ONE-TIME CALIBRATION EYE GAZE DETECTION SYSTEM

Lim Choon Kiat & Surendra Ranganath
Department of Electrical and Computer Engineering
National University of Singapore
{eng01126, elesr}@nus.edu.sg

ABSTRACT

This paper describes a real-time infrared-based system for eye gaze tracking. Most of the existing gaze tracking systems require a cumbersome calibration process every time an user operates the system and work well only with small head movement. Our system is simpler as it requires only one-time calibration for each user and the calibration data is stored for subsequent reuse. Besides that, our system can perform robust and accurate gaze estimation under rather significant head pose variations. This is made possible by a gaze calibration procedure which maps the pupil and glint parameters to screen coordinates using the Radial Basis Function Neural Network (RBFNN). Our system currently has a spatial gaze resolution of about 3 degrees and it operates at 25 fps with an average accuracy of 90% for gaze estimation.

1. INTRODUCTION

Computer users look at various points on the monitor screen during their interaction with the computer. If we can estimate where their gaze is directed, and collect statistics from the estimated gaze point, we may be able to infer useful information such as interest in something displayed on the screen, reading activities, etc. Besides that, we can also create a contact-free eye-mouse system for handicapped users and add excitement to computer games by increasing the interactivity between the user and computer.

The general approaches to gaze detection can be classified into two main categories: contact methods and non-contact methods. Contact methods may use skin electrodes, contact lens and head mounted gear, and require the user to be physically connected to the system. However, the limitations of contact methods such as health issues, restriction to user's head movements and discomfort have led to more focus on non-contact vision-based methods.

Non-contact or vision-based methods can be further classified into two main categories: techniques based on ambient light and techniques based on infrared light. Most of the techniques based on ambient light face the problems of detecting features such as iris and limbus accurately. This is because the color of iris differs between individuals and the limbus is usually partially occluded by eyelids.

Many of the existing eye-tracking systems make use of infrared light source/s and an infrared-sensitive camera [1]. The infrared light source causes pupils to appear bright due to retinal reflection, similar to the "red-eye" effect of a camera flash. This simplifies the process of detecting the pupils in the images. Hence, a combination of the retinal and corneal reflection (glint) can be used to determine the glint and pupil centers respectively.

Given the detected glint and pupil centers, a mapping function can be used to map the pupil-to-glint vector to screen coordinates. The mapping function of most reported gaze tracking system is determined by a calibration procedure each time a user operates the system [2][3]. This conventional approach for gaze calibration suffers from two shortcomings. 1) The system works well only with small head movement. 2) It is a cumbersome process to carry out a gaze calibration each time the user operates the system.

Several calibration free systems [4][5][6] that have been reported. However, some of these systems are unable to give accurate gaze estimation when the user's head moves freely and quickly. There are also others that require complicated multiple cameras and light source setups which slows down the speed of the overall system.

2. SYTEM SPECIFICATION

An infrared sensitive CCD B/W camera JAI CV-M50IR with a 16mm fixed focus lens is used to acquire images. This camera is mounted at the bottom-center of the screen and points upwards into the user's face. Two sets of LEDs (LED 1 and LED 2) as shown in Figure 1 are mounted coaxially with the camera lens and located directly below the screen on the vertical axis defined by the screen

0-7803-8554-3/04/$20.00 ©2004 IEEE.

center. The inner ring of LEDs (LED 1) is mounted in front of the camera lens to acquire the bright pupil image. The outer ring of LEDs (LED 2) is fixed coaxially with the lens but mounted further from the optical center to obtain the dark pupil image.

