Development of Low-Cost Vision-based Eye Tracking Algorithm for Information Augmented Interactive System

Seo-Jeon Park, Byung-Gyu Kim*

Abstract

Deep Learning has become the most important technology in the field of artificial intelligence machine learning, with its high performance overwhelming existing methods in various applications. In this paper, an interactive window service based on object recognition technology is proposed. The main goal is to implement an object recognition technology using this deep learning technology to remove the existing eye tracking technology, which requires users to wear eye tracking devices themselves, and to implement an eye tracking technology that uses only usual cameras to track users' eye. We design an interactive system based on efficient eye detection and pupil tracking method that can verify the user's eye movement. To estimate the view-direction of user's eye, we initialize to make the reference (origin) coordinate. Then the view direction is estimated from the extracted eye pupils from the origin coordinate. Also, we propose a blink detection technique based on the eye apply ratio (EAR). With the extracted view direction and eye action, we provide some augmented information of interest without the existing complex and expensive eye-tracking systems with various service topics and situations. For verification, the user guiding service is implemented as a proto-type model with the school map to inform the location information of the desired location or building.

Key Words: Deep learning, Object detection, Eye tracking, Eye action, Augmented information, Interactive service.

I. INTRODUCTION

Along with the fourth industrial revolution, the deep learning is one of the key technologies that will drastically change our human life in near future. In fact, the concept of deep learning technology has been proposed since the 1950s, and the basic technology was already established in the 1990s of the 1990. The performance of graphics processing units (GPUs) and central processing units (CPUs) that process data will be developed, and big data, which is the material for machine learning, will be in the spotlight after the mid of 2000s.

Big data is pouring into the spotlight. Google ran two deep learning-related teams until 2012, but now runs more than 1,000 artificial intelligence-related projects in search, Gmail, YouTube, Android and Google Maps, while IBM also operates only two projects by 2011. It has recently split into more than 30 teams and is strengthening Watson [1].

As for the creative interactive advertisements that elicit public response, the digital ad industry said, "Many experts are exposed to brands and products for the advertising effects of uniform advertising images and images, but the consumer-participating interactive ads are only more effective." So how can we get information within content that peoples actually look at carefully and care about? While tracking a person's eyes is the most accurate method, the current eye-tracking function is quite complex and requires a tracker [2], [3].

In addition to the development of deep learning technology and the new industry that is occurring by combining various service sectors, the company has set the direction of the project with a service that introduces deep running technology to the digital advertising industry, which has been steadily gaining popularity in the advertising industry, and aims to use it to provide interactive window services that increase users' interest and interest [4], [6], [8], [9].

To detect the eye region, the face detection is very important task such as [11], [12], [13]. From detected face region, we can extract the eye part by using Haar feature. Also, some object tracking methods have been reported in [14], [15], [16]. They employed some deep learning structures to trace the object region more accurately. But if

the system can process the detection task in real-time, these tracking approaches are not needed usually.

This paper aims to implement the eye-tracking technology using only the cheap camera without the eye-tracking machine and applies the object recognition technology required for the project using the previously introduced deep running. Also, we aim to provide an interactive window service based on efficient eye detection and pupil tracking technology that recognizes the user eye's movement and enhances the information of interest.

This paper will be organized as follows: In Section II, we will introduce the developed eye detection and tracking method in detail. Section III will give a proto-type service and some results which has been implemented. In Section V, we will give a concluding remark in final.

