Robust Gaze Estimation via Normalized Iris Center-Eye Corner Vector

Haibin Cai[1] *, Hui Yu[1], Xiaolong Zhou[3], and Honghai Liu[1,2]*

[1]School of Computing, University of Portsmouth,UK.
[2]State Key Laboratory of Mechanical System and Vibration,Shanghai Jiao Tong University,China.
[3]Shenzhen Research Institute,City University of Hong Kong,China.

Abstract. Gaze estimation plays an important role in many practical scenarios such as human robot interaction. Although high accurate gaze estimation can be obtained in constrained settings with additional IR sources or depth sensors, single web-cam based gaze estimation still remains challenging. The purpose of this study is to improve the robustness of pupil center-eye corner(PC-EC) vector based gaze estimation method against slightly head movement with a single, low cost web-cam. Firstly, reliable facial features and pupil centers are extracted. Then, a normalised iris center-eye corner(NIC-EC) vector is proposed to improve the robust and accuracy of gaze estimation. Finally, an interpolation method is employed for the mapping between constructed vectors and points of regard(PoG). Experimental results showed that our method can achieve better performance than the PC-EC based gaze estimation method under slight head movements. The average accuracy of gaze estimaion with calibration and testing is around 1.66^o and 2.68^o respectively.

Keywords: Gaze estimation, eye tracking, normalized iris center-eye corner vector, interpolation

1 Introduction

Gaze estimation is a crucial technology in many applications such as human machine interaction and cognitive processes. For example, Gaze can be applied in therapy for children with Autism Spectrum Disorders [1]. Gaze of people is also a key issue for humaniod robots to acquire human-like intelligence [2]. Although accurate gaze estimation can be acquired using additional hardware such as IR sources and depth sensors, the complex setup hinders its application to daily interaction situations. In this paper, we explore the gaze estimation method under a desktop environment using a single, low cost web camera.

In the last three decades, active research has been carried out for eye tracking and gaze estimation[3]. In general, gaze estimation methods can be mainly categorized into two types: namely appearance based methods and features based

---

* corresponding author: Honghai Liu

methods. Appearance based gaze estimation methods take the whole eye image as an input and then construct mapping functions to gaze directions or the points of regard on a screen. Lu et al.[4] propose an adaptive linear regression method to sparsely select the training samples. The eye images are divided into many subregions and the summary of pixels intensities of each subregion is used to generalize a feature vector for mapping to the points on screen. Zhang et al.[5] use multimodal convolutional neural networks to estimation the gaze in the wild environment. A large training data from 15 participants using laptop over three months is collected for training. Funese et al.[6] fit a 3D Morphable Model to the depth data captured by a depth sensor, the eye images are then cropped to frontal face for gaze estimation. The final gaze is determined by combining frontal gaze and the head pose. Sugano et al.[7] propose a multi-camera based system. The 3D shape of eye region is reconstructed and random regression forest is used for gaze estimation.

Unlike appearance based methods, feature based methods estimate the gaze via extracting remarkable facial points such as eye corners, eye lids, pupil centers or corneal glints. The pupil center-corneal center(PCCR) based gaze estimation methods is one of the most popular feature based methods due to its simplicity and reasonable accuracy [8],[9],[10],[11]. However, these methods require an additional IR source or strong visible light sources. Sesma et al. [12] propose to replace the corneal glints with eye corners thus the light sources can be removed. A PC-EC vector is calculated for gaze estimation where the pupil centers and eye corners are manually labeled to avoid image processing errors. Cheung et al. [13] further combine the PC-EC vector with estimated head pose to deal with the head movement.

Although the PC-EC vector can achieve reasonable accuracy when the head keeps still, it lacks the tolerance to deal with slightly head movement as stated in [12], which is an important ability when combining with estimated head pose for head pose free gaze estimation. In this paper we develop a normalized Iris center-eye corner (NIC-EC) vector to gain the robustness against slight head movements. Unlike the PC-EC vector in [12] where both outer and inner corners are used, the developed NIC-EC vector only uses the inner corners and is further normalized via the length of the inner corners and height of nose to overcome the slight head movements. Further more, this paper also shows that an acceptable gaze estimation accuracy can still be acquired by employing automatic facial feature localization and pupil center localization methods to track the features without manually labelling features.

In the proposed method, we firstly extract reliable facial features by using supervised decent method (SDM). Secondly, the accurate iris centers are localized by using the convolution based integer-differential eye localization method. Then the iris centers and inner eye corners are used to construct the iris center-eye corner vector. It is then further normalized according to the length of two inner eye corners and the hight of the nose. Finally, we use an interpolation method to map the NPC-EC to the points on the screen. Experimental results

show that although there exist image processing errors in finding pupil centers and eye corners, the gaze estimation result is still acceptable.

