cameras, $\mathbf{u}_i$ can be estimated with much better accuracy than $\mathbf{s}_i$. However, simple ruler measurements on $\mathbf{s}_i$ may not provide the required accuracy, and it is important to calibrate the system for accurate light source positions. Therefore, we devised a calibration method to estimate correct light source position $\mathbf{s}_i$ from corresponding image position $\mathbf{u}_i$.

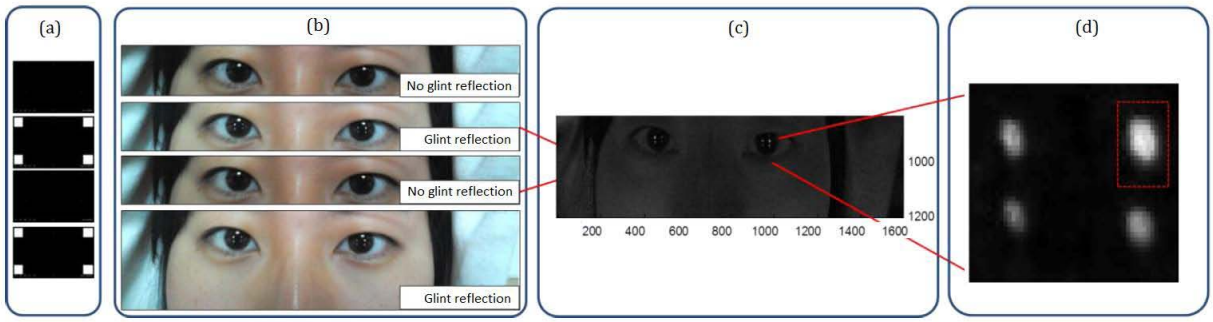


Fig. 2. Image processing for detecting glint location $\mathbf{u}_i$. (a) Four white boxes of edge size 75px are illuminating at each corner of the screen. (A large number of pixels is used here just to show the glint images clearly.) (b) The display is blinking at a high frame rate and the camera captures both images with and without the glints. (c) Absolute values of the frame difference become robust to background illumination. (d) Enlarged images of the left eye.

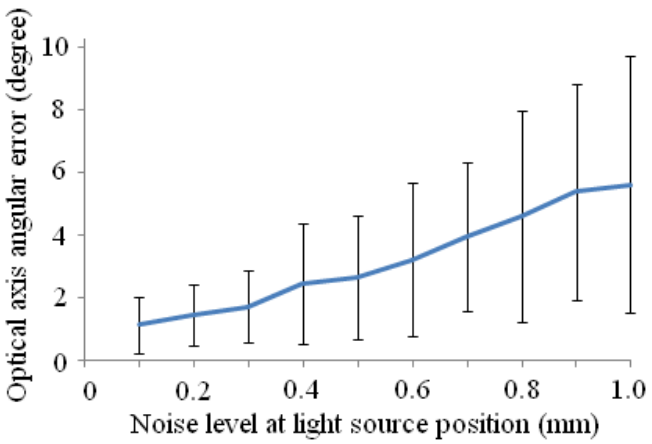


Fig. 3. Angular error of the gazing optical axis as a function of light source position error. The mean and standard deviation are plotted with 50 Gaussian random realizations.

A. Formulation as an Optimization Problem

Since the light sources are not within the viewing angle of the camera, a mirror is placed in front of the system for the calibration. It is worth noting that a mirror is equivalent to a cornea with infinite radius. Therefore, as shown in Figure 4, one can modify the geometrical relations described in Section II. The relation between the light source position $\mathbf{s}_i$ and corresponding image position $\mathbf{u}_i$ is utilized with multiple mirror positions and/or rotation angles. Here, the main simplification comes from $\beta_i = \theta_i$.

