Number 5, Volume VI, December 2011

GAZE TRACKING SYSTEMS FOR HUMAN-COMPUTER INTERFACE

Matěj Černý[1]

Summary: This article deals with currently used gaze tracking systems, their classification and application possibilities. Further are presented selected gaze tracking systems suitable for human-computer interface. Finally are proposed improvements of these systems based on analysis of their weak and strength points.

Key words: Gaze tracking, Human-computer interface, image processing.

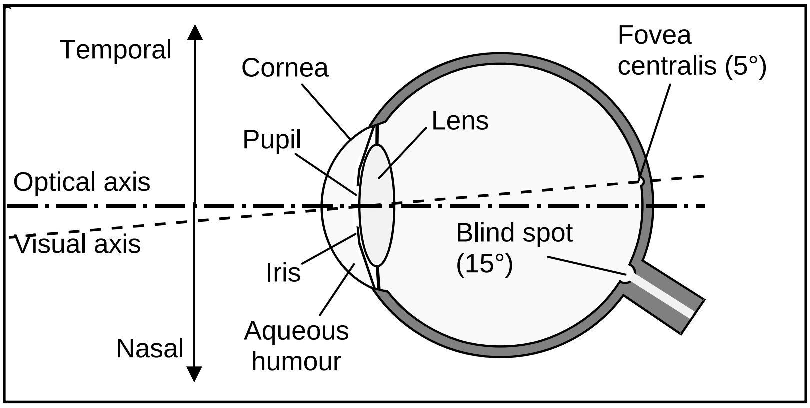
INTRODUCTION

In late years as computing power of personal computers and embedded devices is raising, image processing applications is experiencing rapid development. Also video sensors quality and resolution is improving. These aspects in combination with lowering prices of these devices is challenging opportunity for gaze tracking methods to be practically usable as human machine interface (HMI), specially human computer interface (HCI). Gaze tracking methods can be applied in many fields. In civil field, these methods can be used, for example in medical diagnostics, to help disabled people to interact with world, or to improve comfort of controlling computer, also they can be used for preventing drivers from falling asleep. In combination with head up display (HUD), these methods can be used in military applications, for example weapons aiming for vehicles and soldiers.

Gaze tracking systems can be intrusive (10, 4, 5, 9) or non-intrusive (13, 14, 10, 2). Intrusive systems is limiting user with head fixation or head mounted equipment. In opposite are non-intrusive systems, which are providing much more comfort and mobility. Intrusive HCI systems are more or less historical, for modern and practically usable HCI system is necessary to be non-intrusive. For understanding methods and principles used by these systems is necessary basic knowledge of human eye anatomy and physiology (1).

Arrangement of human eye is displayed on Fig. 1. Most important parts of human eye used for gaze tracking systems are Cornea and Pupil. Also important characteristic is difference between optical and visual axis of eye, which is approximately 5°.

[1] Ing. Matěj Černý, University of Pardubice, Faculty of Electrical Engineering and Informatics, Department of Electrical Engineering, nám. Čs. legií 565, 53210 Pardubice, Tel.: +420 466 037 109,
E-mail: matej.cerny@student.upce.cz



Source: author

Fig. 1 - Human eye arrangement

1. CURRENTLY USED GAZE TRACKING METHODS

1.1 Overview

Currently published systems and algorithms differs in many things, mainly in methods of determination gaze direction which are related to accuracy and complexity of design. Main approaches can be categorized, depending on features they use (6):

- the glint (13, 14),
- a 3D model (8),
- a local linear map network (7).

The first and the most commonly used approach is to calculate direction of gaze by tracking the relative position of the pupil and light reflection on the cornea. These systems can use visible or infra-red light sources. Second approach is using 3D model of face, based on mouth, eyes and nostrils position to evaluate face position and gaze direction. Third approach is marginal and uses neural network, which is learned from examples for each user. According to wide spread of methods using glint features, further in this paper will be discussed these methods.