The rest of this paper is organized as follows. Section 2 describes the details of the gaze estimation method. Section 3 presents experiments including hardware setup, evaluation method and results of gaze estimation. The paper is concluded with discussions in Section 4.

2   METHODOLOGY

Many gaze estimation methods have been proposed in the recent years. Among them, the conventional PCCR based gazed estimation has attracted much attention due to its easy implementation and acceptable accuracy [9]. However an additional IR source or visible light source is required in these methods. Sesma et al. [12] evaluated the possiblity of using eye corner to replace the glint caused by light sources and showed acceptable result can be obtained in a still head environment. In this paper, we further improve the PC-EC vector based gaze estimation method by constructing a NIC-EC vector using the located facial features and iris centers. The framework of the gaze estimation system is shown in Fig.1.
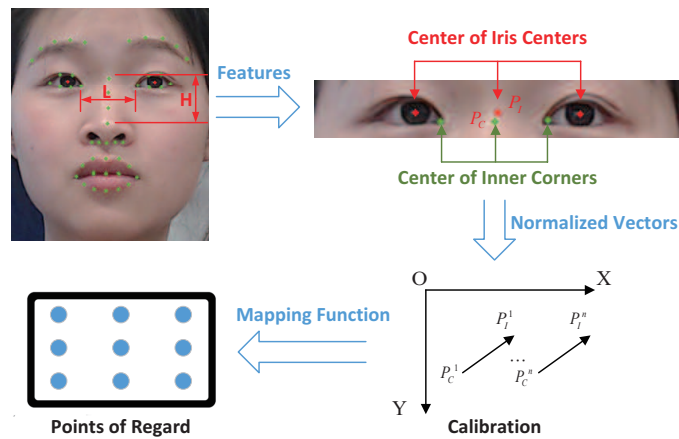

Fig. 1. The framework of gaze estimation system

As shown in Fig.1, the system firstly detects the facial features and eye centers. Then the length of two inner eye corners and the hight of the nose can be determined. The iris center-eye corner is constructed by the center of two inner corners and the iris centers. The final NIC-EC can be obtained after applying normalization for the IC-EC vector. Finally, a calibration procedure is performed to solve the parameters of the mapping function and the final gaze can be determined.

### 2.1	Facial Features and Iris Centers Localization

Many reliable facial features detection methods have been proposed in recent years. In this paper, we employ the SDM[14] to localize the inner eye corners and detect the length of nose. The SDM method can fast and accurately locate 48 facial features. For the iris center localization, we employ the convolution based integer-differential eye localization method [15]. This method mainly uses the large intensity difference of boundary between iris and sclera to fast locate the iris.

In practical, the boosted cascade face detector is firstly employed with default parameters to obtain the approximate location of the face [16]. Then the facial features detection and iris center localization are performed according to the face region.

### 2.2	Normalised Pupil Center-Eye Corner Vector

The conventional PC-CR methods use the pupil center and corneal glints vector to infer the point of regard (PoG) on screens. However it has the disadvantage of complex hardware setup. Recently research shows that the PC-EC vector has the potential to replace the PC-CR vector [13],[12]. However, this vector is sensitive to head movements. This paper proposes a NIC-EC vector to gain the robustness against slight head movements for gaze estimation. The NIC-EC vector is defined as follows:

$$\begin{cases} P_C = P_l^{corner} + P_r^{corner} \\ P_I = P_l^{iris} + P_r^{iris} \\ V_g = P_I - P_C \\ v_x = V_{gx}/L \\ v_y = V_{gy}/H \end{cases} \tag{1}$$

where P_l^{corner} and P_r^{corner} stand for the located left eye inner corner and right eye inner corners, respectively. The center of the inner corners is represent by P_C. P_l^{iris} and P_r^{iris} stand for the located left and right iris centers, respectively. The center of two iris centers is represent by P_I. Then the iris center-eye corner vector V_g can be calculated by using P_I and P_C. To gain the robustness towards slight head movements, we further perform normalization for the vector. The vector V_g can be denoted as (V_{gx}, V_{gy}) in the image coordinate. We use the length of two inner corners to normalize the x direction of V_g. The y direction is normalized by dividing the hight of the nose that can be calculated by the two of the located facial points.

### 2.3	Mapping Function

Many different types of mapping functions for PCCR have been proposed in the literature such as simple 2D linear interpolation[17] and polynomial model [8].

