Single Eye Gaze Tracking With Active Pan-Tilt Camera

Stanislav V. Panev[1], Ognian L. Boumbarov[2] and Plamen P. Petrov[3]

Abstract – **In this paper an algorithm is proposed for estimation of human gaze direction by means of a geometrical model of the eye based on distances between the pupil centre and the two eye corners and artificial infra-red switching lightning. For user comfort an active pan-tilt camera is employed, which makes possible free head movements instead requiring the user to be still.**

Keywords – **eye, gaze, tracking, active camera, pan tilt camera.**

I. INTRODUCTION

Nowadays human-computer interaction takes one of the main positions of scientific researches and developments. Estimation of gaze direction is one very interesting part of it, thus a lot of effort is made in attempting to make computers see where we look at. This way interacting with computer will be much easier for the user because we perceive most of the surrounding information by our eyes.

A large number of systems for eye tracking are available. Many of them are based on contact lenses, electrodes, specialized hardware, and infrared emitters. Such systems could easily be separated into three groups, according to the approach used: methods, using the electric potential of the human skin [1], methods that involve contact lenses [2] and methods that involve image analysis.

In general image based approaches are divided into two groups – appearance-based and model-based.

Appearance-based approaches directly treat an eye image as a high dimensional feature. Baluja and Pomerleau use a neural network to learn a mapping function between eye images and gaze points (display coordinates) using 2,000 training samples [3]. Tan et al. take a local interpolation approach to estimate unknown gaze point from 252 relatively sparse samples [4]. Recently, Williams et al. proposed a novel regression method called S3GP (Sparse, Semi-Supervised Gaussian Process), and applied it to the gaze estimation task with partially labeled (16 of 80) training samples [5]. Appearance-based approaches can make the system less restrictive, and can also be very robust even when used with relatively low-resolution cameras.

Model-based approaches use an explicit geometric model of the eye, and estimate its gaze direction using geometric eye features. For example one typical feature is the pupil glint vector ([6], [7]), the relative position of the pupil centre and the specular reflection of a light source. Model-based approaches typically need to precisely locate small features on the eye using a high-resolution image and often require additional light sources but can be very accurate.

Several approaches exist for eye-tracking by using images. Some of them are largely dependent on active light-sources, such as infra-red emitters [8]. Other approaches, on the other hand, do not use the information from an explicit light-source, but rather use information for improving the quality of the image. Still other approaches even exclude the usage of active lighting, but rather rely on natural light-sources.

The rest of the article is organized as follows: in Section 2 an approach is described for human gaze tracking which relies on detecting pupil and the two eye corners locations. In order to estimate the pupil's centre coordinates a double switching infra-red lightning is used which turns on and off consecutively in two neighboring frames the camera on- and off-optical axis LEDs. In order to estimate the location of the two eye corners a deformable template of the eye is used. This template consists of two parabolas which depicts the two eye lids. The two cross points of the two parabolas are the inner and outer eye corners. Assuming that the eye is a spherical body we can estimate the gaze direction. In Section 3 is described the control system of a Pan-Tilt unit which is employed to centre the image of the eye in the frame. Thus the user has the freedom to move its head instead of standing still. In Section 4 we present our experimental results. In Section 5 we conclude and discuss the future development of our algorithm.

II. GAZE TRACKING ALGORITHM DESCRIPTION

A. Assumptions

Before presenting the operations implemented to realize the gaze tracking procedure some basic assumptions should be taken in mind.

First of all the working light spectrum is in the near-visible infra-red domain. The visible light is cut by a infra-red filter. This way a lot of noise is eliminated coming from the surrounding light sources and the contrast of the pupil is much better.

Only one eye should be presented in the frame. This way a higher resolution of the eye image is achieved with low resolution cameras, which increases the accuracy of tracking and algorithm confusion is avoided. The rest of the image is filled with the face skin.

All the images taken from the camera are in the grayscale space.

