Journal of Convergence for Information Technology e-ISSN 2586-4440
Vol. 11. No. 1, pp. 20-27, 2021 DOI : https://doi.org/10.22156/CS4SMB.2021.11.01.020

# HCI를 위한 시선추적 시스템에서 분해능의 추정기법

김기봉[*], 최현호
대전보건대학교 의료IT융합과 교수

# Resolution Estimation Technique in Gaze Tracking System for HCI

Ki-Bong Kim[*], Hyun-Ho Choi
Professor, Dept. of Health and IT Convergence, Daejeon Health Institute of Technology

**요 약** 시선추적은 NUI 기술 중의 하나로 사용자가 응시하는 곳을 추적을 통해 알아낸다. 이 기술은 텍스트를 입력하거나 GUI를 제어할 수 있도록 하고 더 나아가 사용자의 시선 분석도 가능하게 하여 상업 광고 등에 응용될 수 있도록 한다. 시선추적 시스템은 영상의 품질과 사용자 움직임의 자유도에 따라 허용범위가 달라진다. 따라서 시선추적의 정밀도를 미리 추정하는 방법이 필요하다. 시선추적의 정확도는 하드웨어적인 변수 외에도 시선추적 알고리즘을 어떻게 구현하느냐에 따라 많은 영향을 받는다. 이에 따라 본 논문에서는 영상에서 동공 중심의 가능한 최대 이동 거리의 추정으로 동공 중심이 한 픽셀 움직일 때 시선은 몇 도가 바뀌는지 즉, 이론적 최대 분해능이 얼마인지를 추정하는 방법을 제시한다.

**주제어 :** HCI, 시선 추적, 허프 변환, 시야각, 중심와

**Abstract** Eye tracking is one of the NUI technologies, and it finds out where the user is gazing. This technology allows users to input text or control GUI, and further analyzes the user's gaze so that it can be applied to commercial advertisements. In the eye tracking system, the allowable range varies depending on the quality of the image and the degree of freedom of movement of the user. Therefore, there is a need for a method of estimating the accuracy of eye tracking in advance. The accuracy of eye tracking is greatly affected by how the eye tracking algorithm is implemented in addition to hardware variables. Accordingly, in this paper, we propose a method to estimate how many degrees of gaze changes when the pupil center moves by one pixel by estimating the maximum possible movement distance of the pupil center in the image.

**Key Words :** HCI(human computer interaction), Eye detection, Hough transform, FOV(Field of View), Fovea

## 1. Introduction

The visual-based HCI includes facial expression analysis, gesture recognition, and gaze tracking, and audio-based HCI includes voice recognition and speaker recognition[1]. In addition, cases in which more than one physical sensor such as mouse, keyboard, joystick, and haptic sensor are used are classified as sensor-based HCI. From another point of view, HCI is an interface that inputs commands with a keyboard in the early days of the computer's emergence, through the GUI where the user

---

[*] Corresponding Author : Ki-Bong Kim(kbkim@hit.ac.kr)

Received December 8, 2020 Revised January 6, 2021
Accepted January 20, 2021 Published January 28, 2021

selects and inputs the location of objects and commands displayed on the screen. It can be said that it is developing into a NUI that uses as a command [2]. Eye tracking is one of the NUI's, a technology that finds out where the user is looking. By tracking the user's gaze, it is possible not only to input text or control the GUI, but also to analyze the user's gaze, which was not possible with the existing interface. In other words, it is an important means of knowing what users are interested in and how much they are focusing on.

In fact, the eye tracking application is being used as an important technology for the disabled who have difficulty using the existing input method in terms of text input or GUI control [3]. In order to be used as a means of communication for the disabled, not only the gaze of the disabled must be tracked, but also a GUI for communication must be provided. Recently, related research and product development are also in progress. In addition, eye tracking technology is often required in places where user gaze analysis is required, and is used to improve web page design by analyzing gaze patterns when users view web pages, or to analyze user immersion in advertising content. It is also used for marketing research, such as analyzing consumers' shopping patterns and usage patterns by making them as small devices that can be worn like glasses.

Most eye tracking systems consist of near-infrared illumination and an image acquisition device capable of photographing the wavelength band of the illumination to track the user's gaze. The reason for using near-infrared illumination is not only to suppress various obstacles that may cause errors in image processing compared to images in the visible wavelength range, but also to compare the location of the reflected light reflected on the cornea by near-infrared illumination with the location of the center of the pupil. This is because the user's gaze can be accurately tracked. Therefore, near-infrared illumination is also called reference illumination.