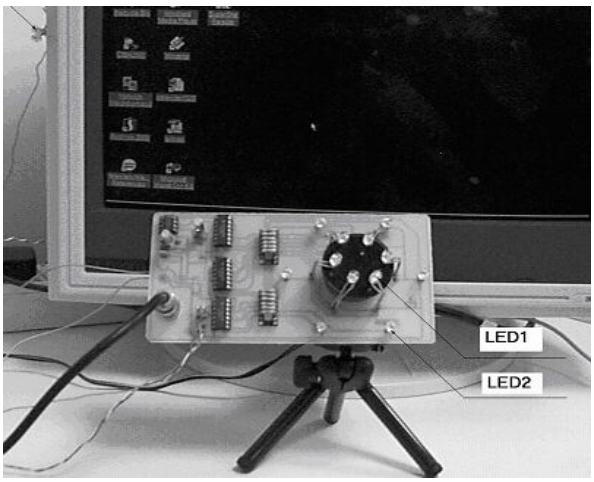

Figure 1. Hardware setup


(a)

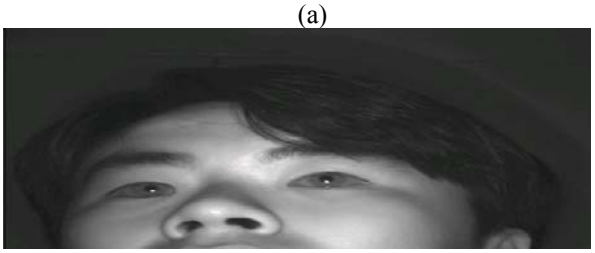
(b)

Figure 2: (a) Bright Pupil Image (Odd Field), (b) Dark Pupil Image (Even Field)

3. EYE TRACKING

3.1. Pupil Detection

In the hardware setup described in Section 2, the bright and dark pupil images are acquired from the odd and even field of each frame respectively as shown in Figure 2. In order to locate the pupils, the dark pupil image is subtracted from the bright pupil image to obtain a difference image. The difference image is thresholded to retain the brightest 0.5% of the pixels, and connected component labeling is then used to obtain connected blobs. From observation, pupils are usually larger than the other blobs in the thresholded image. Hence, the largest 5 blobs are retained as possible pupil candidates, for further analysis.

Since the height of each field is half of that of a frame, the ideal pupil blobs will appear as ellipses with a height to width ratio (HTWR) of 0.5. The 5 candidates are fit to ellipses using an ellipse fitting algorithm [7]. The two with the largest value of |HTWR – 0.5| are discarded. Of the remaining 3 candidates, 3 possible pairs of pupils can be formed. The most likely pair is determined by (i) the sum of their individual |HTWR – 0.5|, (ii) the difference between their sizes and (iii) the difference between the orientations of their major axis. The pair which has the smallest values for two or three of the above criteria is chosen as the most likely pair of pupil blobs. In cases where there is a tie, the pair which has the smallest difference between their orientations is selected.

Often, the pupil centers are estimated by the ellipse centers that are obtained by fitting ellipses to the final pair of pupil blobs. There may be concavities in the pupil blobs due to the glints as shown in Figure 3(a). Thus, directly fitting ellipses to the pupil blobs may be inaccurate as shown in Figure 3(b). Hence, the effect of concavities is first discounted to obtain precise ellipse fitting [8] as shown in Figure 3(c).

(a) Test image (b) With normal ellipse fitting (c) With precise ellipse fitting

Figure 3. Pupil blobs

After estimating the pupil centers, the glint centers are detected. The glint (a bright spot that appears near the pupil) is caused by the light reflected by the cornea and appears in both the dark and bright pupil images. However, they are easier to detect in the dark pupil image as they appear much brighter than the pupil and iris. A search window which is 1.4 times the width and height of pupil, centered at the pupil center is used to locate the glint of each eye. This search window size was determined heuristically.

3.2. Pupil Tracking

Finding the location of pupils in each frame is computationally expensive. Hence, an eye tracking algorithm is implemented to speed up the process of pupil detection. Kalman filtering is used to predict the pupils' positions in consecutive frames thereby greatly limiting the search space.

In this system, a constant velocity $[v_{xn}, v_{yn}]^T$ model has been used for pupil motion from frame to frame. Any changes in velocity are taken as acceleration noise $a_n = [a_{xn}, a_{yn}]^T$. The state vector at time instant n is denoted by $s_n = [x_n, y_n, v_{xn}, v_{yn}]^T$, where $[x_n, y_n]^T$ is the pupil center. The process is governed by the linear stochastic difference equation

$$s_{n+1} = F.s_n + G.a_n \quad (1)$$

where F is the transition matrix and a_n is the process noise. The observed measurement z_n is given by

$$z_n = H.s_n + w_n \quad (2)$$

where H is the measurement matrix and w_n is the measurement noise.