---

[1] Stanislav V. Panev is with the Faculty of Telecommunications, 8 Kl.Ohridski, 1000 Sofia, Bulgaria, E-mail: s_panev@tu-sofia.bg

[2] Ognian L. Boumbarov is with the Faculty of Telecommunications, 8 Kl.Ohridski, 1000 Sofia, Bulgaria, E-mail: olb@tu-sofia.bg

[3] Plamen Petrov is with the Faculty of Machine Technology, 8 Kl.Ohridski, 1000 Sofia, Bulgaria, E-mail: ppetrov@tu-sofia.bg

B. Pupil segmentation in image

Pupil segmentation is implemented according to the method described by Y. Ebisawa in [9]. It uses the ability of the eye's retina to reflect the light beam which penetrates in it exactly in the same direction where it came from. Hence if two light sources are employed, one on and one off the camera optical axis, the pupil in frame when the on-axis light is on will be bright. Respectively in the frame shot during the off-axis light is on the pupil will be dark. The rest of the image will have almost the same intensity in the two frames because the lights are close enough to the camera and object. So if the pixel values of the two frames are subtracted the greatest difference will have these pixels which belongs to the pupil (Fig. 1).

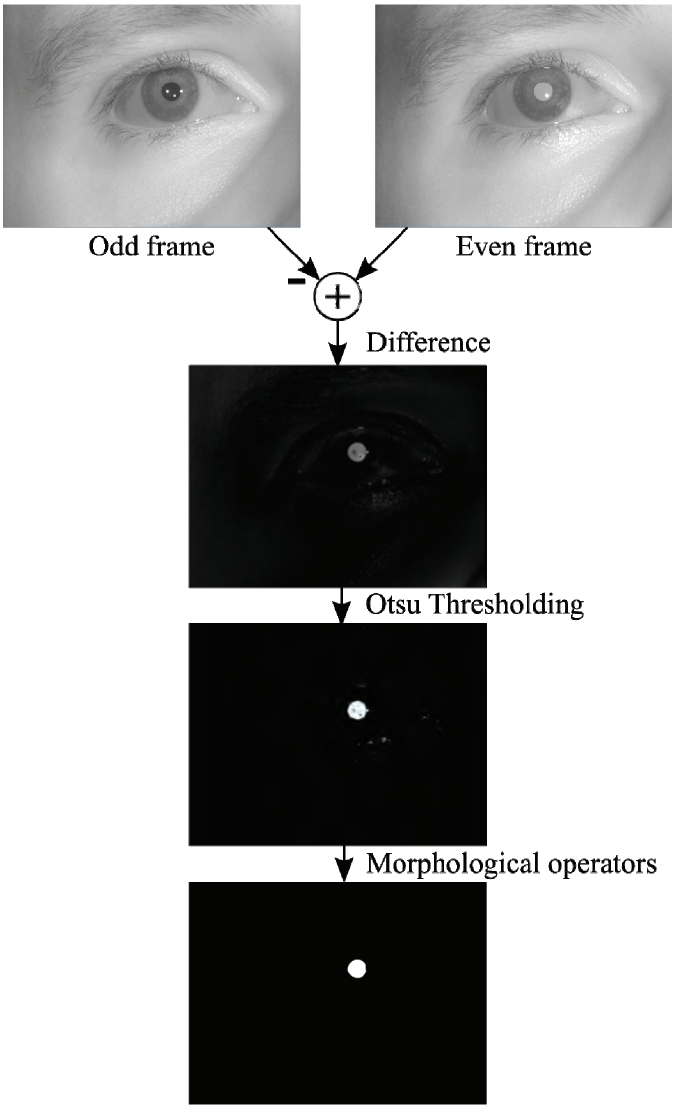


Fig. 1. Pupil segmentation procedure.

As can be seen from Fig. 1 after binarizing the difference image with optimal threshold based on Otsu algorithm, there are some little white dots which can be assumed as noise. By using combination of morphological operators erosion and dilation all these noise is suppressed and the white spot in the final image shows where exactly the pupil is in the image. This way finding the position of the pupil in the image is very fast and doesn't require complex tracking algorithms. As a result the center of the white spot is estimated.

C. Eye corners detection

The detection of the eye corners is achieved by using deformable template of the eye lids showed on Fig. 2. The template which has been used consists of two parabolas which represent the top and bottom eye lids. Hence the two points where the two parabolas intersect with the other one are the two eye corners.

