

A Gaze Tracking Method as an IPTV Interface

Chul Woo Cho*, Ji Woo Lee*, Hyeon Chang Lee*, Won Oh Lee*, Eul Chul Lee**, Kang Ryoung Park*⁺,

Minsik Park***, Heekyung Lee***, and Han-Kyu Lee***

**Dept. of Electronics Eng., Dongguk University, 26 Pil-dong 3-ga, Chung-gu, Seoul, Repblic of Korea*

***Dept. of Computer Science, Sangmyung University, 7 Hongji-dong, Jongro-gu, Seoul, Republic of Korea*

****Electronics and Telecommunications Research Institute, 161 Gajeong-Dong, Yuseong-gu, Daejeon, Republic of Korea*

⁺Corresponding author

cho4400@dgu.edu, ljwgs@dgu.edu, leehc@dgu.edu, 215p8@hanmail.net, oryong@smu.ac.kr, parkgr@dgu.edu, pms@etri.re.kr,
lhk95@etri.re.kr, hkl@etri.re.kr

Abstract—In this paper, we propose a new wearable gaze tracking system based on the glasses-frame-based device which includes an eyecapturing camera. Gaze positions are calculated based on the geometric transform. Also, errors caused by facial movements are compensated by using the geometric relation of two corneal specular reflections. Menu selection is implemented based on the dwell time of gaze position. Experiments are performed in order to confirm the feasibility for applying the proposed method to internet protocol television (IPTV) with large scaled screen. Experimental results showed that the average rootmean-square (RMS) error of gaze detection under allowing facial movements was 1.38 degrees.

Keywords— Gaze Tracking, IPTV

I. INTRODUCTION

IPTV provides user with various kinds of multimedia services (television, video, audio, text, graphics and data) that are delivered over managed IPbased networks with appropriate QoS/QoE, security, interactivity and reliability [1]. To control various TV menus, a remote controller is commonly used. However, according to diversification of TV's function, it becomes very difficult to include such great functionalities on a remote controller. Moreover, televiewers look forward to more comfortable and convenient input device without moving her hands. Therefore, we propose gazetracking system that can be adopted in the IPTV environment. Gaze tracking is a technology to obtain the direction or position which a user is looking at. Gaze tracking has been widely researched because human gaze position can provide important information and cues.

Gaze tracking technology can be categorized into two classes such as 'remote gaze tracking method' and 'head-mounted gaze tracking method'. In the remote gaze tracking method, a narrow view camera with high magnification lens captures user's eye region based on the information from the wide view camera which tracks user's eye position from the face image. Both cameras are operated by panning & tilting mechanism. This method can give the convenience to user because of allowing comparatively natural facial movements.

However, the method has the drawback that the cost and processing complexity is high. That is because it requires the additional mechanism for zooming, panning, and tilting of narrow view camera. And it requires the complicated algorithms for detecting face and eye regions by wide view camera including the calibration procedure between narrow and wide view cameras. In some methods, expensive stereo camera is used for estimating three dimensional eye positions [2].

In the head mounted gaze tracking method, a user should wear a head mounted device including a camera and near infrared (NIR) illuminators which can capture the eye image of user [3-6]. In the researches, the gaze tracking methods based on head mounted display (HMD) are introduced [3][4][6]. It was not necessary to consider the head movements because the virtual screen in the HMD is also moved simultaneously with a user's head movements. However, gaze estimation accuracy can be degraded by the slipping down of HMD after the initial user-dependent calibration. In another method, they used a gaze tracker including only a head mounted camera [8]. However, since an eye and a frontal scene images are captured at a time in one image, gaze tracking accuracy is not reliable.

To overcome such problems, we propose a new wearable gaze tracking method based on the glasses-frame-based device which includes an eye capturing camera. It allows the natural facial movements of user, which can be possible by the geometric relation of two specular reflections generated by two infrared illuminators attached on the upper part of IPTV frame.

This paper consists of following parts. In section II, explanations of the proposed device and method are described. Experimental results, applications and conclusion are shown in section III, IV and V, respectively.

II. PROPOSED METHOD

A. Overview of the proposed method

Fig. 1 shows an overview of the proposed method. Firstly, user-dependent calibration is performed in order to define pupil movable area in captured eye image as Fig. 1 (1). Then, positions of a pupil center and two specular reflections are extracted from a captured eye image as Fig. 1 (2) ~ (4). Based

on the above defined pupil movable area, the coarse gaze position on the IPTV plane is calculated as Fig. 1 (5). However, since this position can include errors caused by facial movements after the user-dependent calibration, the errors are compensated by using the geometry of the above extracted two specular reflections as Fig. 1 (6). Consequently, as shown in Fig. 1 (7) and (8), the compensated gaze position is determined on the IPTV plane including a selection event according to a measured dwell time of fixed gaze position.

