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 20×15×15 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 reacquaint 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
The head movement at the same time. Although the vision estimation, we utilize the wide view vision system to estimate the calculation of the head movement. Since there are no facilities or the compensation should be built on the accurate rectified by compensating for the effects of head movement, therefore, the gaze direction calculation model should be 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 is in distinctive for any individual at the ordinary interaction task. Meanwhile, experiment results show the reasons why the mapping function is linear are from the head movement calculation.

A. 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 \(P_0\), a new head position \(P_1\) relative to \(P_0\) can be represented as \(L_{h1} = L_{h0} + \Delta L_h\); \(L_{v1} = L_{v0} + \Delta L_v\); \(d_1 = d_0 + \Delta d\). The corresponding \(I_1\) can be represented as \(I_{1x} = I_{0x} + \Delta I_x\), \(I_{1y} = I_{0y} + \Delta I_y\). The position changes \(\Delta P_1(\Delta L_h, \Delta L_v, \Delta d)\) has a mapping relation with the image changes \(\Delta I(I_{1x}, I_{1y})\) as follows:

\[
\Delta I = f_1(\Delta P_1)
\]

where \(\Delta P_1\) is the relative position changes. The forms of the equation (3), (4) and their coefficients \(b_{ij}(i,j=x,y)\), \(a_{ij}(i=0,1,2;j=x,y)\) are generated from a set of pairs of \(\Delta I(I_{1x}, I_{1y})\) and \(\Delta P_1(\Delta L_h, \Delta L_v, \Delta d)\) at the reference position with the relative depth \(\Delta d\) to the \(d_0\). These pairs are collected in a modeling experiment which is called the procedure of vector to angle.

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

2. 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 indistinguishable 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 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-glint vector \(V_{PG}\).

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 \(\theta\), and the mapping function \(\theta = f(V_{PG})\) is specified as the following equations:

\[
\theta_h = b_{hb}\theta_v + a_{hb}
\]

\[
\theta_v = b_{vb}\theta_v + a_{vb}
\]

Where \(\theta_h\) represents the angle between the gaze direction and the horizontal direction, \(\theta_v\) represents angle between the gaze direction and the vertical direction. The coefficients \(a_{hb}, b_{hb}\) and \(a_{vb}, b_{vb}\) are estimated from a set of pairs of pupil-glint 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:

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

2. 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 indistinguishable for any individual at the same position, so it can be a general model for different users.

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

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

2. 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 indistinguishable for any individual at the same position, so it can be a general model for different users.

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

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

2. 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 indistinguishable for any individual at the same position, so it can be a general model for different users.

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

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

2. 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 indistinguishable for any individual at the same position, so it can be a general model for different users.

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

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

2. 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 indistinguishable for any individual at the same position, so it can be a general model for different users.
position O, take the mark’s position on the image plane is record as \( I_d(I_o, I_o, I_o) \). Here, the corresponding relations between the X-Y-Z coordinate system in the Fig.3 and the Xc-Yc-Zc coordinate system in the Fig.2 is: \( X = -\Delta L_x / \cos \theta, Y = -\Delta L_y / \cos \theta, Z = -\Delta d / \cos \theta \). The relations between the recorded coordinate \((I_o, I_t)\) and the relative coordinates \((\Delta I_o, \Delta I_t)\) can be expressed as: \( I_o = I_o + \Delta I_o, I_t = I_t + \Delta I_t \). According to the nonlinear regression analysis on the sample pairs of \((X,Y,Z)\) and \((I_o, I_t)\) recorded, we got the mapping function between \((X,Y,Z)\) and \((I_o, I_t)\). And by substituting the two relations analyzed above into the mapping function, we got the equations (3) and (4).

### Dynamic head movement compensation

Combining the analysis as mentioned above with account, the compensation of the head movement in all three directions \((X, Y, Z)\) can be integrated as the following model:

\[
\theta_h' = \frac{k_{dh}}{k_{dh}} (\theta_h + b_{dh} T_h) \quad (7)
\]

\[
\theta_v' = \frac{k_{dv}}{k_{dv}} (\theta_v + b_{dv} T_v) \quad (8)
\]

Where \( \theta_h', \theta_v' \) are the real angles in horizontal and vertical directions respectively. The \( b_{dh}, b_{dv} \) come from Equation (1) and (2); \( T_h \) and \( T_v \) represent the translations of the head position in horizontal and vertical directions respectively. The image magnifying rates in the horizontal and vertical direction at a certain depth of \( \Delta d \) are represented as \( K_{dh}, K_{dv} \) respectively.

