

A Comparison of Pupil-Locating Algorithms in Video-based Eyetrackers

Feng Li, Jeff B. Pelz

Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology

54 Lomb Memorial Drive, Bldg 76, Rochester, NY 14623 USA

Email: {lxf7267, pelz}@cis.rit.edu

Abstract

Video-based eye tracking techniques have become increasingly attractive in many research fields, such as visual perception and human-computer interface design. Most techniques rely on the vector difference between the center of the eye's pupil and the first-surface reflection at the cornea, the corneal reflection (CR). The vector difference is well correlated to an observer's point of regard (POR). A crucial problem in this rapidly developing field is the fast, accurate detection and estimation of the pupil center. In this paper, artificial pupil images are modeled with four kinds of artifact corruptions. Then four popular pupil locating algorithms (center of gravity, circular Hough transform, ellipse fitting, and snake greedy algorithm with ellipse fitting) for isolating and measuring the pupil are implemented, and their performances are evaluated. The results will give researchers useful information for building future video-based eyetrackers.

Introduction

The problem of eye tracking has been studied by a number of researchers; its applications include broad areas, such as clinical diagnosis, eye movement study, and human-machine interface design. According to what physiological features are used, there are three general categories: electro-oculography (EOG), scleral search coil, and reflected-light based techniques. All these techniques have inherent merits and flaws. They are filled with compromises in regard to comfort, accuracy, noise, cost, ease of calibration, and suitability for a large population. Because of the low obtrusiveness to subjects, moderate accuracy and precision, and reliance on rapidly developing optical and electronic imaging devices, video-based eyetrackers have become one of the most popular and successful eye-tracking techniques.

Video-based techniques simultaneously record the images of the pupil and corneal reflection (CR) of a light source. Although the pupil and CR are both sensitive to translational eye movements, their displacement vector (P-CR) provides a signal changing primarily with rotational eye movements (Figure 1). As such, the video-based techniques are tolerant of small head movements (such as, headband slippages, camera movements, and muscular tremors), which are serious problems in some other reflected-light based techniques (e.g., limbus eyetrackers). The functional relationship[1] between the displacement vector and the point of regard (POR) is described by a simple equation:

$$POR = k \sin(\theta) \quad (1)$$

where k is the distance between the apparent pupil center and cornea center, and θ is the angular gaze direction with respect to the light source and camera.

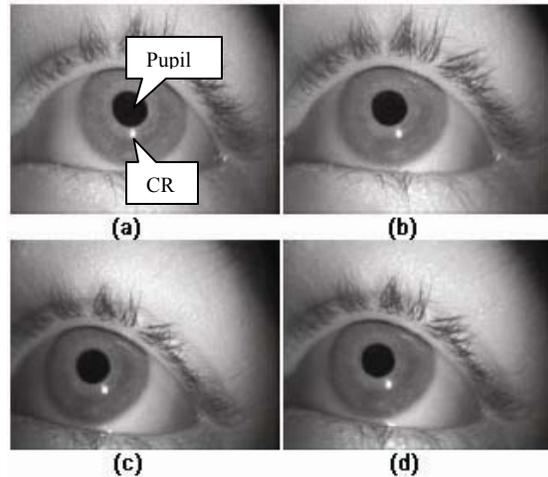


Figure 1. Eye movement and camera movement. (a) Eye image before an eye movement; (b) After the eye movement. The P-CR vector has changed. (c) Before a camera movement; (d) After the camera movement. The P-CR vector has not changed. Courtesy of Kolakowski and Pelz, 2006 [2].

In an eye image from the video-based eyetracker (Figure 1), the corneal reflection is the brightest spot, so it is usually not hard to extract by a thresholding technique. In addition, the CR is quite small (about 0.1% in Figure 1 (a)) which means the mis-locating of its center will not have as substantial an effect as the pupil for determining the P-CR. On the other hand, the pupil has a relative low contrast with its surroundings (e.g., iris) and subtends a relatively large region (about 2% in Figure 1 (a)); thus the accurate locating of its center is crucial. Even a 1mm (4 pixels in a common 17" CRT monitor) head movement parallel to the camera plane will induce an eye rotational error of about 6° (taking the distance between the apparent pupil and the rotational center of the eye as 9.5mm).

