

Non-contact Gaze Tracking with Head Movement Adaptation based on Single Camera

Ying Huang, Zhiliang Wang, and An Ping

Abstract—With advances in computer vision, non-contact gaze tracking systems are heading towards being much easier to operate and more comfortable for use, the technique proposed in this paper is specially designed for achieving these goals. For the convenience in operation, the proposal aims at the system with simple configuration which is composed of a fixed wide angle camera and dual infrared illuminators. Then in order to enhance the usability of the system based on single camera, a self-adjusting method which is called Real-time gaze Tracking Algorithm with head movement Compensation (RTAC) is developed for estimating the gaze direction under natural head movement and simplifying the calibration procedure at the same time. According to the actual evaluations, the average accuracy of about 1° is achieved over a field of $20 \times 15 \times 15$ cm³.

Keywords—computer vision, gaze tracking, human-computer interaction.

I. INTRODUCTION

WITH the improvement of intelligence of the computer vision technology, it has been widely applied in many research areas. Especially in the human-computer interaction domain, the computer can comprehend the user's motion and interact with the user through its vision system. The non-contact gaze tracking technology which is a new interactive technology was inspired from the computer vision technology, and it has become more and more prevalent and mature. Because of the feasibility, reliability, accuracy and ease for use of the non-contact gaze tracking technology, many researchers are focusing on developing new interface for interactions via the user's gaze [1].

It is acknowledged that the non-contact gaze tracking is less intrusive and more comfortable to the user, so various non-contact gaze tracking systems have been proposed as a user interface. However, the low tolerance for head motion is an undeniable drawback of the non-contact gaze tracking. Therefore, improving the ability of accommodating the head movement has become one of the issues in focus on the technique researches. And many techniques have proposed to

tolerate free head motion by utilizing multiple cameras or adjustable vision systems [2], [3], [4], [5]. In general, the substance of the works based on multiple cameras is using a wide angle (WA) vision system to direct a movable narrow angle (NA) vision system. But these kinds of systems share two common deficiencies: Firstly, the complicated configurations of the system increase its operation difficulty and hinder its application prospect; Secondly, it takes a relatively long time to reacquire the gaze information when the eye moves outside the NA camera's view.

Considering the higher operability and better performance in real-time, the gaze tracking technique proposed in this paper is designed for a simple and compact system which bases on a single camera with fixed wide view. As it is analyzed by Hennessey et al. [6], this kind of system is easy to operate and consume little time in reacquisition. Due to the low tolerance for head movement of the single camera system, a set of solutions is presented to make the system adapting to the head movement. The solution is named as Real-time gaze Tracking Algorithm with head movement Compensation (RTAC). The RTAC does not depending on the physical model of eyeball which is widely used at present [5], [6], [7], [8], but estimates the gaze direction based on the relative changes of the eyeball's orientation and the head's position to the prior information captured in the calibration procedure. The RTAC has three primary portions which are the gaze direction calculation relative to face, the head movement calculation, and the head movement compensation respectively.

The remaining portion of the article is organized as follows. In Section 2, the Real-time gaze Tracking Algorithm with head movement Compensation (RTAC) is narrated specifically. Section 3 elaborates on the evaluations of the implementation of the gaze tracking system. Finally, conclusions are summarized in section 4.

II. THE RTAC

A. The gaze direction calculation relative to face

According to the classical Pupil Center Corneal Reflection (PCCR) technique, the gaze direction calculation has two steps: the extraction of the pupil-glint vector and the acquisition of the

Ying Huang is with the University of Science and Technology Beijing, No.30 Xueyuan Road, Haidian district, Beijing, China. (e-mail: huangying0108@126.com).

Zhiliang Wang is with the University of Science and Technology Beijing, No.30 Xueyuan Road, Haidian district, Beijing, China. (e-mail: Zhiliang_W@263.net).

An Ping is with the University of Science and Technology Beijing, No.30 Xueyuan Road, Haidian district, Beijing, China. (e-mail: ppfishpa@gmail.com).

gaze direction mapping function [8].