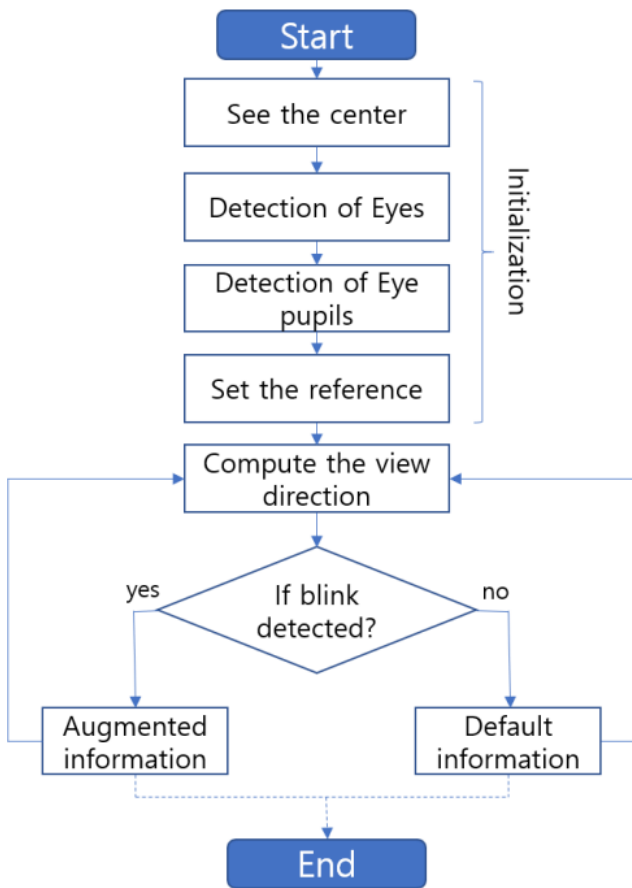## II. PROPOSED EYE TRACKING ALGORITHM

Fig. 1. Overall procedure of the developed scheme.

Figure 1 shows the overall procedure of the developed scheme. First, we initialize to make the reference (origin) coordinate. Then the view direction is estimated from the extracted eye pupils. Then if the blink is detected, the augmented information is presented. Otherwise, the default information is shown.

2.1. Face Detection

The proposed algorithm adopts a face marker detector that extracts and maps 68 point coordinates through deep learning as shown in Fig. 2. Since only the eye area was required for this project, we detect the face area and map it to the input image to identify the area of both eyes using the labeled name instead of the coordinate number of each part.

Fig. 2. Detected feature (face marker) points.

2.2. Eye Pupil Extraction

Once coordinates have been placed on each part of the face through Detect Face Parts, use the labeled name instead of the coordinate number to identify the area of both eyes. Then, we specify each eye as the region-of-interest (ROI). We can use a good and fast segmentation technique to extract the pupil region in [6].

Because this image is blurred, the contrast between light and dark is improved through smoothing and medium-sized filters. After this filtering process, we can find eyes (pupil area) by using the Hough circles function to find a circle. We condition the size of the eye radius and the range of the circle center coordinates so that we find only circles that are likely to be eyeballs. Figure 3 shows the extracted result of the pupil area by the Hough circle transform.

2.3. Eye View Direction Initialization

If an experiment is carried out through an eye tracker, the calibration process of setting the reference point as if adjusting the zero of the scale is carried out first. This is a process that helps to produce accurate results, and the algorithm also tracks the user's eyes after the initialization process is preceded by a baseline [2]. The proposed algorithm also proceeds with initialization, a step in creating a new baseline for a specified number of frames.

During initialization, the user creates a baseline by averaging the center coordinates of the effective circles detected, looking at the center red circle with a pause. Figure 3 shows an example for estimation of view direction
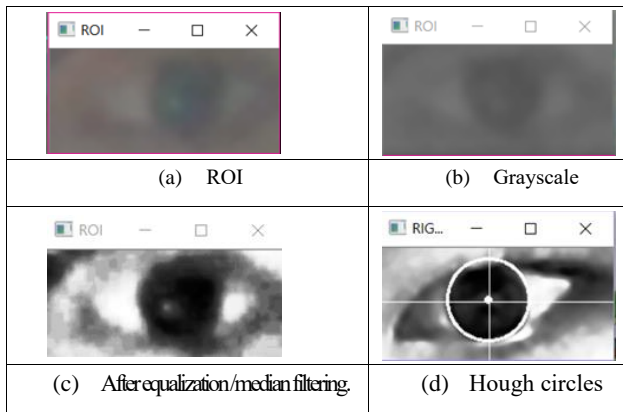
vector from the base line point (cross mark).

(a)　ROI	(b)　Grayscale
(c)　After equalization/median filtering.	(d)　Hough circles

Fig. 3. Extraction of the pupil area in the eye image.

