

Real Time Eye Tracking using Initial Centroid and Gradient Analysis Technique*

Sunu Wibirama*

Supan Tungjitkusolmun*

Chuchart Pintavirooj*

Kazuhiko Hamamoto**

*Department of Electronics, Faculty of Engineering,
King Mongkut's Institute of Technology Ladkrabang
Bangkok 10520, Thailand
kpchucha@kmitl.ac.th

**Department of Information Media Technology,
School of Information and Telecommunication Engineering
Tokai University, Tokyo, Japan
hama@keyaki.cc.u-tokai.ac.jp

Abstract—Robust and accurate algorithm in real time eye tracking system has been a fundamental and challenging problem for computer vision and biomedical engineering area. This paper proposes a new method to estimate eye-motion position and direction based on initial centroid and gradient analysis technique. The proposed method was validated by tracking eye position within high and low occlusion conditions. Experimental results show that the proposed method improves the accuracy of tracking pupil position compared to conventional window-matching method. The proposed method yields 76.07% and 87.89% accuracies while the conventional window-matching performs 44.56% and 67.89% accuracies within high and low occlusion conditions, respectively.

I. INTRODUCTION

Eye tracking is an extensive research area in visual information system and computer vision that encompasses anything from diagnostic tool for assessing vestibular disease and other neurological disorders [1] to human-computer interface device [2, 3].

Normally, eye tracking is performed on two dimensions to measure the horizontal and vertical motions of the eye. Horizontal and vertical eye (2D) positions can be determined from pupil center coordinates, which can be computed using center of mass algorithm [4, 5]. The torsional eye position, rotation about the visual axis, can be determined by tracking the location of landmark on the eye [6, 7] or using the window-matching algorithm [4].

Haro et al. [8] proposed pupil tracking based on combining its appearance, the bright pupil effect and motion characteristics. Ji et al. [9] proposed real time subtraction and a special filter to eliminate the external light interferences. Both methods failed to track eyes when they were closed or occluded and robustness was easily affected by the external illumination interference.

In this paper, we propose a new method based on initial centroid and gradient analysis technique. The method is used to measure spatial and directional pupil position in real time tracking. We attempt to overcome accuracy problem arising in real time tracking as in [8, 9] by adding initial centroid algorithm. The method then was tested in tracking pupil motion involving high and low occlusions.

II. THE PROPOSED METHOD

Our proposed method is pattern recognition based approach as shown in Fig. 1. It comprises of three main steps: pre-processing, pupil detection, and motion direction estimation. Initial centroid algorithm is used to deal with window-matching inaccuracy when detecting pupil motion with large head movement, eye blink, and occlusion caused by eyelashes. Gradient analysis based on motion gradient method proposed by [10] is used to estimate the direction of eye movement. The method is implemented in low cost eye tracking system consists of two infrared cameras attached to special binocular, Pentium PC 2.2 GHz and eye tracking software built by Visual C++ and OpenCV image processing library.

A. Pre-processing

First of all, image frames are captured from the video stream of the camera at rate 25 fps. Then, the pre-processing step is applied. This step consists of binarization and morphological process. Binarization process is done by implementing thresholding process. The thresholding result is shown in Fig. 2(b). Morphological process with basic dilation and erosion is inserted to remove remaining artifact from the binary image. The morphological result is shown in Fig. 2(c).

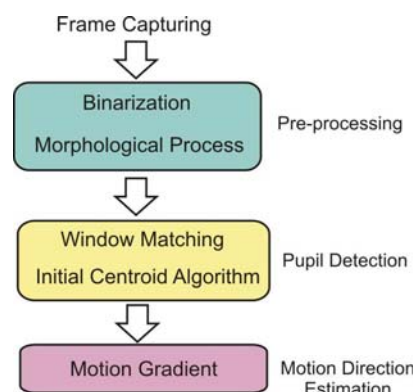


Figure 1. A flowchart diagram for our proposed method.

* This research is supported by AUN/Seed-Net programme from Japan International Cooperation Agency (JICA).