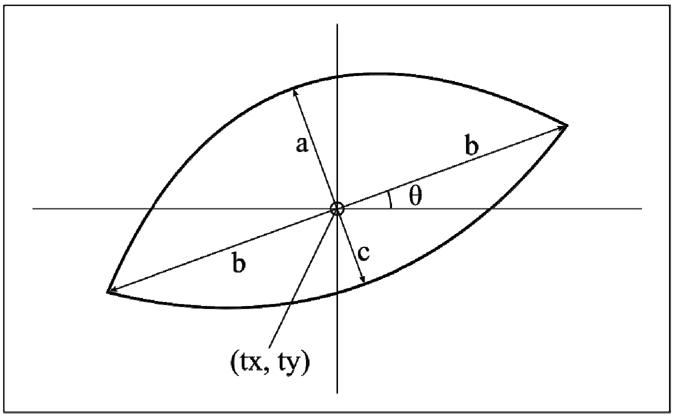


Fig. 2. Deformable template of eye.

The method for template fitting is described by Jyh-Yuan Deng in [10] and uses so called regional forces. These forces can be obtained by calculating the window force f_W (Fig. 3) of a small window from the image according to the following equation:

$$f_W = \frac{\sum each\ pixel\ value\ in\ the\ window}{number\ of\ the\ pixels\ in\ the\ window} - \frac{255}{2} \qquad (1)$$

If we slide this window along a known curve (the two parabolas) and integrate the window force for all the positions, the result will be the *regional force* which “tries” to deform the template such way that the forces will be equal to zero. As the force is has a linear effect this means that the regional forces will try to move or scale the template but not rotate. That's why the concept for the *regional torque* should be introduced (Fig. 3). If a pivot point is chosen as a center of the rotation, for example the center of the pupil which is already known, it is easy to estimate the torque for rotation of the template with the following equation:

$$\begin{aligned}\tau_{clockwise} = & \sum_i \text{each } f_{w_i} w_i \text{ in the left} \\ & - \sum_j \text{each } f_{w_j} w_j \text{ in the right}\end{aligned}, \qquad (1)$$

where w_i is the weighting factor equivalent to the effective distance to the pivot point for calculating the torque.

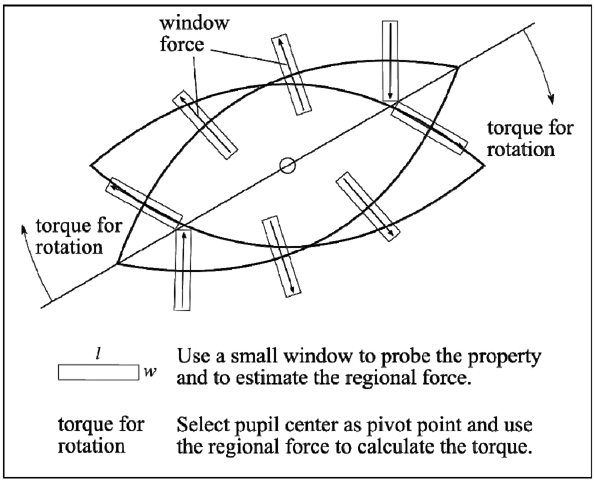


Fig. 3. Regional torque for rotation

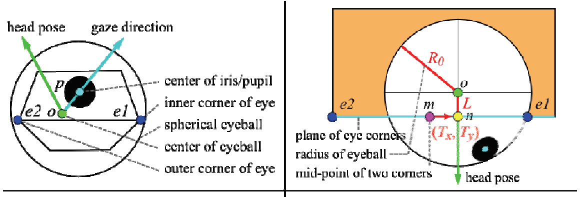
D. Gaze estimation geometric model

In this subsection a description of the geometric eye model which we use to estimate gaze direction is given. There is nothing particularly novel about this model. We assume that the eyeball is spherical and the inner and outer eye corners have been estimated, in our case using a deformable template. The algorithm is split into two steps [11]:

1. Estimate the center and the radius of the eyeball in the image from the eye corners and the head pose.
2. Estimate the gaze direction from pupil location, the center and the radius of the eyeball.

The first of these two steps requires the following anatomical constants (Fig. 4):

- R_0 : The radius of the eyeball in the image when the scale of the face is 1.
- (T_x, T_y): The offset in the image between the mid-point of the two eye corners and the center of the eyeball.
- L : The depth of the center of the eyeball relative to the plane containing the eye corners.