Since a light source $\mathbf{s}_i$ forms a plane $(\Delta \mathbf{s}_i \mathbf{u}_i \mathbf{o})$ any two light sources $\mathbf{s}_i$ and $\mathbf{s}_j$ determine a unit vector $\mathbf{w}_{ij}$, which lies on the intersection line of two planes $(\Delta \mathbf{s}_i \mathbf{u}_i \mathbf{o})$ and $(\Delta \mathbf{s}_j \mathbf{u}_j \mathbf{o})$ and is heading toward the mirror. Both planes and intersection vector $\mathbf{w}_{ij}$ are perpendicular to the mirror. Without any error on the light source position $\mathbf{s}_i$ and its image $\mathbf{u}_i$, $\mathbf{w}_{ij}$'s should be the same for all (ij) pairs. With errors on $\mathbf{s}_i$, $\mathbf{w}_{ij}$'s are not the same and an average value may be defined as

$$\hat{\mathbf{w}}_{av} = \frac{2}{L(L-1)} \sum_{i=1}^{L} \sum_{j=1}^{i-1} \mathbf{w}_{ij}, \quad \mathbf{w}_{av} = \frac{\hat{\mathbf{w}}_{av}}{\|\hat{\mathbf{w}}_{av}\|}. \tag{5}$$

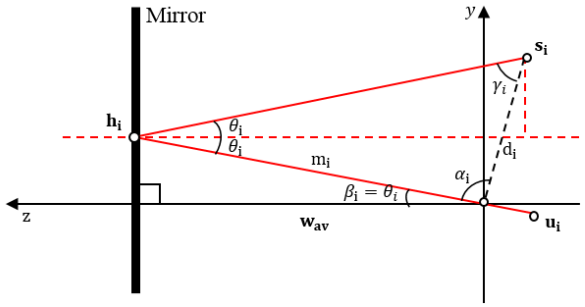


Fig. 4. Reflection of a light source $\boldsymbol{s_i}$ by a mirror and its image $\boldsymbol{u_i}$ on a camera. Here, $\mathbf{w}_{ij}$ is aligned to z-axis at a new rotated coordinate system.

Here, we rotated the original camera coordinate system to align $\mathbf{w}_{av}$ on the z-axis by

$$\mathbf{R}_m = \begin{bmatrix} -\sin\varphi_m & \cos\varphi_m & 0 \\ \cos\phi_m \cos\varphi_m & \cos\phi_m \sin\varphi_m & -\sin\phi_m \\ \sin\phi_m \cos\varphi_m & \sin\phi_m \sin\varphi_m & \cos\phi_m \end{bmatrix}, \tag{6}$$

where ϕ_m and φ_m are polar and azimuthal angles of $\mathbf{w}_{av}$ in the original camera coordinate system, respectively. Then, to locate $\mathbf{u}_i$ on the (yz)-plane, the second rotation along the z-axis is introduced as

$$\mathbf{Q}_{mi} = \begin{bmatrix} -\sin\psi_{mi} & \cos\psi_{mi} & 0 \\ \cos\psi_{mi} & \sin\psi_{mi} & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{7}$$

These rotations result in a transform of position vectors from the original camera-centered coordinated system into a new coordinate system as

$$\hat{\mathbf{u}}_{mi} = \mathbf{Q}_{mi} \mathbf{R}_m \mathbf{u}_i \text{ and } \hat{\mathbf{s}}_{mi} = \mathbf{Q}_{mi} \mathbf{R}_m \mathbf{s}_i \tag{8}$$

Here, a hat $(\hat{})$ denotes the rotated coordinate system. (For notational simplicity the subscript m and hat $(\hat{})$ are omitted in Figure 4.) Then, the reflection point $\mathbf{h}_i$ is determined as

$$\begin{aligned} \hat{h}^y_{mi} &= \frac{1}{2}\left(\hat{s}^y_{mi} + \hat{s}^z_{mi} \tan\theta_{mi}\right), \\ \hat{h}^z_{mi} &= \frac{1}{2}\left(\hat{s}^z_{mi} + \hat{s}^y_{mi} \cot\theta_{mi}\right) \end{aligned} \tag{9}$$

where superscript y and z denote the y and z components, respectively. It is important to note that the reflection point $\mathbf{h}_i$ is located on the mirror and $\hat{h}^z_{mi}$ should be the same for all light sources with a mirror position and rotation.

