High Speed Gaze Tracking with Visible Light

Wen-Chung Kao, Wei-Te Chang, Sheng-Ju Wu, Chien-Hui Liu, and Shih-Yao Yin
Department of Applied Electronics, National Taiwan Normal University, Taipei, Taiwan
jungkao@ntnu.edu.tw

Abstract— **Some gaze tracking systems have been proposed in the field, but very few of them can be directly applied in normal lighting environments. Many image imperfection issues are difficult to be dealt with so that a typical system is usually equipped with a strong infrared ray (IR) illuminant. The user experience is not so good even if the IR illuminant is invisible. In this paper, we present a high speed (240 frames/s) gaze tracking system operating in a normal office lighting condition. A set of image processing algorithms to deal with the issues of poor image contrast and high random noise have been presented. The experimental result shows that the system can achieve the accuracy of $1°$ and $1.7°$ for the horizontal and vertical coordinates of the detected iris center, respectively.**

I. INTRODUCTION

High speed gaze tracking becomes a more and more important research topic due to its applications covering human-machine interface, virtual reality, and the study of human behaviors. A good gaze tracking system should meet several design requirements: high speed, high precision, easy to use, nonintrusive, and applicable on normal illumination [1] – [13]. Many gaze tracking systems have been proposed, but very few of them can meet all requirements mentioned above. A high speed gaze tracking system suffers from a design challenge caused by extremely short exposure, but this can be conquered with a strong infrared ray (IR) illuminant. The IR illumination is invisible, but the user may feel uncomfortable and it may not be applicable in outdoor environments. An attempt to the design of gaze tracking system has been presented by using viable light, but the system was not designed with a high frame rate specification.

In this paper, we present a novel high speed gaze tracking system in normal visible lighting conditions. The design objective is boosting the frame rate to 240 or 480 frames/s, which means the exposure time for each frame is shorter than 1/240s or 1/480s. The captured video frames are typically very dark, noisy, and of low contrast. These new design challenges have been well addressed in the proposed system, which incorporates a robust tone reproduction algorithm [14] – [16], a refinement process for eye region segmentation [17]-[21], and a comprehensive algorithm for limbus circle evaluation [10]. The proposed system allows the user to operate the system in a normal office environment without additional illuminations. The experimental result shows that the system can achieve the accuracy of $1°$ and $1.7°$ for the horizontal and vertical coordinates of the detected iris center, respectively.

In the remaining sections, we will first introduce the signal processing flow in the system in Section II. The detailed algorithms of eye region segmentation and iris center computation are stated in Sections III and IV. Finally, the experimental results and the conclusion are provided in Sections V and VI.

II. SIGNAL PROCESSING/ANALYSIS FLOW

The gaze tracking system we proposed is composed of three major stages: (a) locating the initial eye region, (b) evaluating the coordinates of the iris center, and (c) gaze point calibration. Given a sequence of video frames, the detection of the human eyes is done by several steps. The well-known tool OpenCV can roughly locate the regions of human eyes but the accuracy of the cropped eye region is not good enough for gaze tracking applications. A refinement stage is incorporated with the eye detection function of OpenCV to precisely crop the regions of human eyes. Following the eye detection as well as the location refinement, the image enhancement process is performed to boost the contrast of the eye region. This is because the video frames are taken with a high speed frame rate mode, which means the camera operates at a shorter exposure time condition. The video frames look extremely dark and textureless. Without being processed through a sophisticated tone reproduction algorithm, it is not possible to accurately extract the contour of the iris from the image. In addition, traditional approach to iris contour detection is also sensitive to image noise. To deal with this issue, we have developed a gradient based approach to the optimal circle fitting for the corneal limbus. The center coordinates of the evaluated circles are then recognized as the gaze points of the eyes. Finally, the detected gaze points are remapped to the exact coordinates on the screen by a camera calibration process like traditional gaze tracking systems.

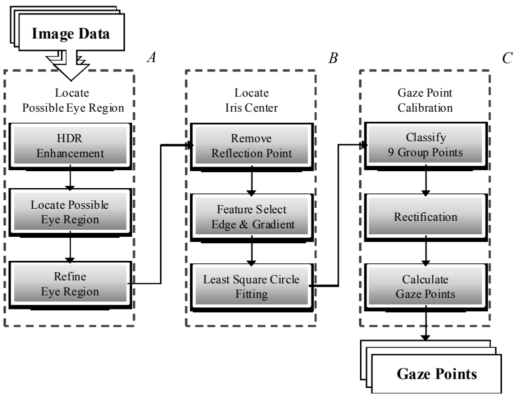


Figure 1. Signal processing/analysis flow chart

Since the algorithms used in the camera calibration are the same as those in the traditional approaches [1], the remaining parts of the paper focus on the algorithm designs for locating eye regions and evaluating the coordinates of the iris center.

