AN EFFICIENT METHOD FOR EYE TRACKING AND EYE-GAZED FOV ESTIMATION

Xiaokun Li[1] *and William G. Wee*[2]

[1]Signal/Image Processing Systems, DCM Research Resources LLC, Germantown, MD 20874, USA
[2]Department of ECE, University of Cincinnati, Cincinnati, OH 45221, USA

ABSTRACT

An eye-tracking integrated head-mounted display (ET-HMD) system can be used in many applications. Its complexity imposes great challenges on designing a compact, portable, and robust system. For an ET-HMD system, having accurate eye-gazed field-of-view (FOV) estimation and performing fast high-fidelity rendering only on the identified FOV is its crucial technique for augmented video/image/graphics display. In this paper, an energy-controlled iterative curve fitting method is proposed for accurate pupil detection and tracking. Based on a 3D eye model and the pupil tracking results, eye-gazed FOV can be identified precisely. The experimental results show that the average tracking error of using the proposed method is less than 0.5 degree for pupil rotation while the other conventional methods can only achieve 1.0 degree.

Index Terms—Eye tracking, eye model, eye-gazed FOV

1. INTRODUCTION

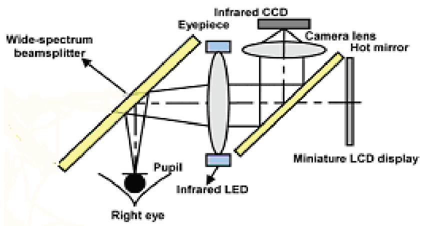
Researchers have been invested techniques to track eye movements for decades and many eye tracking technologies have been developed and used in many applications and fundamental instruments. Among them, a very common application is the eye-tracking integrated head-mounted display (ET-HMD) for virtual and augmented, scientific, and engineering visualization, video game, training simulation, ultra-wide field-of-view (FOV) enhanced vision, etc. For an ET-HMD system, the information of eye movement is crucial for fast FOV rendering and augmented display.

The currently-available eye tracking methods can be broadly classified into two categories: two-dimensional (2D) tracking and three-dimensional (3D) tracking. For 2D-based tracking methods [1], the 3D position of the eyes is projected onto a 2D plane for study. In contrast to 2D tracking, 3D eye tracking methods [2] estimate 3D position of the eyes via stereoscopic models and thus provide 3D information of the line of sight with respect to a fixed reference. These 3D methods are naturally robust to head movements in some extent, but their usage is very limited because of the requirements on the remote-mounted and the large FOV of the eye-tracking cameras for 3D estimation. Therefore, 2D tracking methods are still the most practiced techniques in use. 2D tracking techniques can be roughly categorized as non-image or image-based methods. The non-image approaches include electro-oculography (EOG) methods, which detect eye movements based on the difference around the eyes with electrodes attached around a subject's eye [3]. Image-based tracking methods are the most popular techniques for eye tracking. These methods track eye movement by detecting and analyzing eye features and take the advantages of the spectral properties of the eye under near infrared (NIR) illumination which is robust to the changes of environmental lighting and can work in 24 hours and day/night. In this paper, a systematic approach to a fully integrated ET-HMD instrument is presented to improve and optimize the eye-tracking performance and the estimation of eye-gazed FOV for fast rendering and augmented display.

2. EYE-TRACKING DESIGN IN HMDS

To develop a compact, lightweight, and robust system with improved display quality and tracing accuracy as shown in Fig. 1(a) and (b), one possible strategy is to share the optical paths between the display system and the eye tracker as much as we can [4]. A latest design [5] of ET-HMD is illustrated in Fig. 1(c). In the design, the traces of the optical path of the eye tracking sub-system are the same as the display system of HMD. The necessary components for an optical see-through HMD include miniature displays, display optics, and beam-splitters. To illuminate the eye, one or multiple near infrared LED (NIR-LED) illuminators are mounted around the display optics or at a position between the optics and the eye. NIR light reflected off the eye features is collimated through the display optics, which is a part of display system and shared with the eye-tracking sub-system, reflected by a hot mirror, and imaged through a second group of lenses (e.g. camera lens) onto an infrared CCD sensor. The display optics together with the camera lens serves as the imaging optics of the eye tracking sub-system.