Due to the importance of locating the pupil center accurately, a lot of studies have been done to address the problem. The mainstream work can be divided into three categories: (1) area-based [3, 4]; (2) edge-based [5, 6]; and (3) Hough transform [7]. The area-based method detects the pupil region first and uses all pupil points to calculate the centroid. The edge-based method determines the pupil position by isolating the pupil-iris boundary, and the edge points are averaged or fitted by a circle or ellipse. The Hough transform (circle or ellipse) obtains the pupil parameter by transforming the edge image of the pupil into a parameter space and votes to acquire the optimal pupil center and circumference.

In this paper, the basic ideas of four pupil center locating algorithms, the center of gravity[4], ellipse fitting[6,8], circular Hough transform[7], and snake greedy algorithm[9] with ellipse

fitting, are first introduced. An artificial pupil model is then described followed by a listing of four simulated artifacts in pupil images, Gaussian noise, eye lid occlusion, CR superimposition, and uneven illumination. The performances of the algorithms when dealing with these four group images are evaluated. The paper is finalized by bringing forward conclusions and suggestions for future work.

Algorithms for Locating Pupil Center

Center of Gravity

The center of gravity of the pupil image (X_{COG}, Y_{COG}) is calculated by geometrical moments, and the equations are listed as follows:

$$X_{COG} = \frac{\sum_{i=1}^N I_i \cdot x_i}{\sum_{i=1}^N I_i}, Y_{COG} = \frac{\sum_{i=1}^N I_i \cdot y_i}{\sum_{i=1}^N I_i} \quad (2)$$

where I_i is the intensity, x_i and y_i are the coordinates of the i -pixel.

Direct Ellipse-Specific Fitting

A general conic is given:

$$F(a, p) = M^T N = ax^2 + bxy + cy^2 + dx + ey + f = 0 \quad (3)$$

where $M = [a \ b \ c \ d \ e \ f]^T$, $N = [x^2 \ xy \ y^2 \ x \ y \ 1]$. Finding a fitting ellipse equals minimizing the sum of the squared algebraic distances $\min_a \sum_{i=1}^N |F(a, p_i)|^2$ between points $p_i(x, y)$ and

the conic, given a quadratic constraint $b^2 - 4ac = a'Ca = -1$. The minimization promises to be solved by a generalized eigenvalue problem:

$$D^T D a = S a = \lambda C a \quad (4)$$

where C is a 6×6 constraint matrix and $S = D^T D$ is the scatter matrix. Its only negative eigenvalue provides the exact solution, giving an optimal fitting ellipse[8].

Circular Hough Transform

The Hough transform converts the original spatial information in an image into a parameter space representation. The parametric equation for a circle is

$$x = a + r \cos \theta, y = b + r \sin \theta \quad (5)$$

where a and b are the coordinates of circle center and r is the radius.

In the parameter space, a, b, r become three variables, and a set of cones are yielded for a train of pixels (x, y) . If all the edge pixels belong to the same circle in the spatial image, the corresponding cones in the parameter space will all share a single common intersection point. The cones become two-dimensional circles if the radius of the circle is known.

Snake Greedy Algorithm

Snakes, also called active contours, are an important class of algorithms for finding an object boundary given an initial search curvature. The curvature is represented as $v(s) = (x(s), y(s))$, having the arc length s as the parameter [9]. The process is to optimize an energy function:

$$\mathcal{E} = \int [\alpha(s)E_{cont} + \beta(s)E_{curv} + \gamma(s)E_{img}] ds \quad (6)$$

where α , β and γ are weighting parameters, and E_{cont} represents the continuity (tension) term which encourages the

curvature to shrink to a point; E_{curv} is called the curvature (stiffness) term which controls the smoothness of the curvature; E_{img} is the image force which attracts the curvature toward certain image features (edges in our implementation).

The greedy algorithm only considers closed contours, and its computation is divided into two steps during each iteration [10]:

- (1) Move each contour point to the point which has a minimum \mathcal{E} in a predefined small window;
- (2) Search for a point having a curvature maximum along the contour and set $\beta_i = 0$ at that point for the next iteration.

The greedy algorithm returns contour points whose energy functions reach local minima. Those points are then fed into the ellipse fitting algorithm to find the pupil parameters.

