Realtime 3D Position Detection of Human Pupil

Yoshinobu Ebisawa
Faculty of Engineering, Shizuoka University
Johoku 3-5-1, Hamamatsu, 432-8561 Japan
Phone: +81-53-478-1244, Fax: +81-53-478-1244, Email: ebisawa@sys.eng.shizuoka.ac.jp

Abstract – A real time 3D pupil position detection system necessary for non-contact and remote eye gaze detection allowing head displacement was developed. So far, methodology had been proposed that estimates 3D positions of eyes from their 2D positions detected from the face images captured by two stereo cameras. However, using eye position detection is not prospective for precise and stable eye gaze detection. In the method proposed in the present study, several near-infrared LEDs were arranged around the apertures of two cameras. These two sets of light sources were alternately switched on and off synchronously with the even/odd signal. After differentiating the bright and dark pupil images obtained from each of the two cameras, the center coordinates of the two pupils were detected. An image processing algorithm for precise and stable pupil detection, including the pupil detection method utilizing a blink, is described. In one of the experiments, the 3D position of the pupil of one eye was measured by the two stereo cameras. Another narrow camera, which detects the centers of both pupil and corneal reflection images necessary for eye gaze direction determination, placed on a pan-tilt unit could be automatically directed toward the 3D pupil position.

Keywords – pupil detection, eye gaze detection, 3D position measurement, human interface, human monitoring

I. INTRODUCTION

On many commercial 3D displays, the 3D image can be observed from the limited space. The observers must watch them from a narrow 3D space. Recently, some 3D displays can provide changes in perspective in response to horizontal, frontal, and vertical head movements within the tracking range. Such a system needs a head tracking system, strictly speaking, the information of the 3D eye positions, for perspective control [1].

On the other hand, we have been developing a non-contact and head-free eye gaze detection system. As shown in Fig. 1, we plan to develop the system, where the stereo wide cameras detect the 3D position of one of the two pupils and the narrow camera for gaze detection is directed toward the pupil by a pan-tilt drive unit. The 3D pupil position determines the camera-to-pupil vector. The narrow camera detects the centers of the pupil and the corneal reflection of a light source attached to the narrow camera. The difference between the centers of the pupil and corneal reflection, \( |r| \), in the narrow camera image reflects the angle between the camera-pupil vector and the line of sight, \( \theta \). The declination of the pupil center against the corneal reflection center, \( \phi' \), reflects \( \phi \) shown in Fig. 1. The line of sight forms the angles with the camera-pupil vector, \( \theta \) and \( \phi \), and passes through the pupil center. The gaze point can be determined as the intersection of the line of sight and the formulated visual object (e.g., screen of PC display). Here, \( |r| \) depends on the distance between the narrow camera and the eye. So \( |r| \) must be compensated using the distance. Accordingly, the distance must be accurately estimated for precise eye gaze determination.

In this idea, the precision of 3D pupil position measurement is important also for accurate gaze point detection. The similar idea has been proposed by Ohno et al. [2] and Talmi & Lie [3]. Their idea is to detect the 3D eye position using the eye images from stereo cameras. However, the eye image changes with the eye orientation. In addition, an iris is easily concealed by eyelids. Accordingly, the precision of 3D eye position measurement is not very hopeful. Moreover, in general, the template matching method is used for the eye image detection. The method needs high computer power and the preparation for making the template.

In contrast, the precision of 3D pupil position is prospective because the pupil is small and easy to detect if irradiating light for the face [4][5][6]. If a light source is set near the aperture of the camera, the pupil appears bright in the camera image (bright pupil image). If it is set apart from...
the camera aperture, the pupil appears dark (dark pupil image). However, in both setting conditions, the easiness of pupil detection depends on ambient lighting conditions, the colour of the iris and eyelashes, and so on. However, if subtracting the dark pupil image from the bright pupil image, the pupil is easily detected because the background images cancel each other.

Morimoto et al. [5] and Ji and Yang [6] used the image difference method and multiple infrared LEDs arranged in two rings as the light sources. The multiple LEDs increased the light power and lit up a wide range, producing the advantages for pupil detection, compared to our previous two-light source method [4]. However, the outer ring is large and it disturbs to miniaturize the total system.

The present paper describes a realtime 3D pupil detection system using the stereo cameras, which consists of two wide cameras and LEDs arranged around their apertures. In addition, a new image processing algorithm for stable and precise 2D pupil position determination is described.

II. METHODS

A. Step-up

Two standard, black and white, CCD cameras were set up above a computer display (Fig. 2). The two cameras are synchronized using a synch generator. In front of each camera, band-pass filter (half intensity width: 17nm) and several infrared LEDs were arranged in a ring shape around the aperture of each of the cameras. The LEDs attached to the two cameras were alternately switched on synchronously with the even/odd signal. On each camera, the bright pupil and dark pupil images (Fig. 3(a) and (b)) alternately appears every field. Due to the strong light power, in general, the face appeared brighter than the background.

