Optical Engineering

SPIEDigitalLibrary.org/oe

Improved remote gaze estimation using corneal reflection-adaptive geometric transforms

Chunfei Ma
Seung-Jin Baek
Kang-A Choi
Sung-Jea Ko

Optical Engineering 53(5), 053112 (May 2014)

Improved remote gaze estimation using corneal reflection-adaptive geometric transforms

Chunfei Ma,[a] Seung-Jin Baek,[b] Kang-A Choi,[a] and Sung-Jea Ko[a,*]
[a]Korea University, Department of Electrical Engineering, Seoul 136-713, Republic of Korea
[b]DMC R&D Center, Samsung Electronics Co., Ltd., Suwon 443-742, Republic of Korea

Abstract. Recently, the remote gaze estimation (RGE) technique has been widely applied to consumer devices as a more natural interface. In general, the conventional RGE method estimates a user's point of gaze using a geometric transform, which represents the relationship between several infrared (IR) light sources and their corresponding corneal reflections (CRs) in the eye image. Among various methods, the homography normalization (HN) method achieves state-of-the-art performance. However, the geometric transform of the HN method requiring four CRs is infeasible for the case when fewer than four CRs are available. To solve this problem, this paper proposes a new RGE method based on three alternative geometric transforms, which are adaptive to the number of CRs. Unlike the HN method, the proposed method not only can operate with two or three CRs, but can also provide superior accuracy. To further enhance the performance, an effective error correction method is also proposed. By combining the introduced transforms with the error-correction method, the proposed method not only provides high accuracy and robustness for gaze estimation, but also allows for a more flexible system setup with a different number of IR light sources. Experimental results demonstrate the effectiveness of the proposed method. [DOI: 10.1117/1.OE.53.5.053112]

Keywords: remote gaze estimation; geometric transforms; error correction; flexibility.

Paper 140422 received Mar. 13, 2014; revised manuscript received Apr. 17, 2014; accepted for publication Apr. 22, 2014; published online May 14, 2014.

1 Introduction

Remote gaze estimation (RGE) is a technique to measure a user's point of gaze (POG) on a two dimensional (2-D) surface, such as a monitor screen, without the use of any wearable or head-mounted equipment. RGE methods have been popularly researched and adopted in various applications, such as human-computer interaction (HCI) and virtual reality, due to their nonintrusiveness and comfortableness.[1–9] In particular, these RGE methods have been recently applied to consumer electronics products, such as smart mobile devices,[4,6] Internet protocol television,[5] and infotainment devices,[8] for more friendly interactive application.

Generally, the RGE method can be divided into two classes: (1) 2-D interpolation-based method and (2) 3-D model-based method.[1] The 2-D interpolation-based method estimates the POG by using the calculated mapping function, such as homography or polynomial regression equation. The 3-D model-based method estimates the 3-D gaze direction and intersects it directly into the scene geometry to calculate the POG. Compared to the 2-D interpolation-based method, the 3-D model-based method can provide more freedom of head movement. However, the system setups, such as the camera parameters and the distance between monitor and user, should be fully calibrated for the 3-D model-based method.[10,11] In terms of practical application, we are primarily interested in the remote eye gaze estimation method in uncalibrated setups.[1] In this paper, therefore, our focus is on the 2-D interpolation-based method.

The pupil center corneal reflection (PCCR) method has been known as the most popular 2-D interpolation-based RGE method due to its fairly good accuracy.[1,3,5] This method generally uses the vector between the pupil center (PC) and one or two corneal reflections (CRs) generated on the corneal surface along with a mapping function to estimate the POG. However, the performance of this method can be easily deteriorated by the user's head movement.[12,13] To solve this problem, Yoo et al. presented the crossratio-based method, which achieves improved performance using a single mapping function.[14] However, this method not only requires inconvenient geometric calibration of the environment[15] but also can result in large gaze estimation error, depending on the subject.[16] Another method, called homography normalization (HN)[17,18] was proposed not only to avoid the geometric calibration but also to compensate for the error of the crossratio-based method. Instead of using a single mapping function, this method employs two homography transformations (HTs) to estimate the POG. Specifically, the HN method adopts a normalization procedure, which first maps the PC in the eye image into a virtual plane using an HT prior to the estimation of the POG on the screen. Compared to conventional methods, the HN method provides more accurate and robust gaze estimation performance in uncalibrated setups. However, the aforementioned two methods still require four or more CRs, which make these methods infeasible in the cases where only two or three CRs are available. In this paper, we propose a generalized method that not only can operate with two or three CRs, but can also provide accuracy and robustness that are superior to the HN method. Instead of the first HT used in the HN method, we propose to utilize a similarity transformation (ST) and an affine transformation (AT), respectively, for two- and three-CR-based gaze estimation scenarios. Furthermore,

*Address all correspondence to: Sung-Jea Ko, E-mail: sjko@korea.ac.kr

0091-3286/2014/$25.00

to enhance the performance of gaze estimation, a polynomial function is employed instead of the second HT of the HN method. Experimental results over six subjects demonstrate the effectiveness of the proposed methods under both head fixed and natural head movement cases.