III. Segmentation of Eye Regions

Unlike the traditional gaze tracking system which is equipped with an additional IR light source to provide a strong and invisible illumination, the proposed system can be applied in normal indoor lighting conditions which may cause flickering problems in the captured video frames. For the eye region segmentation of a high speed gaze tracking system, the issues needed to be dealt with contain the compensation of uneven light conditions and the refinement of the detected eye regions. To cope with these issues, we propose a set of robust algorithms and they are described in the following subsections.

A. Image Enhancement by Tone Reproduction

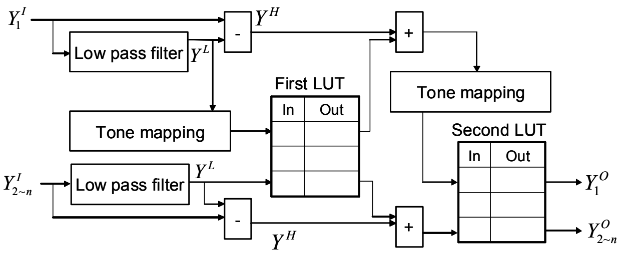
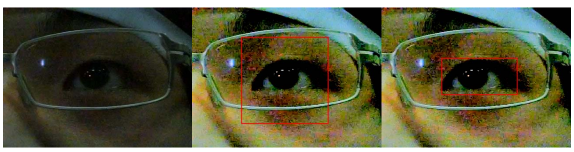


Figure 2. Signal processing flow in tone production

A robust tone reproduction system for video frames has been integrated into the system [3]. As shown in Fig. 2, the pixel luminance values in each frame are further decomposed into high and low frequency planes. The dynamic range of the low frequency components should be attenuated significantly more than that of the high frequency components. Since the lighting condition should not be changed during the operation of the gazing system, the tone mapping parameters are evaluated based on a few frames and a lookup table (LUT) can be constructed. The LUTs are then applied for rescaling the data in the remaining frames. The detailed algorithm for computing the parameters used in tone reproduction are described in [14].

B. Locating an Initial Eye Region



(a) (b) (c)

Figure 3. An example of eye region segmentation

The well-known object detection system OpenCV has been applied to locate possible eye regions. The features used in eye detection are the coefficients of Haar transformation. As shown in Fig. 3, the eye in the original image (Fig. 3-(a)) looks too dark to be seen easily. The visibility of the eye region can be significantly improved after tone reproduction as shown in Fig. 3-(b) which also marks the initial result of the eye detection with OpenCV. The initial guess is still difficult to be handled in the remaining stages, since the cropped region contains some undesired parts such as the glass. An additional step to refine the eye region is necessary and the final result is shown in Fig. 3-(c). The details of the refinement steps are stated in the following subsection.

C. Refine Eye Region

The proposed approach to the refinement of eye regions is based on the information of the skin detection. However, the problem of skin detection in the video frames taken with a high speed video recording system is quite different from the one with a consumer camera. As shown in Fig. 4, the skin colors are varied in different frames. This is because the original video frames are illuminated in different flickering phases of a fluorescent lamp, and the frequency of the flicker is 100 or 120 Hz. If the frame rate is set as 240 frames/s, the two successive frames receive different exposure, which results in various color compensation inside the image processing pipeline of the camera.

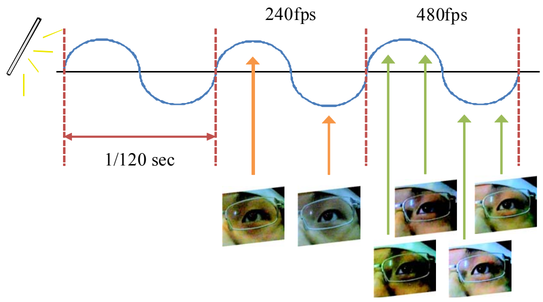


Figure 4. The skin detection problems caused by flickering light sources

In order to conquer the problem mentioned above, two or more frames are processed simultaneously. As shown in Fig. 5, the skin detection results of two successive frames are merged into a single map, which shows a much better result. Based on the skin detection result, the initial region of the eye can be refined to an accurate one, which can be handled easier in the remaining stages.