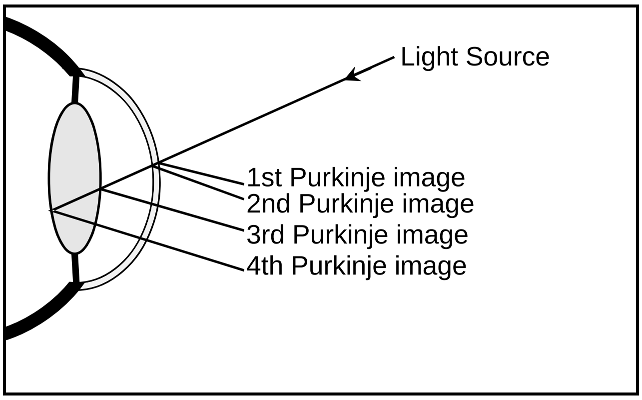
1.2 System using stereoscopic cameras

In (14) is presented system developed at Northwestern Polytechnical University Xi'an in China. Basis of this system consist of pair of stereo cameras and pair of point light sources which are used to estimate 3D gaze of eye. On Fig. 2 is picture of this system. In this context 3D gaze is straight line that passes through fovea centralis, optical center of eye and observing point. This technique is not dependent on head movement.

Source: (14)

Fig. 2 - System with stereo cameras

Determination of 3D gaze is done by connecting 3D cornea center and virtual pupil center. The 3D center of cornea is estimated from first Purkinje image. Purkinje images (Fig. 3) are four reflections of image from optical apparatus of eye.



Source: author

Fig. 3 - Purkinje images

First Purkinje image is reflection from outer side of cornea, is also called glint and is strongest. Second Purkinje image is reflection from inner side of cornea. Third Purkinje image is reflection from outer side of lens and fourth Purkinje image is reflection from inner side of lens. Fourth Purkinje image is inverted.

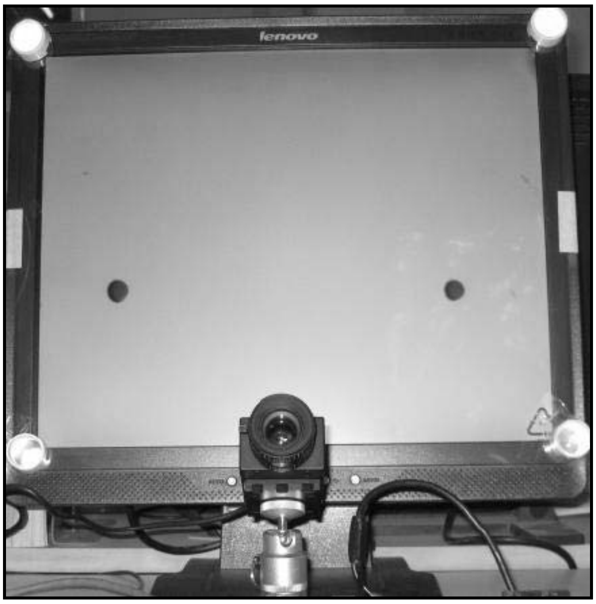
Pupil can be considered to be inside convex lens which consist of cornea and aqueous humor. Due to light refraction in cornea and aqueous humour is image acquired by camera image of virtual pupil, not pupil itself. Considering cornea as spherical lens, image of virtual pupil is still on optical axis of eye. This system requires calibration of stereo cameras, position of LED[2] sources and transformation matrix between cameras and screen coordinates. According to presented results this system has accuracy about three degrees.

[2] Light Emitting Diode

Advantage of this system is simple, user independent calibration. Main disadvantage is limited free head movement caused by narrow view angle of cameras.

1.3 System using single camera

Different approach is presented in (13) which describes system developed at Shandong University in China. This system consists of one standard resolution camera (1.4 Mpx) and four near infrared light sources located in corners of monitor. Picture of this system is on Fig. 4. This system uses reflects of light sources on cornea and pupil center to determine direction of gaze.



Source: (13)

Fig. 4 - System with 4 IR souces

Basic parts of algorithm are:

1. Eye detection.
2. Pupil detection
3. Corneal reflections detection.
4. Pupil center detection
5. Gaze track estimation

For eye location is used difference of two consecutive frames during natural blink of eye. Differential image is binarized with iterative threshold, noise is reduced by morphological erosion. Binarized image has two white blobs and their bounding rectangles are considered as eyes location.

Resulting image is binarized and smoothed to reduce noise. *Pupil* position is determined by the vertical and horizontal integral projection.

Area of *pupil* contains *pupil*, part of *iris* and glints of light sources. Because glints are brightest pixels in image glints are extracted using gray difference and known position relationship between glints.