(a) One-eye head-mounted display system (b) Binocular HMD



(c) Eye-tracking device in HMDs

Figure 1 Eye-tracking integrated HMD

3. TRACKING EYE MOVEMENT

Eye tracking can be thought as the process of measuring the pupil direction - either the point of gaze ("where we are looking at") or the motion of an eye relative to the head. Most modern eye-trackers use contrast to locate the center of the pupil and use infrared and near-infrared non-collumnated light to create a corneal reflection (CR). To receive better quality pupil images in the design as shown in Fig. 1(c), two Infrared LEDs are offset from the optical path for pupil imaging. To detect pupil from eye images, researchers first binarize the eye image, then apply the crossed chord theorem for pupil localization. According to the crossed chord theorem, if two chords of a circle AB and CD intersect at the point P, then we can get the equation: PA×PB=PC×PD. To simplify the problem, two perpendicular chords AB and CD are assumed to intersect at P. If the four points (A, B, C, D) satisfy the crossed chord theorem, then we can choose three of the four points to estimate the radius and the center coordinates of the pupil.

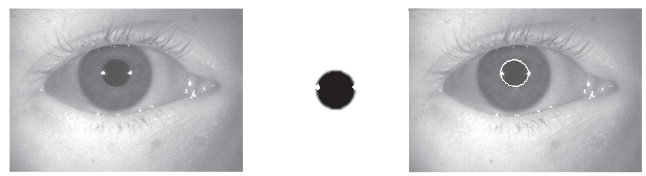
Due to the 3D movements of a human eye, camera is usually either off-angle or in tilted position to the subject's eye during imaging. In these situations, the acquired eye images will experience global geometric distortion - e.g., the circular shape of human pupil often becomes rotated ellipses when pupil rotates. Thus, it will cause errors on pupil center and size estimation if we simply conduct the coroosed chord theorem. Additionally, reflections and out-of-focus further degrade the quality of eye images, which make the problem of non-ideal pupil measurement nontrivial.

To address this challenge, we developed an energy-controlled iterative ellipse fitting (ECIEF) method to estimate the pupil center and the pupil size. In our proposed approach, we first use double thresholding, which is based on morphological reconstruction and quite robust to complex environments, to segment pupil from the eye image. Then, an energy-controlled iterative ellipse fitting will be conducted to the segmentation results to estimate the pupil center and its size precisely.

The double thresholding operator uses a thresholding operator $T_{[t_i,t_j]}$, known as level slicing, which sets all pixel (*x*,*y*) of the ROI whose values are in a pre-specified range $[t_i, t_j]$ to 1 and all other values to 0. Then, the method slices the gray levels of the region of interest by using two ranges of grayscale values, one included in the other. If the two ranges are $[t_1, t_4]$ and $[t_2, t_3]$, such that $[t_1 \le t_2 \le t_3 \le t_4]$. Since $T_{[t_2,t_3]}(f) \in T_{[t_1,t_4]}(f)$, we use $T_{[t_2,t_3]}(f)$ as a marker and $T_{[t_1,t_4]}(f)$ as a mask for binary image reconstruction. The whole process can be described mathematically as:

$$DT_{[t_1 \le t_2 \le t_3 \le t_4]}(f) = R_C(T_{[t_2,t_3]}(f) \mid T_{[t_1,t_4]}(f)) \qquad (1)$$

The wide level slicing result with $T_{[t_1,t_4]}(f)$ will contain both pupil and clutter, while the narrow level slicing result $T_{[t_2,t_3]}(f)$ will contain only information pertaining to pupil. The reconstruction $R_C(T_{[t_2,t_3]}(f) \mid T_{[t_1,t_4]}(f))$ will remove unwanted clutter and reserve pupil. Since the camera is close to human eyes and the light settings of infrared LEDs are already known, the value of t_i is easy to be initialized. A pretty good segmentation result can be obtained. One example of using double thresholding is illustrated in Fig.2.



(a) Original image (b) Binary image (c) Estimated boundary

Figure 2 Pupil localization

Due to the deformed and/or partly-occluded shape of pupil caused by eye rotation and the un-smooth boundary obtained by double thresholding, ellipse fitting is needed to get the precise and smooth boundary of pupil to guarantee the correctness and accuracy of pupil center estimation. To achieve this goal, we develop a learning algorithm to perform an iterative ellipse fitting controlled by a gradient energy function. We use the boundary obtained by double thresholding as the initial to fit an ellipse. Then, we modify the boundary pixels and have ellipse fitting again. We will repeat the modification-and-ellipsefitting procedure until we found the best result.