Now, we are ready to introduce a cost function to minimize for the reduction of the light source position errors. The error on $\mathbf{s}_i$ consists of two components, i.e., one on the $\Delta\mathbf{ozu}_i$ plane and the other perpendicular to the plane. While the latter may be reduced by simply minimizing the perpendicular components ($\hat{s}^x_{mi}$), the former requires more sophisticated processes. Here, we propose to utilize the co-planarity of the reflection point $\mathbf{h}_i$'s. The resulting cost function becomes

$$E(\mathbf{S}) = \lambda E_o(\mathbf{S}) + (1-\lambda)E_p(\mathbf{S}), \tag{10}$$

where E_o and E_p are related to the errors on and perpendicular to the $\Delta\mathbf{ozu}_i$ plane, respectively. Also, $\mathbf{S}$ is a matrix formed from all light source positions $\mathbf{s}_i$'s, and λ is a relative weight between the two terms. The first term is defined as

$$\begin{aligned} E_m(\mathbf{S}) &= \frac{1}{2}\sum_{m=1}^{M}\sum_{i=1}^{L}\sum_{j=1}^{L}\left[\hat{h}^z_{mi}(\mathbf{s}_i) - \hat{h}^z_{mj}(\mathbf{s}_j)\right]^2 \\ &= \frac{1}{2}\sum_{m=1}^{M}\sum_{i=1}^{L}\sum_{j=1}^{L}\left(\hat{\mathbf{g}}^T_{mi}\mathbf{s}_i - \hat{\mathbf{g}}^T_{mj}\mathbf{s}_j\right)^2 \\ &= \sum_{i=1}^{L}\mathbf{s}_i^T\mathbf{G}_{ii}\mathbf{s}_i - \sum_{j=1}^{L}\sum_{i=1}^{L}\mathbf{s}_i^T\mathbf{G}_{ij}\mathbf{s}_j \end{aligned} \tag{11}$$

where

$$\mathbf{g}_{mi} = \frac{1}{2}\begin{bmatrix} 0 & \cot\theta_{mi} & 1 \end{bmatrix}^T, \quad \hat{\mathbf{g}}^T_{mi} = \mathbf{g}^T_{mi}\mathbf{Q}_{mi}\mathbf{R}_m, \tag{12a}$$

$$\mathbf{G}_{ij} = \sum_{m=1}^{M}\hat{\mathbf{g}}_{mi}\hat{\mathbf{g}}^T_{mj} = \sum_{m=1}^{M}\mathbf{R}^T_m\mathbf{Q}^T_{mi}\mathbf{g}_{mi}\mathbf{g}^T_{mj}\mathbf{Q}_{mj}\mathbf{R}_m. \tag{12b}$$

The second term is defined as

$$\begin{aligned} E_i(\mathbf{S}) &= \frac{1}{2}\sum_{m=1}^{M}\sum_{i=1}^{L}\left[\hat{s}^x_{mi}(\mathbf{s}_i)\right]^2 \\ &= \frac{1}{2}\sum_{m=1}^{M}\sum_{i=1}^{L}\left(\hat{\mathbf{v}}^T_{mi}\mathbf{s}_i\right)^2 = \frac{1}{2}\sum_{i=1}^{L}\mathbf{s}_i^T\mathbf{V}_{ii}\mathbf{s}_i \end{aligned} \tag{13}$$

where

$$\mathbf{v} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T, \quad \hat{\mathbf{v}}^T_{mi} = \mathbf{v}^T\mathbf{Q}_{mi}\mathbf{R}_m, \tag{14a}$$

$$\mathbf{V}_{ij} = \sum_{m=1}^{M}\hat{\mathbf{v}}_{mi}\hat{\mathbf{v}}^T_{mj} = \sum_{m=1}^{M}\mathbf{R}^T_m\mathbf{Q}^T_{mi}\mathbf{v}\mathbf{v}^T\mathbf{Q}_{mj}\mathbf{R}_m. \tag{14b}$$

By minimizing the cost function (10), one obtains error-reduced light source positions $\mathbf{s}_i$'s from accurate $\mathbf{u}_i$'s and noisy $\mathbf{s}_i$'s. However, both E_o and E_p are not normalized, and $\mathbf{s}_i = 0$ becomes a trivial solution. Non-trivial solutions are obtainable by imposing an additional constraint such as a fixed value on $\|\mathbf{s}_i\|$ or s^z_i. If the light sources are located on the display surface or some display pixels are used as the light sources, the latter constraint, i.e., a fixed s^z_i is applicable.