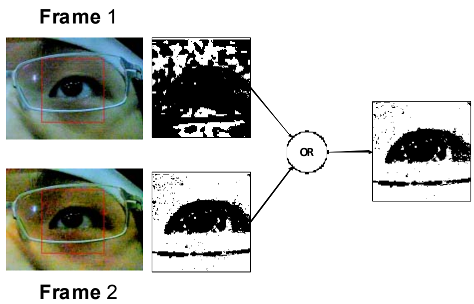


Figure 5. Eye region refinement by skin color information

IV. Computation of Iris Center Coordinates

Following the segmentation of eye regions, the iris center coordinates is evaluated to determine the gaze point. Since the image may be affected by some imperfection factors, some preprocessing steps are required. As shown in Fig. 6, there are some reflection points inside the iris region caused by the illuminants. It is necessary to remove them before evaluating the limbus circle. In addition, traditional approach to fitting the limbus circle is based on the detected image edges. But this approach is quite sensitive to the random noise. To cope with the noise problem, we propose a multi-phase circle fitting algorithm, which is modified from the starburst algorithm with gradient features. The detailed algorithms of removing reflection points and evaluation of limbus circle are described in the following subsections.

A. Removal of Reflection Points

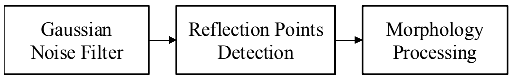
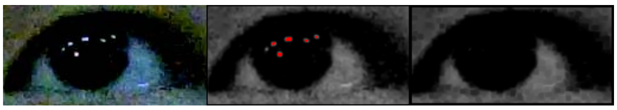


Figure 6. Signal processing flow of removing reflection points.

Fig. 6 shows the signal processing flow for removing reflection points in the eye region. The cropped eye image is first processed by a noise filter to remove the random noise. Then a pixel (i, j) is identified as a reflection point if its luminance value $L(i, j)$ satisfies the following conditions:

$$L(i,j) \geq L_{90} \text{ and } \overline{L}(i,j) \leq L_{50} \tag{1}$$

where $\overline{L}(i,j)$ denotes the local average value at the pixel (i, j). The parameters L_{50} and L_{90} are defined as the luminance values of the pixels which locate at 50% and 90% of all pixels in the cumulative histogram. This heuristic equation comes from the observation that a reflection point inside the iris region is very bright but its neighboring pixels are relative dark. After the reflection points are identified, they can be eliminated by the morphology processing. Fig. 7 shows an example of removing reflection points. The original image shown in Fig. 7-(a) is analyzed, and the reflection points are marked as shown in Fig. 7-(b). Finally, the reflection points are removed through image morphology processing.



(a) (b) (c)

Figure 7. An example of removing reflection points.

B. Computation of Iris Center Coordinates

The proposed approach to computation of the iris center coordinates is modified from starburst algorithm [13]. The original starburst algorithm is composed of two phases: feature point detection and ellipse fitting. It first produces a set of rays from a starting point and then calculates the intensity gradient for each pixel in the rays. If the gradient of a pixel is higher than a threshold, it is marked as a feature point. In the second phase, an ellipse is fitted according to these feature points. This approach achieves promising result for the gaze tracking system with IR illumination. But applying it directly to the system with visible light can not achieve an acceptable result. This is because the image noise will affect the pixel gradient. Thus the feature point detection is quite sensitive to image noise.

We have developed a modified starburst algorithm for the computation of the iris center coordinates. The idea can be explained by a simple example shown in Fig. 8. Several rays are produced from a center point as shown in Fig. 8-(a). The luminance values and luminance gradients for each pixel on a ray is shown in Figs. 8-(b) and (c), respectively. A pixel on a ray is recognized as a feature point if it satisfies the following conditions

$$\nabla L(r, x-k) \leq \theta, 1 \leq k \leq 4 \tag{2}$$

$$\nabla L(r, x+l) \geq \phi, 0 \leq k \leq 3 \tag{3}$$

where $\nabla L(r,x)$ denotes the gradient of the x-th pixel on the r-th ray. The parameters θ and ϕ are two pre-defined values.

Fig. 8-(d) shows the detected feature points which can be fed into the circle fitting algorithm. The center point as well as the radius can be determined accordingly. Note that this computation process can be iteratively executed to achieve a more accurate result. The same processing flow starts again from this new center point, and another set of rays is examined like the first iteration. But some unreliable features points, which are far away from the circle determined in the previous iteration, are removed. The iterative process is executed until the fitting circle reaches to a steady state.