The remainder of this paper is organized as follows. Section 2 describes the preliminaries of the HN method, which are used in the proposed methods. Section 3 presents the details of the proposed methods. The experimental results are presented in Sec. 4, and Sec. 5 concludes this paper.

2 Preliminaries

2.1 *Geometric Model*

Figure 1 illustrates the basic geometric model adopted for both the HN and our proposed methods. This model comprises three planes: the monitor screen plane $\mathbf{\Pi}_S$ with four infrared (IR) light sources ($\boldsymbol{L}_1 - \boldsymbol{L}_4$) mounted on its corners, the corneal plane $\mathbf{\Pi}_C$ formed by four CRs ($\boldsymbol{G}_1 - \boldsymbol{G}_4$), and the camera image plane $\mathbf{\Pi}_I$. With this model, the input eye image for gaze estimation is obtained as follows. First, four IR light sources are projected (projection 1) onto the surface of the cornea, in which center is modeled as the curvature center $\boldsymbol{C}$. The $\boldsymbol{G}_1 - \boldsymbol{G}_4$ and the PC ($\boldsymbol{P}$) of $\mathbf{\Pi}_C$ are then projected (projection 2) onto $\mathbf{\Pi}_I$ by the camera, thereby forming the CRs ($\boldsymbol{g}_1 - \boldsymbol{g}_4$) and PC ($\boldsymbol{p}$) of the image.

2.2 *Homography Normalization Method*

In the HN method, the POG is estimated by using two mapping functions; $\mathbf{\Pi}_I$ -to-$\mathbf{\Pi}_C$ and $\mathbf{\Pi}_C$-to-$\mathbf{\Pi}_S$ mappings as shown in Fig. 1. To accurately estimate these mapping functions, theoretically, 3-D locations of the CRs on the corneal plane should be known. However, under the uncalibrated system setup, actual 3-D locations of the CRs cannot be obtained. Thus, to solve this problem, the HN method assumes the corneal plane $\mathbf{\Pi}_C$ to be a normalized plane, $\mathbf{\Pi}_N$, which is a unit square with the predefined coordinates (Fig. 2). Based on the assumption that the PC and CRs are coplanar, the mappings from $\mathbf{\Pi}_I$ to $\mathbf{\Pi}_N$ and from $\mathbf{\Pi}_N$ to $\mathbf{\Pi}_S$ are established through this normalized plane.

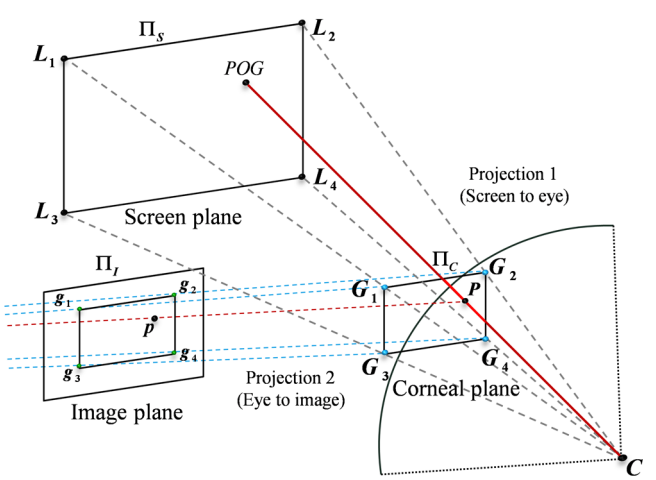

Fig. 1 Geometric model of the pupil center corneal reflection (PCCR) method for remote eye gaze estimation.

The detailed POG estimation process of the HN method is as follows. First, the detected PC of the image plane, $\boldsymbol{p}_I$, is mapped into $\mathbf{\Pi}_N$ by the homography $\boldsymbol{H}_I^N$, which is estimated by four pairs of point correspondences between the detected CRs $\boldsymbol{g}_1 - \boldsymbol{g}_4$ and the corners of $\mathbf{\Pi}_N$, $\boldsymbol{G}_1 - \boldsymbol{G}_4$. The PC of $\mathbf{\Pi}_N$, $\boldsymbol{p}_N$, is then mapped into $\mathbf{\Pi}_S$ to obtain the final POG using another homography ($\boldsymbol{H}_N^S$) estimated via the calibration procedure, which will be illustrated in the following section. The whole mapping $\boldsymbol{H}_I^S$, from $\mathbf{\Pi}_I$ to $\mathbf{\Pi}_S$, is obtained by concatenating $\boldsymbol{H}_I^N$ and $\boldsymbol{H}_N^S$. As reported by Hansen et al.[18] and Coutinho et al.,[15] $\boldsymbol{H}_I^N$ and $\boldsymbol{H}_N^S$ can, respectively, normalize the head movement before estimating the POG and compensate for the angular deviation between the visual and optical axes. Therefore, this method is more robust and less sensitive to head pose changes than other PCCR methods.

