

Acquisition of Eye Images in the Gaze Tracker based on Depth Sensor and Zoom Camera

HeeKyung LEE*, Jeongil SEO*, Hoon Jo**

*ETRI(Electronics and Telecommunications Research Institute), Korea

**Hanyang University, Korea

lhk95@etri.re.kr, seoji@etri.re.kr, hjo@vision.hanyang.ac.kr

Abstract— We designed and developed a gaze tracking device to provide comfortability in gaze tracking and freedom of head movement of a user to be applicable to the large display environment with a long Z distance. We also propose the method of stereo calibration and pixel mapping between the wide-angle camera and the depth sensor, and the method of PTZ calibration between the wide-angle camera and the zoom camera to make the device continuously and effectively track and capture the eye image even though a user's irregular movements in 3D space

Keywords— Gaze, Tracking, Device, Calibration, PTZ

I. INTRODUCTION

Gaze tracking device can be divided into the wearable camera based device and the remote camera based device. The former uses special contact lenses or headsets [1], so it is inconvenient and inaccurate if it is moved or slipped down. The latter is adopted in most of commercial gaze trackers. But some can be just used in the PC environment with a short Z distance about 80cm [2][3][4], and others can be used in the large display environment with the restricted freedom of head movement and high price [5].

So, we propose a gaze tracking device (HW) that is comfortable owing to adopting a remote camera based method, and allows the freedom of head movement of a user in the large display environment with a long Z distance about 3m.

For the sake of the goal, it tracks user's eye in the distance, and then acquires high resolution near-infrared (NIR) eye images with automatic panning, tilting, and zooming. Therefore, it consists of a wide-angle camera and a depth sensor for the eye tracking, and a zoom camera mounted on the pan-tilt units, and 2 NIR illuminators for the high resolution eye image acquisition.

As the wide-angle camera, the depth sensor and the zoom camera in the gaze tracking device have different position and field of view, the pixel coordinate of user's eye on the images captured by each sensor need to be matched up.

So, we carry out the stereo calibration and pixel mapping between the wide-angle camera and the depth sensor for the computation of 3D world coordinate. And we also propose the method of PTZ calibration between the wide-angle camera and the zoom camera to calculate the pan-tilt angles and zoom pulses to capture the high resolution NIR eye images.

Then, the gaze tracker detects the face and eye of a user on the images captured by the wide-angle camera. And the 3D

world coordinate of detected eye is calculated based on the stereo calibration and pixel mapping data. Subsequently, the zoom camera is panned and tilted to head for the detected eye, and then zoomed to capture high resolution NIR eye image based on the PTZ calibration data. This method makes the zoom camera continuously and effectively track and capture the eye image even though a user's irregular movements in 3D space.

II. GAZE TRACKING DEVICE & CALIBRATION

A. Remote Gaze Tracking Device

The proposed gaze tracking device consists of a wide-angle camera and a depth sensor for the eye tracking, and a zoom camera mounted on the pan-tilt units, and 2 NIR illuminators for the high resolution eye image acquisition. The NIR illuminators are also utilized as a reference point for gaze position calculation because they create corneal reflections in the IR image called as glints.

The following table shows a list of devices actually used for each type of device

TABLE 1. ACTUALLY USED DEVICES

Type	Device	Feature	Value
wide-angle camera	MS LifeCam Studio	Resolution	1280X720
depth sensor	Occipital Structure Sensor	Distance	0.4~3.5m
		Resolution	640X480
		Precision	0.5mm at 0.4m 30mm at 3m (1%)
zoom camera	WONWOO MM-302B	Zoom Ratio	Opticalx30
		Resolution	1080p 30fps
pan-tilt units	Piezoelectric Technology PUMR60(E)	Max. Touque	5.0 [Kg+cm]
		Resolution	4,000 [P/R]
NIR illuminators	EPILEDs BN-D4242J-A3	Wavelength	860nm
		LED Array	9X8

The overall devices are integrated into a remote gaze tracking device shown at Figure 1. It has two layers. The upper layer just has a zoom camera and tile unit, and the lower layer has the others.

