Key Techniques of Eye Gaze Tracking Based on Pupil Corneal Reflection

CHI Jian-nan[1], ZHANG Peng-yi[1], ZHENG Si-yi[1,2], ZHANG Chuang[1] and HUANG Ying[1]
[1] *School of information engineering, University of Science and Technology Beijing, 100083, Beijing, China*
[2] *Beijing Command College of Chinese People's Armed Police Force, 100012, Beijing, China*
Email: sy_jnchi@126.com, sisizheng@163.com

Abstract

Eye Gaze Tracking (EGT) is a kind of techniques that can estimate gaze direction of a person via various methods such as optics, electronics, mechanics etc. Pupil corneal reflection is one of those methods that using the vector from Purkinje image center location to pupil center location to estimate the gaze direction. A review on key techniques of EGT based on pupil corneal reflection is given in three aspects of eye feature detection, mapping model from gaze parameters to gaze direction and face pose estimation. The characteristics of two kind mapping models are analyzed and the detection method of gaze direction based on stereovision is discussed. According to a large number of recent references, we propose some major problems and their possible solutions in EGT techniques as well as its research orientation in future.

1. Introduction

Eye gaze tracking (EGT) is a kind of technique that can estimate gaze direction of a person via various methods such as optics, electronics, mechanics etc. EGT techniques can be classified as intrusive and non-intrusive in light of detection method and system structure. Intrusive techniques require some equipment to be put in physical contact with users such as head mounted devices fixed with optical system so that it is not very convenient. Non-intrusive techniques are mostly vision based, i.e., they use cameras to capture images of face and eyes, using image analysis and processing to get eye feature points, which are then conversed to three-dimensional data based on eye-imaging model or mapping model, and gaze falling position or direction can be estimated. Compared with intrusive techniques, non-intrusive techniques are less interference, easier operation and higher precision, and it has been widely adopted by researchers and has large development space in IT field. Because of its advantages of convenience, directness, naturalness and interactivity, EGT has broad application prospects in many areas such as Human Computer Interaction (HCI), handicap assistance, visual reality, driver assistance system, Human Factors Analysis (HFA) etc.

Purkinje images are reflections from the external surface of the cornea created by infrared (IR) source. Pupil corneal reflection is one of EGT methods that using the vector from Purkinje image center location to pupil center location to estimate the gaze direction.. The key techniques it used includes segmentation and center location of cornea and Purkinje images, pupil tracking, non-linear mapping modeling from plane parameters to gaze directions, 3D facial feature detection, multiple facial feature tracking, pose estimation etc.

This paper is organized as follows. In Section 2, a review on key techniques of EGT based on pupil corneal reflection is given in three aspects of eye feature detection, mapping model from gaze parameters to gaze direction and face pose estimation. We propose some hot and difficult issues in study of eye gaze tracking in Section 3. The paper ends in Section 4 with a summary and a generalization of our work.

2. Key techniques of EGT

Some eye's structure and characteristics, which are relatively fixed when the eyeball is moving, are utilized as the reference frame in EGT techniques, and then the gaze variation parameters, the primary evidence to identify the gaze direction, are extracted from the varying positions and those fixed ones. EGT techniques are composed of two parts: information acquisition and gaze direction identification (see Fig. 1).