The noise covariances were determined experimentally, using the centroids of the pupils which were located manually in 20 image sequences with 50 images each. The velocity of the pupil was assumed to be constant within each sequence and any changes were treated as acceleration noise. The implemented eye tracking system is able to track the eyes accurately at 25 fps.

4. GAZE MAPPING

Here, we introduce a one-time calibration procedure whereby the calibration data of each user is stored for reuse every subsequent time the user operates the system. Neural networks are used in our system to map the complex and non-linear relationship between the pupil and glint parameters to the gaze point on the screen under varying head poses. In our system, gaze direction is quantized to 20 regions on the screen, 5 columns by 4 rows as shown in Figure 4. The resolution of the screen is 800 x 600 pixels so that each region is of size 160 x 150 pixels.

A	B	C	D	E
F	G	H	I	J
K	L	M	N	O
P	Q	R	S	T

Figure 4. Quantized Eye Gaze Regions

The radial basis function neural network (RBFNN) is relatively fast to train and gives good accuracy. Two RBFNNs are used for mapping of pupil and glint parameters to screen coordinates. One RBFNN consisting of 4 output nodes, is used for mapping the input parameters to one of the rows of the screen. The other RBFNN consisting of 5 output nodes, is used for mapping the input parameters to one of the columns of the screen.

During the calibration process, the user is asked to fixate his gaze at the center of each of the 20 regions with 27 different head poses, resulting in 540 training samples. Each of these training samples is used as the center of the Gaussian basis function of hidden nodes.

Both of the networks have 11 inputs. The inputs to the network are the x and y coordinates of left and right pupils (4 inputs), the areas of left and right pupils (2 inputs), the orientation of the pupils (1 input), and the x and y pupil-to-glint vectors of the left and right eyes (4 inputs). The weights of the network are stored as calibration data for every subsequent time the user operates the system.

The choice of these input parameters is based on the following rationale. The location of the left and right pupils is used to determine whether the user is sitting to the left or right of the camera or directly in front of the camera. The area of the pupils, together with the location of the pupils, is used to account for the distance of the user from the monitor and the pan & tilt of the face. The orientation of the pupils is used to account for the in-plane rotation of the face around the camera optical axis. Lastly, together with the above parameters, the pupil-glint displacements are used to determine the gaze point of the user.

5. EXPERIMENTAL RESULTS

To evaluate the performance of our eye gaze estimator, we performed a series of experiments. Firstly, the system is tested with gaze samples of 6 users with their respective calibration data. From the results shown in Table 1, it can be seen that our system is able to classify the gaze samples with an average accuracy of 97.9% on training data and 85.4% on unseen test data.

User	Training data	Unseen data
1	99.3%	90.0%
2	98.7%	88.7%
3	99.0%	91.3%
4	97.3%	84.7%
5	94.3%	76.3%
6	98.7%	81.3%
Average	97.9%	85.4%

Table 1. Classification results of gaze samples from 6 users.

However, it was noted that the system was not uniformly accurate for all the 20 regions on the screen. The 5 regions on the top row had a lower average accuracy of 75.3% while the remaining 3 rows had a higher average accuracy of 88.8%. This is possibly due to the use of the localized infrared light source which is mounted below

the screen and pointing upwards into the user's face. Hence, when the user tilts his head up to look at the top row of the screen, some of the infrared rays may not be reflected by the retina of the user thereby causing errors in mapping the gaze point correctly. Gaze detection systems that are using similar setup might face this problem too.

In another experiment, the calibration data of each user was used to test with gaze samples from the other two users. The results in Table 2 show that the calibration data of one user is able to determine the gaze point of other users at an average accuracy of 69.5%. This shows that the use of individual calibration data in the first experiment gave a higher accuracy of gaze estimation.