2.3 *Analysis of the Homography Normalization Method*

To generalize the normalization procedure of the HN method, we first explain how the head movement can be compensated for by mapping the detected PC into the normalized plane. Assume that a person gazes at a specific point on the screen while the head moves naturally within a certain volume in 3-D space. Because the spatial position of the center of the eyeball varies under head movement, the person has to rotate the eyeball to keep gazing at the same point. Therefore, all the absolute positions of the PC and CRs are changed. In this case, the traditional PCCR methods are based on the assumption that the vectors between the PC and the CRs remain unchanged under the head movement. Meanwhile, the HN method utilizes the property of geometric invariance under the HT. Essentially, the normalization procedure of the HN method assumes that the relative position relationship, between the PCs and its corresponding CRs in $\mathbf{\Pi}_I$ under the head movement, remains invariant as long as the person keeps gazing at the same point on the screen plane $\mathbf{\Pi}_S$. Based on this assumption, the PCs under different head poses can be mapped to the same point in the normalized plane $\mathbf{\Pi}_N$. Thus, the one-to-one correspondence

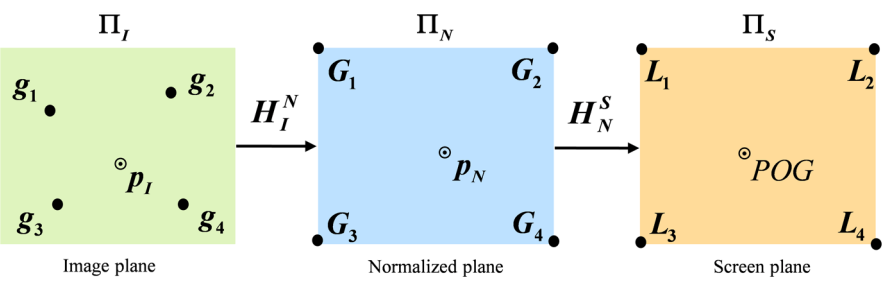

Fig. 2 Illustration of the homography normalization (HN) method.

between $\boldsymbol{p}_N$ and POG can be established successfully without the effect of the head movement.

3 Proposed Method

The motivation of our proposed method is as follows: first, as shown in Figs. 3(a) and 3(b), unconstrained head and eye movement can lead to the loss and distortion of the CRs.[19,20] In this case, the HN method cannot be utilized due to the lack of enough CRs for normalization. Second, as shown in Figs. 3(c) and 3(d), the HN method is inherently infeasible for systems with only two or three IR light sources. To solve these problems, we propose a generalized normalization-based gaze estimation method that allows for a more flexible system setup with a different number of IR light sources.

Figure 4 shows the proposed gaze estimation method, in which inputs are the detected PC and CRs:

(1) Normalization: the proposed normalization procedure compensates for the head movement by using ST, AT, or HT according to the number of available CRs. Thus, the proposed method can be viewed as a generalization of the conventional four-CR-based HN methods. Without loss of generality, the gaze estimation methods that use HT, AT, and ST for normalization are, respectively, called HTN-, ATN-, and STN-based methods.

Figure 5 illustrates the principle of the proposed generalized normalization-based method. First, for the case where four CRs are available [see Fig. 5(a)], just like the original HN method, $\boldsymbol{p}_I^4$ of $\boldsymbol{\Pi}_I^4$ can be mapped into $\boldsymbol{\Pi}_N^4$ using the estimated HT $\boldsymbol{H}_I^N$ to obtain $\boldsymbol{p}_N^4$. Second, Figs. 5(b) and 5(c) show that, in the ATN- and STN-based methods, the AT $\boldsymbol{A}_I^N$ and ST $\boldsymbol{S}_I^N$, respectively, are estimated using three pairs of point correspondences between the detected CR $\boldsymbol{g}_1 - \boldsymbol{g}_3$ of $\boldsymbol{\Pi}_I^3$ and its corresponding predefined corner points $\boldsymbol{G}_1 - \boldsymbol{G}_3$ of $\boldsymbol{\Pi}_N^3$, and two pairs of point correspondences between $\boldsymbol{g}_1 - \boldsymbol{g}_2$ of $\boldsymbol{\Pi}_I^2$ and its corresponding $\boldsymbol{G}_1 - \boldsymbol{G}_2$ of $\boldsymbol{\Pi}_N^2$. In the AN-based method, $\boldsymbol{\Pi}_N^3$ is an isosceles right triangle having two perpendicular sides of unit lengths, which have three corner points $\boldsymbol{G}_1 - \boldsymbol{G}_2$. In the SN-based method, $\boldsymbol{\Pi}_N^2$ is a unit line segment with only two corner points $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$. Finally, to obtain $\boldsymbol{p}_N^3$ and $\boldsymbol{p}_N^2$, $\boldsymbol{p}_I^3$ of $\boldsymbol{\Pi}_I^3$ and $\boldsymbol{p}_I^2$ of $\boldsymbol{\Pi}_I^2$, respectively, are mapped into its corresponding $\boldsymbol{\Pi}_N^3$ and $\boldsymbol{\Pi}_N^2$ by using the estimated AT $\boldsymbol{A}_I^N$ and ST $\boldsymbol{S}_I^N$.

