Proceedings of the 26th Annual International Conference of the IEEE EMBS
San Francisco, CA, USA • September 1-5, 2004

Non-intrusive Eye Gaze Tracking Under Natural Head Movements

S. C. Kim, M. Sked, and Q. Ji
Department of Electrical Engineering System Engineering, Rensselaer Polytechnic Institute, NY, USA

Abstract—**We propose an eye gaze tracking system under natural head movements. The system consists of one CCD camera and two mirrors. Based on geometric and linear algebra calculations, the mirrors rotate to follow head movements in order to keep the eyes within the view of the camera. Our system allows the subjects head to move 30 cm horizontally and 20 cm vertically, with spatial gaze resolutions about 6 degree and 7 degree, respectively and a frame rate about 10 Hz. We also introduce a hierarchical generalized regression neural networks (H-GRNN) scheme to map eye and mirror parameters to gaze, achieving a gaze estimation accuracy of 92% under head movements. The use of H-GRNN also eliminates the need for personal calibration for new subjects since H-GRNN can generalize. Preliminary experiments show our system is accurate and robust in gaze tracking under large head movements.**

Keywords—**Eye movement, Gaze tracking, Generalized Neural Network, Head movement**

I. Introduction

Eye gaze means where a person is looking and gives much useful information to aid in understanding of a person's intention, predictions of a person's actions, or reactions to external stimulation [1, 2]. Especially, fixation sequences or scan-paths are interesting for the study of visual cognition, cognitive ability for language, speech learning studies, and other applications such as the design of human-computer interfaces and virtual reality [3-6].

Many methods for gaze tracking of eye images have been introduced in the past by several researchers [7-11]. These can be categorized into two types. One is intrusive method, which is that user should wear a kind of goggle or hat with an attached the camera, and another is non-intrusive method which does not have to wear any devices. In the intrusive method, we can obtain high resolution eye images because a camera can be attached close to eye. However, we can not know the information about head position and face rotation from eye images. Therefore, additional sensors are required to estimate gaze points such as magnetic sensor. And these equipments obstruct the visual field, make the users uncomfortable, and give some restriction in action. Although it is very useful on the purpose of clinic or research, but it is not convenient for estimation of gaze in the natural environment or long time monitoring such as human computer interaction and cognition research. On the other hand, in the non-intrusive method, it is close to being the ideal method as it does not need any devices to be attached to the person. However, there are also several shortcomings. We use a zoom-in lens to magnify the eye region on the face. But it causes narrow view field of the camera to restrict head movements, and sometimes requires two or three cameras for tracking the eye region on the face.

In this paper, we propose a non-intrusive gaze tracking system with only one camera and two mirrors to overcome the above mentioned limitations. The information of pupil image and mirror angles is mapped by a non-linear mapping function to the actual gaze points.

II. Methodology

A. System Configuration

The system consists of a near infra-red CCD camera (TV ZOOM LENS M6Z, Hitachi, Japan) and two conventional back-surface mirrors as shown in Fig. 1. It is located in front of the computer monitor facing the user. The camera obtains the eye images reflecting from Mirror 1 and Mirror 2. Mirror 1 rotates about the y-axis to compensate for head movements in the vertical direction, and Mirror 2 rotates about the z-axis to compensate for head movements in the horizontal direction. The diameter of each mirror is about 15cm. These angles are controlled by the two step motors (Futaba FP-S148 RC, Pontech, USA) through a computer serial porter (RS232C). The motors output torque of 3kg-cm with a precision of 0.7°.

We use the active illumination to obtain pupil position in the images reflected by mirrors with less light effect of surround environment. Infrared LEDs for light source have a peak wavelength of 880nm. To minimize interference from light sources other than the IR illuminator, we used an optical band pass filter, which has a wavelength pass band only 10nm wide at 880nm [10]. When the inner LEDs of Fig. 1(b) along the camera optical axis are turned on, the bright pupil image, so-called "red eye", is produced such as Fig. 2(a). When the outer LEDs are turned on, the dark pupil image is produced such as Fig. 2(b). The on-and-off switching of the inner and outer rings of infra-red LEDs is synchronized with the even and odd field separator signals which are obtained from the CCD camera.

