Pupil center detection with a single webcam for gaze tracking

Ralph Oyini Mbouna, Seong G Kong

(*Dept. of Electrical and Computer Engineering, Temple University, Philadelphia, PA 19122, USA*)

Abstract: This paper presents a user friendly approach to localize the pupil center with a single web camera. Several methods have been proposed to determine the coordinates of the pupil center in an image, but with practical limitations. The proposed method can track the user's eye movements in real time under normal image resolution and lighting conditions using a regular webcam, without special equipment such as infrared illuminators. After the pre-processing steps used to deal with illumination variations, the pupil center is detected using iterative thresholding by applying geometric constraints. Experimental results show that robustness and speed in determining the pupil's location in real time for users of various ethnicities, under various lighting conditions, at different distances from the webcam and with standard resolution images.

Key words: pupil detection; eye gaze tracking; feature extraction; facial features tracking

CLD number: TP391 **Document code:** A

Article ID: 1674-8042(2012)02-0133-04 **doi:** 10.3969/j.issn.1674-8042.2012.02.008

Pupil center detection, a process of determining the location and trajectory of the eye in real time, offers an effective means of human-machine interface. Pupil center detection plays an important role in various applications such as monitoring driver alertness or tracking the eyes of a person with a psychological disorder who cannot communicate effectively. Detecting the center of the pupil has been used for disabled persons who have lost control of their limbs through gaze estimation[1].

Because of the high variety of applications there is a need for more and more advanced pupil detection systems. As more pupil detection systems are developed in the recent years, two main groups emerged. The first one is called intrusive systems. These systems require specialized hardware on the person to locate the pupil center accurately and therefore are not very practical for natural scenarios. For the best result in not only user comfort, but system portability and versatility, pupil detection has to occur without any elaborate installations. The second one focuses on non-intrusive systems, which are classified further into three: feature, shape and appearance based methods. those are presented in this paper.

This paper presents a non-intrusive infrared illumination free hybrid method for detecting the pupil center, which combines the advantages of both appearance and features based models. Our algorithm accurately locates the left and right eye pupil centers in real time using a regular webcam. The proposed method blends well-established techniques such as the Viola-Jones object detection framework and morphology operators to detect the face and eye region. To estimate the pupil center, the following contributions are made: in preprocessing stage detection obstacles that are often caused by eyebrows, extra skin and illumination variations around the eye are reduced; in a second stage an adaptive and iterative thresholding technique is used to reduce as much as possible the area where the pupil center is situated; and in a final stage the pupil center is identified as the normalized center of gravity of the remaining region. Additionally, the accuracy and robustness of our algorithm is later tested to change in ethnicity, light conditions, viewing angle, position of the pupil, background complexity and head pose using the complete BioID database and real time situations.

1 Previous work

An array of methods has been investigated previously for pinpointing the pupil center in natural scenes in real time. The existing research includes intrusive and non-intrusive systems.

Intrusive systems uses special gear such as electrodes, radar range finder or special glasses with a small webcam mounted on them to track the user's eyes[2]. Those systems tend to be expensive and un-

※ **Received data:** 2011-10-18
Corresponding author: Seong G Kong (skong@temple.edu)

friendly as they require special hardware. To make pupil detection more natural, non-intrusive techniques have been proposed.

Non-intrusive systems fall into two main categories: pupil center corneal reflection technique which uses infrared LED illumination, and video based techniques that do not require illuminators. Most commercial eye gaze systems use pupil corneal reflection algorithms, which uses at least one infrared light to illuminate the eye. A camera then captures images of the eye for analysis, and the pupil and the infrared corneal reflection positions are detected. Based on the distance between the pupil and the corneal reflection, the gaze direction is calculated using geometry[3]. The problem with this technique is that it only works in an indoor-controlled environment which may limit practical applications. Indeed, these systems are very sensitive to light and impractical to outdoor scenarios because they are based solely on the position between the pupil and the infrared illumination.