Thus, the aforementioned normalization method on a case-by-case basis leads to a generalized normalization method that can relax the restriction requiring four CRs for normalization. Moreover, the proposed method can speed up the computation with a reduced number of CRs. This is because only two or three IR light sources are, respectively, theoretically required to estimate the ST with four degrees of freedoms (DOFs) or the AT with six DOFs, both of which have lower DOFs than the HT with its eight DOFs.

(2) Calibration: For calibration, a person is requested to look at several predefined target points (nine points for methods) sequentially appeared from top to bottom and left to right on the screen plane $\boldsymbol{\Pi}_S$ (red points). Then, the coordinates of the normalized PCs (nine gray points in $\boldsymbol{\Pi}_N^4$, $\boldsymbol{\Pi}_N^3$ and $\boldsymbol{\Pi}_N^2$) and the corresponding target points are used to estimate the correction functions ($\boldsymbol{P}_N^S$), which relate $\boldsymbol{\Pi}_N^4$, $\boldsymbol{\Pi}_N^3$ and $\boldsymbol{\Pi}_N^2$ with $\boldsymbol{\Pi}_S$. Note that the calibration only needs to be done once.

(3) Correction: Coutinho et al.[15] pointed out that the gaze estimation error of the PCCR methods is caused by two simple assumptions: first, the optical visual axes

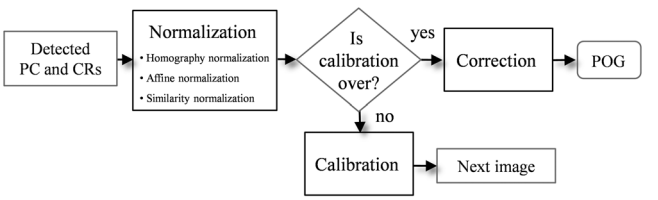

Fig. 4 Flowchart of the proposed method.

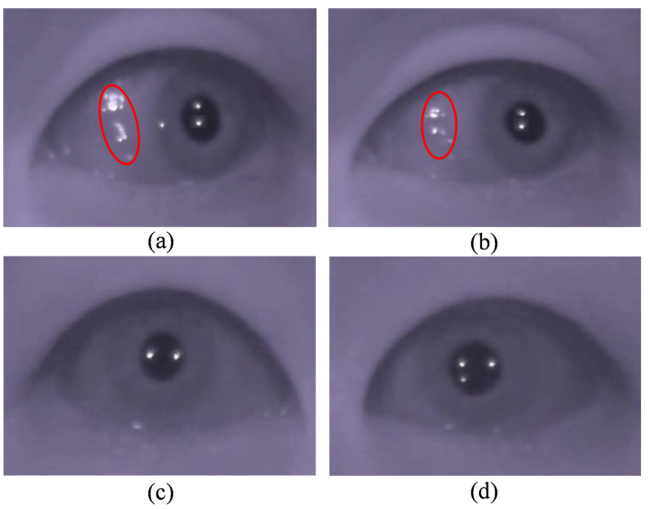

Fig. 3 Example of the infrared (IR) eye images. Red ellipse indicates that (a) one or (b) two corneal reflections (CRs) are distorted or lost. Images captured from the gaze estimation system where only (c) two or (d) three IR light sources are available.

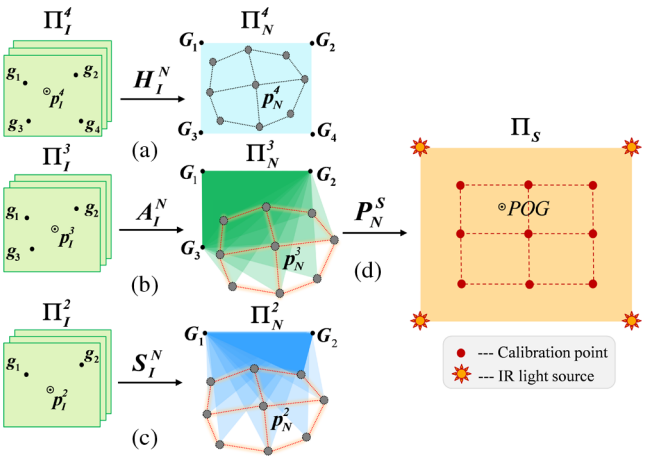

Fig. 5 Illustration of the proposed method. It includes three normalization methods: (a) homography transformation normalization (HTN), (b) affine transformation normalization (ATN), (c) similarity transformation normalization (STN), and (d) the correction method.

are identical. Second, the PC and CRs are coplanar. The first assumption results in an estimation error primarily because the visual axis (the actual axis of sight) deviates from the optical axis by a certain angle, called the kappa angle.[1,15] The second assumption adopted by the HN method ignores the fact that the actual PC is located on a different plane, away from the corneal plane by a certain distance.

