Aided Eyes: Eye Activity Sensing for Daily Life

Yoshio Ishiguro†, Adiyan Mujibiya†, Takashi Miyaki‡, and Jun Rekimoto‡,§

†Graduate School of Interdisciplinary Information Studies,
The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo Japan
‡Interfaculty Initiative in Information Studies,
The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo Japan
§Sony Computer Science Laboratories, 3-14-13 Higashigotanda, Shinagawa, Tokyo, Japan
{ishiy, adiyan, miyaki, rekimoto}@acm.org

ABSTRACT

Our eyes collect a considerable amount of information when we use them to look at objects. In particular, eye movement allows us to gaze at an object and shows our level of interest in the object. In this research, we propose a method that involves real-time measurement of eye movement for human memory enhancement; the method employs gaze-indexed images captured using a video camera that is attached to the user's glasses. We present a prototype system with an infrared-based corneal limbus tracking method. Although the existing eye tracker systems track eye movement with high accuracy, they are not suitable for daily use because the mobility of these systems is incompatible with a high sampling rate. Our prototype has small phototransistors, infrared LEDs, and a video camera, which make it possible to attach the entire system to the glasses. Additionally, the accuracy of this method is compensated by combining image processing methods and contextual information, such as eye direction, for information extraction. We develop an information extraction system with real-time object recognition in the user's visual attention area by using the prototype of an eye tracker and a head-mounted camera. We apply this system to (1) fast object recognition by using a SURF descriptor that is limited to the gaze area and (2) descriptor matching of a past-images database. Face recognition by using haar-like object features and text logging by using OCR technology is also implemented. The combination of a low-resolution camera and a high-resolution, wide-angle camera is studied for high daily usability. The possibility of gaze-guided computer vision is discussed in this paper, as is the topic of communication by the photo transistor in the eye tracker and the development of a sensor system that has a high transparency.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*Theory and methods*

General Terms

Information extracting for lifelog

Keywords

Eye tracking, Lifelog computing, Gaze information

1. INTRODUCTION

Lifelog systems have been a topic of considerable research [3]. The development of a lifelog computing system has led to the hope that the human memory can be augmented. For extracting beneficial information from augmented human memory, we consider the "five W's and one H" (Who, What, When, Where, Why, and How). These provide very important contextual information. Location estimation methods can answer "Where" [12], and a wearable camera can provide the other information. However, we cannot accurately detect a person's actions by using only image information. According to visual lifelog researches, it is definitely necessary to extract important parts of life events from enormous amounts of data, such as people, objects, and texts that we pay attention to. Therefore, we consider using eye activity for obtaining contextual information.

Eye tracking has been extensively studied in medical, psychological, and user interface (UI) researches [5, 6] for more than a century. The study of eye tracking has provided us with a considerable amount of information such as gazed object, stress, concentration ratio, and degree of interest in the objects [4]. Interaction research using eye tracking has been studied. In particular, in this field, wearable computing research has been actively studied using eye movements (gaze information) because wearable devices allow intuitive and free-hand control [17].

Even though the current eye tracking method was developed several decades ago, it still involves the use of headgear with horns that are embedded camera or requires electrodes to be pasted on the user's face and/or other large-scale systems to be used for a psychological experiment. In other words, this system currently cannot be used for daily activities. In this case, a "daily usable system" means a commonly acceptable system that can be used in public in daily life. Moreover, making the system accurate as well as portable is a complicated task. A daily usable system for eye activity sensing could be utilized in many research areas such as wearable computing.