Pupil Model

The synthetic pupil is modeled as a circular disk [3]. Its mathematical representation is given by

$$I(x, y) = 255 \frac{255}{\left(\frac{\sqrt{(x-x_c)^2 + (y-y_c)^2}}{R} \right)^s + 1} \quad (7)$$

where (x_c, y_c) are the center coordinates of the pupil; R is the radius; and s controls the edge sharpness. A moderate s of 60 is used in all test images. The modeled pupil image is a 256 grayscale image where black pixels describe the pupil and white pixels represent the background. The pupil diameter is 50 pixels occupying about 2.5% in a 320×240 image. The images are added in "pepper and salt" noise with the density of 0.02 and then lowpass filtered by a 3×3 matrix to correlate the independent noise with the image [3]. This operation is to simulate images with shot noise from an imaging sensor. For evaluating the influence of different artifacts, four groups of pupil images are generated.

Image Corrupted by Gaussian Noise

The added-in Gaussian noise has a zero mean; its variance v changes from 0 to 0.08 by the step size of 0.02. Three sample images are shown in Figure 2.

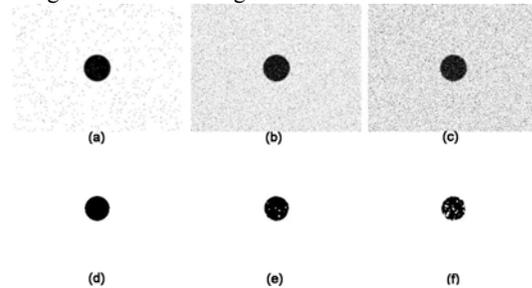


Figure 2. Pupil images with different add-in Gaussian noise v . (a) $v=0$; (b) $v=0.04$; (c) $v=0.08$; (d)(e)(f) are the corresponding thresholded images.

Image Corrupted by Eyelid Occlusion

The pupil is unavoidably occluded by the eyelid in some circumstances, especially when the subject looks to extreme positions (e.g., looks down). Pupil images in the second group are covered partially by 5%, 25%, 50% at the upper edge and 5%, 75% at the upper-right corner (Figure 3). The variance of added-in Gaussian noise is 0.04.

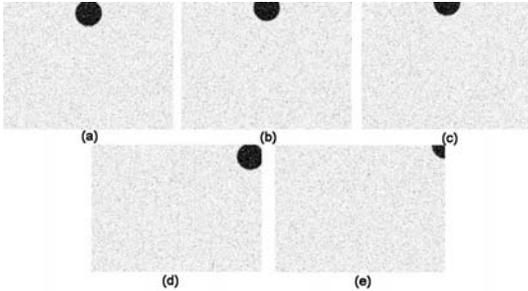


Figure 3. Pupil images with different degree occlusions. (a) 5% upper edge; (b) 25% upper edge; (c) 50% upper edge; (d) 5% corner; (e) 75% corner

Image Corrupted by the Corneal Reflection

In some approaches for locating the pupil center, the corneal reflection is removed through image processing [5, 11]. However, the removal of the CR and other specular reflections (e.g., glare on the glasses) may not be perfect. It is worth testing the algorithms when such unwanted reflections (for locating the pupil) exist. Two different sizes of the CRs, generated by the same model as the pupil, are superimposed on the images at three different locations (Figure 4). The variance of added-in Gaussian noise is 0.04.

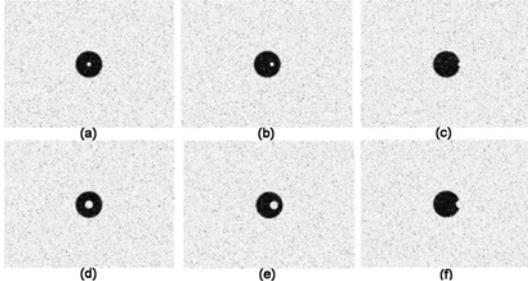


Figure 4. Pupil images with the corneal reflection at three different positions. (a)(b)(c) The diameter of the CR is 7.5 pixels; (d)(e)(f) The diameter is 15 pixels.