2.3. Estimation of View Direction

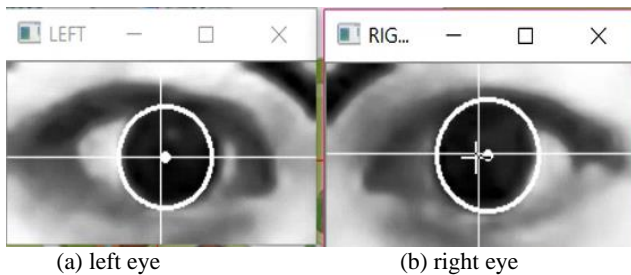| (a) left eye	(b) right eye

Fig. 4. Estimation of the view direction vectors.

In this stage, we compute the direction vectors from the reference point. The proposed algorithm determines which area of the image is being viewed through how far the center coordinates of the circle detected after the initialization is away from the baseline. When a human looking at an object, the human eye does not stop but constantly moves its eye pupil [5]. Therefore, the area determined based on the center coordinates of the circle detected in the pupil went up and down the area represented on the screen. It can be shaken.

Therefore, the last five numbers are averaged to determine the number of the area being viewed to stabilize the area. Figure 4 shows an example for estimation of view direction vectors from the base line point (cross mark). In actual, we need to compute the view directions from both the detected eye pupils. We take the average value of the estimated direction vectors.

2.4. Eye blink detection

Eye link detection uses only the left and right eye coordinates of the coordinates mapped to the critical area of the face through Detect face parts. The main principle of detecting eye flicker is to determine that the eyes are closed

13

when the difference in the coordinate values of the vertical axis coordinates (P2 and P6, P3 and P5) of both eyes is not great [10]. To express this in a formula, there is the eye apply ratio (EAR) (top of Figure 5) as:

$$EAR = \frac{||p_2 - p_6|| + ||p_3 - p_5||}{2||p_1 - p_4||}, \tag{1}$$

where p_i is the specified position of the detected eye region. This is a fraction of the difference in coordinate values between horizontal and vertical axis coordinates (P1 and P4). In EAR, the difference between the values of the denominators P1 and P4 are very little, so it can be seen that the EAR values are similar when the eye opened, but when the eye is closed, the EAR suddenly becomes smaller as shown in Fig. 5 (bottom). Through experiments, we have set 0.15 as the threshold value to detect the eye blink status.
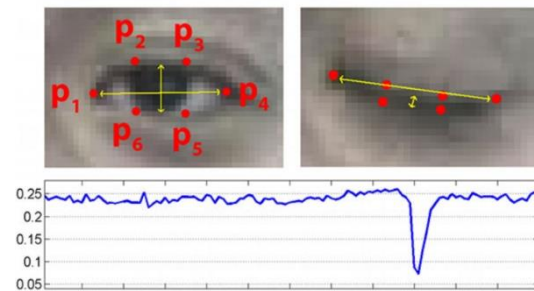

Fig. 5. Coordinate and distance measure for the blink detection.

This eye blink operation is mapped into the click event in the system. That is, the system run an event when the user's eye is blinked once.

III.　　PROTO-TYPE　　SERVICE IMPLEMENTATION

3.1. System Structure

We developed using python and IDLE as an editor. Various libraries were used to receive, process, and display images, but OpenCV was typical to locate each part of the face in real time and refer to the detect face parts using deep learning algorithm for both eyes. The database server was built using the RDS on AWS and MySQL was used as the server engine

3.2. Service Scenario

Examples of actual use scenarios are as follows: If the program is running and people are detected on the webcam, a greeting message is posted. Then, to indicate that the

initialization phase is carried out first at the beginning, a notice is posted as shown in Fig. 6 to indicate that the initialization phase will go through the initialization phase soon. When the guide instruction disappears, a red circle appears in the middle of the screen, as shown in Fig. 7, and the user looks at the circle in front of the screen for a certain period of time.

Fig. 6. A guide instruction for service initialization.

Fig. 7. The service initialization process.

Fig. 8. An illustration of the selected building by the user's view direction.

The location of the user's eye was divided into large areas and processed to show the area in different colors. Figure 8

shows the building where the eyes are.