978-0-7695-3571-5/09 $25.00 © 2009 IEEE
DOI 10.1109/GCIS.2009.338

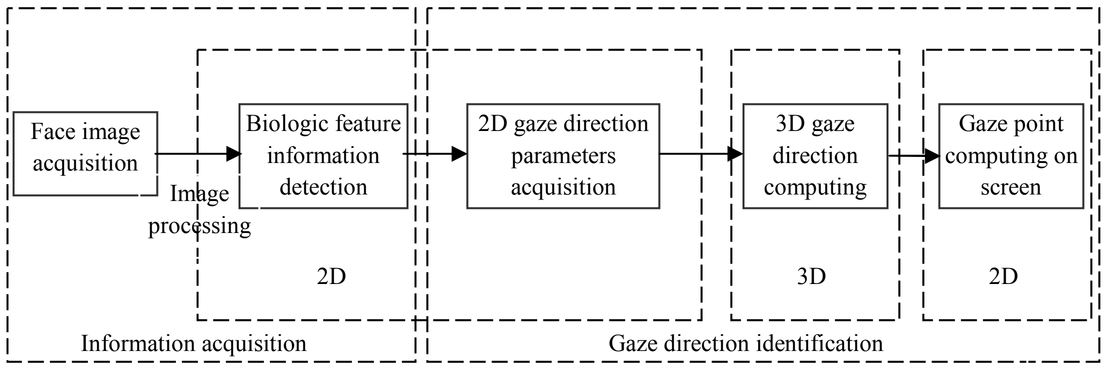


Figure 1. The schematic diagram of EGT techniques

2.1. Eye feature detection

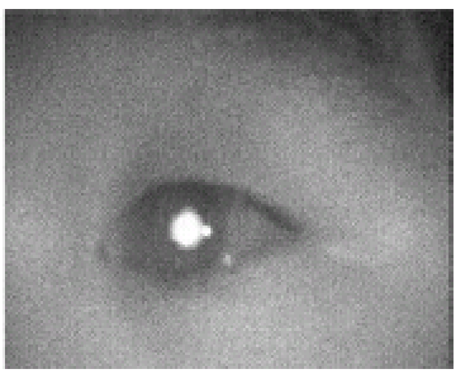


Figure 2. The bright pupil image

Eye feature detection can be divided into two types: appearance-based detection [1-4] and active IR illumination [5-8]. To the latter, the IR source is laid near the optical axis of the camera. Compared with a regular dark pupil, a so-called "bright pupil" (see Fig. 2) now is able to ''see'' in the camera, similar to the red eye effect in night photography using a bright flash light. We can catch the eyes quickly in a whole face image and segment the pupil accurately in eye image in terms of the bright pupil that makes the pupil highlight. Image subtraction presently is mostly used in pupil detection [9-13], that is to subtract bright pupil image from dark pupil image, which can leave the pupil completely. Then the pupil location can be determined by threshold segmentation. Lay IR source near the camera optical axis and off the axis respectively, a bright pupil can be created with the near one while the Purkinje images are produced with the other one in the surface of cornea. That is called active IR illumination, always used in EGT based on pupil corneal reflection. The Purkinje images are the reflections created at different layers of the eye structure. The first Purkinje image, corresponding to the reflection from the external surface of cornea, is the brightest and easiest reflection to detect, and so it is also considered as corneal reflection. The second Purkinje image corresponds to the reflection from the internal surface of cornea, and the third and fourth Purkinje images are from the surface of lens, external and internal respectively. The corresponding parameters can be extracted by the fires Purkinje image that regarded as a fixed location when gaze direction moving. The vector from the center of Purkinje image to the pupil is just the direction parameter in 2D gaze estimation. The relative locations of pupil and Purkinje image in different gaze directions are shown in Fig. 3.

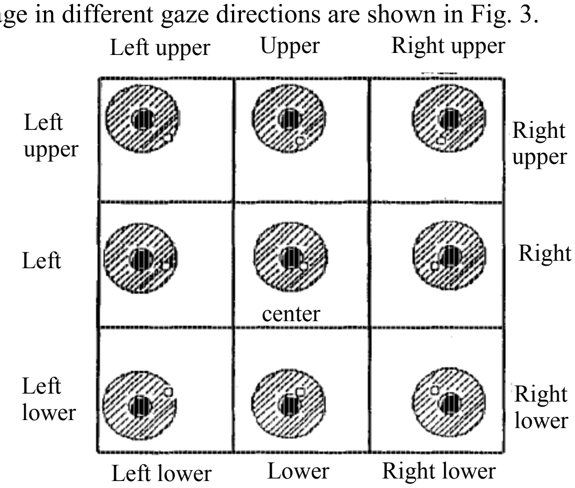


Figure 3. The relative locations of pupil and Purkinje image in different gaze direction

To get better detection, the configuration of light source and system structure are modified many times for testing by researchers. Morimoto et al. [9] used a light source which is made up of two loops IR LED as shown in Fig. 4. The inner loop that is near the optical axis of camera is used to generate bright pupil, and the outer loop that is far from the optical axis of camera is used to generate corneal reflection (or glint). The advantage of looped light source is to generate the Purkinje image which is relative symmetrical in each direction, and the pupil centre location can be computed accurately after segmentation.