During the modification, we apply a function, f, to drive the boundary pixels toward dark areas by utilizing the gradient $G = \nabla I_m$ of a smoothed image $I_m = K \otimes I$, where I denotes the eye image and K is a Gaussian convolution kernel. The gradient in horizontal and vertical direction is estimated with Sobel masks. The moving direction of a boundary pixel is decided by f_1 which can be expressed as:

$$f_1 = -G(x_i, y_i) \qquad (2)$$

The moving value of the boundary pixel is decided by f_2 which can be expressed as:

$$f_2 = (\nabla |G|)((x_i, y_i) \qquad (3)$$

Once the force has been applied to all selected boundary pixels, a new location of pupil boundary is set. The pupil boundary needs to be re-estimated by ellipse fitting. The process will be iteratively repeated until the following condition is satisfied. (Minimum energy is achieved.)

$$E = \min(\sum_{i=1}^{N} f_{i,2}) \quad (4)$$

where N is the size of the selected boundary points for ellipse fitting.

During ellipse fitting, we first select some boundary points from the pupil boundary estimated in the previous step. If this is the first iteration, the sampling points are selected from the boundary identified by double thresholding. Assume the general ellipse equation is given in the following equation:

$$ax^2 + bxy + cy^2 + dx + ey = r \quad (5)$$

where x, y is the coordinate values of on pixel. Its simple form can be expressed as:

$$H\Psi = 1 \quad (6)$$

where Ψ is the vector of the parameters to be estimated $\Psi = (\frac{a}{r}, \frac{b}{r}, \frac{c}{r}, \frac{d}{r}, \frac{e}{r})'$, H is the vector of (x^2, xy, y^2, x, y).

We put the coordinate values of all selected points (size=n) into the above function and obtain the solution of Ψ by the following equation solved via SVD method [6]:

$$\Psi = (H'H)^{-1}H'Z \quad (7)$$

where Z is the n x 1 vector $(1,1,...,1)'$. Note that if b is not equal to zero, the ellipse has an orientation to the camera coordinate system. If we want to get the long axis value and the short axis value of the ellipse, conic representation of the ellipse is needed. By using the square completion method [7], we can build the following standard representation from the conic representation.

$$\frac{(x-x_0)^2}{r_a^2} + \frac{(y-y_0)^2}{r_b^2} = 1 \quad (8)$$

From the standard representation, we can easily get all coefficients of the ellipse. If $r_a > r_b$, r_a is the length of the long axis, otherwise, r_b is the length of the long axis. Fig. 3 gives three examples on pupil detection and measurement by using the proposed energy-controlled iterative ellipse fitting method. From these pupils with different rotation, we can see that the three "white" ellipses circle these pupils very well and the pupil's boundary can be identified precisely with these fitted ellipses.

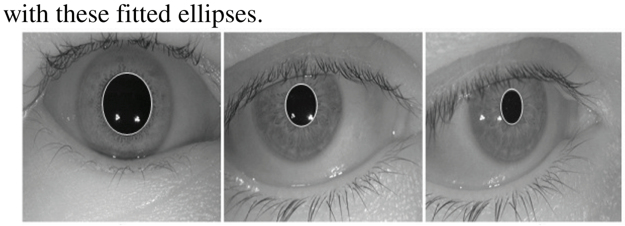


Figure 3 Pupil detection and measurement

4. 3D MODEL AND EYE-GAZED FOV ESTIMATION

To identify eye-gazed FOV with the result of eye tracking, we need to build up a 3D eye model (e.g. Arizona Eye model) first to estimate the gaze direction. In the model, eye optics can be simplified by suppressing the conic coefficients of the surfaces [8]. The iris, a pigmented diaphragm with a centered hole, known as pupil, is the aperture stop of the eye optics. The pupil diameter varies from 2 to 8mm in response to illumination levels as well as psychological factors. Although the eye is not rotationally symmetric in a strict sense, a line passing through the centers of curvature of the optical surfaces in a least square is assumed to be the optical axis of the eye. The axis passes through the rotation center, E, of the eyeball. The radius of the eyeball, ranging from 12 to 13mm, is assumed to be 12.5mm. Eye movement can be thought as a ball rotation centered at E. Gaze direction is overlapped with the visual axis (Zeye) of the eye ball and passes the pupil center as shown in Fig. 4. Therefore, gaze direction can be estimated once the 3D position of the pupil center is identified on the surface of the eye ball.