In this paper, we employ the mapping function proposed by [8]. The detail of the mapping function is as follows:

$$\begin{cases} u_x = a_0 + a_1 * v_x + a_2 * v_y + a_3 * v_x * v_y + a_4 * v_x^2 + a_5 * v_y^2 \\ u_y = b_0 + b_1 * v_x + b_2 * v_y + b_3 * v_x * v_y + b_4 * v_x^2 + b_5 * v_y^2 \end{cases} \quad (2)$$

where the coefficients $(a_0, a_1, a_2, a_3, a_4, a_5, b_0, b_1, b_2, b_3, b_4, b_5)$ are determined during a calibration stage. In the calibration stage, users are asked to look at several fixed points on the screen, thus the coefficients can be solved using a least square method.

3   Evaluation

In this section, we firstly introduce the hardware setup for the experiment. After explaining the evaluation method for gaze estimation accuracy, we then compare the results of our NIC-EC method with the PC-EC method proposed in [12].

3.1   Experiment Configuration

The experiment configuration includes a screen and a normal web camera as shown in fig.2. We use a 24 inch screen with a resolution of $1920 * 1080$, the length and hight of the screen are around 52.1cm and 29.3cm, respectively. The adopted Logitech C910 camera has a resolution of $1920*1080$ and a view angle of 78 degrees. During the experiment, the users are asked to click 21 buttons on the screen to collect eye image data for gaze calibration and gaze estimation. When clicking buttons, the users are asked to keep their head roughly still without a chin rest. Furthermore, to test the robustness of the proposed NIC-EC against PC-EC method for slight head movements, the users are asked to have a rest and then go back to the roughly same head position to click buttons. By doing so, the testing data will have a slight different head movement compare to the calibration data.

3.2   Accuracy Evaluation

To evaluate the accuracy of the proposed gaze estimation method, we use both of the pixel error e_{pixel} on the screen and the angular degree error e_{angel}.

$$\begin{cases} e_{pixel} = ||P_g - P||_2 \\ e_{angel} = arctan(Dg/Ds) \end{cases} \quad (3)$$

where e_{pixel} and e_{angel} stand for the accuracy in pixels and degrees respectively. The estimated gaze point and real gaze point are represented as P_g and P respectively. Dg stands for the distance of real gaze points with estimated gaze points. Ds stands for the distance between human face and the screen.
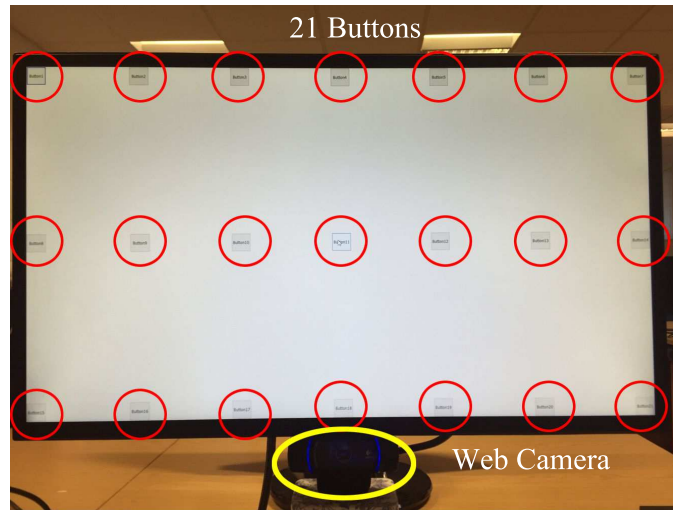
Fig. 2. Hardware setup of our system.

3.3 Experimental Results

In the experiments, images data are captured by manually clicking the 21 buttons on the screen. The users are asked to keep the head still while clicking buttons. No chin rest is used during the clicking thus there might also be small head movement during the procedure. Then the facial features and iris center are localized using the method of [14] and [15] respectively.

Facial Feature Localization Result Fig.3 shows the facial features localization result for all the 21 images when the distance between the subjects and the screen is 71cm. The green points represent for the localized facial points. The red points are the iris centers. As shown in Fig. 3, the first column contains seven eye region images. They are corresponding to the first row buttons as shown in Fig. 2. The second and the third columns are corresponding to the second and the third rows buttons respectively. It should be noted that the low resolution of eye region in Fig.3 is caused by the hardware. Although the selected web-camera has a high resolution of $1920 * 1080$, its wide angle obstructs the way to obtain a high resolution in eye image region. As a result, the length of the located iris radius is 9 pixels.