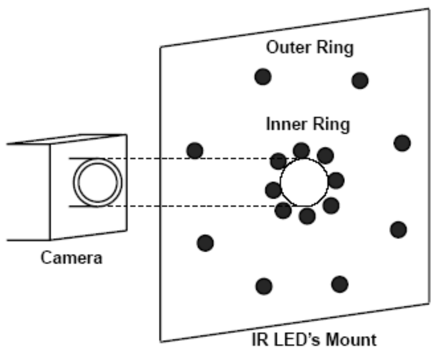
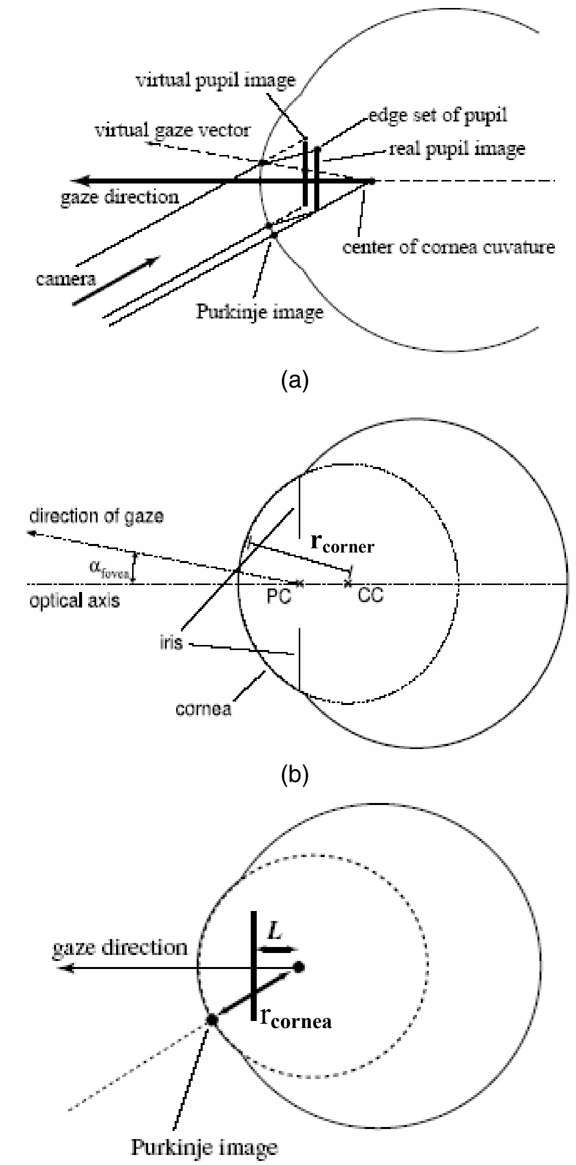


Figure 4. IR source

Ebisawa et al. [1] describe a detection system based on image subtraction of pupil and Purkinje image, which is composed of a single camera with 3 CCDs and 2 near IR sources. The IR sources have different wavelengths (λ_1 and λ_2), IR source with λ_1 placed off the optical axis is to generate the Purkinje image, IR source with λ_2 placed near the axis is to generate the bright pupil. CCD3 is sensitive to λ_2 only, thus it outputs bright pupil image. CCD1 and CCD2 are sensitive to λ_1 (λ_2 is filtered out), and CCD1 also has a polarizing filter in order to receive only the diffuse light components, i.e., the corneal reflection due to λ_1 does not appear in the images from CCD1. Once the three images are available, the pupil is segmented from CCD3 and CCD2 via image subtraction and threshold processing, and the corneal reflection used for gaze estimation is segmented by the images from CCD2 and CCD1.

2.2 Computation of gaze direction in space

In space computation of gaze direction, two kinds of mathematical models are mostly used. One is 3D geometric model with physiological structure of eyeball, the other is mapping model of plane coordinate from 2D parameters to gaze falling location with emphasis on the effect.



(a)

(b)

(c)

Figure 5. The eyeball model

The eyeball model [14-16] used in 3D geometric model is shown in Fig. 5. Suppose the surface of cornea is spherical, CC is the center of curvature, r_{cornea} is the radius of curvature, PC is the center of pupil, and gaze direction is just the vector from CC to PC. In fact, however, PC is not very precise since the refraction index difference between the air and the corneal surface, as shown in Fig. 5(a), which causes the pupil distort and change in location. The angle named as α_{fovea} is formed at the nodal point by the intersection of the visual axis, which links the fovea to the pupil centre, and the optic axis determined by the eyeball model. The difference between the observed pupil center and the real one is shown in Fig. 5(a), as well as the error of gaze estimation induced by this difference. The pupil center PC, the corneal curvature center CC, and the angle α_{fovea} are shown in Fig. 5(b). The two individual different parameters in this model, the radius of curvature r_{cornea} and the distance L from the corneal curvature center to the pupil center are shown in Fig. 5(c). In practical application, this 3D geometric modeling to calculate gaze direction or falling points based on the eyeball model has the following three parts: (i) estimate the center location of corneal curvature and the pupil to determine gaze direction according to the Purkinje image; (ii) deduce 3D geometric model based on the relationship of coordinate conversion among the eyes, the camera and the screen; (iii) the initial calibration should be done to the user due to the two individual different parameters L and C in this eyeball model. How to set up the 3D geometric model has been discussed in detail in [14]. According to the eyeball model, the unit vector of gaze direction is as following:

$$\overline{E} = \frac{P - C}{\| P - C \|} \quad (1)$$

Where, p is the 3D coordinate of pupil center, C is the 3D coordinate of the corneal curvature center.

There are two ways to get the nonlinear mapping model from parameters to direction in 2D gaze. One is to consider the mapping relationship as a second or higher order polynomial, and then determine the polynomial coefficients via calibration procedure. The other is to fit the mapping function used neural network (NN). For example, a second order polynomial transformation [17], defined as:

$$x_s = a_0 + a_1 x_e + a_2 y_e + a_3 x_e y_e + a_4 x_e^2 + a_5 y_e^2 \quad (2)$$

$$y_s = a_6 + a_7 x_e + a_8 y_e + a_9 x_e y_e + a_{10} x_e^2 + a_{11} y_e^2 \quad (3)$$

Where, (x_s, y_s) is the coordinate of falling point on the screen; (x_e, y_e) is the pupil-glint vector; $a_0 \sim a_{11}$ are the coefficients of this second order polynomial. A set of corresponding points can be obtained by the calibration procedure, where the user is asked to fixate her/his gaze at some certain known targets on the screen. For 6 different known targets, 12 equations are produced and a determined linear system is obtained. Beside, it is not the case that the far higher order polynomial is more useful for determining the mapping function [13], since it needs more points to be calibrated, which is complicated to the user.

The establishment of mapping from 2D parameters to gaze direction is to fit the nonlinear model more accurately as well as to reduce the burden of users in initial calibration, thus users' movement can be more freely in certain extent. Many researchers tried to set up the nonlinear model from 2D to 3D gaze direction used NN [18], collecting much data of different users and one user in different poses to train network, so that the user doesn't need to do the initial calibration and his/her head would be more freely in certain degree of accuracy.

2.3 Face pose estimation

Methods for face pose estimation are generally classified into two types: appearance-based [19,20] and model-based [21,22]. Appearance-based approaches use lots of samples to set up a given corresponding relationship between 3D face pose and some features of face images (such as gray scale, color, image gradient etc.) via statistical method. Model-based approaches can realize the estimation of 3D face pose by establishing the corresponding relationship used some features between the image and the geometric model, which represents face framework and shape. In recent years, with the extensive use of the device for obtaining depth data, the 3D model of human face has good prospects in application. Lopez et al. [23] used 3D model and 28 dotted pairs of feature points in images to realize face pose estimation. A large template library, including various poses, should be set up previously in this method with high computing cost. Yang et al. [24] introduced 3D model into a real-time stereovision system, which consisted of two cameras via semi-interactive method previously, in order to realize face pose estimation with comparatively higher precision. Wang Yingjie et al. [25] defined a general face model via statistic method and realized face pose estimation by the optimal iterative process and the corresponding relationship between 3D model and the given curves in images. Using this method, some particular curves need to be extracted in the face images and the matching between the curves of 3D model and the curves of face images should be achieved.

2.4 EGT detection based on stereovision

The final purpose of EGT is to find 3D gaze direction or landing point location on the screen and extract gaze direction accurately in the case of head movement or face movement. Therefore, many researchers are studying how to detect 3D feature of eyeball in real-time and estimate the spatial movement of head and face. Among several solutions, stereovision technique is a good one. A stereovision detection system for head post and gaze direction is introduced in [26]. In this system, head post is determined by matching the stereo features such as corners of mouth and the canthus, and gaze direction is determined finally by 3D eyeball model to map the position of head post and iris to gaze landing point. The system has a good result in natural light and has been used in driver assistance system with the precision of 3 degree. In [27], 3D parameters of eyeball can be obtained by stereovision system, then the optic axis vector, just a line connecting the corneal curvature center and the pupil center, can be determined. Because the mapping function is linear, not only the gaze point on the screen but also the 3D point in gaze space can be found by the calibration of camera. Using two stereovision systems, one is to detect 3D face post, once the 3D face position is determined, which can help the other with narrow eyeshot to get the 3D eye image in high resolution, then 3D feature parameters can be detected by the eye image. Moreover, the intrinsic parameters of eyeball such as curvature center and radius of cornea are estimated based on calibration program with the precision of 0.6 degree in gaze direction.