Image Corrupted by Uneven Illumination

The last group mimics uneven illumination in the eye image, which may make the detection of the pupil area or edges very challenging. The images are generated by using the “gradient fill” tool in Photoshop 8.0. The pixels of the circle at the begin-fill position and end-fill position are chosen as 255 and 0, respectively. The midpoint fill point is selected at ten levels between 15% and 60% with a five-unit increase in each step (Figure 5). The images are added in the “salt and pepper” noise using the same routine as that used in the other groups. No Gaussian noise is added.

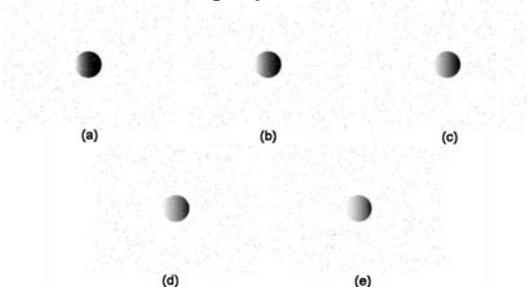


Figure 5. Pupil images with uneven illumination. The midpoint fill point is selected at (a)20%; (b)30%; (c)40%; (d)50%; (e)60%.

Results and Discussions

Fifty images for each level in every group are generated. The mean deviation between the estimated center and the real position is used to evaluate the performance. The noise is reduced by convolving the image with a 5×5 Gaussian filter (standard deviation - STD equals 2). A consistent threshold is then applied to the filtered images followed by a Sobel edge detector if necessary (e.g., for ellipse fitting). The images with heavy noise (Figure 2 (f)) may not be thresholded perfectly so that it mimics a real situation. The Hough transform is given a known radius of 40. The images in testing the snake greedy algorithm are filtered by a STD 20 Gaussian filter so that the resultant fat edges are taken as initial snake points. The final snake points after five loops are fed into the ellipse fitting method.

The algorithms, except for the Hough transform, give a consistent performance when dealing with different levels of Gaussian noise, while the snake greedy algorithm with the ellipse fitting provides the fewest errors (Figure 6). The mean computation time for each image for every algorithm is 0.18s, 0.2s, 0.19s, and 1.63s, respectively (using MATLAB 7 in a Pentium 4 1.6GHz, 1GB memory laptop). Note the Hough transform is given a known radius that reduces its computation burden dramatically.

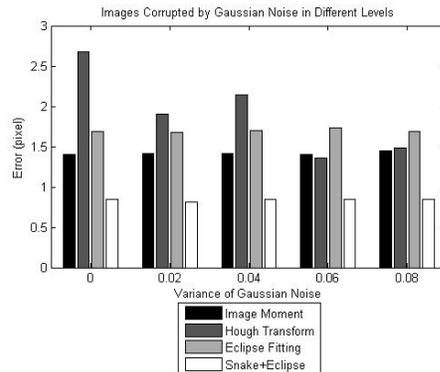


Figure 6. Algorithm performances at different noise levels

For all algorithms, the errors increase when the occlusion levels increase (Figure 7). At low occlusion (5%), they give close results though the Hough transform slightly outperforms the others. All methods fail to provide good measures when heavy cover happens (75%). As such, positioning the eye camera carefully to make it capable of capturing the whole pupil area, even when the eye moves to extreme positions, is critical. The error of image moment increases dramatically when the occlusion levels increase.

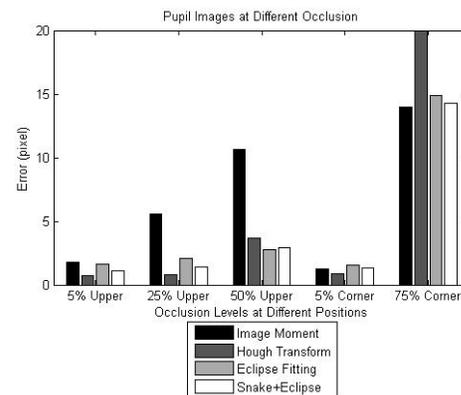


Figure 7. Algorithm performances at different occlusions

Based on Figure 8, the error of the image-moment method also increases significantly with the CR size increases, except when the CR is centered. Those of the ellipse fitting and snake greedy algorithm with the ellipse fitting keep consistent. The snake greedy algorithm with the ellipse fitting is preferable when the CR is centered and at the boundary, but its superiority declines when the CR falls between the pupil center and the boundary.