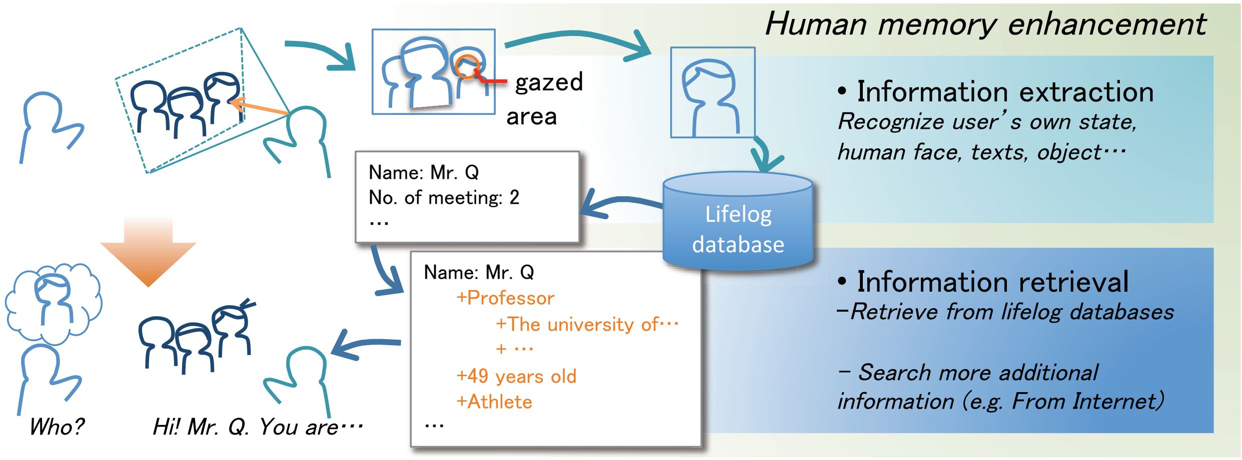

Figure 1: Concept of the eye enhanced lifelog computing

In this research, for human memory enhancement, we examine a method to extract significant information from large-scale lifelog data by using eye activity information. We develop a new method that involves real-time measurement of eye movements for automated information extraction. The method makes use of gaze-indexed images captured by a video camera and an eye tracker with a low accuracy but a high *wearability*; both the camera and the eye tracker are attached to the user's glasses.

2. EYE-ENHANCED LIFELOG COMPUTING

Lifelog with a surrounding video image can give us a considerable amount of information. On the other hand, humans obtain surrounding information from their eyes and gaze at interesting objects. However, it is impossible to record this type of information by using only a camera. Consequently, first, the gazed information is detected from a video. Then, the gazed objects and the user's state are extracted from the video and the eye movement. After this, related information is retrieved by the extracted information. Finally, the results are added to lifelog as shown in Figure 1. For such reasons, we need to record three types of eye activity —**gaze direction**, **eye movement**, and **eye blink frequency**— for using lifelog and UI methodology. Details of each type of eye activity are explained in this section.

2.1 Gaze Direction

It is difficult to extract significant information from the video image of large-scale personal life log data. For example, omnidirectional-camera video images contain a considerable amount of information that is not related to human memories; the camera image may not relate to the gazed object. Therefore, it is difficult to know which objects are being focused on only from the images. In this research, obtaining a video lifelog with gaze information is our objective. Gazed objects such as faces and texts are extracted from a video lifelog, and this information is used for understanding whom you met, what you saw, and what you were interested in.

Gaze direction is used for pointing in the UI research area; however, it is well known that has "Midas touch problem" and that it is difficult to use gaze direction without a trigger such as a key input [8].

2.2 Eye Movement

Not only the gaze direction but also the eye movement has meaning. In particular, **microsaccade** indicates the target of one's potential interest [4]. The microsaccade is a very quick eye movement, almost 600°/s, and it is a spontaneous movement caused when the eye gazes at stable targets. The frequency and direction of this movement change depending on the person's interest in the target. The measurement of this movement makes it possible to know human susceptibilities. The holding time of a gazed object is a conscious movement, and the saccadic movement is an unconscious movement. Therefore, it is possible to extract more information from a susceptible mind by the measurement of saccadic movements.