Let's consider the uniqueness of the optimization solution. With only one mirror position and a known normal vector $\mathbf{w}_{av}$ any point along $\mathbf{s}_i\mathbf{h}_i$ line does not change the cost function (10). Therefore, with one mirror position, the optimization problem is underdetermined and has many possible solutions. Multiple mirror positions are required.

With M mirror positions and L light sources equations (11) and (13) are optimization formulations of $M(L\text{-}1)$ equality constraints ($\hat{h}^z_{mi} = \hat{h}^z_{mj}$) and ML equality constraints ($\hat{s}^x_{mi} = 0$), respectively. The number of unknowns is $2L$with the s^z_i constraint. If no redundancy exists on the mirror and light source positions, two mirror positions are enough to have a unique solution for two light sources or more. In practice it is desirable to have more light sources and mirror positions for better accuracy.

B. Optimization Algorithm

If the normal vector $\mathbf{w}_{av}$ is given, both $\mathbf{R}_m$ and $\mathbf{g}_{mi}$ are independent on the adaptive variable $\mathbf{S}$ and $\mathbf{Q}_{mi}$ becomes the same for all $\mathbf{s}_i$'s on the $y-z$ plane. In this case equation (10) becomes a quadratic function, and the optimization solution may be obtainable by solving coupled linear equations

$$\begin{aligned} \frac{\partial E}{\partial \mathbf{s}_i} &\approx \lambda\left[(L-1)\mathbf{G}_{ii}\mathbf{s}_i - \sum_{j\neq i}^{L}\mathbf{G}_{ij}\mathbf{s}_j\right] + (1-\lambda)\mathbf{V}_{ii}\mathbf{s}_i \\ &= \mathbf{0}. \end{aligned} \tag{15}$$

Unfortunately the normal vector $\mathbf{w}_{av}$ is a function of all $\mathbf{s}_i$'s, and the solution is very sensitive to $\mathbf{w}_{av}$. Only slight changes in $\mathbf{w}_{av}$ result in a big error on the solution. Also, an analytic formula for the derivatives of $\mathbf{w}_{av}$ over $\mathbf{s}_i$ is very complicated.

Two different approaches were implemented, i.e., alternating updates and numerical derivatives [19]. The alternating updates (AU) approach includes two steps at each iterative epoch, i.e., one for the estimation of the normal vector $\mathbf{w}_{av}$ by (5) and the other for updating all $\mathbf{s}_i$'s by

$$\begin{aligned} \mathbf{s}_i &\Leftarrow \mathbf{s}_i - \eta\frac{\partial E}{\partial \mathbf{s}_i} \\ &\approx \mathbf{s}_i - \eta\left\{\lambda\left[(L-1)\mathbf{G}_{ii}\mathbf{s}_i - \sum_{j\neq i}^{L}\mathbf{G}_{ij}\mathbf{s}_j\right]\right. \\ &\qquad \left. +(1-\lambda)\mathbf{V}_{ii}\mathbf{s}_i\right\} \end{aligned} \tag{16}$$

where η is a learning rate. The numerical derivatives (ND) approach utilizes numerical derivatives at each iterative epoch as

$$\mathbf{s}_i \Leftarrow \mathbf{s}_i - \eta\frac{E(\mathbf{s}_i + \Delta\mathbf{s}_i) - E(\mathbf{s}_i)}{\Delta\mathbf{s}_i}. \tag{17}$$

IV. Numerical Results

The performance of the proposed calibration method was evaluated for simulated data with known light source positions. Twenty-five or thirteen light sources are arranged on a display screen, and a camera is located at the top. The display screen has about 220 mm width and 160 mm height. The camera has 8 mega-byte pixels with about 1 μm pixel size. The calibrating