(a) Definition of Terms

(b) Anatomical Constants (R_0, L, T_x, T_y)

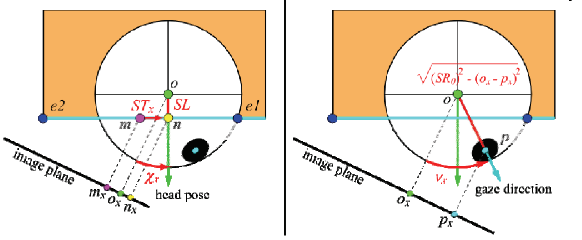


Fig. 4. Gaze estimation geometric model

The center and the radius of the eyeball are computed using the following three steps:

1. The mid-point (m_x, m_y) between the inner corner $(e1_x, e1_y)$ and the outer corner $(e2_x, e2_y)$ is estimated:

$$\begin{pmatrix} m_x \\ m_y \end{pmatrix} = \begin{pmatrix} \dfrac{e1_x + e2_x}{2} \\ \dfrac{e1_y + e2_y}{2} \end{pmatrix}. \tag{2}$$

2. The scale of the face S is computed by

$$S = \frac{\sqrt{(e1_x - e2_x)^2 + (e1_y - e2_y)^2}}{\cos\varphi_x} \tag{3}$$

3. The center of the eyeball (o_x, o_y) is then computed as the mid-point (m_x, m_y) plus two corrections:

$$\begin{pmatrix} o_x \\ o_y \end{pmatrix} = \begin{pmatrix} m_x \\ m_y \end{pmatrix} + S \begin{pmatrix} T_x \cos\varphi_x \\ T_y \cos\varphi_y \end{pmatrix} + SL \begin{pmatrix} \sin\varphi_x \\ \sin\varphi_y \end{pmatrix} \tag{4}$$

Gaze direction (θ_x, θ_y) can then be estimated as follows:

$$\begin{pmatrix} \sin\theta_x \\ \sin\theta_y \end{pmatrix} = \begin{pmatrix} \dfrac{p_x - o_x}{\sqrt{R^2 - (p_y - o_y)^2}} \\ \dfrac{p_y - o_y}{\sqrt{R^2 - (p_x - o_x)^2}} \end{pmatrix} \tag{5}$$

The anatomical constants R_0, (T_x, T_y) and L are pre-computed in an offline training phase.

III. PAN-TILT CAMERA MOTION ALGORITHM DESCRIPTION

The viewing angle of the camera is comparatively small (8,24° in horizontal and 6,18° in vertical direction for ¼" camera sensor and focal length 25 mm) which is a consequence of the assumptions depicted in Section 2.1. This means that there is a real possibility parts of the image of the eye or even the whole of it to go out of the frame if the user moves itself a little in some direction. That's why for its comfort and robustness of the tracking procedure an mechanical Pan-Tilt unit for camera motion is employed. The aim of this mechanical system is to position the image of the eye in the center of the frame.

As the inertia of the camera is comparatively small (its weight is about 100 – 200 gr.), it can be ignored to simplify the law of movement. Besides we assume that the user is seated and the movement limits of the head will be in the interval ±10° from the zero position of the camera and the movement won't be rabid and abrupt. Hence the control law can be simplified to a purely proportional law by eliminating the integral and differential part of a PID controller.

The input data for the control system are the coordinates of the pupil center (x_i, y_i) in the image in relation to a coordinate system which is placed in the center of the frame. Therefore the system has to aspire to move the camera such direction that the pupil coordinates are 0. The angles of rotation of the camera can be estimated from:

$$\begin{aligned} \wp_x &= arctg\left(\frac{w_{ccd}}{w_p f}.x_i\right) \\ \wp_y &= arctg\left(\frac{h_{ccd}}{h_p f}.y_i\right) \end{aligned}, \qquad (6)$$

where w_{ccd} and h_{ccd} are the physical width and height of the camera's CCD sensor respectively, w_p and h_p are the size of the frame in pixels in vertical and horizontal direction respectively and f is the focal length of the lens.

The control of the camera rotation is done through its angular velocity ω. Thus as far the pupil is from the image center as faster the camera will rotate its position. On Fig. 5 a block diagram of the camera control system is depicted.