In practice, the compensation method becomes effective by following the next steps.

1) Calibration. In the calibration procedure, the user is guided to sit at an appointed position which is the standard position defined in the procedure of position changes perception. Because it is difficult to locate the user’s head at the appointed position precisely without any auxiliary mechanism, we make special marks on the surveillance window to direct the user to adjust the head position correctly. By this means, the distance between the user’s face and the camera can be almost kept at \( d_0 \). Then the user is asked to stare at the mark on the monitor center about 3 seconds.

During the fixation, a set of information are extracted as the initial parameter for the next tracking step. The initial parameter is represented as \( \{ P_{0L}, E_{0L}, V_{0L}, E_{0R}, V_{0R}; H_0 \} \). The \( E_{0L}, E_{0R} \) represent the initial positions of left and right eye on the image; the \( V_{0L}, V_{0R} \) represent the initial pupil-glint vectors of left and right eye respectively; and the \( H_0 \) represents the center of two eyes on the image.

2) Real-time gaze tracking. After calibration, the user’s head can move without restrict head limitation when looking at anywhere of the monitor. In this procedure, the parameter \( \{ P_{rL}, E_{rL}, V_{rL}, E_{rR}, V_{rR}, H_r \} \) is captured for every frame. The parameter \( P_{r} \) contains the same information with the initial \( P_{0L} \). The \( E_{rL}, E_{rR} \) represent coordinates of left and right eye position on the image respectively; the \( V_{rL}, V_{rR} \) represent coordinates of pupil-glint vector of left and right eye respectively; the \( H_r \) represents the center of head position on the image.

Obtain the changes of left eye in the camera coordinate system \( \Delta P \) by substituting the \( \Delta E_{rL} \) to the mapping function between \( \Delta P \) and \( \Delta I \); obtain the changes of right eye in the camera coordinate system \( \Delta P \) by substituting the \( \Delta E_{rR} \) to the mapping function between \( \Delta P \) and \( \Delta I \). Then get the image magnifying rates at the changed position for each eye \( K_{rL} \) and \( K_{rR} \) by substituting \( \Delta P_{rL} \) and \( \Delta P_{rR} \) to equations (5) and (6) respectively. Compute changes of eyeball’s

![Fig. 3 The relative position between the target and the camera in the position changes perception procedure. The elevation angle of the camera is 0. The coordinate system is in units of centimeter.](image-url)
orientations $\Delta \theta_L$ and $\Delta \theta_R$ by substituting $\Delta V_L$ and $\Delta V_R$ to the mapping function $\theta = f(V)$ respectively.

Substitute $\Delta \theta_L, \Delta K_L$ and $\Delta P_L$ to compensation model to get the gaze direction of the left eye $\theta_L$. Then get the gaze direction of the right eye $\theta_R$ in the same way. Here, we define the average of the $\theta_L$ and $\theta_R$ as the real gaze direction $\theta$.

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.

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 Table 1. 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 accuracies. Table 1 listed in Table 1. 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 accuracies.

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

<table>
<thead>
<tr>
<th>(0,0,5)</th>
<th>(0,0,10)</th>
<th>(0,0,0)</th>
<th>(5,0,0)</th>
<th>(0,5,0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>V</td>
<td>H</td>
<td>V</td>
<td>H</td>
</tr>
<tr>
<td>1</td>
<td>0.7</td>
<td>0.8</td>
<td>1.1</td>
<td>1.3</td>
</tr>
<tr>
<td>2</td>
<td>0.6</td>
<td>0.8</td>
<td>1.2</td>
<td>1.2</td>
</tr>
<tr>
<td>3</td>
<td>0.7</td>
<td>0.9</td>
<td>1.2</td>
<td>1.3</td>
</tr>
</tbody>
</table>

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