2.3 Eye Blink Frequency

Previous research shows that eye movements can provide information about a person's condition. For example, the eye blink frequency shows the person's degree of concentration on his/her work [15].

The eye blink frequency decreases when a person concentrates on his/her work. In contrast, the frequency increases when he/she is not very focused on his/her work. Therefore, the measurement and detection of the eye blink frequency can estimate the person's level of concentration. The eye blink has several characteristics. It is a motion that is approximately 150 ms fast. An involuntary eye blink is an automatic eye blink that has a shorter motion time than the voluntary eye blink, which is a conscious eye blink.

3. DESIGN OF EYEGLASS-EMBEDDED EYE ACTIVITY SENSOR

3.1 Requested Specification for Eye Sensing

The capability requirement is discussed in Section 2. Eye movements are typically classified as ductions, versions, or

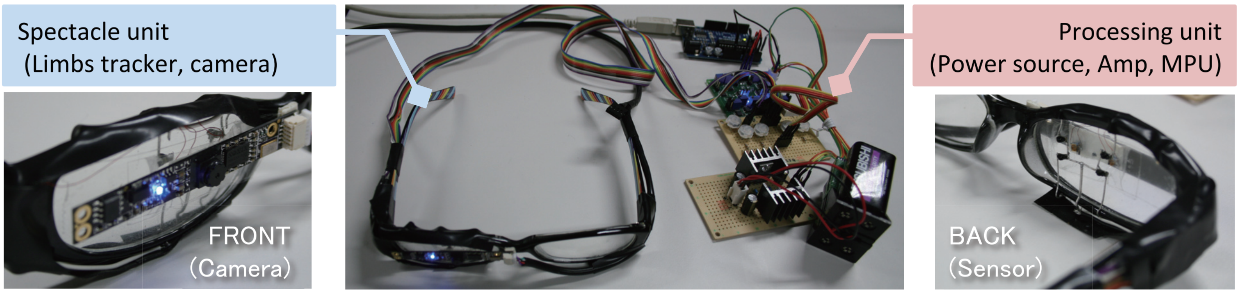

Figure 2: Prototype eye gaze recognizer with camera for lifelog

vergences. Eye movement has several moving speeds. There are several types of high-speed eye movements. For example, the microsaccade frequency is more than 1000 Hz, and the eye blink speed is around 150 ms. The method must distinguish precisely between eye movement and blinks for an accurate detection of eye movements. Further, the human view angle is almost 160° for each eye. Therefore, a 5° resolution is sufficient for information extraction because this system aims not only to achieve a high accuracy but also extract information by a daily usable system using a combination of eye activity information and image processing methods.

3.2 Eye-tracking Technology Candidates

There are several types of eye trackers. In this study, we consider in four different trackers:

Camera based system: The video-based systems [9, 11] can capture a gaze texture image. This is the most commonly used tracker; however, it requires an extremely sophisticated optics system having a light source, lenses, and half mirrors. Additionally, it requires a large (table top size) measurement system for quick eye movements (over 1000 Hz). Scale-wise, it is possible to develop a smaller system; however, currently, such a system cannot measure high-speed eye movements.

Search coil and Optical lever: These methods [13, 18] are used for laboratory experiments in a certain region of space. However, these methods are not user friendly as the users are expected to wear special contact lenses that using a negative pressure on their eyes.

Electrooculogram (EOG): Eyes have a steady electric potential field, and this electric signal can be derived by using two pairs of contact electrodes that are placed on the skin around one eye. This is a very lightweight approach [2] and can work if the eyes are closed. However, it requires an eye blink detection method and has other issues. For example, an electrode is required and is affected by electronoise.

Infrared corneal limbus tracker: An infrared corneal limbus tracker [14] is also a very lightweight tracker. It can be built by using a light source (infrared LED) and light sensors (phototransistor) and only requires very low computational power. This approach is also affected by noise from environmental light. However, this is a very simple approach; no electrodes are required. This approach can sufficiently detect eye blinks. Therefore, it has a high constructability for daily use.