Once the initialization is completed properly, it can be verified that the building is marked red according to the user's gaze. In addition, when user blinks their eyes while looking at the building for additional information, the designed augmented information about the building will appear on the right for a specified period of time as a statement (Figure 9).

Fig. 9. The information augmented case (red colored part) and its detailed information with text.

3.3. The Performance of Accuracy

For measuring the recognition accuracy, 24 subjects were recruited to measure the user's perspective on the service and the accuracy of the service choice. They used the service directly using the developed eyeball tracking algorithm, and each time they determined the selection of a building at a particular location and whether the information was accurately displayed on the screen. Each person tried 3 times for each localization and service event selection. Through this experiment, we measured the accuracy of the exact event.

Table 1 shows the accuracy of the localization and service event selection. The accuracy of the eye pointing was achieved by 88.4% and 85.5% of the service selection accuracy was obtained when a user blinks to select the specific building. The final accuracy of service was 85.5% in the developed system.

Table 1. The performance of the localization and service event selection.

Items	Accuracy (%)
Pointing event	**88.4**
Service selection	**85.5**

Table 2 shows the consumed time as frame per seconds (FPS). Usual commercial webcam supports the full high-

definition (HD) (resolution of 1920x1080). In this resolution, we achieved 22.3 FPS to make final event action. It means that the developed system is able to be applied for real-time system.

Table 2. The performance of the processing time (FPS/consumed time).

Items	FPS	Time per a frame (ms)
Processing time	22.3	44.84

III. CONCLUSION

In this paper, we have developed an interactive and information augmented service based on the efficient eye tracking algorithm which have the potential to develop into a breakthrough technology that can complement the limitations of the existing expensive eye-tracking technology that relied on sensor or physical tracking hardware. We achieved near real-time eye tracking-based interactive system with 85.5% of the service accuracy.

The developed system can be applied by changing the subject matter to a variety of services and situations, as well as mapping services.

REFERENCES

[1] Won-Jun Hwang, "Research Trends in Deep Learning Based Face Detection, Landmark Detection and Face Recognition," *Journal of Broadcast Engineering*, pp. 41-49, 2017.

[2] Lee, Changwook, "Direction for Sustainable Development of Interactive Advertising Media Evolved Cases: The focus on placed Interactive advertisement," Dankook Univ, 2011.

[3] Fadhil Noer Afif, Ahmad Hoirul Basori, "Vision-based Tracking Technology for Augmented Reality: A Survey," *International Conference on Digital Media*, Vol. 1, pp. 46-48, 2012.

[4] Ji-Ho Kim, "Understanding, Present Condition and Suggestion of the Eye-TRACKING Methodology for Visual and Perceptual Study of Advertising", *The Korean Journal of Advertising and Public Relations*, Vol. 19, No. 2, 2017.

[5] Jae-Woo Jung, "A Study on Visual Reaction of TV viewers through the Eye Tracker : Focused on News, Entertainment and Home Shopping Programs", *Ph. D. Thesis*, Graduate School of Hansung University, 2018.

[6] JI-HAE KIM, BYUNG-GYU KIM, PARTHA PRATIM ROY, DA-MI JEONG, "Efficient Facial Expression Recognition Algorithm Based on Hierarchical Deep Neural Network Structure," *IEEE Access*, Vol. 7, Dec. 2019.

[7] Byung-Gyu Kim, J.I. Shim, D. J. Park, "Fast Image Segmentation Based on Multi-resoluition Analysis and Wavelets," *Pattern Recognition Letters*, Vol. 24, No. 16, pp. 2995-3006, 2003.

[8] Ji-Hae Kim, Gwang-Soo Hong, Byung-Gyu Kim, Debi P. Dogra, "deepGesture: Deep Learning-based Gesture Recognition Scheme using Motion Sensors," *Displays*, Vol. 55, pp. 35-48, 2018.

[9] Pradeep Kumar, Subham Mukerjee, Rajkumar Saini, Partha Pratim Roy, Debi Prosad Dogra, and Byung-Gyu Kim, "Plant Disease Identification using Deep Neural Networks," *Journal of Multimedia Information System*, Vol. 4, No. 4, pp. 233-238, December 2017.