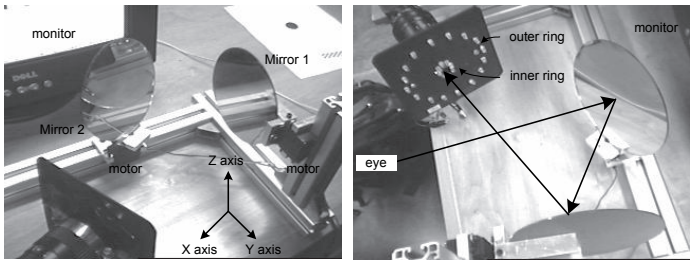
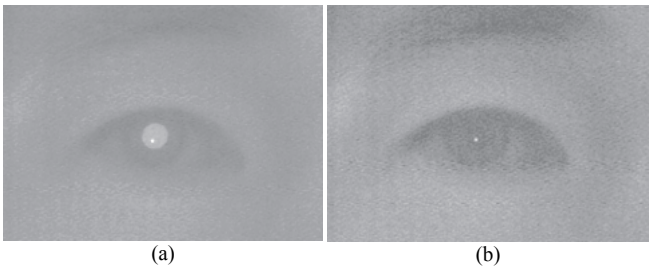


Fig. 1. Eye gaze estimation system consisting of one narrow-field camera, two mirrors, and active light source.

0-7803-8439-3/04/$20.00©2004 IEEE



(a) (b)

Fig. 2. The images reflected from two mirrors (a) bright pupil image, (b) dark pupil image.

B. Compensation of Head Movements Using Two Mirrors

According to head movements, the two mirrors are controlled to align the pupil images with the optical axis of the CCD camera. Because each of the mirrors reflection is independent of the other mirror's rotation, we are able to solve each problem separately.

The problem of moving the image of the pupil to the optic axis of the camera and moving the optical axis of the camera into the visual sight of the user is the same. Fig. 3(a) shows the geometry of the scene in the X-Y plane to solve Mirror 2 rotation angle, Φ_2 (pan angle). Each angle A, B, C, and the total reflection angle D can be found as in (1).

$$\begin{aligned} A &= \frac{\pi}{2} - \Phi_2 = \pi - B - C = \Theta_1 + \Theta_2 \\ B &= \frac{\pi}{2} - \Theta_2 \\ C &= \frac{\pi}{2} - \Theta_1 \\ D &= \pi - 2\Theta_1 - 2\Theta_2 = \pi - 2\left(\Theta_1 + \Theta_2\right) \end{aligned} \tag{1}$$

From (1), we can solve the relationship between Φ_2 and D as in (2).

$$\Phi_2 = \frac{\pi}{2} - \Theta_1 - \Theta_2 = \frac{1}{2} D \tag{2}$$

Finally, we can calculate Φ_{2new} from the difference between D and D' and the previous rotation angle, Φ_{2old} as in (3).

$$\Phi_{2new} = \Phi_{2old} + \frac{D - D'}{2} \tag{3}$$

The problem of solving the Mirror 1 rotation angle, Φ_1 (tilt angle), is easier than the pan angle because it can be done without having to investigate the reflection angle on Mirror 2. The angle of rotation is related to the reflection angle of the pupil from Mirror 2 and the reflection angle between the camera vector and Mirror 1. Fig. 3(b) shows the geometry of the scene in the X-Z plane, and the rotation angle (Φ_1) can be calculated from (4).