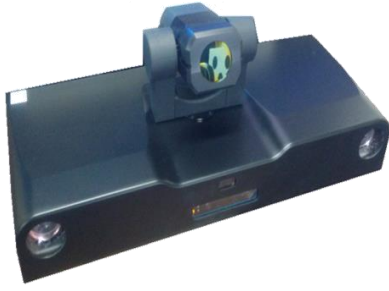


Figure 1. The proposed remote gaze tracking device

B. Stereo Calibration between Wide-angle Camera and Depth Sensor

For the gaze tracking, high resolution NIR eye image have to be acquired based on the 3D world coordinate of a user's eye.

To get the exact 3D world coordinate of a user's eye in the RGB image captured by wide-angle camera, the depth value corresponding to the eye pixel must be identified on a depth image captured by depth sensor. In order to do so, it needs to match up the pixel coordinate of RGB image with the pixel coordinate of depth image. So, we carry out stereo calibration, and then acquire extrinsic parameters (R, t) between two sensors and intrinsic parameters of each sensor. In addition, we carry out rectification, and then acquire rectification maps for each sensor.

Especially, we cover the IR illuminators on the Occipital Structure Sensor and use a tungsten lamp during calibration process as shown at Figure 2(a) to turn off the structured IR light and then get the chessboard patterns clearly as shown at Figure 2(b).

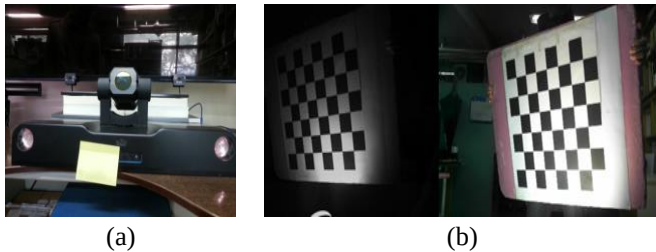


Figure 2. Hidden IR illuminators on depth sensor(a) and chessboard images captured by depth sensor and wide-angle camera (b)

Figure 3 is the result image of pixel mapping between RGB image and depth image using extrinsic parameters (R, t). It shows that the pixel coordinate of RGB image and depth image is exactly matched up.



Figure 3. The result image of pixel mapping between RGB and depth image

C. PTZ Calibration between Wide-angle Camera and Zoom Camera

The calibration between the wide-angle camera and the zoom camera derives the 3D-2D projection matrix and the coefficients of zoom pulse function used later to calculate the pan-tilt angles and zoom pulses for capturing a target object. The process of calibration between wide-angle camera and zoom camera is as in the following:

- 1) Calculate the 3D world coordinates of the points of intersection on chessboard using wide-angle camera and depth sensor
- 2) Locate each point of intersection on chessboard in the center of IR image captured by zoom camera as shown at Figure 4.
- 3) Get the data in forms of (3D world coordinate, pan-tilt angles) and (3D world coordinate, the pulse of the zoom motor).
- 4) Calculate the data in forms of (3D world coordinate, 2D pixel coordinate in the normalized image plane) from the former data. Here, the normalized image plane is a virtual image plan placed on one unit distance in front of the optical center of zoom camera
- 5) Calculate the 3D-2D projection matrix between 3D world coordinate and 2D pixel coordinate in the normalized image plane.
- 6) Calculate the coefficients of the function between the distance of an object and the pulse of the zoom motor from the latter data.