[10] Soukupová, T., & Čech, J., "Real-Time Eye Blink Detection using Facial Landmarks," *The 21-st Computer Vision Winter Workshop*, Czech Technical University in Prague, pp. 1~8, Feb. 2016.

[11] Lu Leng, Jiashu Zhang, Jing Xu, Khaled Alghathbar, "Dynamic weighted discrimination power analysis: a novel approach for face and palmprint recognition in DCT domain," *International Journal of Physical Sciences*, Vol. 5, No. 17, pp. 467-471, 2010.

[12] Lu Leng, Jiashu Zhang, Jing Xu, Muhammad K. Khan, Khaled Alghathbar, "Dynamic weighted discrimination power analysis in DCT domain for face and palmprint recognition," *International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 467-471, Nov. 2010.

[13] Leng L., Zhang J., Chen G., Khan M.K., Alghathbar K., "Two-directional random projection and its variations for face and palmprint recognition," *Computational Science and Its Applications-LNCS*, Vol. 6786, pp. 458-470, 2011.

[14] Yue Yuan, Jun Chu, Lu Leng, Jun Miao, Byung-Gyu Kim, "A scale adaptive object tracking algorithm with occlusion detection," *J. Image Video Proc.*, Vol. 2020, No. 7, pp. 1-14, 2020.

[15] J. Chu, X. Tu, L. Leng and J. Miao, "Double-channel object tracking with position deviation suppression," *IEEE Access*, Vol. 8, pp. 856-866, 2020.

[16] J. Chu, Z. Guo and L. Leng, "Object detection based on multi-layer convolution feature fusion and online hard example mining," *IEEE Access*, Vol. 6, pp. 19959-19967, 2018.

Authors

Seo-Jeon Park received her BS degree in the Department of NanoPhysics from Sookmyung Women's University, Republic of Korea, in 2019. In 2019, she joined the Department of IT Engineering for pursuing her MS degree at Sookmyung Women's University, Republic of Korea.

Her research interests include feature extraction, pattern recognition, and facial expression recognition algorithms in computer vision.

Byung-Gyu Kim has received his BS degree from Pusan National University, Korea, in 1996 and an MS degree from Korea Advanced Institute of Science and Technology (KAIST) in 1998. In 2004, he received a PhD degree in the Department of Electrical Engineering and Computer Science from Korea Advanced Institute of Science and Technology (KAIST). In March 2004, he joined in the real-time multimedia research team at the Electronics and Telecommunications Research Institute (ETRI), Korea where he was a senior researcher. In ETRI, he developed so many real-time video signal processing algorithms and patents and received the Best Paper Award in 2007.

From February 2009 to February 2016, he was associate professor in the Division of Computer Science and Engineering at SunMoon University, Korea. In March 2016, he joined the Department of Information Technology (IT) Engineering at Sookmyung Women's University, Korea where he is currently an associate professor.

In 2007, he served as an editorial board member of the International Journal of Soft Computing, Recent Patents on Signal Processing, Research Journal of Information Technology, Journal of Convergence Information Technology, and Journal of Engineering and Applied Sciences. Also, he is serving as an associate editor of Circuits, Systems and Signal Processing (Springer), The Journal of Supercomputing (Springer), The Journal of Real-Time Image Processing (Springer), Helyion Journal (Elsevier), and International Journal of Image Processing and Visual Communication (IJIPVC). From 2018, he is serving as the Editor-in-Chief (EiC) of the Journal of Multimedia Information System. He also served as Organizing Committee of CSIP 2011 and Program Committee Members of many international conferences. He has received the Special Merit Award for Outstanding Paper from the IEEE Consumer Electronics Society, at IEEE ICCE 2012, Certification Appreciation Award from the SPIE Optical Engineering in 2013, and the Best Academic Award from the CIS in 2014. He has been honored as an IEEE Senior member in 2015.

He has published over 240 international journal and conference papers, patents in his field. His research interests include software-based image and video object segmentation for the content-based image coding, video coding techniques, 3D video signal processing, wireless multimedia sensor network, embedded multimedia communication, and intelligent information system for image signal processing. He is a senior member of IEEE and a professional member of ACM, and IEICE.