Optical Engineering

OpticalEngineering.SPIEDigitalLibrary.org

Robust remote gaze estimation method based on multiple geometric transforms

Chunfei Ma
Kang-A Choi
Byeong-Doo Choi
Sung-Jea Ko

SPIE.

Robust remote gaze estimation method based on multiple geometric transforms

Chunfei Ma,[a] Kang-A Choi,[a] Byeong-Doo Choi,[b] and Sung-Jea Ko[a,*]
[a]Korea University, Department of Electrical Engineering, Anam-Dong 5 ga, Sungbuk-Ku, Seoul 136-713, Republic of Korea
[b]Samsung Electronics Co., Ltd., DMC R&D Center, Maetan-3dong, Yeongtong-gu, Suwon, Gyeonggi-do 443-742, Republic of Korea

Abstract. The remote gaze estimation (RGE) technique has been widely used as a natural interface in consumer electronic devices for decades. Although outstanding outcomes on RGE have been recently reported in the literature, tracking gaze under large head movements is still an unsolved problem. General RGE methods estimate a user's point of gaze (POG) using a mapping function representing the relationship between several infrared light sources and their corresponding corneal reflections (CRs) in the eye image. However, the minimum number of available CRs required for a valid POG estimation cannot be satisfied in those methods because the CRs often tend to be distorted or disappeared inevitably under the unconstrained eye and head movements. To overcome this problem, a multiple-transform-based method is proposed. In the proposed method, through three different geometric transform-based normalization processes, several nonlinear mapping functions are simultaneously obtained in the calibration process and then used to estimate the POG. The geometric transforms and mapping functions can be alternatively employed according to the number of available CRs even under large head movement. Experimental results on six subjects demonstrate the effectiveness of the proposed method. [DOI: 10.1117/1.OE.54.8.083103]

Keywords: remote gaze estimation; multiple geometric transforms; large head movement; robustness; reliability.

Paper 150474 received Apr. 13, 2015; accepted for publication Jul. 1, 2015; published online Aug. 10, 2015.

头部运动

1 Introduction

The remote gaze estimation (RGE) technique has emerged as a human-computer interface (HCI) for decades owing to its nonintrusiveness and simplicity.[1] Recently, the RGE method has been applied to various consumer electronics applications, such as internet protocol television,[2,3] infotainment and game devices,[4–6] and smart mobile devices[7] for human-device interaction.[8] The most commonly used RGE method is the interpolation-based[9,10] method, which estimates the point of gaze (POG) by using a calibrated mapping function such as homography or a polynomial regression equation. Compared to the model-based method,[11,12] the interpolation-based method can achieve similar accuracy while requiring only very light system setup. This paper will mainly focus on the interpolation-based method. The pupil center corneal reflection (PCCR) method[9–22] has been known as the most popular interpolation-based RGE method. 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. The mapping function is generally obtained by a calibration procedure[9] in which the user is asked to gaze at specific screen targets. However, since the PCCR vector is a function of the scene geometry (i.e., camera, eye, and screen),[12,13] it can be very sensitive to the position variations resulting from head movements and extreme eyeball rotations. This means that the PCCR vectors under various head poses can be different even the user looks at the same POG. Consequently, the mapping function obtained through the calibration can be erroneous or invalid during the estimation. Therefore, research of the RGE method that aims to tolerate large head movement still remains to be an open issue.

To solve the head movement problem, Yoo and Chung[14] presented the cross-ratio-based method, which achieves favorable performance using a single mapping function. However, the problem is that this method requires inconvenient geometric calibration of the environment and can result in large gaze estimation error.[15] Hansen et al.[16] proposed a homography normalization (HN) method to avoid the geometric calibration while compensating for the error of the cross-ratio-based method. This method employs two concatenate homography transformations (HTs) to estimate the POG rather than a single mapping function. Coutinho and Morimoto[17] presented the displacement vector correction (CR-D)-based and planarization of CR-features-based method to compensate for the POG estimation errors resulting from two simplified assumptions.[16–18] However, the aforementioned methods still require four or more CRs for gaze estimation, which makes these methods infeasible in the cases where only two or three CRs are available. Recently, a generalized method[18] containing several geometric transforms, which can be operated with two or three CRs, was proposed. Including the HT used in the HN method, this generalized method utilizes a similarity transformation and an affine transformation, respectively, for two- and three-CR-based gaze estimation scenarios. Nevertheless, in real-world gaze estimation scenarios, the numbers of available CRs frequently change due to the erroneous CR detection caused

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

0091-3286 /2015/$25.00

by the unconstrained eyeball rotations and head movements.[19–22] Therefore, the feasibility of these methods can be problematic because these methods can only be separately used in a case-by-case way.