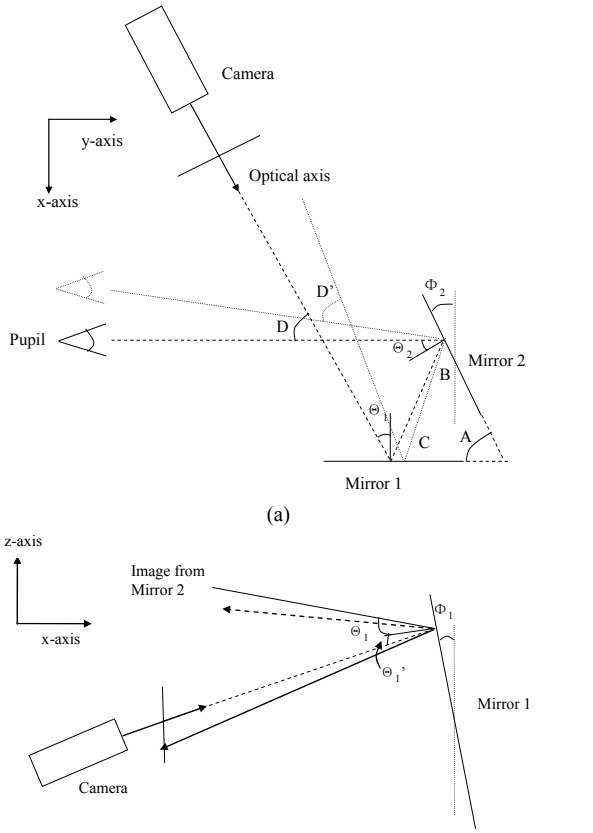


Fig. 3. The relationship among mirrors, pupil, and the optical axis of a camera (a) calculation of reflection angle for Mirror 2 (b) calculation of reflection angle for Mirror 1.

$$\Phi_{1new} = \Phi_{1old} + \frac{\left(\Theta_1{}' - \Theta_1\right)}{2} \tag{4}$$

C. Gaze Estimation

Tracking head position and estimating gaze point is another problem. To estimate gaze point, we need more information related to the pupil as well as the eye position in image and head position. Seven parameters were chosen for gaze estimation: Δx, Δy, r, g_x, g_y, Φ_1, Φ_2. Δx and Δy are the displacements of the pupil and glint in the x and y axis, respectively. These values have high relationship between eye gaze directions. r is the ratio of the major to minor axes of the pupil. The pupil appearances vary by head movements and face orientation. It should be close to one when the face is frontal. g_x and g_y are the glint position in the eye images, which are used to account for the in-plane head translation. Φ_1 and Φ_2 are angle of Mirror 1 (tile mirror) and Mirror 2 (pan mirror), respectively. These have the information about the head position.

We got the clear pupil image by subtracting dark pupil image from bright one, and converting gray image to binary

image with the threshold depends on the boundary of pupil. After removing noise from the binary image, it is fitted with an ellipse and the pupil center is found from the ellipse-fitting results. The ratio of the major to minor of the fitted ellipse is set to *r*. Glint could be found in the dark image. Because it was much brighter than the rest of the eye images compare to bright pupil image.

We mapped the seven parameters to the actual gaze with a mapping function. There are also several other mapping functions such as linear mapping and non-linear mapping. The generalized regression neural networks (GRNN), developed by universal function approximator, are fast in training, can model non-linear function, and have been shown to perform well in noise environments given enough data [12]. Therefore we used the GRNN as a mapping function. To reduce misclassification among neighboring gaze classes, we applied a hierarchical GRNN (H-GRNN) classifier. H-GRNN is again to perform the GRNN with only the gaze region and its neighborhood regions that tend to get misclassified. These are so-called sub-classifiers. The details about that are described in the paper [13]. The mapping function, once determined via a training procedure, does not have to calibrate again because the input parameters are chosen such that they remain relatively invariant for different people. This effectively eliminates the need to re-calibrate for another person.

We used 24(6x4) regions on the screen which has a resolution of 1152x864 pixels. During the training data acquisition, the users should keep their fixing gaze on each gaze region. On each fixation, 20 sets of input parameters were collected for the right eye. We removed all data falling beyond the standard deviations, which are considered as unstable data. The training data under different head positions as many as possible were collected with the above method. We limited the range of head movements to 30 cm in the horizontal direction, and 20 cm in the vertical direction, when we determined the gaze, although the system can track larger head movements, because the range is broad enough to track natural head movements in the front of the computer screen.

Before obtained the weights of the GRNN, the acquired training data were normalized as follows. Δx, Δy, Φ_1 and Φ_2 were normalized with each of those minimum and maximum values. g_x and g_y were normalized by 640 and 480, respectively because the maximum image size is of the monitor is 640x480.