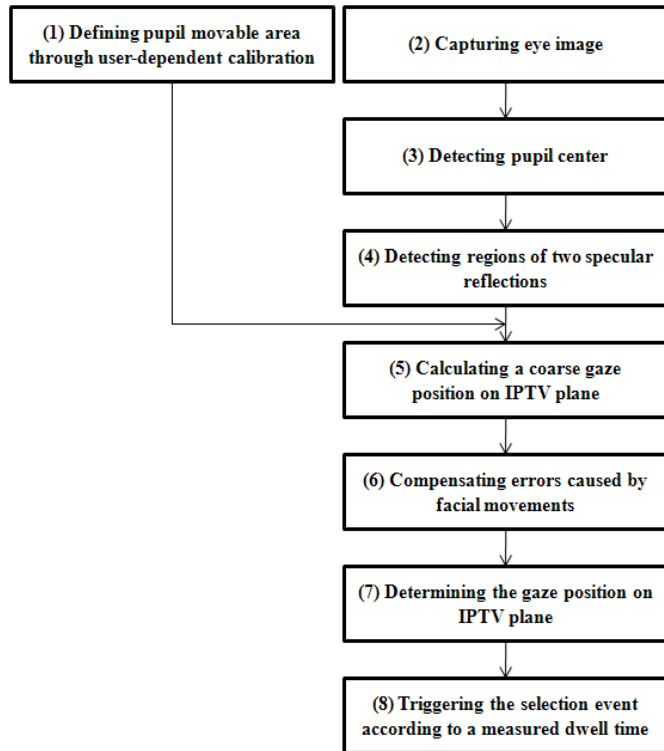


Figure 1. A overall procedure of the proposed gaze tracking method

B. Gaze tracking device

Fig. 2 shows the proposed gaze tracking device. There are two components in the proposed gaze tracking device, which are a glasses frame based eye capturing device including universal serial bus (USB) typed camera, and two infrared light emitting diode (IR-LED) illuminators as shown in Fig. 2 (a) and (b). Eye capturing device and two IR-LEDs are used for analysing eye gaze position and facial movements, respectively.

The eye capturing camera is attached on the lower part of the right eye at the glasses frame as shown in Fig. 2 (a). This camera captures NIR illuminated eye images. Spatial and temporal resolutions of the used camera are 640×480 pixels and 30 frames/s, respectively. This camera is attached on the flexible wire frame which can be freely moved for capturing eye position. This camera contains a zoom lens whose focal length ($\lambda = fc$) is 1,598 pixels for capturing highly magnified eye image. We attached a zoom lens on the eye capturing

camera in order to capture magnified eye image. And an infrared passing filter is attached on the camera lens to reject visible light. Because it maintains the brightness of eye image irrespective of environmental visible light and it can perform the image processing for pupil detection easier.

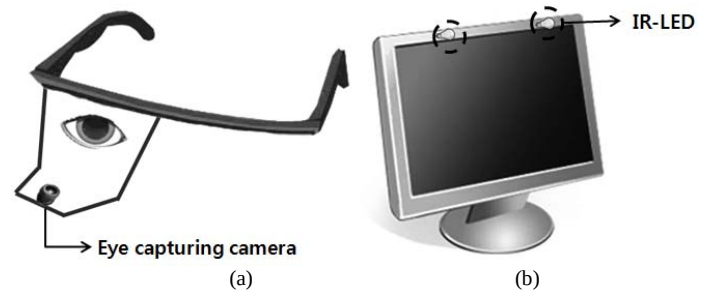


Figure 2. The proposed gaze tracking device. (a) A glasses frame based eye capturing device. (b) Attached two IR-LEDs on the upper frame of IPTV.

NIR illuminators are equipped top of IPTV as shown in figure 2 (b). The wavelength of the illuminator is 850nm, so it does not make any dazzling to user's eye. Since the eyelid easily occludes the upper and the lower part of iris region, we use a pupil region instead of iris region for gaze estimation [3][4][6][14].

C. Gaze Tracking Algorithm

1) User-dependent calibration

To calculate the gaze position on an IPTV from the eye image, a user should carry out user-dependent calibration, which consisted of four stages, in which the user gazes at four corner positions of the IPTV screen (upper left, upper right, lower left, and lower right). After that, our system calculated each user's eye gaze position on an IPTV screen based on the four center positions of pupil.