Comparison With PC-EC Method To show the good performance of the proposed NIC-EC method, we compare it with the PC-EC method proposed in [12]. We perform the comparison in two aspects, one is performed on calibration data and the other is on testing data. Both methods are tested on the same data.

Fig.4 shows the comparison result on calibration data where the red points are the results of NIC-EC method and the blue points are the results of PC-
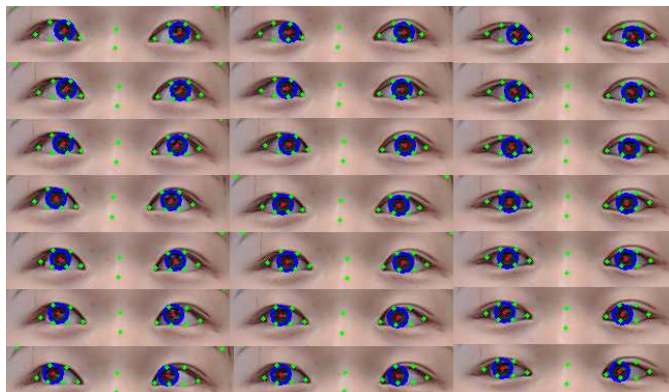
Fig. 3. Samples of facial feature localization results

EC method. The calibration data consists of seven image sets where each set is collected at a different still head position. To compare the accuracy on slight head movements, we perform the calibration on one set of the data and use the rest six data sets for testing. The testing result is shown in Fig.5 where the star points represent the estimated result on calibration data and the dot points represent the estimated result on testing data. The results of NIC-PC and PC-EC are marked with red and blue colors respectively.

The average accuracy is shown in Table 1. The pixel error and degree error are calculated using Eq.3. It can be observed that the calibration accuracy is roughly the same between the two methods. The proposed NIC-EC achieves an accuracy of 66.3 pixels for pixel error and 1.71^o for degree error. The PC-EC method achieves an accuracy with 64.3 pixels for pixel error and 1.66^o for degree error. For the slight head movements environment, the proposed NIC-EC method can achieve around 1° better performance than the PC-EC method. The accuracy of NIC-EC method is 103.7 pixels for pixel error and 2.68^o for degree error. The accuracy of PC-EC methods is 141.4 pixels for pixel error and 3.65^o for degree error.

Table 1. The average processing time of one eye region.

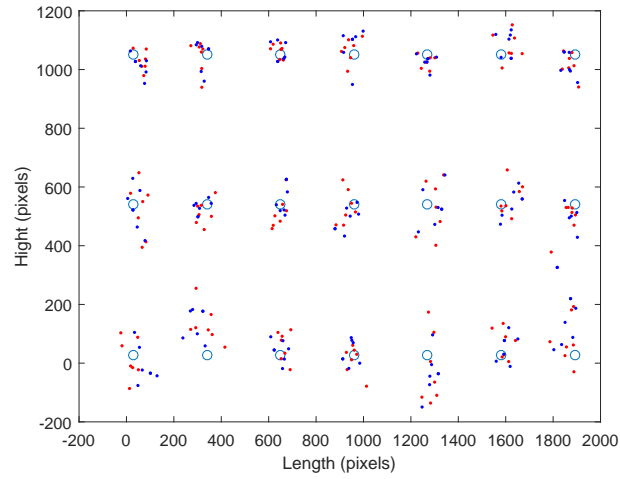| Method	Calibration		Testing	
	Pixel error	Degree error	Pixel error	Degree error
sesma et al.[12](PC-EC)	64.3 pixels	1.66^o	141.4 pixels	3.65^o
proposed method(NIC-EC)	66.3 pixels	1.71^o	103.7 pixels	2.68^o

Fig. 4. Comparison of accuracy on calibration data. The result of the proposed NIC-EC method is marked in red points and result of PC-EC method is marked in blue points.
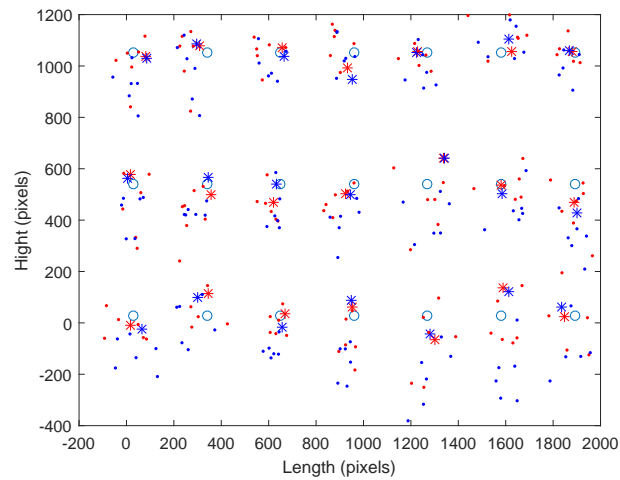

Fig. 5. Comparison of accuracy on testing data. The star points denote the estimated result on calibration data and the dot points denote the estimated result on testing data. The results of NIC-PC and PC-EC are marked with red and blue colors respectively

4   Conclusion

This paper proposes a NIC-PC vector based gaze estimation method. It is an essential subsystem that should be integrated to achieve human level intelligence, especially in the area of humanrobot interaction.