III. Results

A. Tracking Eye Movements under Head Movements

Fig. 4 shows that the proposed system is able to keep the pupil images on the optical axis of the camera under different head positions. Whenever the pupil image moves away from the image center due to head movements, mirrors will also move accordingly to compensate the head movements in order to keep the eyes in the optical axis of the camera. Pupil was not always located at center of image after recovery. That is because the step motors spatial resolution is not sufficiently high (about 0.7°). If we use the step motor having the high spatial resolution, the problem will be easily abated. The system could compensate large head movements such as 80 cm in the horizontal direction, and 60 cm in the vertical direction. The compensation range of the pan mirror is better than the tilt mirror. The system covers sufficient head movements when a user sits down in the front of 70~80cm from a computer screen, because the user has smaller vertical movements than horizontal movements. The vertical speed was also a bit slow compared to the horizontal speed such as about 2cm/sec and 3cm/sec, respectively, but it still works well for natural head motion.

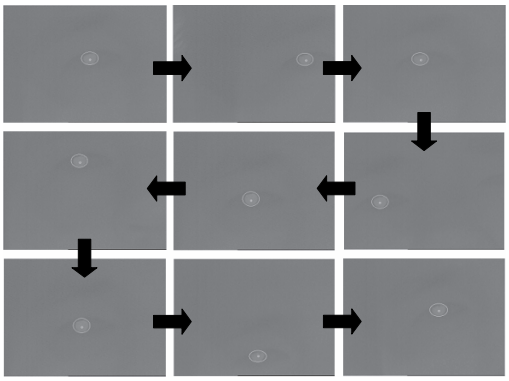


Fig. 4. Eye tracking during head movements.

B. Gaze Estimation under Head Movements

Table I shows the accuracy of gaze classification by GRNN and H-GRNN, when a user gazed one of 24 regions (6x4 grid) on the screen with nature head movements about 30cm horizontally, 20cm vertically. An accuracy of gaze classification by GRNN was 85%, but that by H-GRNN was 92%. H-GRNN showed higher accuracy than GRNN, and especially H-GRNN was more superior on the boundary

TABLE I.
GAZE CLASSIFICATION RESULTS FOR THE GRNN AND THE H-GRNN CLASSIFIER. GRNN (MEAN: 86±8%), H-GRNN (MEAN: 92±5%)

Vertical \ Horizontal		1	2	3	4	5	6
1	GRNN	85	67	85	90	96	92
	H-GRNN	89	98	94	91	96	100
2	GRNN	72	88	87	80	99	78
	H-GRNN	90	88	89	86	99	88
3	GRNN	88	93	89	93	86	80
	H-GRNN	94	93	89	95	88	92
4	GRNN	67	85	88	93	85	84
	H-GRNN	85	89	99	93	85	96

regions than the inner regions of the screen such as 2 (first row and second column), 3 (first row and third column), 6, 7 (second row and first column), 12, 13, 18, 19, 21 and 24 regions. Even though the head or face was a little rotated, the gaze points can be estimated.

IV. Discussion

The obtained image was not as bright as it had been before being implemented into the mirror system. This is most likely attributed to the low reflectivity of the mirrors. Conventional mirrors, like the ones you probably have in your home are typically "rear surface or back surface" mirrors. A back surface mirror typically has between 60-75% of reflection in the visible range. Assume the reflectivity is 70%, the image brightness decreases to 50% because we use two mirrors. The reflected image is reflected back out through the glass surface a second time, the light tends to scatter, resulting in the image blurred. If we use the front surface mirror to overcome this problem, the image bounces directly off the mirrored surface without having to travel through the glass. Therefore, we can expect to get brighter and clearer images. As an alternative method, we can increase the intensity of infra-red light source by increasing the number of LEDs or current flowing into the LEDs, but the high intensity can cause damage to the eyes. In this case, the intensity should not exceed the guideline recommended from international standard (ANSI Z136.1).

Although the tracker worked very well to always keep the pupil in the view of the camera under slow and steady head movements, the system could not track the fast head movements. This problem can be solved by using faster and more accurate motors and controllers because the motors need time delay at least 5ms due to the communication through the serial port. If the sizes of mirrors, which have higher mechanical inertia, are decreased, it makes control of the mirror easy.