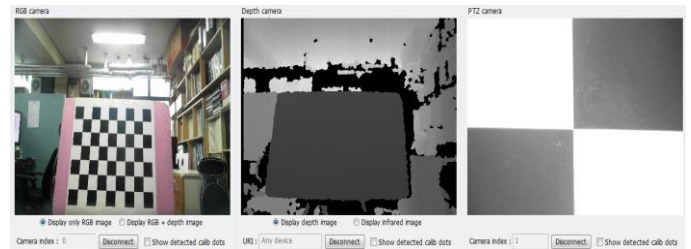


Figure 4. Calibration between wide-angle camera and zoom camera

III. ACQUISITION OF EYE IMAGES

A. Face & Eye Detection

The gaze tracker detects the face and eye of a user on the images captured by the wide-angle camera. It detects them using Viola-Jones algorithm [6], which combines Haar-like feature and Adaboost classifier. In addition, it tracks the face region using Kalman filter which predicts the face position in succeeding images to reduce the processing time.

B. Calculation of 3D World Coordinate of an Eye

After getting the eye point (x, y) on RGB image, we calculate the 3D world coordinate of the eye using the epipolar geometry and extrinsic/intrinsic parameters acquired in the stereo calibration. The process is as in the following:

- 1) Resize the RGB and depth image to be equal each other. In other words, unify both images to the smaller one.

- 2) Adjust the eye point (x, y) according to the image size ratio when the RGB image is reduced.
- 3) Find a point (u, v) on rectified coordinate from the adjusted eye point (x', y') using the rectification maps acquired from the stereo calibration as described earlier.
- 4) Find a point (x'', y'') on the depth image and Z distance from the depth sensor per each point on the v 'th row, called as epipolar line on rectified coordinate of the depth image.
- 5) Calculate the 3D world coordinate (X, Y, Z) of the eye from the wide-angle camera using the point (x'', y'') , extrinsic parameters and intrinsic parameters of the depth sensor.
- 6) Calculate a point (x''', y''') on RGB image using the 3D world coordinate (X, Y, Z) of the eye and the intrinsic parameter of the wide-angle camera
- 7) Output the 3D world coordinate (X, Y, Z) of the eye when the point (x''', y''') is most adjacent to the adjusted eye point (x', y') .

C. Acquisition of High Resolution NIR Eye Images

After acquiring the 3D world coordinate of a user's eye, gaze tracker calculates the pan-tilt angles and zoom pulses to capture high resolution NIR eye image as follows:

- 1) Calculate the 2D pixel coordinate in the normalized image plane from the 3D world coordinate of the eye using the 3D-2D projection matrix acquired in the PTZ calibration.
- 2) Calculate the pan-tilt angles from the 2D pixel coordinate in the normalized image plane.
- 3) Calculate the zoom pulses using the function between the distance of an object and the pulse of the zoom motor.
- 4) Pan, tilt, and zoom the zoom camera according to the pan-tilt angles and the zoom pulses.
- 5) Acquire high resolution NIR eye images.

IV. EXPERIMENT RESULTS

After the gaze tracker is placed on the table underneath a 60 inch TV, and users are located about 3m away from the TV, the high resolution eye image acquisition experiments are performed. Figure 5 shows that the zoom camera can track very well the movement of a user. In detail, the upper left image shows that a target object is located in the gaze tracker's right. So it can be found in the upper right image that the zoom camera is panned right facing toward the object. The lower left image shows that the target object is located in the gaze tracker's left. So it can be found in the lower right image that the zoom camera is panned left facing toward the object.

Figure 6 shows the high resolution NIR eye image of the target object captured by the zoom camera shown in Figure 5. Here, the zoom camera has been set to capture only left eye of the target object.

Through the Figure 5 and 6, we can come to a conclusion that the gaze tracker can track an eye of the target object exactly and capture high resolution NIR eye image for gaze

tracking even when the target object moves freely in front of TV.

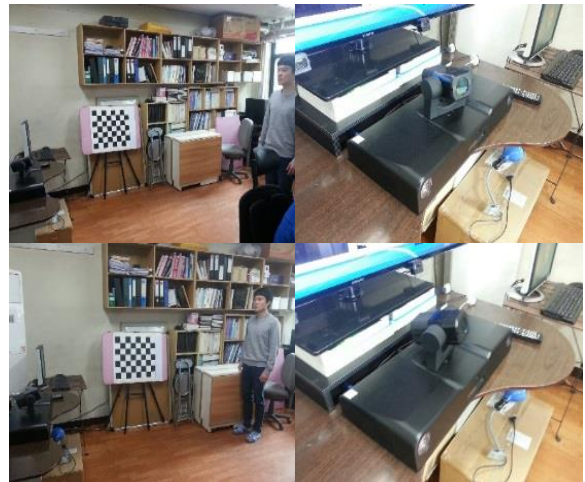


Figure 5. Zoom camera following the movement of a target object.