So far, although several methods have been proposed to improve the accuracy[10–18] and usability[2,3,8,18] of the RGE system, very few studies have been conducted on increasing the robustness and reliability of the RGE method. Specifically, the robustness and reliability mentioned here indicate that the method should have the ability to estimate the POG even when there exist various noises such as the loss and distortion of the CRs caused both by unconstrained eye and head movement and erroneous CR detection[19]. To address these problems, Li et al.[20] proposed a multiple infrared (IR) lights-based CR detection method to handle the CR distortion and loss problems by utilizing the redundant detectable CRs. However, this method requires nine light sources placed in a regular pattern, which increases both the system and computational complexity. Hansen et al.[21] recently also proposed a robust CR detection method by leveraging the geometric constrain of the projective projection. It is reported that this method can achieve promising performance, but it still requires five IR light sources to provide redundancy. Hennessey and Lawrence[22] proposed a pattern-matching-based method to improve both the accuracy and reliability of the RGE system using multiple CRs. Since the CRs are redundant, the RGE method can still utilize the remaining CRs that are not distorted or lost. However, this method requires a time-consuming pattern-match process, which degrades its efficacy.

In this paper, based on the work proposed by Ma et al.,[18] a novel multiple transform-based method is proposed. In this method, through the multiple geometric transform-based normalization processes, several mapping functions are simultaneously obtained in the calibration process and then used to adaptively estimate the POG. The geometric transforms and the mapping functions can be alternatively employed according to the number of the available CRs even under large head movement. Thereby, the robustness and reliability of the RGE system can be improved through the proposed method.

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

2 Preliminaries

2.1 *Geometric Setup and Model of the Pupil Center Corneal Reflection Method*

Figure 1 illustrates a typical geometric setup considered by the PCCR method.[9] It models the cornea surface as a sphere centered at C, and assumes that the corneal reflection G and the projected image g do not move when the eye rotates around C. Therefore, the image g can be used as a reference point. As the eye rotates to gaze at different targets, the PC P moves in space, and G and P define an image vector gp, which is mapped to screen coordinates through a mapping function to get the estimated POG.

To further understand the PCCR method, the basic geometric model adopted for many PCCR methods[16,17] and the proposed method is illustrated in Fig. 2. This model

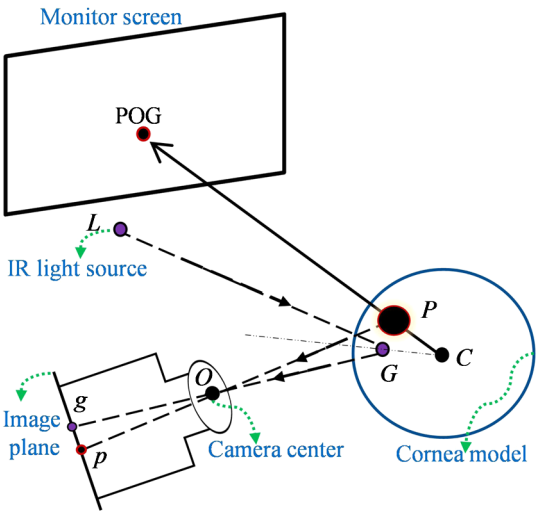

Fig. 1 Geometric setup of the pupil center corneal reflection (PCCR) method.

comprises three planes: the monitor screen plane Π_S with four IR light sources (L_1 to L_4) mounted on its corners, the corneal plane Π_C formed by four CRs (G_1 to G_4), and the camera image plane Π_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 of which the center is modeled as the curvature center C. The G_1 to G_4 and the PC (P) of Π_C are then projected (projection 2) onto Π_I by the camera, thereby forming the CRs (g_1 to g_4) and PC (p) of the image.

2.2 *Corneal Reflection-Adaptive Methods*

Figure 3 illustrates the principle of the corneal reflection-adaptive methods.[18] The normalization function, N_I^N, which is comprised of three alternative geometric transforms (homography, H_I^N; affine, A_I^N; and similarity, S_I^N, transforms) is first employed to normalize the PC p_I in the image plane Π_I into the normalized plane Π_N according to the number of the available CRs.[18] Then, the normalized PCs p_N are mapped to the screen plane $_S$ using the polynomial regression equation, P_N^S, obtained through the personal calibration process. It is demonstrated that this method is superior to the