The pupil center coordinates in the two camera images were obtained by a PC (Pentium 4, 3GHz) and a conventional image grabber board.

B. Image processing

The concept of image processing for pupil detection is as follows: Basically, the dark pupil image is subtracted from the bright pupil images. In general, in the difference images, pupils can be detected as brighter parts. However, bright parts also appear in the background or boundary of face, and so on. In addition, under bad conditions, e.g., when a user wears eyeglasses, the bright spots (glass reflection light) sometimes appear. To avoid mistakenly judging the spots as the pupils, windows are applied to the pupils once the pupils are detected. For reliable window initial setting, image variation caused by a blink is utilized. In the window, precise pupil detection is executed.

The concrete image processing algorithm for each camera image is described below. The explanation consists of three procedures, (A)-(C), and the entire flow, (D).

(A) Making of the difference image and the face area difference image:

Step 1. For eliminating the eyeglass reflection light [4], in the bright pupil image (320 x 240), a threshold was determined by the P-tile method for 10%-brighter part. The pixels of the brighter than the threshold were replaced by the threshold.

Step 2. The dark pupil image was subtracted from the modified bright pupil image.

Step 3. Neighborhood smoothing was executed in the differentiated image. We simply call the image at this stage as the difference image.

Step 4. To avoid mistakenly detecting the pupils outside the face area, the face area was determined as follows:

(i) The bright or dark pupil image was binarized by automatic threshold determination using the discriminant analysis method. (ii) Opening operation was executed to speed up the following labeling method. (iii) The image was labeled, and the widest area was judged as the face area. (iv) By filling the holes in the face area, the face area binary image was made (Fig. 4(a)).

Step 5. By Boolean AND operation between the face area binary image and the difference image obtained in Step 4, the face area difference image was made (Fig. 4(b)).

(B) Making of the eye area difference image utilizing a blink:

(i) The absolute difference image between the current bright or dark pupil image and the same image four frames before the current. (ii) 15% brighter part in the difference image was extracted as the eye area binary image by the P-tile method (fig. 5(a)). (iii) Boolean AND operation between this binary image and the face area difference image obtained in Step 5 was executed and the eye area binary image was made (Fig. 5(b)).
(C) Precise pupil center detection:
(i) When a pupil or two pupils were continuously existed in the previous, more than two frames, the pupil position in the present frame was predicted from the pupil position in the latest two frames using Kalman Filter.
(ii) In the difference image obtained in Step 3, rectangular window(s) (21 x 21) were applied to the predicted pupil position (Fig. 6(a)).
(iii) In the window(s), some moving objects, such as eyelids (Fig. 6(b)) and the relics of the removed eyeglass reflection light images may appear. The use of the separability filter [7] may be a good candidate. However, since the method determines the pupil center by the pixel, 3D pupil position determination with high resolution is not hopeful. So, first, the image in the window was weighted by the 2D normal distribution function whose maximum coincides with the window center. Second, pupil parameter, \( P(x, y) \), in each position \( (x, y) \) in the window was calculated using the following equation.

\[
P(x, y) = I(x, y) \cdot S(x, y)
\]

where \( I(x, y) \) is the level of the difference image and \( S(x, y) \) is the maximum separability, which is calculated as described below. To accelerate a calculation speed, the shape of the mask for the separability filter [7] was given as radiated in eight directions (Fig. 7). Here, the region in the mask was divided into the inside and outside at a pixel number \( L \) from the center of the mask. The boundary is indicated by the square depicted by a broken line. In each position of the mask, the \( L \) where the separability became maximal was determined as \( L_M \). The maximum separability was used as \( S(x, y) \) in equation (1).

By the above mentioned procedure, the pupil difference image (Fig. 6(b)), \( I(x, y) \), is converted into a small spot and the eyelid almost disappears in the \( P(x, y) \) image (Fig. 6(c)). If the maximum of \( P(x, y) \) in the window exceeded a threshold, the pupil was judged to exist. The center of gravity of \( P(x, y) \) in the area of \( 4L_M \times 4L_M \) was calculated as the precise pupil center. Here, the area was determined so that its center coincided with the position indicating the maximal \( P(x, y) \). If the maximal \( P(x, y) \) did not exceed, the pupil was judged not to exist.

(D) The entire flow:
For every frame, after the procedure (A) was executed, the flow was divided into the following four cases, depending on the pupil existence in the previous frame. (i) If two pupils had existed in the previous frame, it proceeded to procedure (C) for precise center detection of the two pupils. (ii) and (iii) If one pupil (primary pupil) had existed in the previous frame, it proceeded to procedure (C) for precise center detection of the pupil. In addition, it proceeded to procedure (B) to make the eye area difference image and then to detect another pupil (secondary pupil). While masking the primary pupil (21 x 21), the center of the secondary pupil candidate was detected as the pixel showing the maximum intensity in the eye area difference image (Fig. 5(b)). If the y-coordinate difference between the primary and secondary pupils is less than 50, it was judged that the secondary pupil existed. If not, it was judged that the secondary pupil did not exist. (iv) If no pupil had existed in the previous frame, it proceeded to procedure (B) to detect the primary pupil candidate as the pixel having the maximum intensity. The pixel and its surrounding area (21 x 21) in the eye area difference image were masked. Then the pixel having the maximum intensity was detected as the secondary pupil candidate center. From the y-coordinate difference between the primary and secondary pupils, it was judged whether they existed or not.