After user-dependent calibration, the tetragon shape generated by four pupil center positions is projected onto the rectangle of the screen plane by using the Geometric transform [14][15]. The geometric transform process refers to the way of defining the relationship between a tetragon and a rectangle. Detail explanations are described in section II-C.D.

2) Detecting the center of pupil

To detect the pupil center in captured eye image, we used two methods such as circular edge detection and local binarization [16]. The template for circular edge detection is composed of two scalable circles. Through circular edge detection, the position at which the maximum difference value between the inner and outer circles is calculated is regarded as the pupil region. That is because the gray level difference between pupil and iris is greatest compared to other regions.

However, in general, the pupil is not an intact circle shape and the shape becomes ellipse or distorted [5] in case that a users gazes at the IPTV's corner area. In such case, the detected center of pupil by circular edge detection is not accurate. Therefore, the post processing using local binarization is

performed based on the coarse pupil center determined using the circular edge detection. Also, morphological closing operation is performed for removing specular reflective regions. Consequently, the pupil center is determined by calculating the center of black area.

3) Detecting regions of two specular reflections

To detect the center positions of two specular reflections caused by two IR-LEDs of Fig. 2 (b), local binarization is performed based on the detected pupil position. Local region is defined as 200×200 pixels based on the approximate iris diameter ($10 \sim 12$ mm) since the cornea curvature diameter is almost same that of iris [15]. After that, the component labelling method [15] is performed for separating two reflections. Consequently, each center position of two reflections is obtained which is used for compensating facial movements.

4) Calculating a coarse gaze position

Gaze position is calculated based on the four pupil center positions acquired at the user-dependent calibration stage and the detected pupil centers from continuously captured eye images.

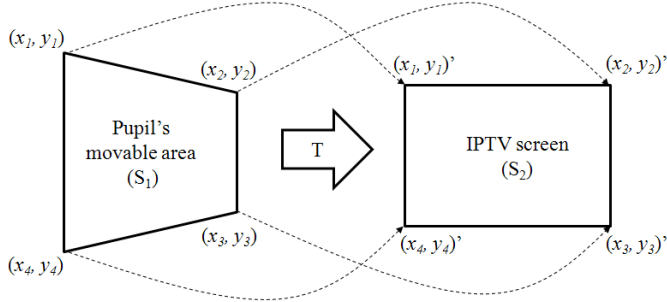


Figure 3. Mapping a center of the pupil in the movable area of the pupil into the IPTV area.

As shown in Fig. 3, the geometric transform is performed for mapping pupil's movable area into the screen [6][14]. The transform represents the relation between the distorted quadrilateral region (S_1) and the rectangular region S_2 . In Fig. 3, the quadrilateral region (S_1) is the defined one by four eye gazing position at the user-dependent calibration stage. Since the IPTV screen (S_2) has a pixel coordination based rectangular shape, a transform matrix (T) between S_1 and S_2 can be acquired by geometric transform method. Based on the obtained geometric transform matrix, the pupil position which is extracted in a captured eye image is corresponded onto one position of S_2 .

Based on the obtained geometric transform matrix (T), the gaze position $G(g_x, g_y)$ on the S_2 can be calculated from the pupil center $P(p_x, p_y)$ in S_1 as following equation:

$$G = TP$$

$$\begin{bmatrix} g_x \\ g_y \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} a & b & c & d \\ e & f & g & h \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} p_x \\ p_y \\ p_x p_y \\ 1 \end{bmatrix} \quad (1)$$

5) Compensating errors caused by facial movements

Generally, the errors caused by facial movements should be compensated because this is naturally occurred after the user-dependent calibration. Therefore, to estimate such facial movements, the geometric change of two reflections is used. An example image of such specular reflections is shown in Fig. 4.

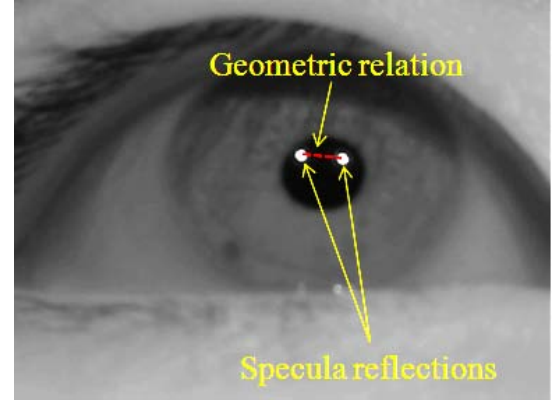


Figure 4. An example of two specular reflections.