To solve these problems, our proposed method uses a polynomial regression function $\boldsymbol{P}_N^S$ to estimate the POG [see Fig. 5(d)] instead of the homography $\boldsymbol{H}_n^s$ adopted by the original HN method. The polynomial function is more effective for two reasons. First, as illustrated in Ref. 15, the angular deviation between the visual and optical axes varies with the rotation of the eye and presents a strong property of nonlinearity. Consequently, it can result in the nonlinear gaze estimation error that cannot be compensated for by a linear mapping function, such as the homography. Therefore, it is expected that a nonlinear polynomial function $\boldsymbol{P}_N^S$ can compensate for the error more effectively. Second, the mapping between the actual PC located behind the real corneal surface (3-D spherical surface) and the POG on the 2-D screen plane is essentially a nonlinear mapping rather than a linear mapping. Hence, as suggested by Hansen et al.,[18] the nonlinear polynomial function $\boldsymbol{P}_N^S$ can be used to compensate for the errors caused by the noncoplanarity of PC and CRs, better than the original linear function $\boldsymbol{H}_N^S$.

Although a Gaussian processes (GP)-based interpolation method described in Ref. 18 can also be adopted to further compensate for the aforementioned errors after the proposed correction procedure, several problems of this method still remain to be solved. First, the estimation of the parameters of the squared-exponential kernel used in the GP-based method involves an inconvenient empirical Bayes optimization process.[21] Second, the performance of this GP-based method is highly dependent on the number of the calibration point.[18] Third, this method has less robustness to the depth change as compared with the original HN method.[18] Therefore, as the way adopted by Coutinho et al. in Ref. 15, in this paper, both the original HN method and the proposed methods are compared without using the GP-based error modeling procedure.[15]

The second-order polynomial function adopted in our proposed methods is defined as follows:

$$X_i = a_0 + a_1 x_i + a_2 y_i + a_3 x_i y_i + a_4 x_i^2 + a_5 y_i^2,$$
$$Y_i = b_0 + b_1 x_i + b_2 y_i + b_3 x_i y_i + b_4 x_i^2 + b_5 y_i^2, \quad (1)$$

where (x_i, y_i) and (X_i, Y_i) represent the horizontal and vertical coordinates of points in $\boldsymbol{\Pi}_N$ and $\boldsymbol{\Pi}_S$, respectively, and $a_0 - a_5$ and $b_0 - b_5$ are the coefficients to be estimated through calibration. Note that, in this paper, all the mapping functions including $\boldsymbol{S}_I^N, \boldsymbol{A}_I^N, \boldsymbol{H}_I^N, \boldsymbol{H}_N^S$ and $\boldsymbol{P}_N^S$ are estimated by solving the linear least-squares problems based on the point correspondences.[22]

4 Experimental Results

4.1 *Experimental Methodology*

We used a 24-in screen with a resolution of 1920 × 1080 pixels. Four IR light sources were placed in the corners of the monitor (for STN- and ATN-based methods, only two and three of them were used). Input images with a resolution of 640 × 480 pixels were obtained from a web camera (from which the IR cutoff filter was removed), to which a 4.5× zooming lens was attached, placed at the center region slightly under the monitor screen. To evaluate the proposed generalized normalization-based methods, six subjects (three females and three males) with ages ranging from 25 to 35 years participated in the experiments. From the six subjects, three of them (subjects 1, 5, and 6) make daily use of corrective lenses but all of them were able to see the target screen points used during calibration and test with the naked eye. The subjects were requested to seat themselves in front of a monitor screen at a distance of approximately 70 cm. Fig. 6 shows the experimental setup.

During the experiment, each subject was requested to perform one calibration procedure (nine calibration points) and the following two trial cases, as adopted by Hansen et al. in Ref. 18. In one trial case, the subjects kept their heads still (as much as possible without head fixtures) and in the other case the subjects moved their heads naturally in a space with an approximate volume of 20 × 20 × 20 cm (width × height × depth) while gazing at the test target points on the screen.