To avoid interference of glints during pupil center location area of glints is filled with color of pupil. Edge of *pupil* is obtained by *Canny* operator. On edge is applied ellipse fitting algorithm, center of ellipse is center of *pupil*.

Gaze tracking is based on cross-ratio-invariant in projective transformation between two planes. Considering one plane of monitor and second plane of camera image, cross-ratios are equal.

Calibration is compensating difference between optical and visual axis of eye. During calibration process user is observing points with known coordinates and these coordinates are compared with coordinates computed with gaze tracking algorithm. Presented accuracy of system is about 0.3 degrees.

Main advantages of system are its accuracy and low hardware requirements. Disadvantage is user-dependent calibration, also eye detection method seems to be weak.

In (12) is presented improvement of this system. Preprocessing using gray difference is used to improve eye detection mainly for users wearing eyeglasses.

2. PROPOSED OWN SOLUTION

These systems are very well representing most precise solutions presented in late years, but in these systems is still a lot of space for further improvements.

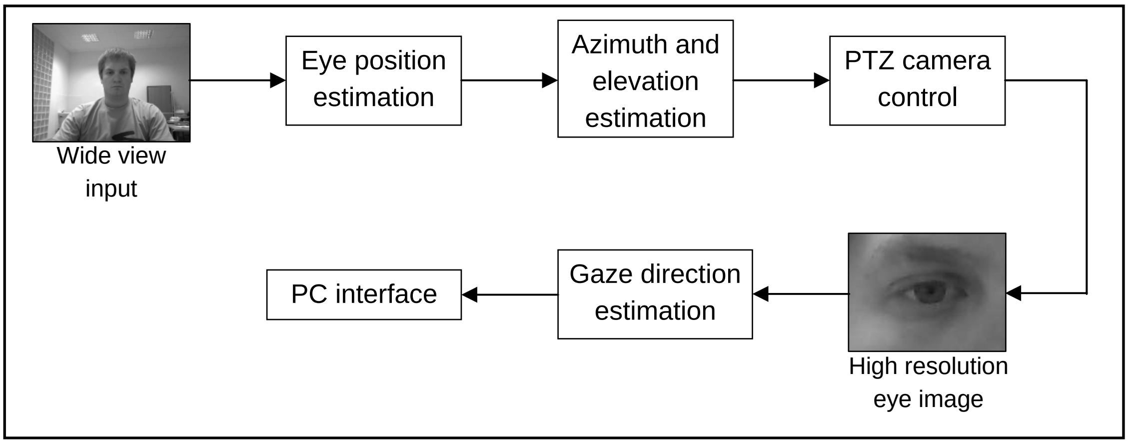
Improvements can be done in two ways:

- Accuracy and reliability.
- Computational complexity.

Under accuracy and reliability we understand improvements in precision and robustness of system. Computational complexity is important for speed (latency) and hardware requirements. These two ways are on opposite sides of scales and depends on specific application which way will be preferred.

Significant disadvantage of these systems is narrow view angle of camera(s) which is limiting free head movements of user. This limit is caused by minimal pixel density of resulting image needed for successful gaze direction estimation. Increasing view angle can be done by increasing of camera resolution, but this leads to longer computational time. This disadvantage can be eliminated by decomposing video acquiring into two partial steps. In first step will be used wide view angle camera to evaluate ROI[3] position. In second step will be used PTZ[4] camera with narrow view angle which will follow ROI using data evaluated from first step. PTZ camera will be providing high resolution images of ROI. ROI can be face, or eye depending on application. ROI will be processed with commonly used methods but according to higher pixel density, this system should achieve higher accuracy with standard resolution camera. This solution solves problem with view angle of camera, and also should improve accuracy of the system. Schematic of proposed system is on Fig 5.

[3] Region Of Interest
[4] Pan Tilt Zoom



Source: author

Fig. 5 - System overview

Eye position estimation will be realized by Viola–Jones object detection framework (11) and Haar cascades. Azimuth and elevation estimation will be based on known camera-display position and head size. Head size will be used to estimate distance of user from display, which is necessary for correct azimuth and elevation estimation. In PTZ camera control, computed azimuth and elevation will be translated into control signals for camera drive. Output from PTZ camera will provide high resolution image of eye area. Gaze direction estimation will be using four IR light sources and their corneal reflexes similar to earlier presented methods (13). Azimuth and elevation can be used for additional improvement of accuracy. Output of this section is display coordinates. PC interface will be connecting system to PC via USB[5] port.