## 2. Eye detection technology

In order to track the user's gaze with a non-wearable device, the user's eye position must first be detected in the input image. As a method of detecting the user's eye position, it can be divided into a method of using active lighting with hardware assistance and a method of detecting eyes using only software image processing.

### 2.1 Active lighting-based technology

The active illumination-based eye detection method detects eyes by flashing near-infrared illumination to generate a bright pupil effect. In this method, one of the near-infrared lights is placed on the optical axis of the camera and the other is placed outside the optical axis of the camera, and the near-infrared lights are continuously flashed in synchronization with the capture period of the camera. An example of an eye tracking system using active lighting is shown in Fig. 1. In an image captured using active lighting, one frame is photographed with a dark pupil and the next frame is photographed with a bright pupil, so the difference can be used to detect the eyes. However, this method incurs a burden on the hardware configuration and is weak against interference from external light sources such as reflected light from glasses.

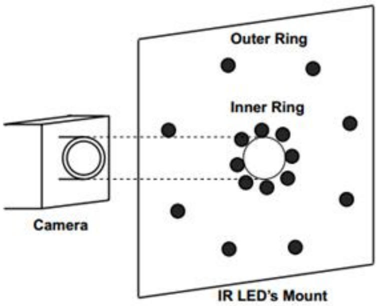

Fig. 1 Example configuration of camera and lighting

### 2.2 Image processing based technology

Eye detection based on image processing is a method of detecting an eye area using only an image processing technique from an input image without hardware assistance. Image processing-based eye detection methods include a method of detecting eyes using a learned classifier, a method of detecting eyes based on color, and a method of detecting using the shape of an eye. In a method of using the learned classifier, after learning the classifier in advance, when an actual image is input, the learned classifier detects the eye region. The classifier is mainly learned by the Adaboost algorithm, which receives a large number of eye image samples and non-eye image samples and trains the classifier.

However, the method using a classifier requires a large amount of samples to learn the eye shape and a lot of time to learn it. In addition, there is a problem in that it is difficult to detect images in situations other than the trained classifier, such as rotation or changes due to glasses. n particular, methods of detecting the face first and then detecting the eyes are difficult to use when the face is covered.

Another image processing-based eye detection method is the color-based eye detection method proposed by Nasiri et al. [14-16]. This method separates the luminance component and the chrominance component in the YCbCr color space, generates two maps for the eye candidate region in each channel, and then finds the eye position in the image using a binarization technique through geometrical test. However, this method is vulnerable to the rotation of the face and the obstruction of the face, and is not particularly suitable for eye-tracking images using the infrared wavelength band.

A method of detecting eyes using the Hough transform based on the feature that the iris is circular was also proposed by Kawaguchi et al. In this method, a face region is first detected, a region having a low brightness value within the face region is detected, and then an iris candidate is selected from the detected region. This is a method of determining the eye area by comparing the cost of the selected iris candidate areas based on Hough transform. However, this method does not work when one eye is covered, and there is a problem with weak rotation. As a method of detecting eyes based on image processing, there is also a method using a template proposed by Yuille et al. However, it is known that the template matching method requires designing the eye model first, and generally requires a lot of processing time.

There is also a study to detect eyes using reflected light in an application for eye tracking. The method proposed by Cho et al. divides the input image into N×M subblocks and calculates the difference between the maximum and minimum brightness values in each subblock. The four subblocks having the largest difference between the maximum and minimum values are selected, and the average value of the selected four subblocks is calculated. If this value is less than the average of all 4 subblocks, it is selected as the eye area. However, the above method has a problem that cannot be applied to a wearer of glasses.

## 3. Eye tracking technology

### 3.1 Structure and characteristics of the eye

The anatomy of the human eye is shown in Fig. 2.

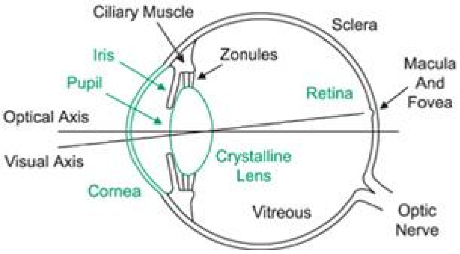

**Fig. 2 Structure of the eye**

Humans perceive an object as light from the eye is condensed onto the retina, because cones and rods, which are light receptors, are distributed in the retina. Cones are color-sensitive receptors, concentrated in fovea, and the number is more than 6 million. Rods are receptors sensitive to brightness and spread throughout the retina, and the number of rods is 75 to 150 million, which is much higher than that of cones. Therefore, the part that is sensitive to both color and brightness is the fovea, and the image formed on the fovea is clearly recognized, and other parts are not clearly recognized. Since the fovea has a diameter of about 1.5mm and an angle of about 1 to 2°, it means that only about 1 to 2° is clearly recognized in the area inputted by the eye.