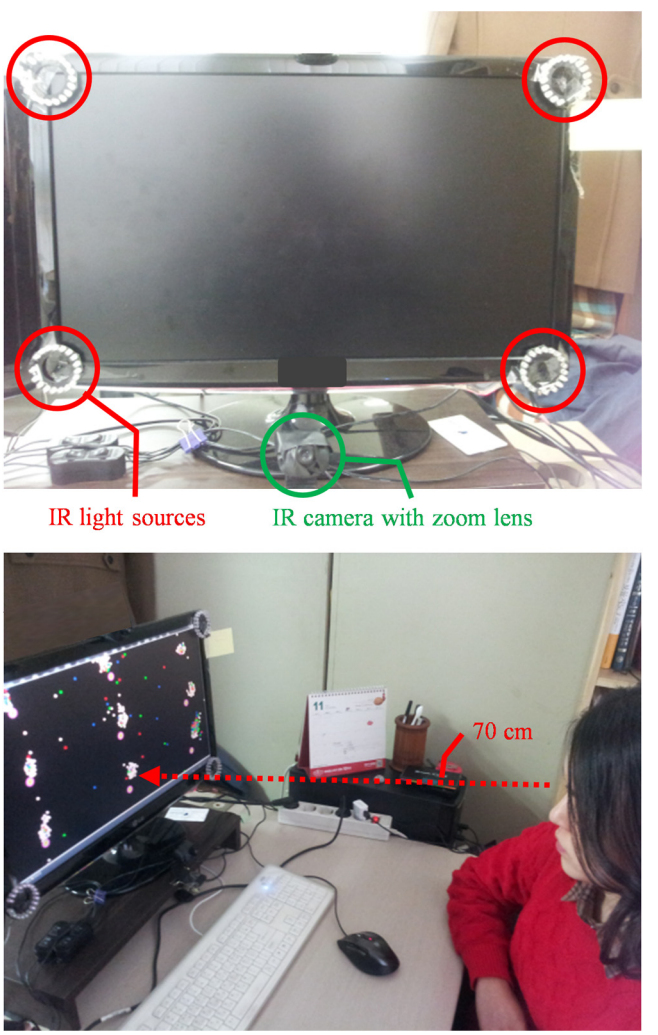

Fig. 6 Experimental setup.

To evaluate the performance of each method quantitatively, the average angular error and standard deviation, as adopted in Refs. 15 and 16, were measured in degrees over several uniformly distributed test points on the screen. For each test point, the angular error μ and the standard deviation σ over N frames are respectively defined as

$$\mu = \tan^{-1}(d_M/d_U), \tag{2}$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mu_i - m)^2}, \quad \text{where } m = \frac{1}{N}\sum_{i=1}^{N}\mu_i, \tag{3}$$

and where d_M indicates the Euclidean distance between the ground truth and the estimated gaze positions on the screen, and d_U denotes the Euclidean distance between the subject and the screen. μ_i and m, respectively, represent the angular error corresponding to i'th frame and the mean error over N frames. In our experiments, nine test points were used and the averaged results over them were reported for both two trials ($N = 60$ and $N = 150$ per test point). Note that under our experimental setup, the angular error of 1–2 degree corresponds to 12–24 millimeters or 43–86 pixels on the screen. Figure 7 shows the target screen points used for both the calibration and test.

The experiment was designed to provide a comparison of: (1) effectiveness: the performance comparison between the proposed methods and the original HN method (Ori_HN) without using the Gaussian processes-based error modeling and (2) substitutability: the performance comparison among three proposed methods, which have different geometric transforms but the same correction functions. For fair comparison, the HTN-, ATN-, and STN-based methods employed the same polynomial function Eq. (1) for correction. The PC and CRs of the eye image were detected using the thresholding and blob detection methods.[23]

4.2 *Performance Evaluation*

Tables 1 and 2 list the experimental results of the proposed generalized normalization-based methods. Note that, for both head fixed and movement cases, all the STN-, ATN- and HTN-based methods achieved higher accuracy than the Ori_HN method. This result indicates that the proposed polynomial correction method is more effective than the original HT. It is also seen that the STN- and ATN-based methods exhibit accuracy and robustness similar to the HTN-based one, which is strong evidence that validates the substitutability of the normalization transformation. Although for the head movement case, a slight decrease can be found as compared with head fixed case, the accuracy of all the methods is fairly high and acceptable for most of the HCI applications. In addition, we confirm that the normalization can not only greatly remove the effects of head movement in the $x - y$ plane (parallel to the screen plane) but also cope with scale changes of the vectors formed by PC and CRs due to head movement in depth (orthogonal to the screen plane) to some extent. This result is accordance with the observations made by Hansen et al. in Ref. 18. The volume of space allowed for head movement is large enough for each subject to interact with the computer comfortably and naturally. On a desktop computer with a 2.8-GHz quad-core CPU and 4-GB RAM, the processing times of the HTN-, ATN-, and STN-based methods are 27, 25, and 23 ms per frame, respectively, which all exhibit the satisfactory real-time property.