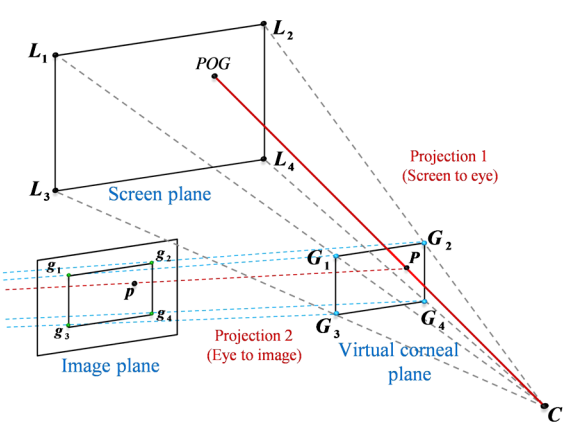

Fig. 2 Geometric model of the PCCR method for remote eye gaze estimation.

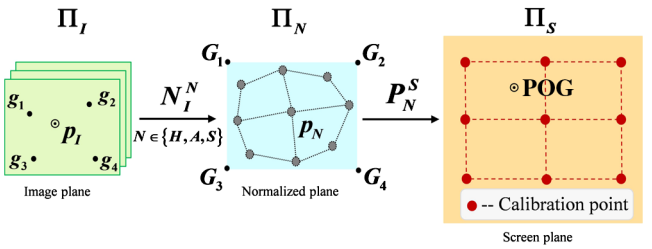

Fig. 3 Illustration of the corneal reflection-adaptive methods.

Table 1 Summary of the methods used in this paper.

Method	CR number	Normalization function	Mapping function
STN	2	Similarity transform	Polynomial regression equation
ATN	3	Affine transform	Polynomial regression equation
HTN	4	Homography transform	Polynomial regression equation

conventional HN method[16] in terms of both accuracy and robustness. As in the work by Ma et al.,[18] the homography, affine, and similarity transform-based gaze estimation methods are called homography transform (HTN), affine transform (ATN), and similarity transform (STN) method, respectively. The characteristics of these methods are summarized in Table 1.

3 Proposed Method

The unconstrained head and eye movement can lead to the loss and distortion of the CRs[19–22], as shown in Fig. 4. In these cases, the conventional RGE methods requiring a fixed number of CRs will be infeasible due to the lack of enough CRs for the POG estimation. On the other hand, although some methods that require fewer numbers of CRs may still work when the head pose is constrained in a certain space volume, the accuracy may degrade because the CR image tends to be distorted as the head moves away from the position where the calibration was performed. Therefore, the RGE method should not only be robust against the large head movement, but also ensure its high accuracy. In this section, a robust and reliable RGE method is presented. It mainly comprises two steps: (1) a simple yet effective CR pattern matching (patternization) method and (2) a multiple transform-based RGE method.

3.1 *Patternization Method*

As illustrated in Fig. 5(a), when a user looks at the screen in an upright frontal view, all four CRs can be detected; however, if the eyeball or head rotate along the vertical or horizontal direction, the four CRs can be undetectable because one or more CRs suffer from distortion or loss. Therefore, it is possible to model these cases as shown in Fig. 5(b).

To cope with the various appearances of the CRs, it is necessary to define a number of representative CR patterns. As shown in Fig. 5(c), pattern 1 corresponds to the case where four CRs are well detected, while patterns 2 to 5 correspond to three CRs, and patterns 6 to 9 are for two CRs. Note that the extreme cases where there exists only one CR, depicted in Fig. 5(b), are excluded from the CR patterns