TABLE I
TRUE POSITIONS OF THE LIGHT SOURCES AND MIRRORS

25 light source positions (*x,y,z*) in *mm*				
(-100,-5,-5)	(-50,-5,-5)	(0,-5,-5)	(50,-5,-5)	(100,-5,-5)
(-100,-45,-5)	(-50,-45,-5)	(0,-45,-5)	(50,-45,-5)	(100,-45,-5)
(-100,-85,-5)	(-50,-85,-5)	(0,-85,-5)	(50,-85,-5)	(100,-85,-5)
(-100,-125,-5)	(-50,-125,-5)	(0,-125,-5)	(50,-125,-5)	(100,-125,-5)
(-100,-165,-5)	(100,-165,-5)	(0,-165,-5)	(50,-165,-5)	(100,-165,-5)
13 light source positions (*x,y,z*) in *mm*				
(50,-5,-5)	(100,-5,-5)	(100,-55,-5)	(100,-105,-5)	(100,-155,-5)
(50,-155,-5)	(0,-155,-5)	(-50,-155,-5)	(-100,-155,-5)	(-100,-105,-5)
(-100,-55,-5)	(-100,-5,-5)	(-50,-5,-5)		
11 mirror positions (φ_m, ϕ_m, z_m) in degrees and *mm*				
(7.5,15,200)	(15,15,200)	(30,15,200)	(15,15,190)	(30,15,210)
(15,30,180)	(30,30,220)	(15,-15,190)	(30,-15,210)	(15,-30,180)
(30,-30,220)				

mirror is positioned between 180 and 220 mm from the camera center with rotation angles between 0 and 30 degrees.

We set 25 light sources and 11 mirror positions as the reference system configuration. Then, to investigate the effect of the numbers of light sources and mirror positions, we also tested for 13 light sources and 5 mirror positions. The exact positions of the light sources in (*xyz*) coordinates are summarized in Table I and shown in Figure 5(a). (The 5 mirror positions are selected as the first five mirror positions in Table I.) The 11 mirror rotation angles and z-axis translations are also shown in Table I. For the measurement errors Gaussian random numbers were generated with standard deviation of 0.1, 0.2, 0.5, and 1.0 mm for the light sources and 0, 0.2, 0.5, 0.7, and 1.0 μm for the glint images, respectively. For the light source positions the errors were chosen as practical values with careful measurements by a ruler. Also, the errors on the glint images correspond to practical accuracy with current image processing technology and about 1 μm camera pixel size. Ten random realizations were tested for each error levels. With each realization the proposed calibration algorithm was applied to reduce the light source position errors. The relative weight λ was set to 0.5, which equally imposes both constraints in (10). Both the alternating updates (AU) and numerical derivatives (ND) were tested, but the results were similar.

Figure 5 shows typical results of the error reduction during the calibration process. In Figure 5(a) twenty-five light sources and a camera are shown. The positions during the iterative process are superimposed. In Figure 5(b) each position is enlarged to show the details for 8 light sources. Here, the red dot denotes the true position, while the blue circles denote the positions during the iterative adaptation process. A few light source positions move to the true positions along almost a straight line, while the others show curved trajectory to the true position. Also, for many light sources the converged positions are not the exact true positions, but become closer to the true positions. It shows the error surface defined in (10) is a complex function of the light source positions.

The calibration performance is summarized in Figure 6. Figure 6(a) shows the mean and standard deviation of