B. Pupil Detection

The second step of our algorithm is pupil detection. It comprises of window-matching and initial centroid algorithm. Window-matching method as in [11] is based upon an analysis of the grey level pattern around a point of interest and the search for the most similar pattern in the subsequent frame. Window-matching method calculates motion between a region centered around a certain point of interest $p(x, y)$ on image I_1 that will be displaced by integer values i, j (along the horizontal and vertical directions, respectively) in the subsequent image I_2 . The region on the image is represented by squared window of $N \times N$ dimension. To measure similarity, SSD (Sum of Squared Differences) function in reference [11] will be used here, which is defined as

$$R_{sq_diff}(x, y) = \sum_{i, j = -\frac{N}{2}}^{\frac{N}{2}} [I_1(x, y) - I_2(x + i, y + j)]^2 \quad (2).$$

Minimizing the result R represents minimizing the distance of similarity, then it means finding, on the subsequent image, the most similar region to the current image. Therefore, a perfect match will be 0 and bad matches will be large.

Some occlusions caused by large head movement or eye blink can be occurred during eye tracking process. Consequently, an ambiguity between the texture of Region of Interest (ROI) and other regions of the image will yield some error in tracking trajectory. To deal with this problem, improving window-matching process by applying initial centroid algorithm can be considered. The main goal of initial centroid algorithm is reducing inaccuracy of window-matching algorithm by inserting an initial position of previously tracked eye center and applying a time-stamp to prevent tracking error when eye blink or other occlusions are happened. The initial centroid algorithm can be explained as follows:

Step 1: Determine distance threshold (γ_T) and time-stamp threshold (τ_T).

Step 2: Define initial centroid $C(x, y)$. Compute pupil center and save as initial centroid $C(x, y)$. Compute ROI from initial centroid $C(x, y)$.

Step 3: Compute current centroid $C'(x, y)$.

Step 4: Compute the distance between $C(x, y)$ and $C'(x, y)$. The distance γ between initial centroid $C(x, y)$ and current centroid $C'(x, y)$ can be described as

$$\gamma = \sqrt{(x_C - x_{C'})^2 + (y_C - y_{C'})^2} \quad (3),$$

where (x_C, y_C) and $(x_{C'}, y_{C'})$ represent the coordinates of $C(x, y)$ and $C'(x, y)$, respectively.

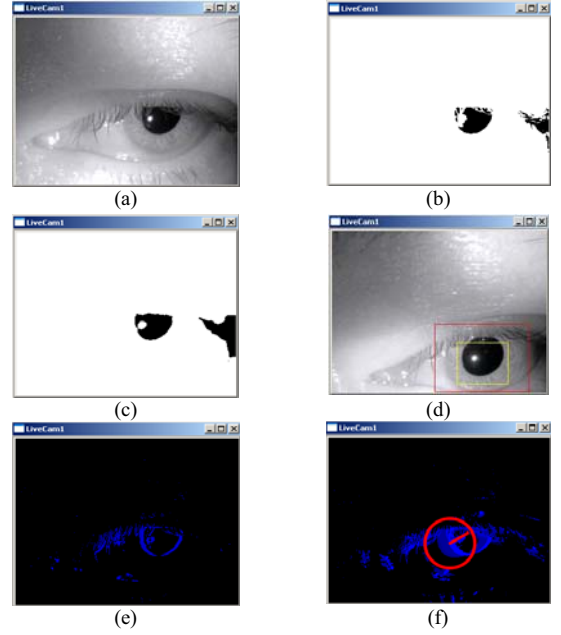


Figure 2. (a) Frame captured at 25 fps. (b) Result after binarization process. (c) Result after morphological process. (d) Implementation of Window Matching and Initial Centroid Algorithm. (e) Motion History Image. (f) Motion direction estimation.

Step 5: Compare between $C(x, y)$ and $C'(x, y)$

- 5.1. while $\gamma < \gamma_T$ then keep the current ROI same as the previous ROI.
- 5.2. while $\gamma > \gamma_T$
 - 5.1.1. Apply the time-stamp τ .
 - 5.1.2. while $\tau < \tau_T$ then keep the current ROI same as the previous ROI.
 - 5.1.3. while $\tau > \tau_T$ then save current centroid $C'(x, y)$ as the new initial centroid and repeat step 2 until 5.