Table 1 Fixed head case (unit: degree).

	Ori_HN	HTN	ATN	STN
Subject 1	0.93 ± 0.48	0.68 ± 0.36	0.70 ± 0.40	0.73 ± 0.43
Subject 2	0.81 ± 0.50	0.72 ± 0.27	0.74 ± 0.31	0.77 ± 0.36
Subject 3	0.95 ± 0.47	0.73 ± 0.40	0.83 ± 0.41	0.90 ± 0.44
Subject 4	0.84 ± 0.50	0.76 ± 0.41	0.78 ± 0.43	0.79 ± 0.47
Subject 5	0.91 ± 0.52	0.75 ± 0.39	0.88 ± 0.44	0.89 ± 0.45
Subject 6	1.11 ± 0.51	0.70 ± 0.25	0.81 ± 0.34	0.90 ± 0.37
Average	0.93 ± 0.50	0.72 ± 0.35	0.79 ± 0.39	0.83 ± 0.42

Table 2 Natural head movement case (unit: degree).

	Ori_HN	HTN	ATN	STN
Subject 1	1.36 ± 0.63	1.13 ± 0.58	1.30 ± 0.59	1.33 ± 0.61
Subject 2	1.29 ± 0.66	1.12 ± 0.59	1.15 ± 0.63	1.23 ± 0.63
Subject 3	1.30 ± 0.64	0.95 ± 0.54	1.01 ± 0.58	1.18 ± 0.64
Subject 4	1.35 ± 0.59	1.29 ± 0.44	1.31 ± 0.49	1.35 ± 0.50
Subject 5	1.55 ± 0.77	1.43 ± 0.54	1.52 ± 0.58	1.54 ± 0.60
Subject 6	1.57 ± 0.69	1.31 ± 0.56	1.43 ± 0.59	1.50 ± 0.65
Average	1.40 ± 0.66	1.23 ± 0.54	1.29 ± 0.58	1.36 ± 0.61

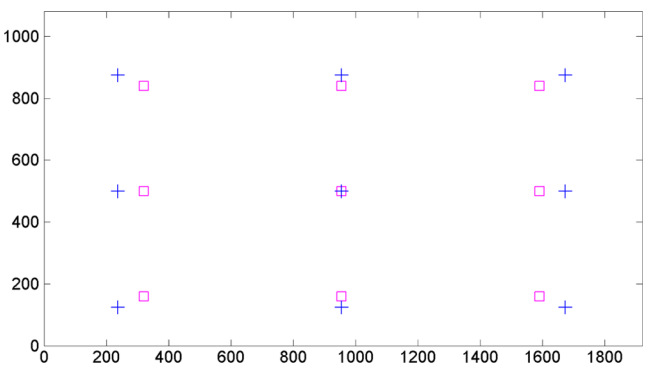

Fig. 7 Target points used for calibration (pink square) and test (blue cross).

5 Conclusion

In this paper, we proposed a novel RGE method based on the alternative geometric transforms, which are adaptive to the number of the CRs. The experimental results indicate that the proposed method can achieve better performance as

compared with the Ori_HN method. Specifically, the significance of the proposed method is multifold. First, in terms of accuracy, the proposed method outperforms the Ori_HN method both under head fixed and natural head movement cases. Second, in terms of robustness, the proposed method is robust to head movement due to the effective normalization procedures adopted; for the cases when one or two CRs in the eye image are distorted or lost, due to either the unconstrained eye and head movement or the erroneous CR image detection, the Ori_HN method is inherently infeasible while the proposed method remains valid. Third, in terms of usability, to the RGE system where only two or three IR light sources are used, the proposed method can be alternatively employed to achieve favorable performance. Our ongoing work is to combine the proposed method with the robust CR pattern detection method[20] to further increase both the accuracy and reliability of the RGE system by handling the distortion and loss of the CR image. Moreover, since the proposed method requires the uncalibrated system setup,[18] the flexibility of the system is fairly high while the cost of the system setup is very light. Therefore, by using the proposed method, it is expected that not only the accuracy and robustness but also the usability and reliability of the RGE system can be further improved.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (Grant No. 2012R1A2A4A01008384). This research was funded and supported by Samsung Electronics Co., Ltd.