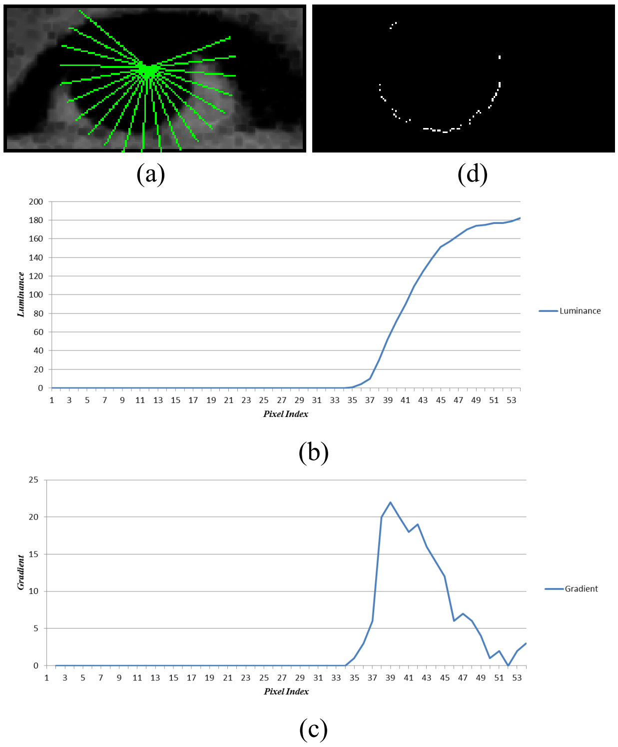


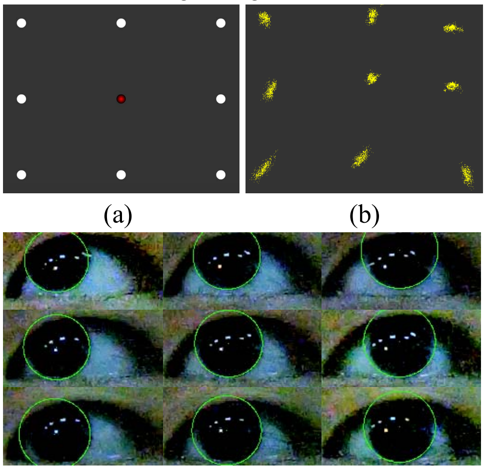
Figure 8. An example of processing results with the modified starburst algorithm

V. EXPERIMENTAL RESULTS

We have set up a complete experimental environment by using a high speed camera to take videos, and the video frames are then processed in the computer. The detected coordinates of the iris center are then mapped to the locations on the screen. As shown in Fig. 9-(a), the accuracy of gaze tracking is evaluated by asking the users to observe nine points displayed on the computer screen. Fig. 9-(b) shows some sample images when the user observes these nine points. The final result of gaze tracking for a user can also be displayed on the screen as shown in Fig. 9-(c).

The accuracy of gaze tracking is reported based on the error measurement shown in Fig. 10. The equation for evaluating angle accuracy can be stated as (4). The accuracies of gaze tracking based on the seven analyzed subjects are shown in Fig. 11. The experimental result shows that the system can achieve the accuracy of 1 degree and 1.7 for the horizontal and vertical coordinates of the detected iris center, respectively.

$$\theta_e = \cos^{-1}\left(\frac{D_{gaze}^2 + E_{gaze}^2 - d_{error}^2}{2 \cdot D_{gaze} \cdot E_{gaze}}\right) \tag{4}$$



(a) (b)

(c)

Figure 9. The testing chart and the experimental results.

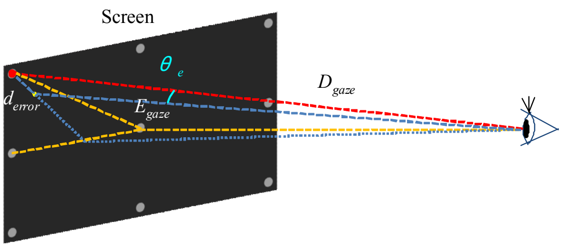


Figure 10. The evaluation of accuracy.

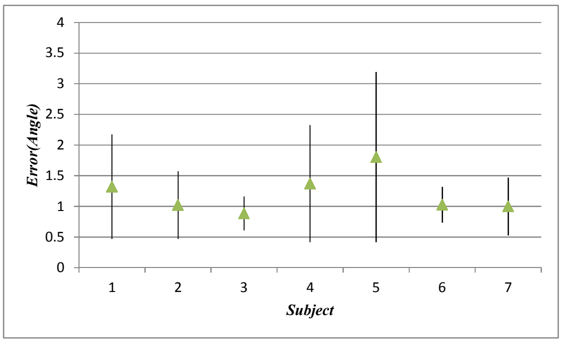


Figure 11. Horizontal accuracy of seven subjects.