![Fig. 4. (a) Face area binary image and (b) face area difference image.](image)

![Fig. 5. (a) Eye area binary image and (b) eye area difference image.](image)

C. Stereo matching

By the image processing described above, in general, the right and left pupils could be rightly distinguished. So stereo matching of the pupils was very easy. The camera calibration was executed using the method of Tsai [8]. In the camera calibration procedure, a while plate, where many black circles (dia. 15mm) were arranged horizontally and vertically at an interval of 2 cm, were put up on an automatic stage (seen in Fig. 8). The distance between the center of the stereo cameras and the plate was changed by using the automatic stage from approximately 61cm to 81 cm every 4 cm. The center positions of the circles were detected by an image processing algorithm different from that for pupil detection.
The error in 3D position measurement was estimated by moving the plate every 5 mm in the same range as for the camera calibration procedure. The 3D errors between the actual and estimated positions of the circle centers were 0.13±0.09 mm in horizontal (x), 0.11±0.08 mm in vertical (y), and 0.32±0.36 mm in depth (z), respectively.

Fig. 6. (a) Windows applied on pupils, (b) the difference image within window, and (c) image representing pupil parameter, P(x, y).

Fig. 7. Shape of mask for calculation of Separability.

III. RESULTS AND DISCUSSION

In the first experiment, the head of a subject was fixed on the automatic stage using a dental bite. His head was moved in the depth direction (z) while he fixed his gaze on an appropriate 3D position so that his eye should not rotate even if his head moves. The measurements points were 11 positions between 64 cm and 84 cm. The mean and standard deviation (SD) of the SD in each position were 2.52 mm and 0.19 mm, respectively. The discrepancy between the measured and actual pupil position displacements was estimated 1.7 mm per 10 cm. In the same fashion, his head was moved in the horizontal direction (x) from the center toward 12 cm in the right (approx. 77 cm in depth). The number of the measurement points was five. The mean and SD of the SD in each position were 0.27 mm and 0.07 mm.

In the second experiment, an artificial eye having a 4-mm diameter of pupil was smoothly moved horizontally at a depth of approximately 65 cm using the automatic stage. The top speed was 10 cm/sec (limitation of the stage). The results showed that position measurement of the artificial pupil was successful for a speed of 10 cm/sec. However, the error for depth was approximately 12 mm. This seems to be caused by the difference of the bright pupil image acquisition timing of the two cameras. However, when the artificial pupil existed at a static position, the SDs of the estimated positions for all of horizontal, vertical, and depth were less than 0.1 mm.

Fig. 8. Pupil tracking experiment.

Fig. 9. (a) and (b) indicate pupil detection from the left and right stereo wide cameras and (c) indicates detection of pupil center and corneal reflection in the narrow camera.
In the third experiment, a narrow camera for gaze detection was directed toward the right eye whose 3D position was measured. The narrow camera was placed below the computer display (Fig. 8) and was rotated two-dimensionally by a pan-tilt unit (PTU-D46). Fig. 9 shows the sample images indicating that the pupils were detected from the left and right stereo wide cameras and that the centers of the pupil and the corneal reflection were detected. Fig. 10(a) indicates the time course of the horizontal pupil position estimated by the stereo cameras when the subject moved his head largely in lateral. (The peak speed of the pupil displacement was estimated 25cm/sec.) Fig. 10(b) indicates the coordinates of the pupil and corneal reflection detected by the narrow camera in the same period of time. The coordinates were detected by the improved method from [9]. The constant coordinates (near 0 or 600) shown in Fig. 10(b) mean that the pupil and corneal reflection had not been detected. The results in Fig. 10 show that the eye image (pupil and corneal reflection) came out from the frame of the narrow camera by quick lateral head movements, and that, however, then immediately the eye image returned back into the frame by camera rotation.

The calculation time for the image processing for the two stereo cameras and for the narrow camera was approximately 20 ms and 5 ms, respectively.

IV. CONCLUSIONS

In the present paper, the noncontact method for realtime 3D pupil detection was proposed and implemented. This system may be easy to miniature. High precision of the 3D pupil detection system would contribute for precise eye gaze detection.

We succeeded in directing the narrow camera for gaze direction detection toward the 3D pupil position detected by the stereo cameras. From now on, using the present system, we will detect the eye gaze points on the computer display under head-free condition.

ACKNOWLEDGMENT

The present research was supported by Cooperative Link of Unique Science and Technology for Economy Revitalization (CLUSTER) in Hamamatsu area (Hamamatsu Optronics Cluster).

REFERENCES