C. Motion Direction Estimation

The last step of our proposed method is applying motion gradient which represents the direction of eye movement. The concept of motion gradient is to estimate global gradient estimation from Motion History Image (MHI) [10]. MHI is a scalar-valued image where intensity is a function of recency motion. MHI image calculation can be described as follows:

$$tMHI_{\delta} = \begin{cases} \tau & \text{if current silhouette at } (x, y) \\ 0 & \text{else if } tMHI_{\delta}(x, y) < (\tau - \delta) \end{cases} \quad (4),$$

where $tMHI$ is timed Motion History Image, τ is current time-stamp, and δ is time duration constant. The result is a scalar-valued image where more recently moving pixels are brighter.

The gradient of MHI can be calculated efficiently by convolution with separable Sobel filters [12] in the X and Y directions yielding the spatial derivatives $F_x(x,y)$ and $F_y(x,y)$. Gradient magnitude $\|F\|$ and orientation $\phi(x,y)$ at each pixel are defined by (5) and (6), respectively:

$$\|F\| = \sqrt{F_x^2 + F_y^2} \quad (5),$$

$$\phi(x,y) = \arctan \frac{F_y(x,y)}{F_x(x,y)} \quad (6).$$

After calculating the motion gradient, we can then compute the global motion orientation [10]. Calculation of the global orientation should be weighted by normalized MHI values $tMHI$ to give more influence to the most current motion within the MHI process. Global motion orientation can be computed as follows:

$$\bar{\phi} = \phi_{ref} + \frac{\sum_{x,y} \text{angDiff}(\phi(x,y), \phi_{ref}) \times \text{norm}(\tau, \delta, tMHI_{\delta}(x,y))}{\sum_{x,y} \text{norm}(\tau, \delta, tMHI_{\delta}(x,y))} \quad (7),$$

where:

- $\bar{\phi}$: the global motion orientation.
 - ϕ_{ref} : base reference angle (peaked value in the histogram of orientation).
 - $\phi(x,y)$: motion orientation map found from gradient convolutions.
 - $\text{norm}(\tau, \delta, tMHI_{\delta}(x,y))$: normalized $tMHI$ value (linearly normalizing the $tMHI$ from 0 until 1 using the current time-stamp τ and duration δ).
 - $\text{angDiff}(\phi(x,y), \phi_{ref})$: minimum, signed angular difference of an orientation from the reference angle.
- Global motion orientation in our eye tracking process is shown in Fig. 2(f).

III. EXPERIMENTAL RESULT

We evaluated performances of our method using videos of two different patients acquired from Otolaryngology (Ear, Nose, and Throat) Center, Bangkok Hospital Medical Center. The first patient had a lot of head movement while the second patient had less head movement. The first patient blinked his eyes more than the second patient. We used 1120 and 900 captured frames from patients with high and low occlusion, respectively. The experimental results show that our proposed algorithm exhibited good performance in both high and low occlusion conditions. We evaluated the performance of our proposed method using the following equation:

$$\text{SuccessRate} = \text{Samples} - \text{error} \quad (8).$$

$$\text{Accuracy}(\%) = \frac{\text{SuccessRate}}{\text{Samples}} \times 100\%$$

TABLE I
PERFORMANCE OF INITIAL CENTROID+WINDOW MATCHING+GRADIENT ANALYSIS

Result Condition	Accuracy (%)	Error (Frames)	Total Sample (Frames)
High Occlusion	76.07	268	1120
Low Occlusion	87.89	109	900

TABLE II
PERFORMANCE OF CONVENTIONAL WINDOW-MATCHING+GRADIENT ANALYSIS

Result Condition	Accuracy (%)	Error (Frames)	Total Sample (Frames)
High Occlusion	44.56	621	1120
Low Occlusion	67.89	289	900

From our experimental results, our proposed algorithm exhibited good performance in both high and low occlusion conditions compared to conventional window-matching algorithm. Major errors in conventional window-matching algorithm are caused by ambiguity between the texture of ROI and other regions of the image when the patients blinked their eyes or moved their heads.

Fig. 3 shows various condition of eye when both proposed algorithm and conventional window-matching were tested. Eyelid occlusions often cause more error than eyelash occlusions in tracking trajectory because of mismatch between the reference pupil image and the captured image.