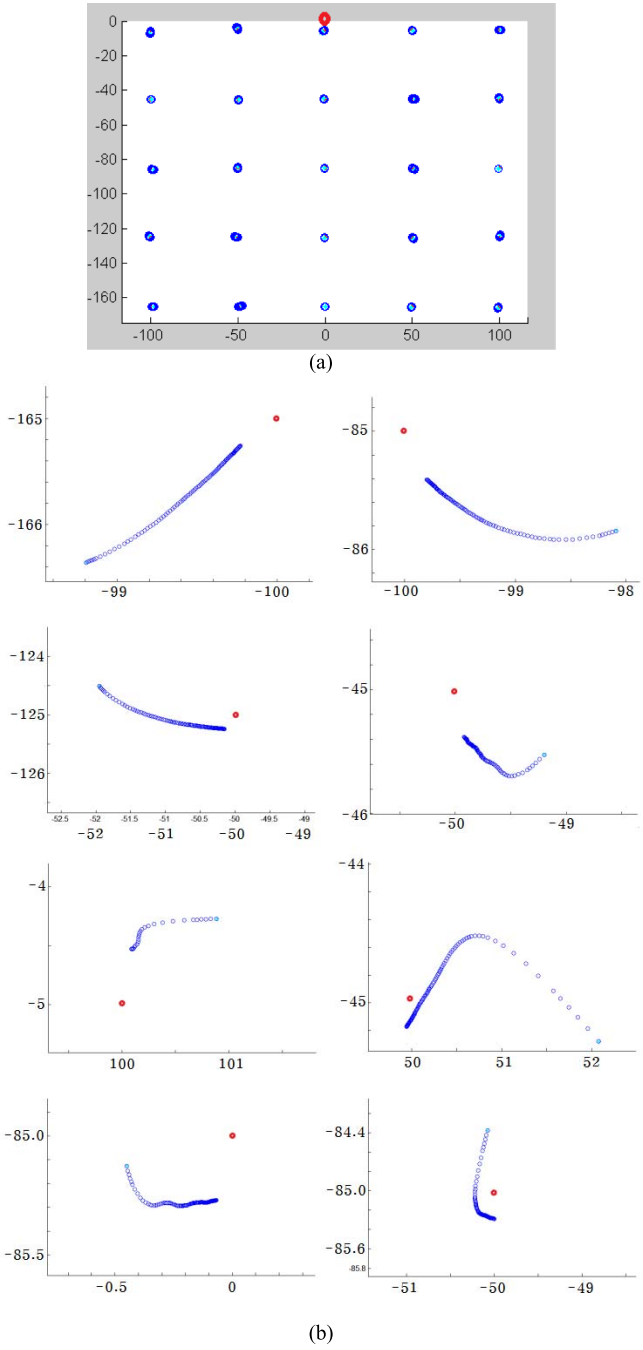


Fig. 5. Typical results of error reduction with 25 light sources and 11 mirror positions. (a) All right sources. (b) Enlarged 8 sources. The estimated source positions (blue circles) start from noisy positions and move to the true position (red dot). All coordinates are in mm units.

converged source position errors for 10 random realizations with 25 lights sources and 11 mirror positions. The light source position errors are reduced to about a half. It is worth noting that the residual errors are not sensitive to the glint imaging position errors, but sensitive to the initial light source position errors. Also, to obtain 0.1 mm residual error for about 1 degree eye-gaze angular error, about 0.2 mm initial error is good enough for the light source position measurements.

When the number of light sources is reduced to 13 in Figure 6(b) or the number of mirror positions is reduced to