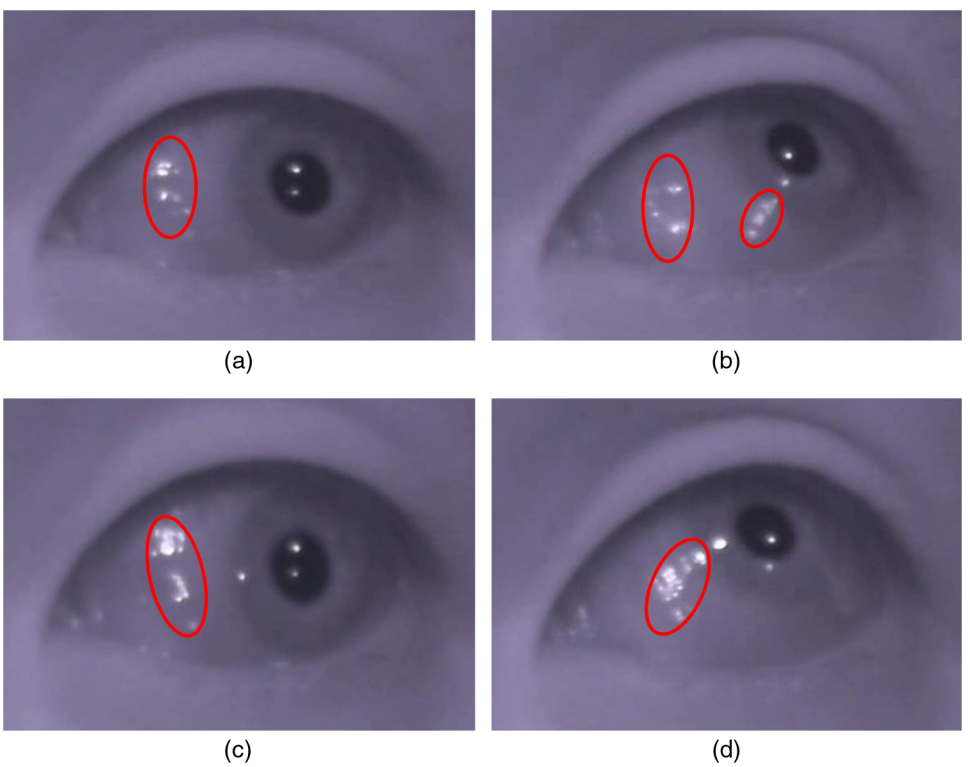

Fig. 4 Examples of the corneal reflection (CR) loss and distortion: (a),(b) two CRs are distorted and lost, and (c),(d) one CR is distorted and lost.

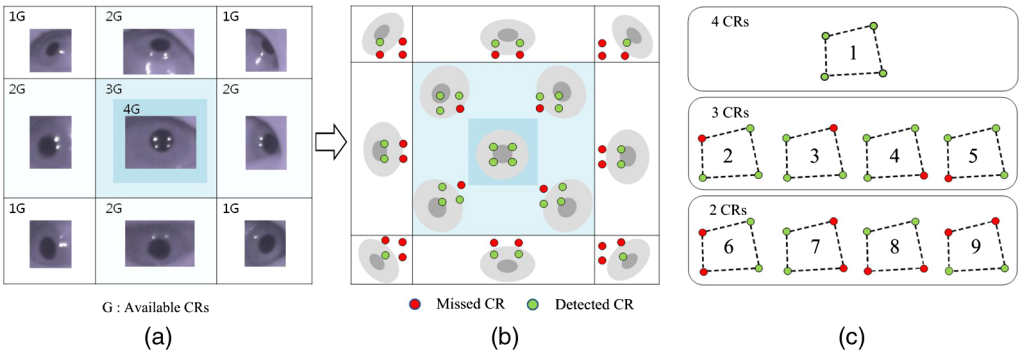

Fig. 5 Patternization of detected CRs: (a) infrared eye images where the number of detected CRs varies; (b) modeling of the CR detection; (c) patterns of the detected CRs.

because they are practically hard to take place in actual scenarios.

For the patternization of both three and four CRs, the centroids of the existing CRs are first obtained and from which four quadrants are defined. It is then possible to differentiate among patterns 1 to 5 by examining the existence of the CRs on each quadrant. Patterns for two CRs are determined using the characteristics of continuity of eyeball movement, i.e., if the pattern found in the previous frame is 5 and there exist two CRs detected in the current frame, the possible pattern number is 6 or 8. Then, the final pattern number can be determined by using the slope between two detected CRs. Specifically, if the measured degree of slope yields near 180 deg with a margin of 10 deg, then the CR pattern number is identified as 8, and the same goes for the case when the slope degree becomes 90 deg, which makes the CR pattern number 6.

模式化匹配根据上一帧情况，或者斜率判断处于哪种情况

Note that the proposed patternization method differs from the pattern-matching method used by Hennessey and Lawrence[22] as follows. First, the proposed method is much simpler yet efficient since it does not needs any elaborative process for the candidate pattern matching. Second, the process for determining the empirical parameter, such as the distortion threshold, is not required in the proposed method because the POG estimation methods adopted are based on the effective normalization procedure, which has been shown to be fairly robust to the shape variation of the CR pattern caused by eyeball or head movements.

3.2 *Multiple Transform-Based Remote Gaze Estimation Method*

As described previously, the unconstrained eye and head movement can lead to the loss and distortion of CRs. The reasons can be observed as follows: first, when an eye is rotated to view different points on the screen, the CRs translate across the surface of the cornea; therefore, the extreme eye rotation leads to the CR loss and distortion. Second, in the RGE system, the translation and rotation of the head also causes the CRs to translate; thereby, the large head pose variations also result in the CR distortion or loss problems. Third, at certain orientations of the head and eye, the CRs may be blocked by eyelashes, distorted, or lost on the boundary between the cornea and sclera. The distortion and loss occur on the rougher surface of the sclera, which is mainly due to the diffuse reflection caused by the different radius of curvature of the sclera and the cornea. Specifically, the rougher surface of the sclera can cause the reflections to be spurious or even disappear, as shown in Figs. 4 and 5. For the aforementioned cases, the conventional RGE methods based on the fixed number of CRs will fail to estimate the POG, making the gaze estimation system less reliable and usable.