Therefore, we use an “infrared corneal limbus tracker” in our study. This method has a lower accuracy than the method of search coil and optical lever. However, our purpose is to extract significant information; hence, the accuracy of this method can be enhanced by combining image processing methods and contextual information such as eye direction.

3.3 Prototype of Eye Activity Sensor

Four phototransistors and two infrared LEDs are mounted on the eye glassed as shown in Figure 2. A small camera is mounted on the glasses for recording surrounding information, and not for eye tracking. An infrared LED and four phototransistors are mounted inside of the glasses.

The infrared light is reflected by the eye surface and is received by the phototransistor. These sensor values throw to instrumentation amplifier and analog/digital (AD) conversion, then input to the microprocessing unit (MPU). In this study, ATmega128 from Atmel is used for the MPU and AD conversion. The MPU clock frequency is 16 MHz, and the AD conversion time is 16μs per channel.

Before the measurement, the head position and the display are fixed for a calibration, and then, the display shows the targets to be gazed in the calibration (Figure 3). The sensor wearer gazes at the target object on the display, and the MPU records the sensor value. One target has 240 points (W 20 points x H 12 points) and each points are gazed for 1 second. After the calibration, the system estimates the gaze direction by using the recorded data. The recorded data and sensor value are compared first. Then, the center of gravity is calculated from the result in order to estimate the gaze direction. Simple method is enough for this research because only gaze area in the picture is needed to know for using information extraction system.

3.4 Life Events Extracting System

When an infrared limbus tracking method is used, the sensor value is changed rapidly by eye blinking. The speed is approximately 150 ms, as shown in Figure 4. Therefore, the system can simply distinguish between blinks and other eye movements. Further, the system extracts information as face, texts, and preregistered objects. Pre-registered objects are recognized in real time by the user’s visual attention area. We use fast object recognition by using the SURF [1]

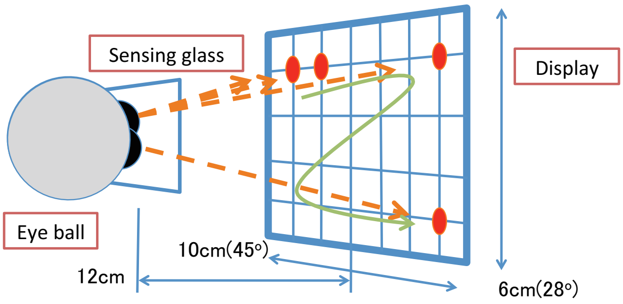

Figure 3: A calibration method for the gaze recognizer system. The head position and the display are fixed for a calibration and then the display shows targets. A user gazes target object on the display and MPU records sensor value.