Fig. 4(a), 4(b), and 4(c) show trajectory results between our proposed algorithm and conventional method. We measured horizontal position as a sample. The conventional window-matching and gradient analysis technique is shown by green colour (WM+MG) while the combination of initial centroid, window-matching, and gradient analysis is shown by red colour (WM+MG+IC).

Fig. 4(d) shows the direction of pupil movement computed from global motion orientation under normal condition with less head movement. The direction was measured in degree unit.

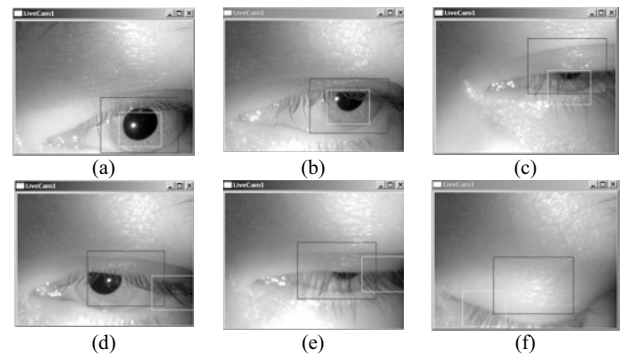


Figure 3. (a) Full circular pupil. (b) Half circular pupil. (c) Almost closed eye image. (d-e) Error pupil tracking using conventional window-matching technique (small rectangle) and success pupil tracking using initial centroid and window-matching (big rectangle). (f) Both methods fail to locate pupil on full closed eye

IV. CONCLUSION

This paper presents a new method to track pupil position and direction in low cost eye tracking application. The initial centroid algorithm is added in window-matching technique to increase the accuracy when tracking in high occlusion. The results were compared to conventional window-matching which used no initial centroid algorithm. Gradient analysis technique is used to estimate eye-motion direction.

Experimental results show that the proposed algorithm offers robustness and flexibility in situations where there are a lot of factors causing unexpected error in real time tracking process. Some error in real time eye tracking process are mainly caused by eyelid and eyelash occlusions.

Future works can be done by improving the accuracy of proposed algorithm and applying the low cost eye tracking system as human computer interaction device or a diagnostic tool for vestibular disease.

ACKNOWLEDGMENT

Our thanks to AUN/Seed-Net programme and JICA (Japan International Cooperation Agency) for supporting research fund and equipments, Otolaryngology (Ear, Nose, and Throat) Center, Bangkok Hospital Medical Center for providing eye movement recordings.

REFERENCES

- [1] B.A. Wilson, R.D. Jones, P.J. Bones, and T.J. Anderson, "A Flying-Spot Laser Scanner for Tracking Eye Movements," in *18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE EMBS 1996)*, Amsterdam, 1996.
- [2] S.D. Magistris, "Eye movement tracking for advanced man-machine interaction," in *Department of Systems and Informatics*. Florence, Italy: University of Florence, 1995.
- [3] Y. Kurio, T. Yagi, and Y. Uchikawa, "Development of eye pointer with free head-motion," in *20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE EMBS 1998)*, 1998.
- [4] S.T. Moore, L.S. Curthoys, and S.G. McCoy, "VTM-an image-processing system for measuring ocular torsion," in *Computer Methods and Programs in Biomedicine*, vol. 35, pp. 219-230, 1991.
- [5] T. Vieville and D. Masse, "Ocular counter-rolling during active head tilting in humans," in *Acta Otolaryngol*, vol. 103, pp. 280-290, 1987.
- [6] J.A. Parker, R.V. Kenyon, and L.R. Young, "Measurement of Torsion from Multi Temporal Images of The Eye Using Digital Signal Processing Techniques," in *IEEE Transactions on Biomedical Engineering*, vol. BME-32, no. 1, pp. 28-36, 1985.
- [7] S. Yamanobe, S. Taira, T. Morizono, T. Yagi, and T. Kamio, "Eye Movement Analysis System Using Computerized Image Recognition," in *Arch Otolaryngol Head Neck Surg.*, vol. 116, pp. 338-341, 1990.
- [8] A. Haro, M. Flickner, and I. Essa, "Detecting and Tracking Eyes by Using Their Physiological Properties, Dynamics, and Appearance," in *Proceedings IEEE CVPR 2000*, Hilton Head Island, South Carolina, June 2000.
- [9] Q. Ji and X. Yang, "Real Time Visual Cues Extraction for Monitoring Driver Vigilance," in *Proceedings of International Workshop on Computer Vision Systems*, Vancouver, Canada, July 7-8, 2001.
- [10] G. Bradski and J. Davis, "Motion Segmentation and Pose Recognition with Motion History Gradients," in *IEEE Workshop on Applications of Computer Vision*, 2000.
- [11] P. Anandan, "A computational framework and an algorithm for the measurement of visual motion", vol.2, 1989, p. 283-310.
- [12] R.C. Gonzalez and R.E. Woods, "Digital Image Processing", 2nd edition, 2002, p. 134-137.