For a standard perspective camera, let Z-axis of the camera reference align with the optical axis of the imaging optics which passes the eye rotation center. All 3D points are expressed in 4D homogeneous coordinates in all reference frames as $q=[\tilde{x}, \tilde{y}, \tilde{z}, \tilde{w}]^T$. Given a 3D point in a camera reference, its projection in an imaging sensor can be obtained through the following equation:

$$\tilde{q} = M_{proj} \times q = \begin{bmatrix} 1/s_x & 0 & u_0/f & 0 \\ 0 & 1/s_y & v_0/f & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1/f & 0 \end{bmatrix} \times q \quad (9)$$

where f(mm) is the effective focal distance of the imaging sensor, s_x and s_y measured in mm/pixel are scale factors of the imaging sensor in horizontal and vertical directions. u_0 (pixels) and v_0 (pixels) are the horizontal and vertical offsets of the optical axis from the sensor origin which is the bottom-left pixel of the sensor. Point (u_0, v_0) corresponds to the intersection of the eye axis at zero-degree of eye rotation as shown in Fig. 4. Given a CCD sensor measured by WxH (mm) with mxn pixels, $s_x = W/m$, $s_y = H/n$. The actual 2D pixel coordinates of the projection point, $p = [u_{\tilde{q}}, v_{\tilde{q}}]^T$, can be otained by:

$$p = [\tilde{x}_{\tilde{q}} / \tilde{w}_{\tilde{q}}, \tilde{y}_{\tilde{q}} / \tilde{w}_{\tilde{q}}]^T \quad (10)$$

Since the camera positioned in a way so that its Z-axis is perpendicular to the forehead and considered with the direction when the eye looks straightforward, the eye rotations can be measured by the angles between the z-axes of the eye and camera references, which can be decomposed into rotations of α deg around the horizontal X-axis and β deg around the vertical Y-axis. The transformation for a point from camera reference to eye reference can be expressed as:

$$T = T_z(-t)R_y(\beta)R_x(\alpha) \quad (11)$$

where t is the distance from the camera projection center to the eye rotation center. Once the pupil center has been transformed from camera reference to eye reference, the intersection of the display screen and the visual axis which is decided by eye-ball center, E, and the pupil center can be calculated straightforward, and the eye-gazed FOV will then be identified as shown in Fig. 4.

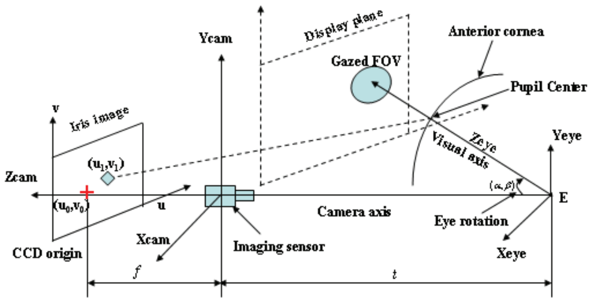


Figure 4 Eye and camera coordinate systems

5. EXPERIMENTAL RESULTS

An initial implementation of the eye tracking and eye-gazed FOV estimation was tested with 10 human subjects. In the tests, eye movements of these people within a range of $\pm 30^0$ in horizontal direction and $\pm 20^0$ in vertical direction respectively were tracked by a NTSC format CCD sensor with 640 by 480 resolution and W=4.8mm, H=3.6mm, f=12mm. Based on the tracking results, the eye-gazed FOV on the display screen is estimated thereafter. In these tests, human subjects wore an ET-HMD and gazed at a stimulus shown on the display screen. The position of each stimulus displayed on the screen corresponds to an eye rotation, which means people need to rotate their eye ball in a certain degrees to gaze at the stimulus. The α and β angle of eye rotation were estimated and their estimation error were calculated and added up as the estimation error of the eye movement. To investigate the performance of the proposed method, the proposed method was also compared with the image segmentation based method [9], and the ellipse fitting (EF) based method but without energy-controlled iteration. The comparison results are illustrated in Fig. 5 and Fig. 6.