ACKNOWLEDGMENT

This work was supported in part by the National Science Council, Taiwan, under the Grant NSC 100-2221-E-003 -012 -MY3, 100-2221-E-003-011-MY2, and the "Aim for the Top University Plan" from National Taiwan Normal University, and the Ministry of Education, Taiwan.

REFERENCES

[1] K. P. White, Jr., T. E. Hutchinson, and J. M. Carley, "Spatially Dynamic Calibration of an Eye-Tracking System," *IEEE Trans. Syst., Man, Cybern.*, vol. 23, no. 4, pp. 1162–1168, Jul./Aug. 1993.

[2] Z. R. Cherif, A. Naït-Ali, J F. Motsch, and M. O. Krebs, "An Adaptive Calibration of an Infrared Light Device Used for Gaze Tracking," in *Proc. IEEE Instrum. Meas. Tech. Conf.*, 2002, pp. 1029–1033.

[3] N. Ramanauskas, "Calibration of Video-Oculographical Eye-Tracking System," *Electron. Electr. Eng.*, vol. 8, no. 72, pp. 65-68, 2006

[4] L. N. Trefethen and D. Bau III, *Numerical Linear Algebra*. Philadelphia: Society for Industrial and Applied Mathematics

[5] OpenCV Computer Vision Library. (2013). [Online]. Available : http://opencv.willowgarage. com/wiki

[6] K. H. Tan, D. J. Kriegman, and N. Ahuja, "Appearance-Based Eye Gaze Estimation," in *Proc. IEEE Workshop Appl. Comput. Vis.*, 2002, pp. 191–195.

[7] D. Beymer and M. Flickner, "Eye Gaze Tracking Using an Active Stereo Head," in *Proc. Int. Conf. Comput. Vis. Pattern Recognit.*, 2003, pp. 451–458.

[8] C. H. Morimoto, D. Koons, A. Amir, and M. Flickner, "Pupil Detection and Tracking Using Multiple Light Sources," *Image Vis. Comput.*, vol. 18, pp. 331–336, 2000.

[9] Z. W. Zhu and Q. Ji, "Eye and Gaze Tracking for Interactive Graphic Display," *Mach. Vis. Appl.*, vol. 15, no. 3, pp. 139–148, 2004.

[10] C. Hennessey B. Norueddin, and P. Lawrence. "Fixation Precision in High-Speed Noncontact Eye-Gaze Tracking," *IEEE Transactions on System, Man, and Cyberneics – Part B : Cybernetics*, vol. 38, No. 2, April 2008

[11] C. Hennessey and P. Lawrence. "Noncontact Binocular Eye-Gaze Tracking for Point-of-Gaze Estimation in Three Dimensions," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 3, March

[12] J. Sigut and S. A. Sidha, "Iris Center Corneal Reflection Method for Gaze Tracking Using Visible Light," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 2, Feb. 2011

[13] D. Li, D. Windield, and D. J. Parkhurst, "Starburst: A Hybrid Algorithm for Video-Based Eye Tracking Combining Feature-Based and Model-Based Approaches," *IEEE Computer Society Conference on Vision and Pattern Recognition (CVPRW'05) Workshops*, pp. 79, 2005

[14] E. Reinhard, "Parameter Estimation for Photographic Tone Reproduction," *Journal of Graphics Tools*, vol. 7, no. 1, pp. 45-52, Nov. 2002.

[15] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic Tone Reproduction for Digital Images," *ACM Trans. on Graphics*, vol. 21, no. 3, pp. 267-276, July 2002

[16] W. C. Kao, X. T. Huang, H. C. Wang, C. C. Pan, and F. C. Yang, "Real-Time Tone Reproduction for Video Recording, " in *Proc. IEEE International Symposium Consumer Electronics*, Harrisburg, US, June 2012.

[17] K. Sobottka and I. Pitas, "A Novel Method for Automatic Face Segmentation, Facial Feature Extraction and Tracking," *Signal Procesing: Image Comm.*, vol. 12, no. 3, pp. 263-281, 1998.

[18] D. Chai and K. N. Ngan, "Face Segmentation Using Skin Color Map in Videophone Applications," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 9, no. 4, pp. 551-564, 1999.

[19] H. Li and R. Forchheimer, "Location of face using color cues," in *Proc. Picture Coding Symp.*, Lausanne, Switzerland, Mar. 1993.

[20] Y. J. Wang and B. Z. Yuan, "A Novel Approach for Human Face Detection from Color Images under Complex Background," *Pattern Recognition*, vol. 34, no. 10, 1983-1992, June 2001.

[21] S. L. Phung, A. Bouzerdoum, and D. Chai "Skin Segmentation Using Color Pixel Classification: Analysis and Comparison," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 1, Jan. 2005.