As shown in Fig. 2, it can be seen that the fovea does not exist on the optical axis passing through the center of the eyeball and the center of the pupil. That is, the optical axis and the optical axis generally do not coincide, and the difference is called the kappa angle. Since this kappa angle varies among individuals, in order to accurately track the gaze, each individual must undergo a calibration process before gaze tracking.

One more important thing you need to know to track your gaze is about eye movement. From the perspective of eye tracking, the movement of the eyeball can be largely divided into the stages of eye fixation and saccade. The rapid movement of the eyeball that occurs in order to gaze at another place from the eyeball fixation stably looking at one place is called the leap movement. However, even at the stage of eye fixation, the eyeball shows movement regardless of the intention of the person. The movement during fixation is tremor, which is a very fine and fast movement in the range of about 0.008°, and drift that moves about 2 to 5Hz in the range of about 0.1°. And it can be divided into micro saccade, which is a fast motion to compensate for slippage. There is a need for a post-processing method that reduces the effect of movement in the fixed eye condition. The cornea of the eye is characterized by reflecting light. In the field of eye tracking technology, research has been conducted to more accurately track the user's gaze by using the corneal reflected light of the reference illumination caused by this property.

### 3.2 Eye tracking based on non-wearable devices

The non-wearable device-based gaze tracking technology refers to a technology that tracks the gaze of a user without wearing a device for photographing eyes. Therefore, a camera and a near-infrared reference light are installed in a place away from the user, and the user's face is photographed to track the gaze. Although some eye-tracking studies in the visible light wavelength band are being conducted, there is a limitation in accuracy, so most eye-tracking techniques use corneal reflected light generated by near-infrared reference illumination. This method, unlike wearable device-based eye tracking technology, can perform eye tracking only by first detecting the user's eye position in an image.

Most of the non-wearable device-based gaze tracking technology targets PC users within 1m of a short distance, but in recent years, researches to extend the distance available for gaze tracking to be used as a TV control interface are also actively underway. In such long-distance eye tracking, there is a problem that the screen does not include the user or the resolution of the image around the eyes is degraded depending on the angle of view of the camera. Therefore, a wide-angle camera to determine the user's location and a narrow-angle camera to acquire high-quality eye images are simultaneously used. Research to do is also ongoing.

### 3.3 Wearable device-based eye tracking

Wearable device-based gaze tracking technology is a technology that tracks gaze by using eye images acquired from a user-wearable camera such as glasses or head mounted display(HMD). When the object of eye tracking moves with an athlete or a product buyer, a wearable eye tracking device that can constantly acquire eye images regardless of the user's movement is required. The wearable device photographs only the eye area by attaching a near-infrared light and a camera under the glasses. The principle of operation of the wearable device-based eye-tracking technology also uses corneal reflected light, similar to the non-wearable eye-tracking technology, and calculates the user's gaze point based on its relative position with the pupil.

## 4. Estimating the resolution of the eye tracking system

The camera used to track the gaze can be selected according to the purpose of gaze tracking, from low-cost models such as webcams to expensive models such as precision industrial cameras. Also, depending on the focal length of the lens, the image quality of the captured image and the allowable range of the user's freedom of movement are different, so it is necessary to select an appropriate one according to the eye tracking application. While there are applications that need to track the user's gaze with great precision, there are applications that require only a general grasp of which area they are viewing. Therefore, there is a need for a method of estimating eye tracking accuracy so that it is possible to grasp in advance whether the developed eye tracking system provides appropriate precision for the application.

The meaning of each symbol is as follows.

- $W_s$: Screen width (mm)
- $W_c$: Image sensor width (mm)
- $A_h$: Camera's horizontal FOV (degree)
- $d_c$: Distance from eye to camera (mm)
- $d_s$: Distance from eye to screen (mm)
- $\theta$: Eye tracking range angle (degree)
- $\alpha$: The angle of the movement range of the pupil center to the camera (degree)
- r: Radius of the eyeball (mm)
- $e_{mm}$: Distance of movement in the center of the pupil in the eye tracking range(mm)

Since the vertical resolution can be estimated by the same method as the horizontal resolution, this paper presents a method of estimating the resolution based on the horizontal resolution. The resolution of the eye tracking system used in the PC means how finely it is possible to distinguish between the center of the pupil when looking at the leftmost of the screen, which is the range of eye tracking, and the center of the pupil when looking at the right.