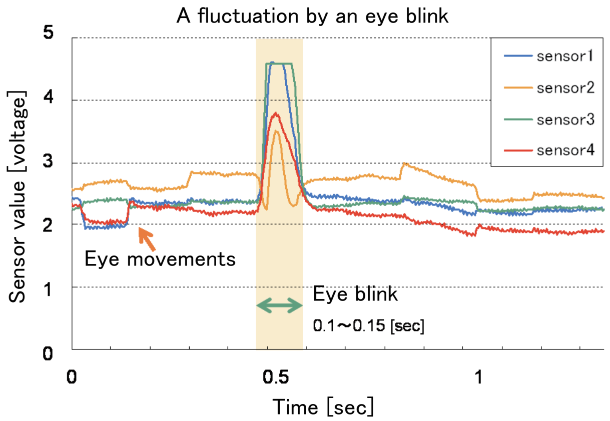

Figure 4: An example of a fluctuation in the sensor data by an eye blink

Figure 5: Image feature extraction by SURF [1], for real time object recognition

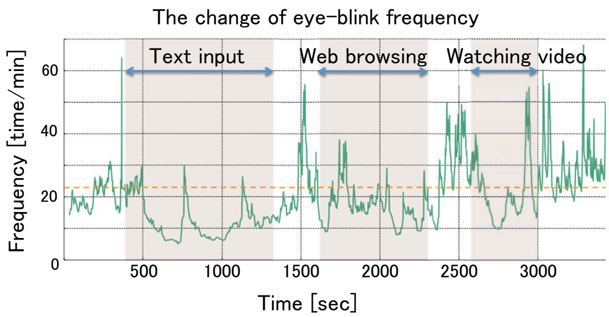

Figure 6: An example graph of eye blink frequency

descriptor for matching images that is limited to the gazed area with the past-images database (Figure 5).

Face recognition using haar-like objects by "OpenCV Library[1]" is implemented for logging of "When I meet someone." This method can extract the human face first, and then the system records the time, location, and face image.

Additionally, text logging with the OCR technology "tesseract-ocr[2]" is implemented. This system can extract a clipped image, wherein it is clipped that gazed area of head-mounted camera image. This system attempts to extract text from these clipped images. Finally, the extracted text is recorded along with time and location data for life logging.

4. CASE STUDY USING PROTOTYPE SYSTEM

4.1 Preliminary Experiment

An infrared limbus tracker is a commonly used tracker; therefore, the detail of hardware evaluation experiment is spared. We checked the specifications of the proposed prototype system. More than 99% of the eye blinks were detected in 3 min. Very slow eye blinks caused the 1% failure of detection. The gaze direction angle was 5°, and the processing rate was set as 160 Hz in the preliminary experiments.

4.2 Concentration State Measurements

Our system can detect eye blinks with a high accuracy. We recorded the eye-blink detection and the user's tasks for approximately 1 hour, as shown in Figure 6. The results showed that the eye blink frequency changed with a change in the tasks. The frequency was slower when the user concentrated on the objects. Therefore, the system could tell the user's concentration states and we consider that can use for human interface technique such as displaying and annotation.

4.3 Life Event Extraction

The proposed method in this study extracts pre-registered objects, human face, and character by using images and eye gaze information. Figures 7 and 8 show the extraction of objects such as posters. In this situation, the user observes each poster of the 100 pre-registered posters in the room.

[1]http://opencv.willowgarage.com/wiki/
[2]http://code.google.com/p/tesseract-ocr/

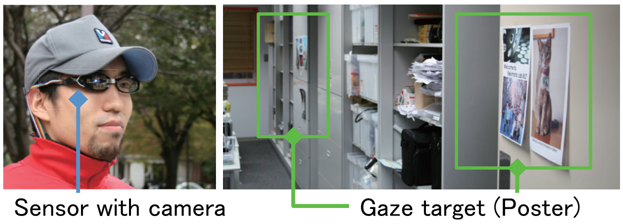

Figure 7: Photographs of experimental environment

Figure 8: Object recognition scene by proposed system. This figure shows that the object recognition system can identify two different objects next each other.

The IDs of these extracted objects are logged with time, actual images, and the eye direction when the system detects the pre-registered objects, as shown in Figure 9. Figure 10 shows the optical character reading of the gazed information. An image of the gazed area is clipped, characters are extracted from the clipped image. Additionally, the face image is extracted along with the actual time, as shown in Figure 11. Usually, when multiple people stand in front of the camera, such as in a city or a meeting room, the normal recorded video image does not tell you who you are looking at. However, this method can pick up who you are looking at by using gaze information. Our system can handled with multiple objects that shown up in head-mounted camera. Finally, these three pieces of data are logged automatically.