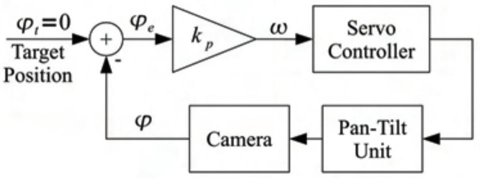


Fig. 5. Block diagram of Pan-Tilt Unit control system

IV. EXPERIMENTAL RESULTS

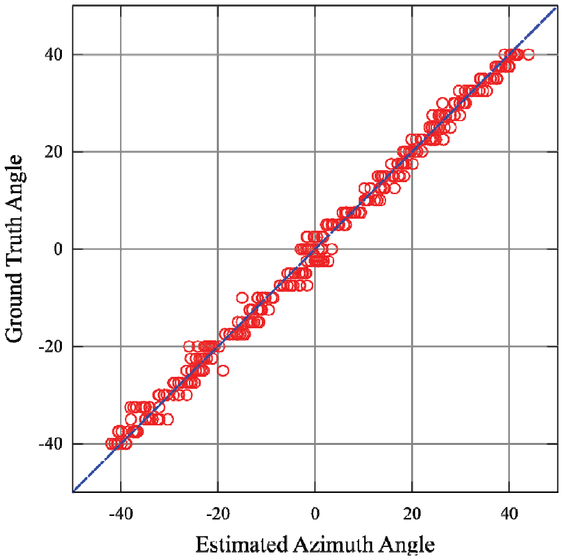


Fig. 6. Azimuth angle estimation of an object

The described above system for gaze estimation has been tested in real conditions and gave quite satisfying results. An experimental estimation of the azimuth angle of the gaze direction is shown on Fig. 6. The calculated mean squares error is 1,56°.

V. CONCLUSION

In this paper an approach to gaze direction estimation of a single eye image was presented. It uses a deformable template and double switching infra-red lighting to estimate the position of the pupil and the eye corners. Hence by the means of an eye geometrical model the direction of the user gaze is estimated. The system behaves notably robust and accurate.

ACKNOWLEDGEMENT

This work was supported by the National Ministry of Science and Education of Bulgaria under contract BY-I-302/2007: “Audio-video information and communication system for active surveillance cooperating with a mobile security robot”.

REFERENCES

[1] J. Gips, P. Olivieri, “EagleEyes: An Eye Control System for Persons with Disabilities,” The Eleventh Intern. Conference on Technology and Persons with Disabilities, 1996, pp. 128-132.

[2] L. Young and D. Sheena, “Methods & Designs: Survey of Eye Movement Recording Methods,” Behavioral Research Methods & Instrumentation 7(5), 1975, pp. 397-429.

[3] S. Baluja and D. Pomerleau, “Non-intrusive gaze tracking using artificial neural networks,” Advances in Neural Information Processing Systems (NIPS) 6, 1994, pp. 753–760.

[4] K. H. Tan, D.J. Kriegman, and N. Ahuja, “Appearance-based eye gaze estimation,” In: Proceedings of the Sixth IEEE Workshop on Applications of Computer Vision (WACV 2002), 2002, pp. 191–195.

[5] O. Williams, A. Blake, and R. Cipolla, “Sparse and semi-supervised visual mapping with the S3GP,” Proceedings of the 2006 IEEE Conference on Computer Vision and Pattern Recognition, 2006, pp. 230–237.

[6] T. E. Hutchinson, K. P. White Jr., W.N. Martin, K.C. Reichert, and L.A. Frey, “Human-computer interaction using eye-gaze input,” IEEE Transactions on Systems, Man and Cybernetics 19(6), 1989, pp. 1527–1534.

[7] R.J. Jacob, “What you look at is what you get: eye movement-based interaction techniques,” In: Proceedings of the SIGCHI conference on Human factors in computing systems, 1990, pp. 11–18.

[8] C. Morimoto, D. Koons, A. Amir, and M. Flickner, “Pupil detection and tracking using multiple light sources,” IVC18(4), 2000, pp. 331-335.

[9] Y. Ebisawa, “Improved Video-Based Eye-Gaze Detection Method”, IEEE Transactions on Instrumentation and Measurement, Vol. 47, No. 4, August 1998

[10] JY. Deng, F. Lai, “Region-based Template Deformation and Masking for Eye-feature Extraction and Description”, Pattern Recognition, The Journal of Pattern Recognition Society, Vol. 30, No. 3, pp. 403-419, 1997

[11] T. Ishikawa, S. Baker, I. Matthews, T. Kanade, “Passive Drive Gaze Tracking with Active Appearance Models”, In Proceedings of the 11th World Congress on Intelligent Transportation Systems, October 2004