To solve these problems, a multiple transform-based method is proposed. In this method, through the multiple geometric transform-based normalization process, several mapping functions are simultaneously obtained in the calibration process and then adaptively used to estimate the POG according to the number of the available CRs. Since the proposed method is based on the effective normalization process, it outperforms the conventional PCCR methods without normalization in terms of both accuracy and robustness.[16,18,22]

Figure 6 illustrates the principle of the proposed multiple geometric transform-based RGE method. As shown in Figs. 6(a) and 6(b), the proposed method consists of two primary procedures: calibration and test.

1. Calibration: the user is requested to orderly gaze at several predefined target points on the screen plane Π_S. Meanwhile, all of the nine possible candidate CR patterns described in Sec. 3.1 undergo their own calibration process, respectively, by using the HTN, ATN, and STN methods.[18] Specifically, for the four-CR case (pattern 1), the HTN method is used to calculate the corresponding polynomial mapping function P_1. For the three-CR cases (pattern 2 to 5) and two-CR cases (pattern 6 to 9), the ATN and

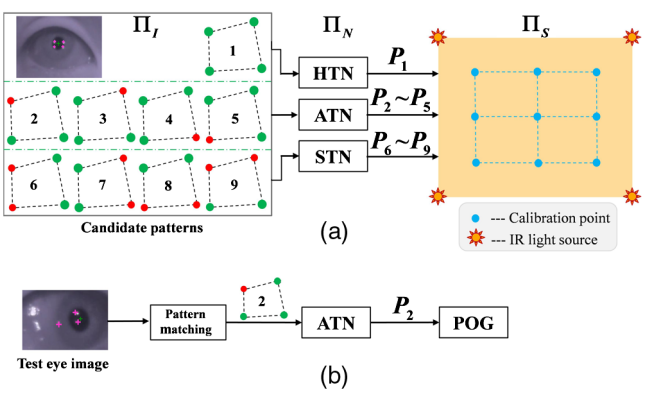

Fig. 6 Illustration of the proposed method: (a) calibration and (b) test.

STN methods are employed respectively to calculate the corresponding mapping functions P_2 to P_5 and P_6 to P_9. Note here that it is implicitly assumed that all of the four CRs are available during the calibration process, which is valid since the user is requested to sit in the central region in front of the screen keeping a frontal upright head pose. Therefore, the four CRs can stably be detected without any occlusion or distortion. Thus, after the calibration, a total of nine polynomial mapping functions are obtained and ready for use.

2. Test: given a test eye image with detected PC and CRs, the proposed patternization is first adopted to identify the pattern number. Taking pattern 2 as an example, the corresponding normalization method ATN method is activated and the corresponding polynomial mapping function, P_2, obtained in the calibration process, is then used to estimate the POG.

Thus, by making use of the advantages of the HTN, ATN, and STN methods, the proposed RGE method cannot only handle the cases where one or two CRs are lost or distorted because of the occlusions (e.g., eyelashes and eyelids), but also allow for large head movements by solving the CR distortion or loss problems caused by the unconstrained head translation and rotation. Therefore, the reliability and usability of the gaze estimation system are further enhanced.

Note that the novelties of this work that are different from the authors' previous work[18] are discussed as follows: first, in terms of the reliability, the three methods proposed in Ref. 18 can only be used separately according to the available number of CRs, which provides simply the possible approaches rather than a comprehensive solution to solve CR distortion and missing problems in the real scenarios. However, the system proposed in this paper significantly improves the reliability of the gaze tracking system by considering all the possible cases in the real scenarios. Second, in terms of usability, as compared with the methods proposed in Ref. 18, the constraints on the eye and head movements are further released since the proposed system in this paper can handle those extreme cases where only two or three CRs are available due to the unconstrained eye and head movement. Third, in terms of robustness, the methods proposed in Ref. 18 are infeasible when the CRs are distorted or lost; instead, the pattern detection (i.e., the patternization) method proposed in this work implicitly encodes the distortion and missing information of the CRs, resulting in an adaptive gaze tracking system that can resist the interferences from the pattern changes of the CRs.