The gaze accuracy in the boundary regions by H-GRNN showed much higher than those by GRNN. That is why H-GRNN uses sub-classifier trained with only the gaze region and its neighborhood regions as the second pass, and besides the number of neighborhoods in the boundary regions is smaller than those in inner regions.

We used only one camera to track eye, it made difficult to find initial eye position when the system started or it could not catch the pupil position due to fast eye movements. It may be also solved using the motors with fast response time and high spatial resolution. However, the problem can occur with eyes closed, it is still critical problem in eye tracking system using eye images.

V. Conclusion

The major deficiency in most gaze estimation systems has been the lack in ability to robustly track the users gaze under natural head movements. We proposed an eye gaze tracking method with only one camera, two mirrors, and active light source under the natural head movements. The mirrors of rotation angles according to the head position were computed by geometric and linear algebra calculations. The system responded very well. This effectively reduced the head movement influence, and could eliminate eye gaze loss. The use of H-GRNN schemes further improved the accuracy of the gaze classification more than GRNN. It has significantly relaxed the constraints such as head and face fixation or recalibration imposed by most existing commercial eye trackers.

Our system can be applied in many applications such as human computer interaction and smart graphics. Especially, it would be very useful for language, speech learning studies to apply to infants to keep from continually stabilizing the head or face position

Acknowledgment

This work was supported by the Post-doctoral Fellowship Program of Korea Science & Engineering Foundation (KOSEF). The authors would also like to thank Zhiwei Zhu for his assistance throughout this project.

References

[1] H. Bekkering and S. F. W. Neggers, "Visual search is modulated by action intentions," *Psychological Science*, vol. 13, pp. 370-374, 2002.

[2] P. Vuilleumier, "Perceived gaze direction in faces and spatial attention: a study in patients with parietal damage and unilateral neglect," *Neuropsychologia*, vol. 40, pp. 1013-26, 2002.

[3] M. L. Phillips and A. S. David, "Visual scan paths are abnormal in deluded schizophrenics," *Neuropsychologia*, vol. 35, pp. 99-105, 1997.

[4] G. Csibra, "Teleological and referential understanding of action in infancy," *Philos Trans R Soc Lond B Biol Sci*, vol. 358, pp. 447-58, 2003.

[5] M. K. Tanenhaus and M. J. SpiveyKnowlton, "Eye-tracking," *Language and Cognitive Processes*, vol. 11, pp. 583-588, 1996.

[6] C. L. Kleinke, "Gaze and eye contact: a research review," *Psychol Bull*, vol. 100, pp. 78-100, 1986.

[7] C. H. Morimoto, D. Koons, A. Amir, and M. Flickner, "Pupil detection and tracking using multiple light sources," *Image and Vision Computing*, vol. 18, pp. 331-335, 2000.

[8] Y. Ebisawa, "Improved video-based eye-gaze detection method," *Instrumentation and Measurement, IEEE Transactions on*, vol. 47, pp. 948-955, 1998.

[9] D. H. Yoo, J. H. Kim, D. H. Kim, and M. J. Chung, "A human-robot interface using vision-based eye gaze estimation system," presented at Intelligent Robots and System, 2002. IEEE/RSJ International Conference on, 2002.

[10] Q. Ji and X. J. Yang, "Real-time eye, gaze, and face pose tracking for monitoring driver vigilance," *Real-Time Imaging*, vol. 8, pp. 357-377, 2002.

[11] K. Ukai, S. Saida, and N. Ishikawa, "Use of infrared TV cameras built into head-mounted display to measure torsional eye movements," *Japanese Journal of Ophthalmology*, vol. 45, pp. 5-12, 2001.

[12] D. F. Specht, "A General Regression Neural Network," *IEEE Transactions on Neural Networks*, vol. 2, pp. 568-576, 1991.

[13] Q. Ji and Z. Zhu, "Non-intrusive Eye and Gaze Tracking for Natural Human Computer Interaction," *MMI-interaktiv Journal*, vol. 6, 2003.