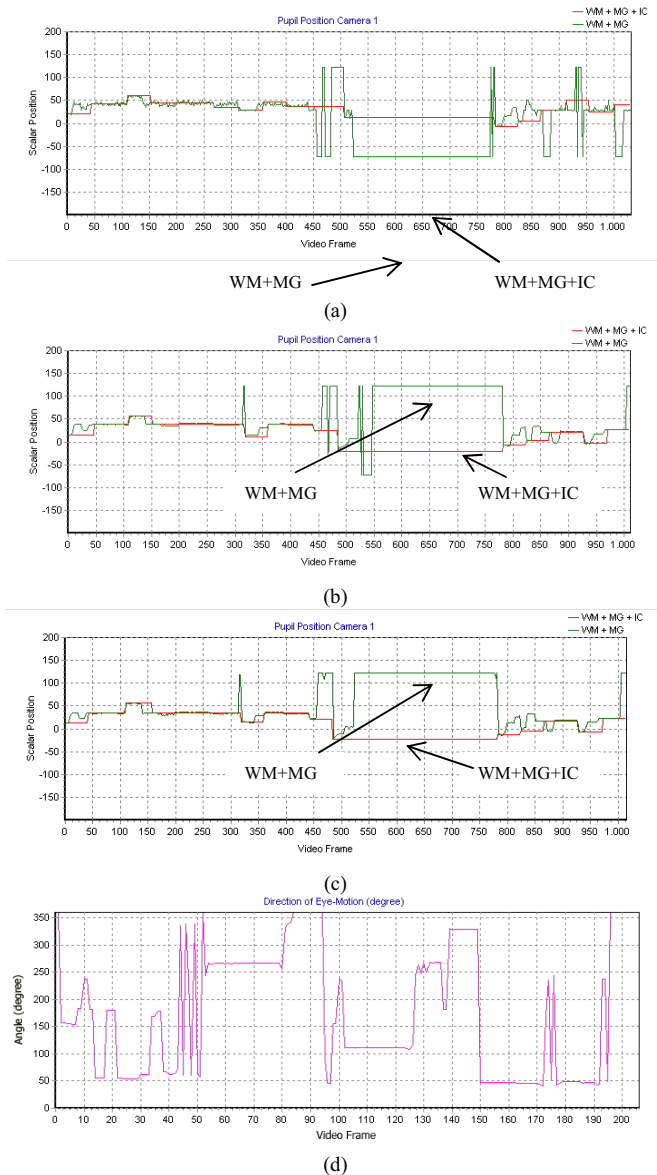


Figure 4. Horizontal trajectory of eye when thresholding value : (a) 20, (b) 40, and (c) 80. (d) Pupil movement direction (degree unit).

Fig. 4(a), 4(b), and 4(c) show that when the thresholding value in binarization process was changed, it affected directly to the accuracy of conventional technique (WM+MG). It can be implied from the drastic change of conventional window-matching and gradient analysis trajectory graph. Nevertheless, the change of threshold value would not caused a lot of error when our proposed algorithm (WM+MG+IC) was applied. The proposed algorithm trajectory graphs show that the tracking trajectory remain stable in low or high threshold value. Thus, the proposed algorithm can handle different thresholding values in reasonable range which is very useful for different grey level image in binarization process.