3. Hot and difficult issues in EGT research

To be applied in general computer interfaces, an ideal eye tracker should [16]:

(i) be accurate, i.e., precise to minutes of arc; (ii) be reliable, i.e., have constant, repetitive behavior; (iii) be robust, i.e., should work under different conditions, such as indoors/outdoors, for people with glasses and contact lenses, etc; (iv) be non-intrusive, i.e., cause no harm or discomfort; (v) allow for free head motion; (vi) not require calibration, i.e., instant setup; (vii) have real-time response.

Of course, no product satisfies these usability requirements now, some products of EGT just satisfy some special application condition. However, how to produce the eye tracker that could meet those above requirements and the solutions of relative problems and difficulties are the objectives to many worldwide researchers. There are some difficulties as follows:

(i) How to match the different systems to variable illumination fitly should be discussed for the purpose of enhancing the accuracy and real-time performance of system as well as reducing the interference in users.

(ii) How to realize the estimation of gaze direction accurately and the calibration program easily in the condition of head motion.

(iii) 3D geometric model is comparatively impersonal, accurate and good self-adaptive, but its establishment is very complex depend on the eyeball configuration. The mapping model is more affected by anthropic factor based on experimental data in large quantities though its establishment is a little simple. Therefore, the difficulties in research are the accuracy and good adaptability in mapping from gaze parameters to gaze direction.

(iv) There are some difficulties in the practical application and correlative assignment in HCI such as how to deal with the saccadic eye movement and dithering in unconsciousness.

If we can solve above problems, the gaze tracker can be achieved under natural head movement, which may make breakthrough to HCI.

4. Conclusion

EGT is a kind of techniques that can estimate gaze direction of a person via various methods such as optics, electronics, mechanics etc. Pupil corneal reflection is one of those methods that using the vector from Purkinje image center location to pupil center location to estimate the gaze direction. A review on key techniques of EGT based on pupil corneal reflection is given in three aspects of eye feature detection, mapping model from gaze parameters to gaze direction and face pose estimation. The characteristics of two kind mapping models are analyzed and the detection method of gaze direction based on stereovision is discussed. According to a large number of recent references, we propose some major problems and their possible solutions in EGT techniques as well as its research orientation in future.

Pupil corneal reflection used by many researchers offers some advantages over the other available alternatives. The difficulty to detect eye features in gaze tracking system is reduced and the precision to extract gaze direction is improved by the development of pupil corneal reflection. We have a good result of pupil corneal reflection in one camera system that the precision of gaze direction can reach one degree allowing for slight head motion, which meets the basic requirement of HCI. However, since 3D information of head and eye features can not be obtained by one camera system, the accurate gaze is unable to be gained allowing for free head motion. To solve this problem, the next step of our team is to design a stereovision system consisted of two cameras based on pupil corneal reflection technique.

Acknowledgements

This work is sponsored by National Natural Science Foundation (60775042) and National High Technology Research and Development Program of China (2007AA01Z160).

References

[1] Y. Ebisawa, S. Satoh, Effectiveness of pupil area detection technique using two light sources and image difference method. *Proc. of 15th Annual Internat Conf. of the IEEE Eng. in Medicine and Biology Society,* San Diego, 1993, pp.1268-1269.

[2] R.Kothari, J.Mitchell, Detection of eye locations in unconstrained visual image. *Proc. of International Conference on Image Processing,* Switzerland, 1996, pp.519-522.

[3] Yuan-Pin Lin, Yi-Ping Chao, Webcam mouse using face and eye tracking in various illumination environments. *Proc. of the 2005 IEEE Eng. in Medicine and Biology 27th Annual Conference,* Shanghai, China, Sep.2005, pp.3738-3741.

[4] Nuria Oliver, Alex Pentland, Francois Berard, Lips and face real time tracker. *Proc. of the 1997 Conference on Computer Vision and Pattern Recogniton,* 1997, pp.123-129.