The proposed method has a significant benefit in that it only requires a web camera to perform gaze estimation. The experimental results also show that the proposed method is more robust against slight head movement which is an important requirement when integrating with head pose information to acquire head pose-free gaze estimation.

Future work will be targeted to explore head pose-free and calibration-free gaze estimation method for humanrobot interaction and humanrobot skill transfer [18],[19] by incorporating effective visual tracking methods [20], [21].

References

1. Cai, Haibin, et al. Gaze estimation driven solution for interacting children with ASD[C]. International Symposium on Micro-NanoMechatronics and Human Science (MHS). IEEE, 2015.
2. Brooks R A. Prospects for human level intelligence for humanoid robots[C]//Proceedings of the First International Symposium on Humanoid Robots (HURO-96). 1996: 17-24.
3. Hansen D W, Ji Q.: In the eye of the beholder: A survey of models for eyes and gaze[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2010, 32(3): 478-500.
4. Lu F, Sugano Y, Okabe T, et al. Adaptive linear regression for appearance-based gaze estimation[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2014, 36(10): 2033-2046.
5. Zhang X, Sugano Y, Fritz M, et al. Appearance-based gaze estimation in the wild[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 4511-4520.
6. Mora K, Odobez J M. Geometric generative gaze estimation (G3E) for remote RGB-D cameras[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 1773-1780.
7. Sugano Y, Matsushita Y, Sato Y. Learning-by-synthesis for appearance-based 3d gaze estimation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 1821-1828.
8. Morimoto C H, Mimica M R M. Eye gaze tracking techniques for interactive applications[J]. Computer Vision and Image Understanding, 2005, 98(1): 4-24.
9. Topal C, Gunal S, Kodeviren O, et al. A low-computational approach on gaze estimation with eye touch system[J]. Cybernetics, IEEE Transactions on, 2014, 44(2): 228-239.

10. Sigut J, Sidha S A. Iris center corneal reflection method for gaze tracking using visible light[J]. Biomedical Engineering, IEEE Transactions on, 2011, 58(2): 411-419.
11. Cho D C, Kim W Y. Long-range gaze tracking system for large movements[J]. Biomedical Engineering, IEEE Transactions on, 2013, 60(12): 3432-3440.
12. Sesma L, Villanueva A, Cabeza R. Evaluation of pupil center-eye corner vector for gaze estimation using a web cam[C]//Proceedings of the symposium on eye tracking research and applications. ACM, 2012: 217-220.
13. Cheung Y, Peng Q. Eye gaze tracking with a web camera in a desktop environment[J]. Human-Machine Systems, IEEE Transactions on, 2015, 45(4): 419-430.
14. Xiong X, De la Torre F. Supervised descent method for solving nonlinear least squares problems in computer vision[J]. arXiv preprint arXiv:1405.0601, 2014.
15. Cai H, Liu B, Zhang J, et al. Visual Focus of Attention Estimation Using Eye Center Localization[J].
16. Viola P, Jones M J. Robust real-time face detection[J]. International journal of computer vision, 2004, 57(2): 137-154.
17. Zhu, Jie, and Jie Yang. "Subpixel eye gaze tracking." Automatic face and gesture recognition, 2002. proceedings. fifth ieee international conference on. IEEE, 2002.
18. Liu, Honghai. "Exploring human hand capabilities into embedded multifingered object manipulation." Industrial Informatics, IEEE Transactions on 7.3 (2011): 389-398.
19. Ju, Zhaojie, and Honghai Liu. "Human hand motion analysis with multisensory information." Mechatronics, IEEE/ASME Transactions on 19.2 (2014): 456-466.
20. Zhou X, Yu H, Liu H, et al. Tracking multiple video targets with an improved GM-PHD tracker[J]. Sensors, 2015, 15(12): 30240-30260.
21. Zhou X, Li Y, He B, et al. GM-PHD-based multi-target visual tracking using entropy distribution and game theory[J]. Industrial Informatics, IEEE Transactions on, 2014, 10(2): 1064-1076.