When facial movement occurs in three dimensional space, the distances between two reflections are changed. If a distance between user's eye and the IPTV is near, the distances between two reflections is longer. In opposite case, the distance between two reflections becomes shorter. From that, we obtained the weighted average value about the change of Z-distance. Also, the facial movement can be estimated by analysing the position of two reflections. Consequently, the final gaze position ($P(X, Y)$), considering eye and facial movements is calculated as shown in Equation (2):

$$P(X, Y) = (\overline{g_x} + w_x \Delta d_x, \overline{g_y} + w_y \Delta d_y)$$

$$\overline{g_x} = \frac{g_x}{w_1 \cdot \Delta d_z}, \quad \overline{g_y} = \frac{g_y}{w_2 \cdot \Delta d_z} \quad (2)$$

$$\Delta d_z = \text{sgn}((d_t - d_1)), \quad \text{sgn}(x) = \begin{cases} 1, & x > 0 \\ 0, & x = 0 \\ -1, & x < 0 \end{cases}$$

In Equation (2), (g_x, g_y) is the coarse gaze position, calculated by Equation (1), Δd_z is the distance between two reflections, which can represent the distance between the IPTV screen and a user's eye position. ($\Delta d_x, \Delta d_y$) is the facial movement value, which can be obtained by measuring translational offset of the reflections. w_x, w_y, w_1 , and w_2 are the empirically determined weight values which represent the

confidence levels among (Δd_x , Δd_y) and Δd_z . Consequently, the errors caused by facial movements after user-dependent calibration can be compensated by analyzing two reflection's geometry.

6) Selection method based on dwell time

Above explanations show the procedure of detecting gaze position. By using them, menu navigation or exploration can be performed. However, there is no adequate method of triggering selection event such as mouse button clicking. To solve this problem, we used the scheme of measuring the dwell time of fixed gaze position [10]. Based on that, 3 seconds are empirically determined as a threshold of the dwell time of fixed gaze position.

III. EXPERIMENTAL RESULTS

The proposed method was tested on a desktop computer with Intel Core2 Quad 2.33GHz CPU and 4 GB RAM. The algorithm has been implemented by using microsoft foundation class (MFC) based C++, DirectShow of DirectX 9.0 SDK. 30 inch LCD monitor is used for simulating IPTV environment, which has the monitor resolution of $1,366 \times 768$ pixels.

Firstly, the gaze detection accuracy was measured under no allowing facial movements. In order to measure the root mean square (RMS) error of the gaze detection, subjects were requested to gaze at 12 reference points on the monitor plane. Five subjects participated in the experiment, and each subject performed five tests. The experimental results showed that the average RMS error between estimated gaze positions and ground-truth ones was 0.98 degrees as shown in case 1 of table 1, which corresponds to errors of 59 pixels in the used LCD monitor with resolution of $1,366 \times 768$ pixels.

Table 1. RMS errors of gaze detection. Cases 1, 2, and 3 are the results of no facial movements, allowing facial movements without compensation method and allowing facial movements with compensation method, respectively.

	Case 1	Case 2	Case 3
Accuracy (degrees)	0.98	2.53	1.48

In the next experiment, we measured the gaze detection accuracy in terms of the RMS error when allowing facial movements without and with the proposed compensation procedure of facial movements. Experimental schemes were same with the first experiments. As shown in the cases 2 and 3 of table 1, the average RMS errors in two cases were 2.53 and 1.48 degrees, respectively.

Through above experimental results, we found that the gaze detection accuracy was absolutely degraded by facial movements and the proposed method fairly compensated the errors.

Fig. 5 shows an example of using the proposed system on IPTV environment.

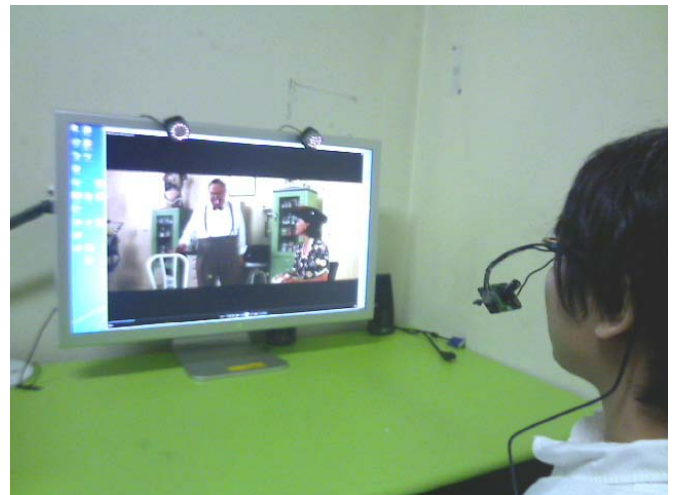


Figure 5. An example of using the proposed system on IPTV environment.