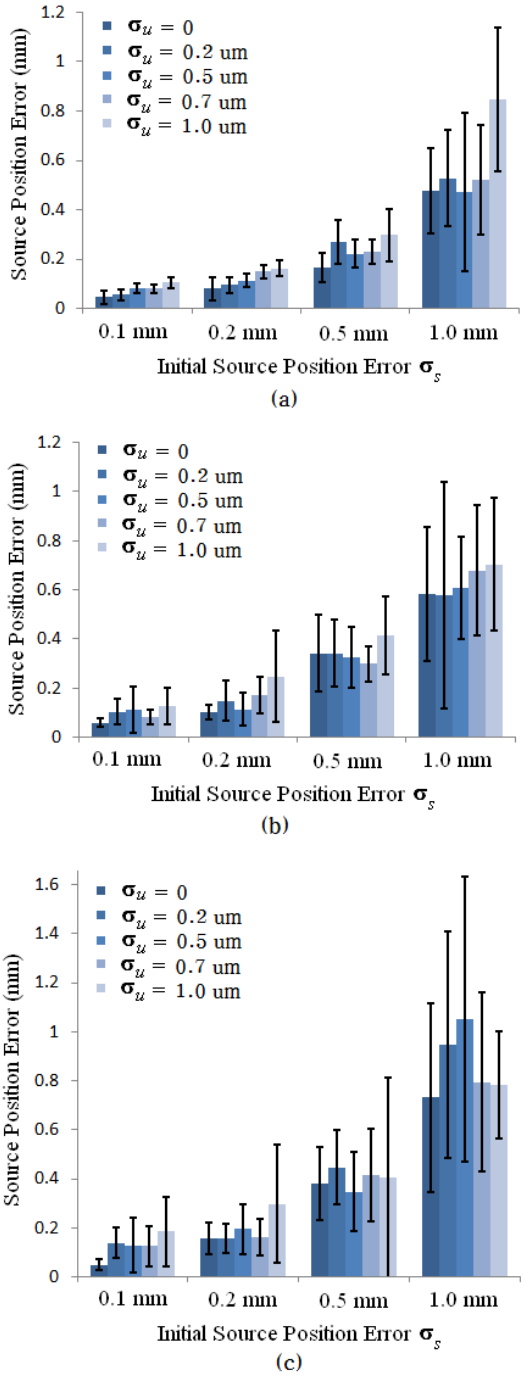


Fig. 6. Residual light source position errors after calibration for (a) 25 light sources and 11 mirror position/rotations, (b) 13 light sources and 11 mirror position/rotations, (c) 25 light sources and 5 mirror position/rotations. Both the mean and standard deviation from 10 random position errors are shown. Here, σ_u and σ_s denote errors of the imaging positions $\mathbf{u}_i$ and initial light source position $\mathbf{s}_i$, respectively.

5 in Figure 6(c), the error reduction becomes less effective. The number of mirrors is more critical to the calibration accuracy than the number of light sources.

V. Conclusion

We proposed a new calibration algorithm to estimate better light source positions from measured data, which utilizes glint images reflected from a mirror at several positions and rotations. With 25 light source positions and 11 mirror positions the light source position errors are reduced to about a half. Although the calibration process is simple, the error reduction from practical measurement errors results in 0.1 mm accuracy which is good enough to obtain 1 degree gaze angular error. Therefore, it becomes possible to use commercial mobile devices and customized devices as robust eye-gaze estimation systems for the understanding of human mind.

References

[1] A. T. Duchowski, *Eye Tracking Methodology: Theory and Practice*, 2nd ed. New York, NY, USA: Springer-Verlag, 2007.

[2] K. M. Dalton, B. M. Nacewicz, T. Johnstone, H. S. Schaefer, M. A. Gernsbacher, H. H. Goldsmith, A. L. Alexander, and R. J. Davidson, "Gaze fixation and the neural circuitry of face processing in autism," *Nature Neurosci.*, vol. 8, no. 4, pp. 519–526, 2005.

[3] Z. Boraston and S. J. Blakemore, "The application of eye-tracking technology in the study of autism," *J. Physiol.*, vol. 581, no. 3, pp. 893–898, Jun. 2007

[4] C. Karatekin, "Eye tracking studies of normative and atypical development," *Develop. Rev.*, vol. 27, no. 3, pp. 283–348, 2007.

[5] G. Gredebäck, S. Johnson, and C. von Hofsten, "Eye tracking in infancy research," *Develop. Neuropsychol.*, vol. 35, no. 1, pp. 1–19, 2009.

[6] M. L. Spezio, P. S. Huang, F. Castelli, and R. Adolphs, "Amygdala damage impairs eye contact during conversations with real people," *J. Neurosci.*, vol. 27, no. 15, pp. 3994–3997, Apr. 2007

[7] A. T. Duchowski, "A breadth-first survey of eye tracking applications," *Behavior Res. Methods, Instrum., Comput.*, vol. 34, no. 4, pp. 455–470, 2002.

[8] D. W. Hansen and Q. Ji, "In the eye of the beholder: A survey of models for eyes and gaze," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 478–500, Mar. 2010.

[9] C. H. Morimoto and M. R. M. Mimica, "Eye gaze tracking techniques for interactive applications," *Comput. Vis. Image Understand.*, vol. 98, no. 1, pp. 4–24, Apr. 2005.

[10] T. Nagamatsu, R. Sugano, Y. Iwamoto, J. Kamahara, and N. Tanaka, "User-calibration-free gaze estimation method using a binocular 3D eye model," *IEICE Trans. Inf. Syst.*, vol. E94-D, no. 9, pp. 1817–1829, Sep. 2011.

[11] D. Model and M. Eizenman, "User-calibration-free remote gaze estimation system," in *Proc. Symp. Eye-Tracking Res. Appl.*, Mar. 2010, pp. 29–36.