References

1. 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.* **32**(3), 478–500 (2010).
2. T. E. Hutchinson et al., "Human-computer interaction using eye-gaze input," *IEEE Trans. Syst. Man. Cybern.* **19**(6), 1527–1534 (1989).
3. P. M. Corcoran et al., "Real-time eye gaze tracking for gaming design and consumer electronics systems," *IEEE Trans. Consumer Electron.* **58**(2), 347–355 (2012).
4. E. C. Lee, Y. J. Ko, and K. R. Park, "Gaze tracking based on active appearance model and multiple support vector regression on mobile devices," *Opt. Eng.* **48**(7), 077002 (2009).
5. H. C. Lee et al., "Gaze tracking system at a distance for controlling IPTV," *IEEE Trans. Consumer Electron.* **56**(4), 2577–2583 (2010).
6. N. Iqbal, H. Lee, and S. Y. Lee, "Smart user interface for mobile consumer devices using model-based eye-gaze estimation," *IEEE Trans. Consumer Electron.* **59**(1), 161–166 (2013).
7. W. Zhang, T. N. Zhang, and S. J. Chang, "Eye gaze estimation from the elliptical features of one iris," *Opt. Eng.* **50**(4), 047003 (2011).
8. T. Nawaz, M. S. Mian, and H. A. Habib, "Infotainment devices control by eye gaze and gesture recognition fusion," *IEEE Trans. Consumer Electron.* **54**(2), 277–282 (2008).
9. C. W. Cho et al., "Robust gaze-tracking method by using frontal-viewing and eye-tracking cameras," *Opt. Eng.* **48**(12), 127202 (2009).
10. S. W. Shih and J. Liu, "A novel approach to 3D gaze tracking using stereo cameras," *IEEE Trans. Syst. Man Cybern.* **34**(1), 234–245 (2004).
11. E. D. Guestrin and M. Eizenman, "General theory of remote gaze estimation using the pupil center and corneal reflections," *IEEE Trans. Biomed. Eng.* **53**(6), 1124–1133 (2006).
12. Z. Zhu and Q. Ji, "Novel eye gaze tracking techniques under natural head movement," *IEEE Trans. Biomed. Eng.* **54**(12), 2246–2260 (2007).
13. L. Sesma-Sanchez, A. Villanueva, and R. Cabeza, "Gaze estimation interpolation methods based on binocular data," *IEEE Trans. Biomed. Eng.* **59**(8), 2235–2243 (2012).
14. D. H. Yoo and M. J. Chung, "A novel non-intrusive eye gaze estimation using cross-ratio under large head motion," *Comput. Vis. Image Und.* **98**(1), 25–51 (2005).
15. F. L. Coutinho and C. H. Morimoto, "Improving head movement tolerance of cross-ratio based eye trackers," *Int. J. Comput. Vis.* **101**(3), 459–481 (2013).
16. J. J. Kang et al., "Investigation of the cross-ratio method for point-of-gaze estimation," *IEEE Trans. Biomed. Eng.* **55**(9), 2293–2302 (2008).
17. J. S. Agustin, "Off-the-shelf gaze interaction," Ph.D. Dissertation, IT Univ., Copenhagen, Denmark (2009).
18. D. W. Hansen, J. S. Agustin, and A. Villanueva, "Homography normalization for robust gaze estimation in uncalibrated setups," in *Proc. ACM Symp. Eye-Tracking Research and Applications*, pp. 13–20 (2010).
19. Y. J. Ko, E. C. Lee, and K. R. Park, "A robust gaze detection method by compensating for facial movements based on corneal specularities," *Pattern Recognit. Lett.* **29**(10), 1474–1485 (2008).
20. C. A. Hennessey and P. D. Lawrence, "Improving the accuracy and reliability of remote system-calibration-free eye-gaze tracking," *IEEE Trans. Biomed. Eng.* **56**(7), 1891–1900 (2009).
21. K. P. Murphy, *Machine Learning: a Probabilistic Perspective*, MIT Press, Cambridge (2012).
22. R. Hartley and A. Zisserman, Eds., *Multiple View Geometry in Computer Vision*, 2nd ed., Cambridge University Press, New York, pp. 87–126 (2003).
23. F. Chang, C.-J. Chen, and C.-J. Lu, "A linear-time component-labeling algorithm using contour tracing technique," *Comput. Vis. Image Und.* **93**(2), 206–220 (2004).

Chunfei Ma received his BS degree in automation from Southeast University in 2008 and his MS degree in flight vehicle design from Harbin Institute of Technology in 2010. He is currently pursuing his PhD degree in electrical engineering at Korea University. His research interests are in the areas of pattern recognition, computer vision, and human computer interface.

Seung-Jin Baek received his PhD degree in 2013 and his BS degree in 2007, both in electrical engineering, from Korea University. In 2013, he joined the DMC R&D Center, Samsung Electronics Co., Ltd. His research interests are in the areas of video signal processing, video compression codec, and computer vision.

Kang-A Choi received his BS degree in electrical engineering from Korea University in 2010. He is currently pursuing a PhD degree in electrical engineering at Korea University. His research interests are in the areas of video signal processing, image processing, and computer vision.

Sung-Jea Ko received his PhD degree in 1988 and his MS degree in 1986, both in electrical and computer engineering, from State University of New York at Buffalo, and his BS degree in electronic engineering at Korea University in 1980. In 1992, he joined the Department of Electronic Engineering at Korea University, where he is currently a professor. He is the vice president of the IEEE CE Society, and an IEEE fellow.