CONCLUSION

Proposed system has more complicated hardware design, compare to presented methods, but should achieve better performances. In proposed system will be necessary to solve dependency between accuracy and resolution of cameras to find compromise between accuracy and computational complexity. Also latency issue will be crucial for practical application of this system. To achieve expected accuracy will be necessary to filtrate spontaneous jerky eye movement (saccades). Separate chapter will be interconnection with modern GUI, but this is not object of this work. Further improvements can be done by using multilayer neural networks for eye or face position estimation.

ACKNOWLEDGEMENT

The results presented in this article were supported by the specific research project of the IGA (SGFEI05/2011), University of Pardubice.

[5] Universal Serial Bus

REFERENCES

(1) DESPOPOULOS, A.; SILBERNAGL, S. Color atlas of physiology. 2nd. New York: Thieme New York, 2010. 433 p. ISBN: 978-3-13-116552-7.

(2) EBISAWA, Y. *Robust pupil detection by image difference with positional compensation*. IEEE International Conference on Virtual Environments, Human-Computer Interfaces and Measurements Systems. Hong Kong, China, 2009-05-11, p. 143 - 148. ISBN: 978-1-4244-3808-2.

(3) GLENSTRUP, A. J.; ENGELL-NIELSEN, T. *Eye Controlled Media: Present and Future State*[online]. University of Copenhagen, 1995-06-01 [cited 10 Sepember 2011]. Available from World Wide Web: <http://www.diku.dk/hjemmesider/ansatte/panic/eyegaze/article.html>.

(4) KUMAR, N. KOHLBECHER, S.; SCHNEIDER, E. *A Novel Approach To Video-Based Pupil Tracking*. IEEE International Conference on Systems, Man and Cybernetics. 2009-10-11, p. 1255 - 1262. ISBN: 978-1-4244-2793-2.

(5) LI, X.; WEE, W. G. *An efficient method for eye tracking and eye-gazed FOV estimation*. 16th IEEE International Conference on Image Processing. 2009-11-07, p. 2597 - 2600. ISBN: 978-1-4244-5653-6.

(6) Mohamad, A., O., Da Silva, M., P., Courboulay, V. *History of gaze tracking.* in HAL : hal-00215967, version 1 . 2007.

(7) RAE, R.; RITTER, H. J. *Recognition of human head orientation based on artificial neural networks*. IEEE Transactions on Neural Networks. 1998-03, p. 257 - 265. ISSN: 1045-9227.

(8) STIEFELHAGEN, R. YANG, J.; WAIBEL, A. *A model-based gaze tracking system*. IEEE International Joint Symposia on Intelligence and Systems. 1996-11-04, p. 304 - 310. ISBN: 0-8186-7728-7.

(9) TOPAL, C. DOGAN, A.; GEREK, O.N. *An Eye-Glasses-Like Wearable Eye Gaze Tracking System*. 16th IEEE Signal Processing, Communication and Applications Conference. 2008-04-20, p. 1 - 4. ISBN: 978-1-4244-1998-2.

(10) TUNHUA, W. et al. *Real-time Non-Intrusive Eye Tracking for Human-Computer Interaction*. 5th International Conference on Computer Science and Education. Hefei China, 2010-08-24, p. 1092 - 1096. ISBN: 978-1-4244-6002-1.

(11) VIOLA, Paul. JONES, Michael. Robust Real-time Object Detection . In *Second international workshop on statistical and computational theories of vision – modeling, learning, computing, and sampling*. Vancouver: 2001.

(12) YANG, C. et al. *A gray difference-based pre-processing for gaze tracking*. IEEE 10th International Conference on Signal Processing. 2010-10-24, p. 1293 - 1296. ISBN: 978-1-4244-5897-4.

(13) YANG, X. et al. *A Gaze Tracking Scheme for Eye-Based Intelligent Control*. 8th World Congress on Intelligent Control and Automation. Jinan China, 2010-07-07, p. 50 - 55. ISBN: 978-1-4244-6712-9.

(14) ZHANG, K. et al. *A simplified 3D gaze tracking technology with stereo vision*. International Conference on Optoelectronics and Image Processing. 2010, p. 131-134. ISBN: 978-1-4244-8683-0.