4 Experimental Results

4.1 *Experimental Methodology*

The experiments were conducted using a 24-in. screen with a resolution of 1920×1080 pixels. Four IR light sources with a wavelength of 850 nm were placed in the corners of the monitor. Input images with a resolution of 640×480 pixels were obtained from a web-camera (from which the IR cut-off filter was removed), with a 4.5× zoom lens attached, placed at the center region slightly under the monitor screen as shown in Fig. 7(a).

The experiment was designed to provide a demonstration of the following: (1) reliability—the comparison between the original four-CR-based method and the proposed method in terms of the percentage of the loss of the POG estimation caused by the loss or distortion of the CRs; and (2) accuracy—the accuracies of the proposed method under different depths and the total accuracy of the proposed method.

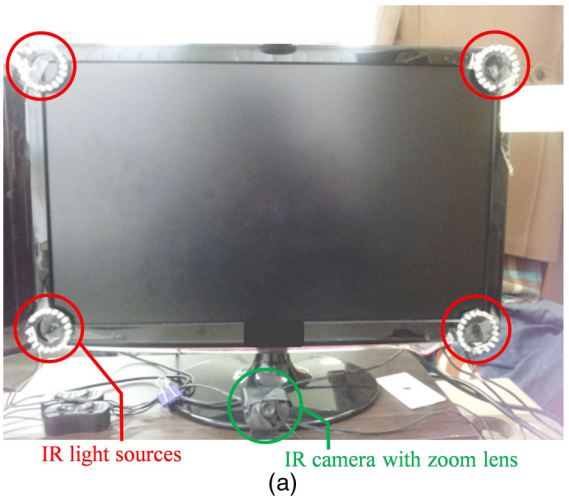

Fig. 7 (a) System setup and (b) calibration position.

To test the proposed method, six subjects (three females and three males) ranging from 20 to 35 years in age participated in the experiments. The experimental procedure required each subject to first take a nine-point calibration at the midpoint of the depth of focus of the camera lens, i.e., approximately 65 cm from the screen as shown in Fig. 7(b). Then, the test procedure was continuously conducted at three different depths which were, respectively, mid-depth 65 cm, near-depth 55 cm, and far-depth 75 cm. In the test procedure, each subject was asked to sequentially gaze at the nine target points displayed on the screen plane while moving his/her head in the three-dimensional (3-D) space volume of $10 \times 10 \times 10$ cm (width × height × depth). For simplicity, the test was first conducted in the mid-depth where the calibration was performed. Then, the subject was asked to move to the near-depth and then finally the far-depth to perform the test. The accuracy was measured as the average angular error and standard deviation in degrees over nine uniformly distributed test positions on the screen.[18] Figure 8 shows the target screen points used for the calibration and test. For each test point, the angular error μ and the standard

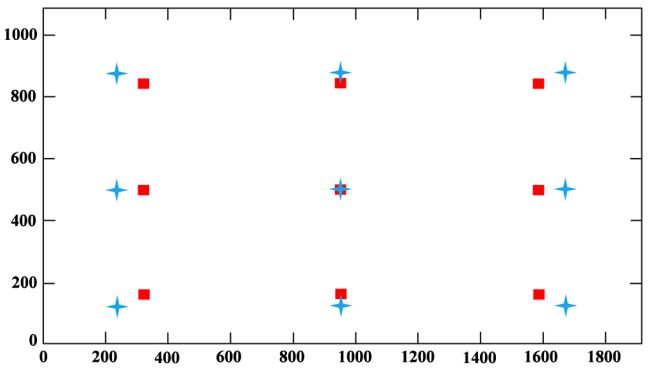

Fig. 8 Target points used for calibration (square) and test (cross).

deviation σ over N frames ($N = 150$) are, respectively, calculated as Ref. 18. The same as Ref. 18, under the experimental setup in this paper, the angular error of 1 to 2 deg corresponds to 12 to 24 mm or 43 to 86 pixels on the screen. For each of three depths, i.e., near-, mid-, and far-depth, the average gaze estimation error is measured on the nine test points; thus, a total of three resultant error values are obtained. Note that, as compared with the conventional methods,[1–8] the available space for head movement in the proposed method covers a fairly wider 3-D range, of which the depth is from 45 to 85 cm, and both the width and height are 10 cm.