Since the image is recorded in pixel units in the captured image, the theoretical maximum resolution of the eye tracking system can be defined as in Equation (1).

$$R = \theta / e_{pxl} \tag{1}$$

In Equation (1), $e_{pxl}$ is a value expressed in pixels as the maximum distance that the pupil center can move in the horizontal direction in the captured image. $e_{pxl}$ can be calculated according to the total number of pixels in the horizontal direction of the captured image as shown in Equation (2).

$$e_{pxl} = \alpha \cdot p_x / A_h = 2\tan^{-1}\left(\frac{e_{mm}}{2d_c}\right) \cdot p_x / A_h \tag{2}$$

The linear distance the pupil's center moves can be approximated as in (3), so $e_{pxl}$ can be expressed as in (4).

$$e_{mm} \cong 2r \cdot \sin\frac{\theta}{2} = 2r \cdot \sin\left(\tan^{-1}\left(\frac{W_s}{2(d_s+r)}\right)\right) \tag{3}$$

$$e_{pxl} \cong 2\tan^{-1}\left(\frac{r}{d_c}\sin\left(\tan^{-1}\left(\frac{W_s}{2(d_s+r)}\right)\right)\right)P_x / A_h \tag{4}$$

On the other hand, the FOV can be calculated using Equation (5) with the focal length of the lens and the size of the image sensor.

$$A_h = 2\tan^{-1}\left(\frac{W_c}{2f}\right) \tag{5}$$

Therefore, Equation (4) can be expressed as a function of the size and focal length of the image sensor as shown in Equation (6).

$$e_{pxl} \cong \tan^{-1}\left(\frac{r}{d_c}\sin\left(\tan^{-1}\left(\frac{W_s}{2(d_s+r)}\right)\right)\right)P_x / \tan^{-1}\left(\frac{W_c}{2f}\right) \tag{6}$$

According to the result of equation (6), the resolution of equation (1) can be expressed by equation (7).

$$R \cong \frac{2\tan^{-1}\left(\frac{W_c}{2f}\right) \cdot \tan^{-1}\left(\frac{W_s}{2(d_s+r)}\right)}{\tan^{-1}\left(\frac{r}{d_c}\sin\left(\tan^{-1}\left(\frac{W_s}{2(d_s+r)}\right)\right)\right) \cdot P_x} \tag{7}$$

As can be seen from Equation (7), the parameter for calculating the resolution is composed of the focal length of the image sensor and lens, the size of the monitor, and the distance to the user, excluding the radius r of the eyeball. Therefore, the resolution of the eye tracking system can be estimated by using the known average eyeball radius (11-12mm) as an estimate of r that is difficult to actually measure.

The experiment assumes a case of tracking the gaze of a user staring at a 19-inch monitor sitting in front of about 70cm with a point gray grasshopper3 camera and a 35mm focal length lens that acquires 2048×2048 images. It can be seen that the resolution of the system decreases as the range of gaze tracking increases. In addition, as the number of pixels of the captured image increases, the resolution improves, and it can be seen that the degree of improvement in the resolution decreases as the image size increases. And as the focal length increases, the resolution improves rapidly at the beginning, and the degree of improvement decreases gradually. And it can be estimated that the resolution will improve linearly as the distance between cameras gets closer.

Through actual experiments, we looked at how accurately the proposed resolution is estimated. As can be seen from Equation (1), since the accuracy of the resolution estimation is determined by the accuracy of $e_{pxl}$, in the experiment, it was confirmed that the pupil movement distance in pixels was properly estimated. $e_{pxl}$ was measured using three focal lengths of 25, 35, and 50mm. In order to get the pupil movement without moving the face as much as possible, a face holder was used, and the pixels of the actual pupil movement distance were calculated by overlapping the images when staring at the left and right ends of the screen.

## 5. Conclusion

The first problem in actual non-wearable eye tracking applications is the failure of eye detection due to unintended interference from external light sources, the user's wearing glasses, face rotation, or face occlusion. If eye detection fails, eye tracking itself becomes impossible, so an eye detection algorithm that is robust against environmental changes as described above is required. In addition, in the eye tracking application, a camera with a narrow field of view (FOV) is generally used to acquire a user's eye image. The narrower the FOV, the higher the resolution of the acquired image, so that more accurate gaze tracking is possible, but there is a problem that the user's movement is limited because the spatial range included in the image is narrow. Therefore, in order to allow the user's movement more freely in the eye-tracking application, it is necessary to detect eyes even in an image where only one eye is captured outside of the shooting range of a part of the face.