Figure 6. High resolution NIR eye image of the target object

V. CONCLUSIONS

In conclusion, we designed and developed a gaze tracking device to provide comfortability in gaze tracking and freedom of head movement of a user to be applicable to the large display environment with a long Z distance.

In addition we carry out stereo calibration and pixel mapping between the wide-angle camera and the depth sensor, and propose the PTZ calibration method between the wide-angle camera and the zoom camera to make the device continuously and effectively track and capture the eye image even though a user's irregular movements in 3D space

The experiment result of the stereo calibration and pixel mapping shows that the pixel coordinate of RGB image and depth image is exactly matched up. The experiment result of the PTZ calibration shows that the device can calculate the pan-tilt angles and zoom pulses to capture the high resolution NIR eye images, and then track very well the movement of a user based on the data.

In future, we will detect the pupil center and glints on the NIR eye image by applying ellipse fitting and adaptive thresholding method. And then, we will calculate a gaze position using 2D mapping function on the basis of the vector between the pupil center and the glint.

ACKNOWLEDGMENT

This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIP) [R0101-15-293, Development of Object-based Knowledge Convergence Service Platform using Image Recognition in Broadcasting Contents]

REFERENCES

- [1] <http://www.tobii.com/en/eye-tracking-research/global/products/hardware/tobii-glasses-eye-tracker/>
- [2] <http://www.tobii.com/eyex>
- [3] <https://theeyetribe.com/>
- [4] <http://www.umoove.me/>
- [5] <http://www.eyetechds.com/specialty-hardware.html>
- [6] P. Viola, M. Jones, "Robust real-time face detection," Computer Vision, Vol. 2, 747-747 (2001).



HeeKyung Lee was born in Daegu, Korea, in 1976. She received her BS degree in Computer Engineering in 1999 from Yeungnam University, and MS in Information & Communication Engineering in 2002 from KAIST-ICC.

In 2002, she joined Electronics and Telecommunications Research Institute (ETRI) of Korea and now serving as a senior member of engineering staff. She participated in "TV-Anytime" standardization and IPTV GSI Metadata

standardization. She currently involved in the development of gaze tracking technology. Her research interests include personalized service via metadata, HCI, Gaze Tracking, Bi-directional advertisement and video content analysis.



Jeongil Seo was born in Goryoung, Korea, in 1971. He received the Ph.D. degree in electronics from Kyoungpook National University (KNU), Daegu, Korea, in 2005 for his work on audio signal processing systems.

He was worked as a member of engineering staff at the Laboratory of Semiconductor, LG-semicon, Cheongju, Korea, from 1998 until 2000.

He has worked as a director at the Immersive Media Research Section, Electronics and

Telecommunications Research Institute (ETRI), Daejeon, Korea, since 2000. His research activities include image processing, audio processing, multi-modal user interface, and realistic broadcasting systems.



Hoon Jo was born in Ulsan, Korea in 1986. He received B.S degree in 2012 at Hanyang University in Seoul, and has been studying on a master's and doctorate integrated program at same university. The major field of study of the author is computer vision.

He has been researched vision sensor based gaze tracking in the graduate school. He published three related conference papers: A robust gaze tracking method for users wearing glasses (Jeju-si, Jeju-do:

Workshop on Multimedia, 2013), Fast and accurate locating method of pupil center and glints for gaze tracking (Bangkok, Thailand: IWAIT, 2014), and Eye feature localization under visible light (Seoul, Korea: IEIE Signal processing society, 2014) His current research interest is stereo vision