[5] C.H. Morimoto, M. Flickner, Real-time multiple face detection using active illumination. *Proc. of the 4th IEEE International Conference on Automatic Face and Gesture Recognition,* Grenoble, France, 2000, pp.234-245.

[6] A Haro, M. Flickner, I. Essa, Detecting and tracking eyes by using their physiological properties, dynamics and appearance. *Proc. of IEEE CVPR 2000,* South Carolina, 2000, pp.125-129.

[7] Paul Smith, Mubarak Shah., X. da Vitoria Lobo, Monitoring head/eye motion for driver alertness with one camera[C], Proceedings of the 2000 International Conference on Pattern Recognition, 2000, pp.1-3.

[8] Q.Ji, X.Yang, Real time visual cues extraction for monitoring driver vigilance. *Proc. of International Workshop on Computer Vision Systems*, Canada, 2001, pp. 233-245.

[9] C.H.Morimoto, Free head motion eye gaze tracking using a single camera and multiple light sources. *The 19th Brazilian symposium on computer graphics and image processing,* 2006, pp.171-178.

[10] Alexander Toet, Gaze directed displays as an enabling technology for attention aware systems. *Computers in Human Behavior,* 2006, 22, pp. 615-647.

[11] B.Noureddin, P.D.Lawrence, A non-contact device for tracking gaze in a human computer interface. *Computer Vision and Image Understanding*, 2005, 98, pp,52-82.

[12] Duchowski A. T., *Eye Tracking Methodology: Theory and Practice.* Berlin: Springer-Verlag, London, 2003.

[13] Morimoto C H., Mimica, Eye gaze tracking techniques for interactive applications. *Computer Vision and Image Understanding*, 2005, 98, pp.4–24.

[14] B.Noureddin, P.D.Lawrence, A non-contact device for tracking gaze in a human computer interface. Computer Vision and Image Understanding, 2005, 98, pp.52-82.

[15] Arantxa Villanueva, Rafael Cabeza, Sonia Porta. Eye tracking: Pupil orientation geometrical modeling. *Image and Vision Computer*, 2006, 24, pp.663-679.

[16] D.H.Yoo, M.J.Chung, A novel non-intrusive eye gaze estimation using cross-ratio under large head motion. *Computer Vision and Image Understanding,* 2005, 98, pp.25-51.

[17] Mimica M.R.M, Morimoto C.H., A computer vision framework for eye gaze tracking. Proc. of 16th Brazilian Symposium on Computer Graphics and Image Processing, 2003.

[18] Coughlin M.J, Cumore T.R.H, Automated eye tracking system calibration using artificial neural networks. *Computer Methods and Programs in Biomedicine*, 2004, 76, pp.207-220.

[19] S. Mekenna, S. Gong, Real-time face pose estimation. *Real-Time Imaging, Special Issue on Visual Monitoring and Inspection*, 1998, 4, pp.333-347.

[20] Mukesh M, Qiang Ji, 3D face pose discrimination using wavelets. *Proc. of IEEE International Conference on Image Processing,* Thessaloniki, Greece, 2001.

[21] Glyn E, Pingping Y, Using affine correspondence to estimate 3D facial pose. *Proc. of the IEEE International Conference on Image Processing,* 2001, pp.919-922.

[22] Qiang Ji, 3D face pose estimation and tracking from a monocular camera. *Image and vision computing*, 2002, pp.235-247.

[23] Lopez R., Huang T.S. 3D head pose computation from 2D images: Template versus features. Proc. of the IEEE International Conference on image Processing, Washington, 1995, 2, pp.599-602.

[24] Yang R., Zhang Z. Model-Based Head Pose Tracking With Stereovision. *Proc. of the IEEE International Conference on Automatic Face and Gesture Recognition,* Washington, 2002, pp.255-260.

[25] Yingjie W, Chin-Seng C, Facial feature detection and face recognition from 2D and 3D image. *Pattern Recognition Letters*, 2002, 23, pp.1191-1202.

[26] Matsumoto, Y., Zelinsky A, An algorithm for real time stereo vision implementation of head pose and gaze direction measurement. *Proc. of IEEE fourth international conference on face and gesture recognition,* Piscataway, 2000, pp. 499-505.

[27] Shih S.W., Liu J, A novel approach to 3D gaze tracking using stereo cameras, *IEEE Transactions on Systems, Man and Cybernetics*, 2004,34(1), pp.234-245.