To overcome those difficulties, techniques relying solely on video cameras have been developed: appearance, shape and features based methods[4]. Appearance based methods estimates the pupil center from comparing a new image of an eye to a set of eye images contained in a database to find the best match. In Ref. [5], by detecting user's eye blinks and measuring their duration, the eyes are spotted using an appearance based approach, which compares the current eye with an open eye template and as a result the amount of closure of the current eye is measured. Their blinks detection system uses a set of correlation coefficients as a measure of match between the eye template and current eye. For appearance based methods, extracting the pupil position precisely is very difficult. The accuracy of such a system is often based on how much data is used. Therefore, these systems require a lot of data to be accurate and the system tends to be too slow and fails in real time tracking. Shape based methods use the shape of the eye to build a geometric eye model. The position of the eye features is utilized to update the eye model and estimate the current pupil center location. For instance, the circle algorithm[6] fits the iris into an ellipse shape from selected features. WANG and SUNG use the special shape characteristics of the iris which are the transition from white to dark then dark to white to detect the iris. Because the iris can be partially occluded by the upper and lower eye lids, it is difficult to detect and fit the iris accurately into an ellipse. Therefore, shape based methods along with the circle algorithm depend highly on the resolution of the eye images to shape the ellipse correctly. When used with a regular webcam, the circle algorithm performs poorly. Moreover, there are features based methods that explore the characteristics of the human eye to identify the center of the pupil. Feature based approaches aim for features that are less sensitive to variations in illuminations. A generalized projection function (GPF) that combines integral and variance projection functions together has been proposed by ZHOU and GENG to detect the pupil center[7]. The aim of their method is to merge the robustness of the integral projection function and the sensitiveness of the variance projection function. Their method exploits the fact that the eye area is darker than its neighboring areas and the intensity of the eye area rapidly changes. However, these techniques show poor results in very illuminated environments because the pupils become very bright and smaller after thresholding.

2 Pupil center detection

We propose a hybrid method that combines the strengths of appearance and feature based algorithms while limiting their weaknesses. This pupil center detection system uses the precision of appearance technique. To make sure our program runs in real time, the amount of training data is reduced to only detect the face and rough position of the eyes in the image. Then, we exploit feature based technique processing steps to locate the pupil center by using an adaptive thresholding that is robust against illumination changes. Fig. 1 shows the overall procedures of the eye pupil center detection. The face and the eyes region are detected using the Viola-Jones appearance technique. Because the Viola-Jones method does not detect the eyes accurately, some pre-processing steps including color space conversion and Otsu thresholding are applied to each eye image to remove the skin and eyebrows parts around the eyes. Once the eyes are localized more precisely, an adaptive thresholding technique that uses the physical properties of the eye is utilized to localize the pupil center.

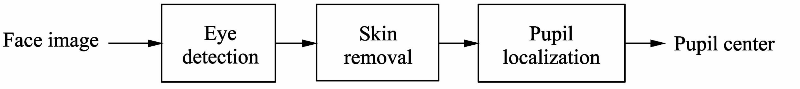


Fig. 1 Pupil center detection block diagram

2.1 Face and eyes detection

The face is first detected using the Viola-Jones detector also called the Haar classifier which uses the Haar transform features to add and subtract rectangular image regions before thresholding the result. The Haar classifier detector is trained in this case to recognize faces. Once the face is detected, it is divided horizontally into three parts. The top two parts are selected to be the region of interest that contains the eyes. By reducing the search area, the computation time is reduced. Using the same Viola-Jones detector, both eyes are detected. The face is then divided into two parts vertically. The left eye is chosen to be the eye on the left side of the face and the right eye on the right side. At this stage, the eye image detected by the Haar classifier still contains a lot of useless information such as the skin and eyebrows. Because the skin is illuminated intensively, it influences the thresholding when detecting the pupil center. To remove the skin, the eye image is converted into hue saturation value (HSV) color space. The HSV coordinate system is cylindrical. The angle around the central vertical axis represents "hue", the distance from the axis is the "saturation", and finally the vertical position defines the "value" also referred the "brightness" (Fig. 2). Knowing that almost all humans have the same hue, the skin is removed by Otsu automatic thresholding the hue component of the image. Thus, the eye image only contains a little amount of skin and no eyebrows. The image is cropped to the nearly dimensions of the eye for further processing.