IV. CONCLUSIONS

In this paper, we proposed a glasses frame based gaze tracking device and method which can be applied to the function control of IPTV. Our system tracks the user's gaze position under allowing facial movements after the initial user-dependent calibration. An eye gaze position is calculated by using geometric transformation between two rectangular regions. Errors caused by facial movements after user-dependent calibration were compensated based on geometric relation of two specular reflections. To trigger selection events, the dwell time of fixed gaze position is used. Through the experimental results, we confirmed that our proposed method could be applied to the IPTV environments.

In future work, we will research remote typed gaze tracking system without wearing any devices based on the proposed algorithm. Also, more reliable and convenient triggering method of selection event will be considered to improve the throughput of IPTV control.

ACKNOWLEDGMENT

This work was supported by the IT R&D program of MKE (Ministry of Knowledge Economy) and KCC(Korea Communications Commission) in Korea. [2009-S-019-01, Development of UHD/3D Realistic Video and Interactive View Control Technologies for IPTV].

REFERENCES

- [1] <http://www.software.or.kr/ICSFiles/afildfile/2008/08/27/IPTV.pdf> (accessed on 2009.10. 9).
- [2] S. W. Shih, and J. Liu, "A Novel Approach to 3-D Gaze Tracking Using Stereo Cameras," IEEE Transaction on Systems, Man and Cybernetics, Part B, vol. 34, no. 1, pp. 234-245, (Feb. 2004).
- [3] E. C. Lee, and K. R. Park, "A Robust Eye Gaze Tracking Method Based On Virtual Eyeball Model," Machine Vision and Applications, vol. 20, no. 5, pp. 319-337, (2009).

- [4] E. C. Lee, and K. R. Park, "A Study on Eye Gaze Estimation Method Based on Cornea Model of Human Eye," MIRAGE 2007 on Lecture Notes in Computer Science, vol. 4418, pp. 307-317 (2007).
- [5] J. Daugman, "New Methods in Iris Recognition," IEEE Trans. System, Man, and Cybernetics-Part B, vol. 37, no. 5, pp. 1167-1175, (Oct. 2007).
- [6] E. C. Lee, and K. R. Park, "3D First Person Shooting Game by Using Eye Gaze Tracking," Journal of Korea Information Processing Society, vol. 12-B, no. 4, pp. 465-472, (Aug. 2005).
- [7] X. He, and P. Shi, "A Novel Iris Segmentation Method for Hand-Held Capture Device," ICB 2006 on Lecture Notes in Computer Science, vol. 3832, pp. 479-485, (2005).
- [8] P. Chaudhuri, G. Papagiannakis, and N. Magnenat-Thalmann, "Camera-Based Gaze Control for Virtual Characters," in Proc. of ENACTIVE/07, 4th International Conference on Enactive Interfaces, (Nov. 2007).
- [9] H. Kato, and M. Billinghurst, "Marker Tracking and HMD Calibration For a Video-based Augmented Reality Conferencing System," in Proc. of the 2nd Int. Workshop on Augmented Reality, pp.85-94, (1999).
- [10] D. Salvucci, and H. Joseph, "Identifying Fixations And Saccades in Eye-Tracking Protocols", in Proc. of the 2000 Symposium on Eye tracking Research and Applications, pp. 71-78.
- [11] ZC&R Coatings for Optics, Inc., <http://www.zcrcoatings.com> (accessed on 2009.10.9)
- [12] <http://www.logitech.com> (accessed on 2009.10.9).
- [13] Y. K. Jang, B. J. Kang, and K. R. Park, "A Study on Eyelid Localization Considering Image Focus for Iris Recognition", Pattern Recognition Letters, vol. 29, no. 11, pp. 1698-1704, 1 (Aug. 2008).
- [14] C. W. Cho, J. W. Lee, E. C. Lee, and K. R. Park, "Robust Facial and Eye Gaze Estimation Method Based on Eye and Monitor Tracking", Optical Engineering, submitted, (2009).
- [15] R. C. Gonzalez, and R. E. Woods, Digital Image Processing, Prentice-Hall, Inc., Second Edition, pp 587-591, 2002.
- [16] J. Daugman, "How Iris Recognition Works," IEEE Trans. Circuits and Systems for Video Technology, vol. 14, no. 1, pp. 21-30, 2004.