[12] D. H. Yoo and M. J. Chung, "A novel non-intrusive eye gaze estimation using cross-ratio under large head motion," *Comput. Vis. Image Understand.*, vol. 98, no. 1, pp. 25–51, Apr. 2005.

[13] A. Villanueva, R. Cabeza, and S. Porta, "Eye tracking: Pupil orientation geometrical modeling," *Image Vis. Comput.*, vol. 24, no. 7, pp. 663–679, Jul. 2006.

[14] A. Villanueva, R. Cabeza, and S. Porta, "Gaze tracking system model based on physical parameter," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 21, no. 5, pp. 855–977, 2007.

[15] K. P. White, Jr., T. E. Hutchinson, and J. M. Carley, "Spatially dynamic calibration of an eye-tracking system," *IEEE Trans. Syst., Man, Cybern.*, vol. 23, no. 4, pp. 1162–1168, Jul. 1993.

[16] 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.

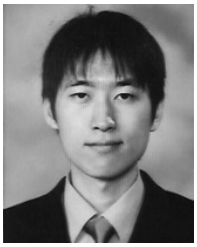
[17] A. Villanueva and R. Cabeza, "A novel gaze estimation system calibration point," *IEEE Trans. Syst., Man, Cybern.*, vol. 38, no. 4, pp. 1123–1138, Aug. 2008.

[18] J. Sun, C. Yang, J. Liu, and X. Yang, "Gaze tracking based on similarity between spatial triangles and two-stage calibration," *Electron. Lett.*, vol. 47, no. 4, pp. 254–255, Feb. 2011.

[19] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, 2nd ed. Cambridge, U.K.: Cambridge University Press, 2007.

Hawarn Lee received the B.S. degree in mathematical science from the Korea Advanced Institute of Science and Technology, Deajeon, Korea, in 2012, where she is currently pursuing the M.S. degree in electrical engineering. Her current research interests include computational neuroscience, machine learning, and human-computer interface.

Nadeem Iqbal received the M.S. degree in computer system engineering from the Ghulam Ishaq Khan Institute, Topi, Pakistan, in 2003. He is currently pursuing the Ph.D. degree in bio- and brian engineering at the Korea Advanced Institute of Science and Technology, Daejeon, Korea. His current research interests include image processing, eye gaze tracking, image-based eye analysis, supervised and unsupervised machine learning methods, and computer vision.



Wonil Chang was born in Seoul, Korea, in 1981. He received the B.S. degree in electrical engineering and the Ph.D. degree in bio- and brian engineering from the Korea Advanced Institute of Science and Technology, Daejeon, Korea, in 2003 and 2012, respectively. He was a Visiting Scholar with the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2004, and Pennsylvania State University, University Park, PA, USA, in 2006 and 2007. His current research interests include audio and image signal processing, pattern recognition, human-computer interface, and artificial neural networks.

Soo-Young Lee received the B.S. and Ph.D. degrees from Seoul National University, Seoul, Korea, and the Polytechnic Institute of New York, Brooklyn, NY, USA, in 1975 and 1984, respectively. He was with the Korea Advanced Institute of Science and Technology, Daejeon, Korea. He is a Full Professor with the Department of Electrical Engineering and Computer Science. He received the INNS Leadership Award and the INNS Presidential Award in 1994 and 2001, and joined the INNS Governing Board in 2012. He was a President of the Asia-Pacific Neural Network Assembly (APNNA) and received the APPNA Service Award in 2004. He is the Editor-in-Chief of *Natural Intelligence – The INNS Magazine* and is on Editorial Boards of several other journals. In 1997, he established the Brain Science Research Center, and from 1998 to 2008, he served as the Director and Principal Investigator of the Brain Neuroinformatics Research Program, which is the first and truly interdisciplinary research program in Korea for the brain-inspired intelligent systems with perception, learning, inference, and human-like behavior.

His current research interests include artificial brains, i.e., human-like intelligent systems based on biological information processing mechanisms. He has worked on auditory models, information-theoretic processing, proactive knowledge development, and top-down selective attention. He is interested in combining cognitive neuroscience and information theory for artificial cognitive systems.