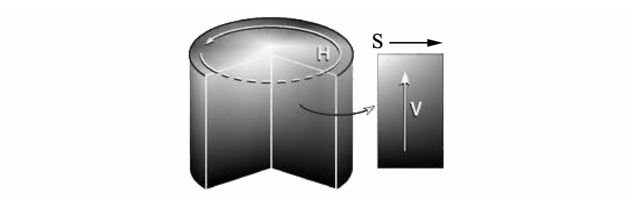


Fig. 2 HSV color space

2.2 Pupil center detection

The pupil center is detected by the following method. First, the eye image without skin is upsampled to twice its original size. Then, the resulting image is eroded and normalized to reduce the effect of the variations in illumination. At that point, the eye image is thresholded using an adaptive approach. The pupil is assumed to be darker than the background which is mainly the white of the eye. Therefore any pixel below a certain threshold value is labeled as the pupil. Because the environment especially illumination changes all the time, a fixed threshold cannot be selected. So, an initial threshold value t_i is selected such as t_i is large enough to contain at least the pupil and some of the background as well. Then using the ratio between the eye and pupil widths as well as the ratio between the eye and pupil heights, the image is iteratively thresholded until the geometric constraints are satisfied and the desired threshold t_f is chosen. At last, only the pupil remains. The pupil center is then estimated by computing the center of mass also called the center of gravity. Knowing $I(x, y)$ is the pupil image, the coordinates of the center of gravity x_c and y_c are defined as the spatial moments of first order $m_{1,0}$ and $m_{0,1}$ divided by the area A of the object

$$m_{p,q} = \sum_{i=1}^{n} I(x, y) x^p y^q . \qquad (1)$$

The center of gravity has shown better results in determining the center of the pupil than computing the center of the contour area.

3 Experimental results

The experiments are made using the BioID face database because it provides with images with a variety of illumination, background and face size. The database contains 1 521 gray scale images with a 384×286 resolution. The images in the database were created using 23 different test persons in real time conditions. Subjects having their eyes closed and wearing eyeglasses were not tested. Table 1 shows the results obtained using our algorithm compared with state of art methods that have been applied to the BioID images too. The relative error d_{eye}[8] between the expected and the estimated pupil center positions is used to generate the results

$$d_{eye} = \frac{\max(d_l, d_r)}{\| C_l - C_r \|}, \qquad (2)$$

where $d_i = \| C_i - \bar{C}_1 \|$ and $C_i = \sqrt{x_i^2 + y_i^2}$.

Table 1 Pupil center detection rate

Author	Algorithm	Detection rate	Remarks
Jesorsky et al., 2001[8]	Ada Boost	91.8%	
Zhou and Geng, 2004[7]	Generalized projection function(GPF)	94.8%	
Asadifard and shanbezadeh, 2010[9]	Cumulative distribution function (CDF)	96.0%	*
Oyini Mbona and Kong	Hybrid	97.8%	*

(*) Images with closed eyes and eyeglasses are not considered

The distances d_l and d_r are the left and right distances between the true pupil center positions C_l, C_r and the estimated pupil center positions $\bar{C}_l$, $\bar{C}_r$. The subscript represents the left l or right r side (Fig. 3). Note that a relative error of less than 0.25 is considered as a successful detection so that the detection rate is computed as

$$Detection\ rate = \frac{Number\ of\ success}{Total} * 100. \quad (3)$$

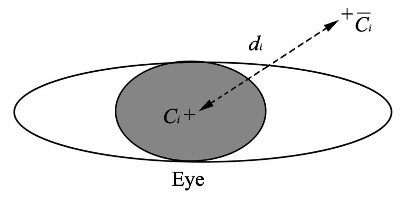
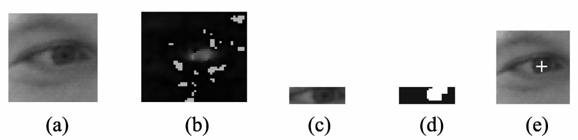


Fig. 3 Pupil Center Detection Block diagram

To validate our results, the algorithm was tested in real time. It is important for the pupil center detection algorithm to be fast since it only represents the first step of many of the human-computer interaction applications. Our pupil detection speed was about 12 frames per sec using a PC Intel Core 2 Duo 3 GHz.