The eye and pupil center point detection algorithm that is robust to environmental changes in this paper has the advantage of not requiring a lot of learning data and a long learning time, unlike detecting eyes through learning because it uses the common features of eye tracking images. In addition, the proposed method can detect the center of the pupil more accurately by suppressing the influence of the eyelashes and corneal reflection light, which leads to improved accuracy of eye tracking.

## REFERENCES

[1] F. Karray, M. Alemzadeh, J. A. Saleh, and M. N. Arab. (2008). Human-Computer Interaction: Overview on State of the Art. *International Journal on Smart Sensing and Intelligent Systems, 1(1),* 137-159.

[2] A. J. Hornof and A. Cavender. (2005). EyeDraw: enabling children with severe motor impairments to draw with their eyes. *In Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 161-170).

[3] G. Buscher, E. Cutrell & M. R. Morris. (2009). What do you see when you're surfing?: using eye tracking to predict salient regions of web pages, *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 21-30).

[4] S. Zhao & R. R. Grigat. (2006). Robust eye detection under active infrared illumination, *In Proceedings of the International Conference on Pattern Recognition* (Vol. 4, pp. 481-484). IEEE.

[5] Z. Zhu, K. Fujimura & Q. Ji. (2002). Real-time eye detection and tracking under various light conditions. *In proceedings of the ACM SIGCHI symposium on eye tracking research and applications* (pp. 139-144).

[6] X. Zhang, Q. Tang, H. Jin, Y. Qiu & Y. Guo. (2012). Eye location based on adaboost and random forests. *Journal of Software, 7(10),* 2365-2371.

[7] H. C. Lee, W. O. Lee, C. W. Cho, S. Y. Gwon, K. R. Park, H. Lee, and J. Cha. (2013). Remote Gaze Tracking System on a Large Display, *Sensors*, 13 (10), 13439-13463.

[8] P. Wang, M. B. Green & Q. Ji. (2005). Automatic eye detection and its validation. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp.164-164).

[9] T. Kawaguchi & M. Rizon. (2003). Iris detection using intensity and edge information. *Pattern Recognition, 36(2),* 549-562.

[10] B. Kim, H. Lee & W. Kim. (2010). Rapid eye detection method for non-Glasses type 3D display

on portable devices. *IEEE Transactions on Consumer Electronics, 56(4),* 2498-2505.

[11] A. L. Yuille, P. W. Hallinan & D. S. Cohen. (1992). Feature extraction from faces using deformable templates. *International Journal of Computer Vision, 8(2),* 99-111.
DOI : 10.1007/BF00127169

[12] P. Artal, A. Benito & J. Tabernero. (2006). The human eye is an example of robust optical design. *Journal of Vision, 6(1),* 1-7.
DOI : 10.1167/6.1.1

[13] R. C. Gonzalez & R. E. Woods. (2001). Digital Image Processing. *Addison-Wesley Longman Publishing Co.,* Inc., Boston, MA.

[14] H. C. Lee, D. T. Luong, C. W. Cho, E. C. Lee & K. R. Park. (2010). Gaze tracking system at a distance for controlling IPTV. *Consumer Electronics, IEEE Transactions on, 56(4),* 2577-2583.
DOI : 10.1109/TCE.2010.5681143

[15] S. T. Jang, J. H. Lee, J. Y. Jang & W. D. Jang. (2018). Gaze Tracking with Low-cost EOG Measuring Device. *Journal of Korea Convergence Society, 9(11), 53-60.*
DOI : 10.15207/JKCS.2018.9.11.053

[16] H. Kim, S. Lee, H. H. Han & J. Kim. (2020). Saliency Attention Method for Salient Object Detection Based on Deep Learning. *Journal of the Korea Convergence Society, 11(12), 39-47.*

김 기 봉(Ki-Bong Kim) [종신회원]

- 1991년 2월 : 충남대학교 전산학과 (전산학학사)
- 1993년 2월 : 충남대학교 전산학과 (전산학석사)
- 1998년 8월 : 충남대학교 전산학과 (전산학박사)
- 1997년 3월 ~ 현재 : 대전보건대학교 의료IT융합과 교수
- 관심분야 : 데이터베이스, 의료정보시스템
- E-mail : kbkim@hit.ac.kr

최 현 호(Hyun-Ho Choi) [정회원]

- 1994년 2월 : 충남대학교 전산학과 (전산학학사)
- 1997년 2월 : 충남대학교 전산학과 (전산학석사)
- 2000년 2월 : 충남대학교 컴퓨터과학과(박사수료)
- 2002년 3월 ~ 현재 : 대전보건대학교 의료IT융합과 교수
- 관심분야 : 병렬처리, 정보통신
- E-mail : hyuno@hit.ac.kr