The pupil-glnt vector was extracted from the eye image under infrared illumination. As shown in Fig.1, the vector from the glints center to the pupil center is defined as the pupil-glnt vector (V_{PG}).

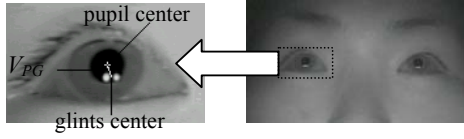


Fig.1 Image of the eyes under infrared illumination. Since we used two infrared illuminators, there were two glints on the cornea (Purkinje images).

The vector V_{PG} is a 2-D parameter extracted from the image. A specific mapping function was used to map the V_{PG} to the 3-D gaze direction in space. The gaze direction which is the eyeball orientation can be interpreted as the visual angle θ , and the mapping function $\theta=f(V_{PG})$ is specified as the following equations:

$$\theta_h = b_{\theta_h} V_x + a_{\theta_h} \quad (1)$$

$$\theta_v = b_{\theta_v} V_y + a_{\theta_v} \quad (2)$$

Where θ_h represents the angle between the gaze direction and the horizontal direction, θ_v represents angle between the gaze direction and the vertical direction. The coefficients a_{θ_h} , b_{θ_h} and a_{θ_v} , b_{θ_v} are estimated from a set of pairs of pupil-glnt vectors and the corresponding visual angles. These pairs are collected in a modeling experiment which is called the procedure of vector to angle.

The reasons why the mapping function is linear are from the following two aspects:

Firstly, the gaze direction θ is calculated relative to the face only. And the non-linear factor brought by the head pose is not taken into account here.

Secondly, because the wide view of the camera, the image resolution of the eye is relatively lower. Therefore, little non-linear information is manifested in detail.

In practice, the linear functions are applicable for the ordinary interaction task. Meanwhile, experiment results show that the linear model is indistinctive for any individual at the same position, so it can be a general model for different users.

B. The head movement calculation

The gaze direction mapping function do not have adaptability to the significant head movement, it will fail to estimate the gaze direction accurately when the user moves his head. Therefore, the gaze direction calculation model should be rectified by compensating for the effects of head movement, and the compensation should be built on the accurate calculation of the head movement. Since there no facilities or vision systems are especially for the head movement estimation, we utilize the wide view vision system to estimate the head movement at the same time. Although the vision system based on single camera cannot localize the head position in space exactly, especially in depth. But it can perceive the relative position changes by computing changes of

the corresponding image. It is the relative changes that are fit for making the rectification of the gaze direction mapping function.

As shown in Fig.2, position changes in front of the camera can basically be projected along three directions which are X_c , Y_c and Z_c axes in the camera coordinate system. Define the position where we obtain the sample data for forming the gaze direction mapping function as a relative original position $P_0(L_{h0}, L_{v0}, d_0)$ in the X_c - Y_c - Z_c coordinate system and the position of glnt on the image plane at the reference standard position is defined as I_0 . A new head position $P_1(L_{h1}, L_{v1}, d_1)$ relative to P_0 can be represented as $L_{h1} = L_{h0} + \Delta L_h$; $L_{v1} = L_{h0} + \Delta L_v$; $d_1 = d_0 + \Delta d$. The corresponding $I_1(I_{x1}, I_{y1})$ can be represented as $I_{x1} = I_{x0} + \Delta I_x$, $I_{y1} = I_{y0} + \Delta I_y$. The position changes $\Delta P_c(\Delta L_h, \Delta L_v, \Delta d)$ has a mapping relation with the image changes $\Delta I(\Delta I_x, \Delta I_y)$ as follows:

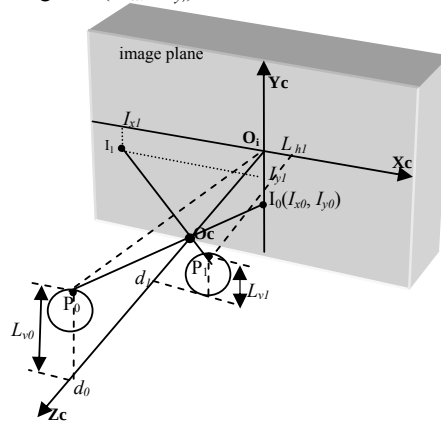


Fig. 2 Glnt image formation when the eye at different positions in front of the camera. The origin of the camera coordinate system (O_c) is the focus of the lens and the origin of the image coordinate system (O_i) is the center of the image plane.

$$\Delta I_x = (b_{2x} \Delta L_h + a_{2x}) \Delta d^2 + (b_{1x} \Delta L_h + a_{1x}) \Delta d + (b_{0x} \Delta L_h + a_{0x}) \quad (3)$$

$$\Delta I_y = (b_{2y} \Delta L_v + a_{2y}) \Delta d^2 + (b_{1y} \Delta L_v + a_{1y}) \Delta d + (b_{0y} \Delta L_v + a_{0y}) \quad (4)$$

The forms of the equation(3), (4) and their coefficients $b_{ij}(i=0,1,2;j=x,y)$, $a_{ij}(i=0,1,2;j=x,y)$ are generated from a set of pairs of $(\Delta I_x, \Delta I_y)$ and $(\Delta L_h, \Delta L_v)$ at the reference position with the relative depth Δd to the d_0 . These pairs are collected in a modeling experiment which is called the procedure of position changes perception. In the procedure we used a target with a special mark to imitate the eye with the glnt, because the target can be located at the appointed position much easier and more precise than human eyes. The sample positions we selected distribute equally over the camera's view. As shown in Fig.3 which is a lateral view, select 21 sample positions on the Z axis in the range of $[-10, +10]$ with the interval of 1cm. At every z coordinate, select 18 sample positions on the X axis in the range of $[-9, +9]$ with the interval of 1cm, and select 14 positions on the Y axis in the range of $[-7, +7]$ in the same way.

At every sample position, the mark's coordinate (X, Y, Z) and its corresponding image's coordinate (I_x, I_y) are recorded. At

The diagram illustrates the experimental setup. A camera is positioned at a distance of 30 cm from a target. The camera's optical axis is aligned with the Z-axis. The target is located at a distance of 10 cm from the camera. The camera is tilted at an angle θ relative to the Z-axis. A fixation screen or monitor is positioned at the origin of the coordinate system. The coordinate system has axes X_w , Y_w , and Z_w .

2665

orientations $\Delta\theta_L$ and $\Delta\theta_R$ by substituting ΔV_L and ΔV_R to the mapping function $\theta=f(V_{PG})$ respectively.

Substitute $\Delta\theta_L, \Delta K_L$ and ΔP_{cl} to compensation model to get the gaze direction of the left eye θ'_L . Then get the gaze direction of the right eye θ'_R in the same way. Here, we define the average of the θ'_L and θ'_R as the real gaze direction θ .

III. EVALUATION

A physical implementation of the gaze tracking system is shown in Fig.4. The system is composed of a single camera and dual infrared illuminators. The camera is wide view and low luminance sensitive. The wavelengths of the infrared illuminators are both 850 nm, and they sit by the side of the camera symmetrically. The resolution of the image grabber is 768×576, and the frame rate is 25fps.



Fig. 4 The physical system implementation.

An evaluation experiment was taken to measure the accuracy of the system at several different head positions. Three subjects were invited to take the experiment sequentially. In the experiment, the subject was seated at five different positions after the calibration. As shown in Fig.3, the head positions in the X-Y-Z coordinate system are (0,0,0), (0,0,5), (0,0, -10), (5,0,0) and (0,5,0) respectively. And there were 10 objects arrayed in 2 rows by 5 columns on the monitor.

The accuracies of each subject at each head position are listed in Table1. As shown in Table 1, where its first row represents the different positions of head, and its first column represents the different subjects, 'H', 'V' represent the accuracies in units of degree in the horizontal and vertical directions respectively. Each data represents the tracking accuracy, which is the average value of the errors when the subject looks all the 10 objects on the monitor at the position.