User's calibration data	Calibration data tested on user					
	1	2	3	4	5	6
1		80.0%	74.7%	68.7%	65.3%	72.3%
2	78.0%		74.0%	67.7%	68.3%	69.0%
3	75.3%	72.3%		69.3%	67.3%	75.3%
4	68.0%	73.3%	66.7%		71.3%	66.3%
5	59.3%	63.7%	58.0%	64.7%		59.7%
6	73.7%	71.7%	72.7%	70.3%	68.7%	

Table 2. Classification results for each user, using calibration data of another user.

In the last experiment, we combined the calibration data of all 6 users and grouped them into 540 clusters using K-means clustering algorithm. The center of each cluster was used as the center of the Gaussian basis function of hidden nodes. As shown in Table 3, this yielded an average gaze estimation accuracy of 43.7%.

	Calibration data tested on user						
	1	2	3	4	5	6	Avg
Accuracy	45.3%	48.7%	46.3%	37.7%	38.7%	45.7%	43.7%

Table 3. Classification results on test data based on combined training data from all 6 users.

6. CONCLUSION

In this paper, we propose a technique for eye gaze estimation using infrared light source. Our system has a spatial gaze resolution of about 3 degrees as compared to other systems which are accurate to 1-2 degrees, but it requires only a one-time calibration procedure, allows natural head movement and yet provides a relatively robust and accurate gaze tracking. Currently, our system operates at 25 fps with an average accuracy of 85.4% for gaze estimation using individual calibration data.

It is possible to increase the spatial gaze resolution by increasing the number of quantized regions on the screen. However, this will result in increasing the number of training samples and hidden nodes of the RBFNN thus causing the calibration procedure to be more cumbersome. Besides that, the experimental results in Tables 1, 2 and 3 show that our system gives better gaze estimation accuracy using calibration data which is specific for each user. However, it was noted that existing techniques that use localized infrared light source may not work well when the user's head tilts above a certain angle. This is because some of the reflected infrared rays from the retina may not be captured by the camera.

We are currently considering ways to further increase the spatial gaze resolution. Besides that, we are also in the process of making the system to be more user-independent.

7. REFERENCES

[1] Y. Ebisawa and S. Satoh, "Effectiveness of Pupil Area Detection Technique using Two Light Sources and The Image Difference Method", *Proceedings of The 15th Annual International Conference of the IEEE Eng. In Med. & Biol. Soc.*, pp. 1268-1269. 1993

[2] Marcio R. M. Mimica and Carlos H. Morimoto, "A Computer Vision Framework for Remote Eye Gaze Tracking", *Proceedings of the XVI Brazilian Symposium on Computer Graphics and Image Processing, paper 105*, 2003

[3] A.Garcia, F.M. Sanchez, A. Perez, J. L. Pedraza, R. Mendez, M. L. Cordoba, M. L. Munoz, "A Distributed Real Time Eye-Gaze Tracking System," *Proceedings on ETFA '03. IEEE Conference, Volume: 2 pp. 545-551,* Sept 2003

[4] Dong Hyun Yo, Jae Heon Kim, Bang Rae Lee, Myoung Jin Chung, "Non-contact eye gaze tracking system by mapping of corneal reflections", *Proceedings Fifth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 94-99*, May 2002

[5] Carlos H. Morimoto, Arnon Amir, & Myron Flickner, **"**Free head motion eye gaze tracking without calibration", *Extended abstracts on Human Factors in Computing Systems CHI '02, ACM Press New York, NY, USA , pp. 586-587*, April 2002

[6] Sheng-Wen Shih, Yu-Te Wu, Jin Liu, "A calibration-free gaze tracking technique", *Proceedings of the 15th International Conference on Pattern Recognition, pp. 201-204*, sept 2000

[7] A.W. Fitzgibbon, M. Pilu and R.B. Fisher, "Direct Least Square Fitting of Ellipses", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 21, No. 5, pp. 476 – 480, May 1999

[8] Wen Gang, "Eye Gaze Aided Human Computer Interface", *Master of Engineering Thesis*, *NUS*, 2003