(a) Eye detection, (b) Hue channel conversion, (c) Skin and eyebrow removal, (d) Iterative thresholding, (e) Pupil center localization

Fig. 4 Eye processing results

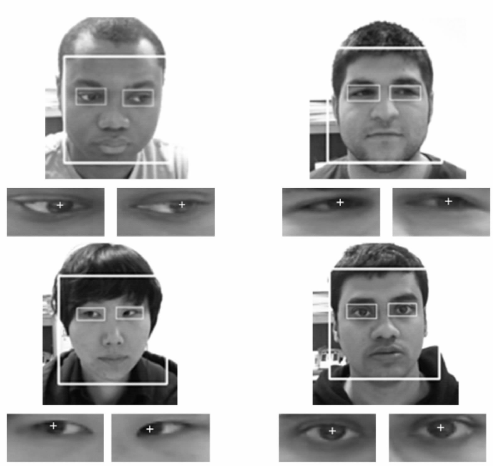


Fig. 5 Real time detection results for different subjects

4 Conclusion

A hybrid pupil center detection system which combines the advantages of both appearance and feature based method is presented in this paper. The pupil detection algorithm detects the face and rough position of the eyes using the Viola-Jones technique. Then, it applies morphology operators such as erosion and normalization to reduce the effect of illumination on eye images. To avoid false errors, the extra skin around the eyes and eyebrows are removed using HSV color space conversion. An iterative thresholding technique is computed to the resulting image and finally the pupil center position is estimated by using the center of mass. Comparing with conventional pupil center detection systems, the proposed method requires low computational difficulty and works in real time with various users from different ethnicity background. In addition, it is scalable and does not necessitate any special light source such as IR illumination. Because of its practicality, this system has a lot of potential that could be used in countless applications.

References

[1] Morimoto C H, Mimica M R. Eye gaze tracking techniques for interactive applications. computer vision and image understanding. Special Issue on Eye Detection and Tracking, 2005, 98(1): 4-24.

[2] Matsumoto Y, Zelinsky A. An algorithm for real-time stereo vision implementation of head pose and gaze direction measurement. Proc. of 4th IEEE Internatioanl Conference on Automatic Face and Gesture Recognition, Grenoble, France, 2000: 499-504.

[3] Grover D. Progress on an eye tracking system using multiple near-infrared emitter/detector pairs with special application to efficient eye-gaze communication. Proc. of the 2nd COGAIN Conference on Communication, Turin, Italy, 2006: 21-25.

[4] Hansen D, JI Qiang. In the eye of the beholder: a survey of models for eyes and gaze. IEEE Trans. on PAMI, 2010, 32(3): 478-500.

[5] Grauman K, Betke M, Gips J, et al. Communication via eye blinks: detection and duration analysis in real time. Proc. of IEEE Conference on Computer Vision and Pattern Recognition, 2001, 1: 1010-1017.

[6] Wang J G, Sung E. Gaze determination via images of irises. Image and Vision Computing, 2001, 19(12): 891-911.

[7] Zhou Z, Geng X. Projection functions for eye detection. Pattern Recognition, 2004, 37(5): 1049-1056.

[8] Jesorsky O, Kirchberg K, Frischholz R. Robust face detection using the Hausdorff distance. Proc. of the 3rd Audio-and Video-Based Biometric Person Authentication, Halmstad, Sweden, 2001: 90-95.

[9] Asadifard M, Shanbezadeh J. Automatic adaptive center of pupil detection using face detection and CDF analysis. Proc. of the International MultiConference of Engineers and Computer Scientists, Hongkong, China, 2010, 1: 130-133.