According to Table 1, little difference of the gaze tracking accuracy level is observed in different subjects. So the RTAC is proved to be applicable to different people. In addition, for each subject, the average accuracy has been achieved at about 1°. Therefore, the RTAC is proved to be applicable to natural head movement, and the accuracy is suitable for the ordinary human-computer interaction.

TABLE 1
ERRORS OF THREE SUBJECTS AT FIVE POSITIONS

	(0,0,5)		(0,0, 10)		(0,0,0)		(5,0,0)		(0,5,0)	
	H	V	H	V	H	V	H	V	H	V
1	0.7	0.8	1.1	1.3	0.5	0.6	0.9	0.9	0.6	1.1
2	0.6	0.8	1.2	1.2	0.5	0.6	0.9	0.8	0.7	1.1
3	0.7	0.9	1.2	1.3	0.6	0.7	1.0	0.8	0.8	1.2

IV. CONCLUSIONS

In this paper, a design of proposal is presented to improve the usability and practicality of the non-contact gaze tracking system as the user interface. The proposal bases on simple configuration to make the system easy to set up. For offsetting the disadvantage due to the simple hardware, an algorithm called the Real-time gaze Tracking Algorithm with head movement Compensation (RTAC) is developed to accommodate the head movement automatically. The RTAC is different from the conventional methods based on the physical model of the eyeball. It calculates the gaze direction by extracting the relative changes of the eyeball orientation with addition of the relative head movement compensation. And in the RTAC, the calibration procedure is simplified as a procedure of one point at one position calibration which only lasts about 3 seconds. An evaluation test confirmed that the average accuracy of gaze tracking is about 1° for different people at different positions within the camera's view. The allowable range of the head movement achieves approximately 20×15×15cm³ at the distance of 30cm to the camera, which is comparable to that of other reported systems with complicated configurations. The system operates at about 20fps, and the reacquisition time after the rapid and significant head movement is about 80ms for the 25Hz camera.

ACKNOWLEDGMENT

This project is supported by a grant from the National High-tech R&D Foundations of China (863 Program) (No.2007AA01Z160).

REFERENCES

- [1] C. H. Morimoto, and M. R. M. Mimica, "Eye gaze tracking techniques for interactive applications," *Computer Vision and Image Understanding* 98, 2005, pp. 4-24.
- [2] B. Nouredin, P. D. Lawrence, and C. F. Man, "A non-contact device for tracking gaze in a human computer interface," *Computer Vision and Image Understanding* 98, 2005, pp.52-82.
- [3] D. Beymer, and M. Flickner, "Eye Gaze Tracking Using an Active Stereo Head," In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Jun.16-22, Madison), USA, 2003, Vol.2, pp.451-458.
- [4] T. Ohno, N. Mukawa, and S. Kawato, "Just Blink Your Eyes: A Head-Free Gaze Tracking System," In *proceedings of CHI' 03* (Apr.5-10, Florida), ACM Press, USA, 2003, 950-951.
- [5] Z. Zhu, and Q. Ji, "Eye Gaze Tracking Under Natural Head Movements," In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and pattern Recognition* (Jun.20-25, San Diego, CA), USA, 2005, pp. 918-923.
- [6] C. Hennessey, B. Nouredin, and P. Lawrence, "A Single Camera Eye-Gaze Tracking System with Free Head Motion," In *Proceedings of the Symposium on Eye Tracking Research and Applications* (Mar. 27-29, San Diego, CA), USA, 2006, pp.87-94.
- [7] A. Villanueva, R. Cabeza, and S. Porta, "Gaze tracking system model based on physical parameters," *International Journal of Pattern Recognition and Artificial Intelligence* 21, 5, 2007, pp.855-877.
- [8] T. Nagamatsu, J. Kamahara, T. Iko, and N. Tanaka, "One-point calibration gaze tracking based on eyeball kinematics," In *Proceedings of the Symposium on Eye Tracking Research and Applications* (Mar.26-28, Georgia), USA, 2008, pp.95-98.