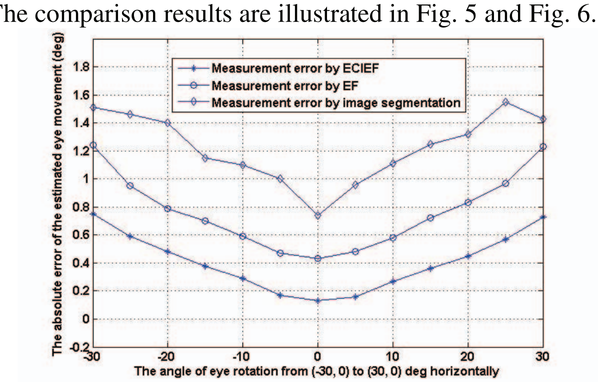


Figure 5 Eye-tracking error for horizontal movements

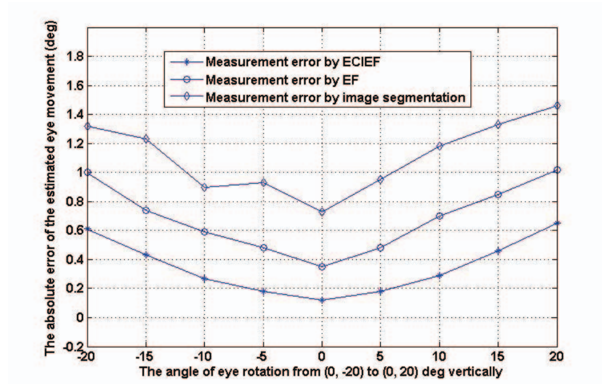


Figure 6 Eye-tracking error for vertical movements

Besides its superior performance compared to the two selected methods as shown in Fig. 5 and Fig. 6, the average angular error for eye movement estimation is less than $\pm 0.5^0$ while the average angular error of the most of conventional methods was above $\pm 1^0$. The tests on the eye movements along diagonal direction were also conducted. Similar results were obtained.

6. CONCLUSIONS

An energy-controlled iterative ellipse fitting method has been presented for pupil detection and tracking. The eye-gazed FOV can then be estimated precisely based on the pupil-tracking results and a 3D eye model. Compared with the other eye tracking methods, the average error on pupil tracking can be reduced to half (in the range of 0.3 to 0.5 degree) by using the proposed algorithm.

REFERENCES

[1] J. P. Rolland, L. Davis, and Y. Baillot, “A survey of tracking technology for virtual environments,” in *Fundamentals of Wearable Computers and Augmented Reality*. 2000.

[2] M. Hayhoe, D. Ballard, J. Triesch, and H. Shinoda, “Vision in natural and virtual environments,” *ACM Symposium on Eye Tracking Research and Applications,* pp. 7-13, 2002.

[3] C. W. Hess, R. Muri, O. Meienberg, “Recording of horizontal saccadic eye movements: methodological comparison between electro-oculography and infrared reflection oculography,” *Neuro-Ophthalmology* Vol.6, pp. 264-272, 1986.

[4] J. Rolland, H. Hua, “Head-mounted display systems,” *Encylopedia of Optical Engineering*, DOI: 10.1081, 2005.

[5] H. Hua, P. Krishnaswamy, J. Rolland, “Video-based eyetracking methods and algorithms in head-mounted displays,” *Optics Express*, Vol. 14, No. 10, pp. 4328-4350, 2006.

[6] M. E. Hochstenbach, “A Jacobi-Davidson Type SVD Method,” *SIAM Journal on Scientific Computing,* Vol. 23, No. 2, pp. 606-628, 2001.

[7] M. L. Bittinger, D. Ellenbogen, B. L. Johnson, D. J. Ellenbogen, *Elementary and Intermediate Algebra: concepts and Applications a Combined Approach*, Addison-Wesley, 2002.

[8] J. Schwiegerling, *Field Guide to Visual and Opthalmic Optics* (SPIE Press), Bellingham, WA, 2004.

[9] X. He, P-F Shi, “An efficient iris segmentation method for recognition,” *ICAPR 2005*, LNCS 3687, pp. 120-126, 2005.