5. HIGHER REALIZATION OF DAILY USABILITY AND FUTURE POSSIBILITES

From these case studies, it is concluded that information extraction by means of image processing requires the use of a wide-angle, high-resolution camera for providing more accurate information. However, it is difficult to mount such a device on a person's head. Moreover, the prototype of the infrared limbus tracker is very small, but the phototransistors obstruct the user's view. In this section, a combination of a wide-angle, high-resolution camera and a head-mounted camera along with the limbus tracker structure without phototransistors is discussed.

5.1 Combination with High Resolution Wide Angle Camera

Having a large-size camera such as a commercially used

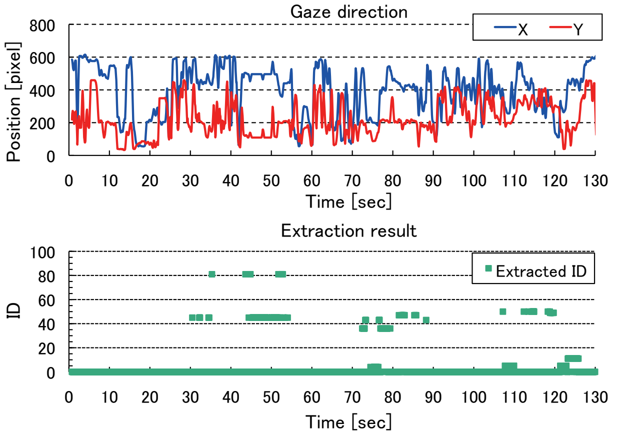

Figure 9: Gaze direction and extraction results (ID 0 means no objects was extracted)

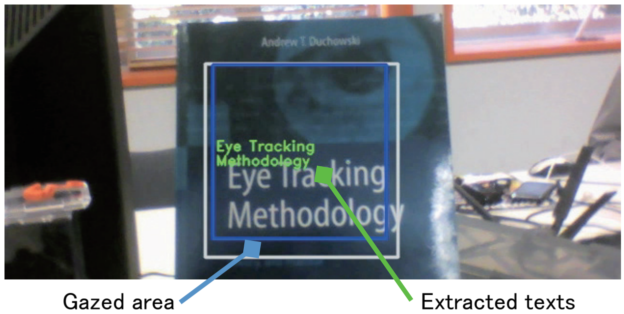

Figure 10: An example image of OCR extraction for clipped image by gaze information using tesseract-ocr

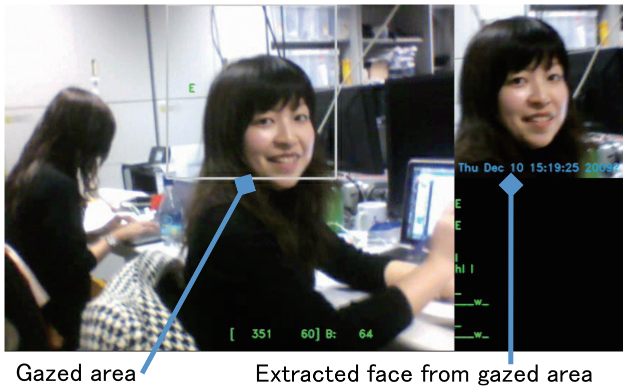

Figure 11: An example image of face extraction. Faces are extracted from clipped image of head-mpunted camera by gaze information.

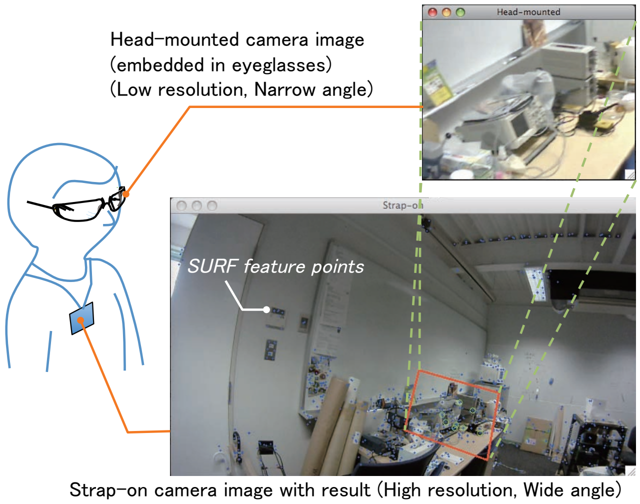

Figure 12: An example image of view point in wide-angle camera by head-mounted camera. Gazed position in head-mounted camera is known, thus it is possible to project gaze position to high resolution camera image by using position relation of two images.