The PC and CRs of the image were, respectively, detected using the method proposed by Choi et al.[23] and thresholding and blob detection methods proposed by Chang et al.[24]

4.2 *Performance Evaluation in Terms of Reliability and Accuracy*

Table 2 lists the experimental results of the proposed method in terms of reliability and accuracy defined in Sec. 4.1. It can be seen that the accuracy at mid-depth position where the calibration is performed is the best among the three depths for all six subjects, which is consistent with the expectation since the accuracy will decrease with the depth change. The final averaged accuracy of the proposed method evaluated over three depths, ranging from 45 to 85 cm, is 1.29 ± 0.83 deg. The accuracy is favorable and acceptable for most of the HCI applications allowing large head movement. Moreover, it can be observed that the two- and three-CR-based ATN and STN methods achieve a similar accuracy as the four-CR-based HTN method. In terms of reliability, it is also can be seen that among all of the frames used during the experiment, approximately only 73% of them could be applied for the conventional four-CR-based methods, such as the HN[16] and HTN methods. However, for the rest of the cases where one or two CRs are lost or distorted, i.e., more than 26% of the total frames, it is impossible to handle them using the conventional methods. The proposed method, on the other hand, cannot only deal with this problem successfully and reliably, but also provides a fairly good accuracy. Therefore, through the proposed method, the restrictions on the eye and head movement are further released and the space allowed for natural eye and head movement is greatly increased. In addition, on a desktop computer with a 2.8-GHz quad-core CPU and 4 GB RAM, the processing times of HTN, ATN, and STN methods are 27, 25, and 23 ms per frame, respectively. The averaged processing time of the proposed method is 26 ms per frame; therefore, the real-time property can be guaranteed.

5 Conclusion

In this paper, a novel multiple transform-based RGE method is proposed. In the proposed method, through the multiple geometric transform-based normalization processes, several mapping functions are obtained in the calibration process and then used to estimate the POG. The geometric transforms and mapping functions can be alternatively employed according to the number of the available CRs. The contributions of the proposed method are, first, the usability of the RGE system is further enhanced since much more space is available for head movements; second, the robustness and

Table 2 Performance comparison of the proposed and conventional methods in terms of accuracy (unit: deg) and reliability (unit: %).

	Mid-depth (cm) (55-65-75)	Near-depth (cm) (45-55-65)	Far-depth (cm) (65-75-85)	Total (cm) (45-65-85)	HTN (4 CRs) (percentage)	ATN (3 CRs) (percentage)	STN (2 CRs) (percentage)	3+2 CRs (percentage)
Subject 1	1.15 ± 0.71 deg	1.30 ± 0.89 deg	1.28 ± 0.77 deg	1.24 ± 0.79 deg	1.09 ± 0.68 deg 69.51 (%)	1.34 ± 0.76 deg 15.87 (%)	1.68 ± 0.83 deg 14.62 (%)	1.51 ± 0.79 deg 30.49 (%)
Subject 2	0.92 ± 0.61 deg	1.26 ± 0.78 deg	1.07 ± 0.67 deg	1.08 ± 0.69 deg	0.98 ± 0.60 deg 73.51 (%)	1.32 ± 0.73 deg 19.49 (%)	1.36 ± 0.68 deg 7.00 (%)	1.33 ± 0.72 deg 26.49(%)
Subject 3	1.21 ± 0.85 deg	1.68 ± 1.04 deg	1.35 ± 0.91 deg	1.41 ± 0.94 deg	1.34 ± 0.84 deg 67.06 (%)	1.38 ± 0.93 deg 17.70 (%)	1.56 ± 0.88 deg 15.24 (%)	1.46 ± 0.90 deg 32.94 (%)
Subject 4	1.20 ± 0.75	1.37 ± 0.93 deg	1.31 ± 0.92 deg	1.30 ± 0.86 deg	1.33 ± 0.85 deg 80.67 (%)	1.05 ± 0.84 deg 16.64 (%)	1.47 ± 0.96 deg 2.69 (%)	1.11 ± 0.72 deg 19.33 (%)
Subject 5	1.24 ± 0.87 deg	1.68 ± 1.18 deg	1.27 ± 0.78 deg	1.39 ± 0.94 deg	1.29 ± 0.90 deg 74.63 (%)	1.48 ± 1.13 deg 18.10 (%)	1.87 ± 0.96 deg 7.27 (%)	1.60 ± 1.08 deg 25.37 (%)
Subject 6	1.19 ± 0.64 deg	1.42 ± 0.75 deg	1.40 ± 0.84 deg	1.34 ± 0.74 deg	1.32 ± 0.68 deg 76.50 (%)	1.29 ± 0.73 deg 19.72 (%)	1.68 ± 1.00 deg 3.78 (%)	1.35 ± 0.77 deg 23.50 (%)
Average	1.15 ± 0.74 deg	1.45 ± 0.93 deg	1.28 ± 0.82 deg	1.29 ± 0.83 deg	1.23 ± 0.76 deg 73.65 (%)	1.31 ± 0.85 deg 17.92 (%)	1.60 ± 0.89 deg 8.43 (%)	1.39 ± 0.83 deg 26.35 (%)

the reliability of the RGE system can be significantly improved as the proposed method is capable of handling various noises, such as the loss and distortion of the CRs caused both by unconstrained eye and head movement and erroneous CR detection.