USB camera mounted on the head interferes in daily communication. Therefore, we embed a very small, poor-resolution camera for capturing the surrounding information in the eye tracker. Hence, this camera can be integrated into the user's eye glasses and can capture the user's actual view. On the other hand, the small camera has a very poor performance, and it is difficult to obtain a high frame rate and a high resolution by using such a camera. Therefore, the image processing of information extraction methods is at times not possible. Therefore, we consider a *strap-on* camera (such as SenseCam [7] that can be dangled around one's neck) that has fewer problems than a head-mounted camera. Strap-on cameras do not disturb any communication and can be attached to the body more easily than a head-mounted camera. Therefore, we can use a high-resolution camera with wide-angle lens. This prototype system compares the SURF descriptor between the head-mounted camera and the strap-on camera and then calculates the homography matrix. From the results, we can identify the focus of the head-mounted camera from the strap-on camera's images. As a result, a high-resolution image can be used for the information extraction, as shown in Figure 12.

5.2 Improving Transparency of Eye-tracker

Developing a new system for daily use that is so comfortable that the user is not even aware of wearing it is our long-term objective. The infrared limbus tracker has a very simple mechanism; therefore, it has highly possibility that modification of camera based system. This tracker does not require lens and focal distance. The camera-based system can use a half mirror to see the eye image; however, the system has to be in front of the eyes, as shown in Figure 13.

Because of the above-mentioned reasons, we consider a *transmissive sensor system*. The infrared limbus tracker does not have a focal point unlike a camera, and it is easy

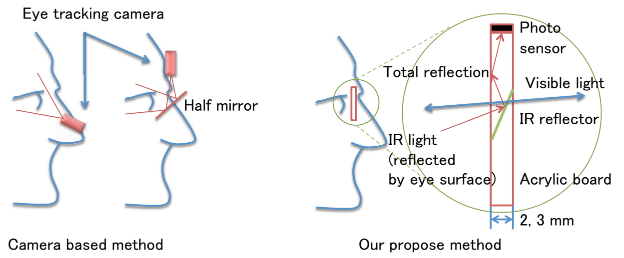

Figure 13: Illustrations of transparent infrared corneal limbus tracker

to design the light path, as explained in Figure 13. In this figure, acrylic boards (refractive index = 1.49) are chamfered at approximately 30°, and an infrared reflection filter is placed in between. The infrared light reflected by the eye completely reflected in the acrylic material, and then, the light is received by the phototransistor that is placed out of the user's view.

5.3 Modulated Light for Robustness Improvement and Using for Information Transmission

Since the infrared cornea limbus tracker is affected by the environmental light, this method needs to be devised such that the infrared light can be modulated for a lock-in amplifier (also known as a phase-sensitive detector) [16]. In other words, this tracker allows the measurement of environmental light through a reflecting eye surface. In fact, the embedded phototransistor received the modulated backlight of the normal display from the user's view during the experiments. This phenomenon with a lock-in amplifier can isolate the reflected light from the modulated tracker light source that measures eye movements and the modulated environmental light. It is also possible to get information from objects when the user gaze light sources as studied in [10].

6. CONCLUSIONS

In this research, we have described an infrared corneal limbus tracker system to measure the eye activity for contextual information obtained by information extraction from the lifelog database. It is possible to use the proposed method in daily life. In fact, we combined the low-accuracy, high-wearability eye tracker and image processing methods in our system. In the case study, we could detect the eye blinks with a high accuracy and estimate the participant's concentration state. Then, we combined this tracker and an image processing method such as face detection, OCR, and object recognition. Our eye tracking system and eye activity information successfully extracted significant information from the lifelog database.