Acknowledgments

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

References

1. T. E. Hutchinson et al., "Human-computer interaction using eye-gaze input," *IEEE Trans. Syst. Man Cybern.* **19**(6), 1527–1534 (1989).
2. D. C. Cho et al., "Long range eye gaze tracking system for a large screen," *IEEE Trans. Consum. Electron.* **58**(4), 1119–1128 (2012).
3. H. C. Lee et al., "Gaze tracking system at a distance for controlling IPTV," *IEEE Trans. Consum. Electron.* **56**(4), 2577–2583 (2010).
4. H. Heo et al., "A realistic game system using multi-modal user interfaces," *IEEE Trans. Consum. Electron.* **56**(4), 1364–1372 (2010).
5. P. M. Corcoran et al., "Real-time eye gaze tracking for gaming design and consumer electronics systems," *IEEE Trans. Consum. Electron.* **58**(2), 347–355 (2012).
6. T. Nawaz, M. S. Mian, and H. A. Habib, "Infotainment devices control by eye gaze and gesture recognition fusion," *IEEE Trans. Consum. Electron.* **54**(2), 277–282 (2008).
7. N. Iqbal, H. Lee, and S. Y. Lee, "Smart user interface for mobile consumer devices using model-based eye-gaze estimation," *IEEE Trans. Consum. Electron.* **59**(1), 161–166 (2013).
8. K. A. Choi, C. Ma, and S. J. Ko, "Improving the usability of remote eye gaze tracking for human-device interaction," *IEEE Trans. Consum. Electron.* **60**(3), 493–498 (2014).
9. 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).
10. 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).
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. S. W. Shih and J. Liu, "A novel approach to 3D gaze tracking using stereo cameras," *IEEE Trans. Syst., Man, Cybern. B* **34**(1), 234–245 (2004).
13. Z. Zhu and Q. Ji, "Novel eye gaze tracking techniques under natural head movement," *IEEE Trans. Biomed. Eng.* **54**(12), 2246–2260 (2007).
14. D. H. Yoo and M. J. Chung, "A novel non-intrusive eye gaze estimation using cross-ratio under large head motion," *Comput. Vision Image Understanding* **98**(1), 25–51 (2005).
15. 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).
16. 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*, Austin, Texas, pp. 13–20 (2010).
17. F. L. Coutinho and C. H. Morimoto, "Improving head movement tolerance of cross-ratio based eye trackers," *Int. J. Comput. Vision* **101**(3), 459–481 (2013).
18. C. Ma et al., "Improved remote gaze estimation using corneal reflection-adaptive geometric transforms," *Opt. Eng.* **53**(5), 053112 (2014).
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. F. Li, S. M. Kolakowski, and J. B. Pelz, "Using structured illumination to enhance video-based eye tracking," in *Proc. IEEE Int. Conf. on Image Processing*, San Antonio, Texas, pp. 373–376 (2007).
21. D. W. Hansen, L. Roholm, and I. G. Ferreiros, "Robust glint detection through homography normalization," in *Proc. ACM Symp. Eye-Tracking Research and Applications*, Safety Harbor, Florida, pp. 91–94 (2014).
22. 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).
23. K. A. Choi et al., "Improved pupil center localization method for eye-gaze tracking-based human-device interaction," in *Proc. IEEE Int. Conf. on Consumer Electronics*, Las Vegas, Nevada, pp. 524–525 (2014).
24. F. Chang, C. J. Chen, and C. J. Lu, "A linear-time component-labeling algorithm using contour tracing technique," *Comput. Vision Image Understanding* **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. Currently, he is 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.

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

Byeong-Doo Choi received his BS, MS, and PhD degrees, all in electronics engineering from Korea University, in 2001, 2003, and 2007, respectively. He joined the Fraunhofer Institute for Tele-communication, Heinrich-Hertz-Institut, Berlin, Germany, as a visiting scholar from 2007 to 2008. Currently, he is working at DMC R&D center in Samsung Electronics Co., Ltd., as a senior research staff member. His current research topics are video compression algorithms, multimedia transmission systems, and visual quality enhancement processing.

Sung-Jea Ko received his PhD in 1988 and his MS degree in 1986, both in electrical and computer engineering, from the 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.