Finally, we discussed the possibility of developing a transmissive sensor system with an infrared corneal limbus tracker and two cameras having different resolutions for our long-term objective of designing a system suitable for daily use. In addition, since the eyes follow objects even when the user's body moves, information about the eye direction can be used for image stabilization and it can be effective utilized in image extraction methods. We believe this research can contribute to the utilization of augmented human memory.

7. ACKNOWLEDGMENTS

This research was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for JSPS Fellows, 21-8596, 2009.

8. REFERENCES

[1] H. Bay, T. Tuytelaars, and L. V. Gool. Surf: Speeded up robust features. In *9th European Conf. on Computer Vision*, May 2006.

[2] A. Bulling, D. Roggen, and G. Tröster. Wearable eog goggles: eye-based interaction in everyday environments. In *Proc. of the 27th int. conf. extended abstracts on Human factors in computing systems*, pages 3259–3264, 2009.

[3] B. P. Clarkson. Life Patterns: structure from wearable sensors. *Ph.D thesis*, 2002.

[4] S. M. Conde and S. L. Macknik. Windows onthe mind. *Scientific American*, 297(2):56–63, 2007.

[5] A. Duchowski. *Eye Tracking Methodology*. Springer, 2007.

[6] J. M. Findlay and I. D. Gilchrist. *Active Vision: The Psychology of Looking and Seeing*. Oxford University Press, 2003.

[7] J. Gemmell, G. Bell, and R. Lueder. MyLifeBits: a personal database for everything. *Commun. ACM*, 49(1):88–95, 2006.

[8] R. J. K. Jacob. Eye movement-based human-computer interaction techniques: Toward non-command interfaces. In *Advances in Human-Computer Interaction*, pages 151–190. Ablex Publishing Co, 1993.

[9] D. Li, J. Babcock, and D. J. Parkhurst. openEyes: a low-cost head-mounted eye-tracking solution. In *Proc. of the 2006 symp. on Eye tracking research & applications*, pages 95–100, 2006.

[10] Y. Mitsudo. A real-world pointing device based on an optical communication system. In *Proc. of the 3rd Int. Conf. on Virtual and Mixed Reality*, pages 70–79, Berlin, Heidelberg, 2009. Springer-Verlag.

[11] T. Ohno. Freegaze : a gaze tracking system for everyday gaze interaction. *Proc. of the symp. on eye tracking research & applications symposium*, 2002.

[12] J. Rekimoto, T. Miyaki, and T. Ishizawa. Life-Tag: WiFi-based continuous location logging for life pattern analysis. In *3rd Int. Symp. on Location- and Context-Awareness*, pages 35–49, 2007.

[13] D. Robinson. A method of measuring eye movement using a scleral search coil in a magnetic field. In *IEEE Trans. on Bio-Medical Electrics*, number 10, pages 137–145, 1963.

[14] W. M. Smith and J. Peter J. Warter. Eye movement and stimulus movement; new photoelectric electromechanical system for recording and measuring tracking motions of the eye. *J. Opt. Soc. Am.*, 50(3):245, 1960.

[15] J. A. Stern, L. C. Walrath, and R. Goldstein. The endogenous eyeblink. *Psychophysiology*, 21(1):22–33, 1983.

[16] P. A. Temple. An introduction to phase-sensitive amplifiers: An inexpensive student instrument. *American Journal of Physics*, 43(9):801–807, 1975.

[17] D. J. Ward and D. J. C. MacKay. Artificial intelligence: Fast hands-free writing by gaze direction. *Nature*, 418:838, 2002.

[18] A. Yarbus. *Eye movements and vision*. Plenum Press, 1967.