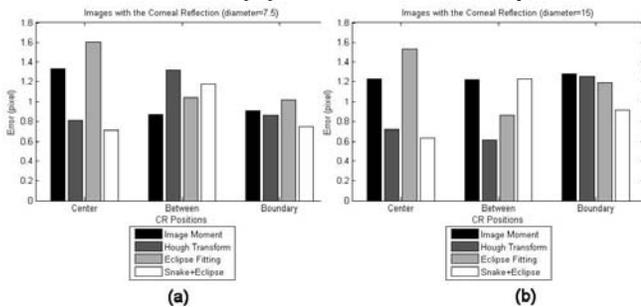


Figure 8. Algorithm performances when dealing with the corneal reflection. (a) The diameter of the CR is 7.5; (b) The diameter of the CR is 15.

Obviously, the Hough transform excels the others in this trial (Figure 9). The method does a good job when the eye image is unevenly illuminated; its performance has not been worse when the lighting uniformity deteriorates. The snake greedy algorithm with the ellipse fitting is superior to the image moment method and direct ellipse fitting, especially when the image is slightly unevenly-illuminated.

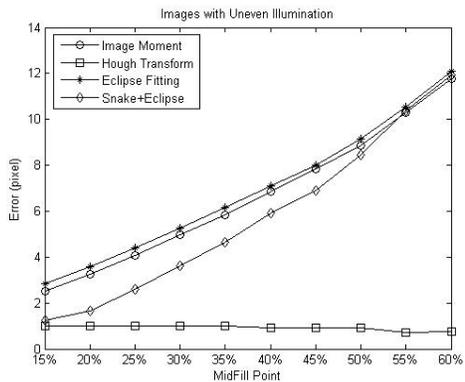


Figure 9. Algorithm performances for unevenly illuminated images.

Conclusion

Among these commonly used pupil locating algorithms, the snake greedy algorithm with the ellipse fitting has an overall best performance if the eye camera is positioned carefully and the uneven illumination is avoided. However, the method would only be a good choice when working on off-line analysis. The Hough transform gives fewer errors than the other methods when the eye images are unevenly illuminated. The image moment method has significantly increased errors when the occlusions, CR size and/or uneven illumination worsen. This method should be rejected in these situations, but may still be useful for “normal” pupil images (e.g., no artifacts like above), those which happen most frequently, because of its real-time efficiency. Another approach for locating the pupil accurately could be a mixed scheme which seeks an optimal method for one or several kinds of artifacts, while applies other methods to take care of other undesirable situations.

In this paper, the artificial pupil images are used for testing the performance of the algorithms. Future work will encompass real eye images from a video-based eyetracker. The author also suspects that the performance of the snake greedy algorithm with the ellipse fitting can be improved further if other force constraint techniques are adopted instead of the greedy algorithm. These works are left for future research.

References

- [1] Merchant, J., R. Morrissette, and J.L. Porterfield, Remote measurement of eye direction allowing subject motion over one cubic foot of space. *IEEE Trans Biomed Eng.* **21**(4): p. 309-17. (1974).
- [2] Kolakowski, S. and J. Pelz. Compensating for eye tracker camera movement. To be appear on Eye Tracking Research & Applications Symposium. San Diego, CA. (2006).
- [3] Sung, K. and M.F. Reschke, A Model-Based Approach for the Measurement of Eye Movements Using Image Processing, in Technical Report, TP-1997-3680, NASA., (1997).
- [4] Clarke, A.H., Image processing techniques for the measurement of eye movement, in *Eye Movements in Reading*, J. Ygge and G. Lennerstrand, Editors., Elsevier Science. p. 21-38 (1994).
- [5] Li, D., D. Winfield, and D.J. Parkhurst. Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches. in *Proceedings of the IEEE Vision for Human-Computer Interaction Workshop at CVPR.* (2005).
- [6] Yoo, D.H. and M.J. Chung, A novel non-intrusive eye gaze estimation using cross-ratio under large head motion. *Computer Vision and Image Understanding.* **98**(1): p. 25-51. (2005).
- [7] Clarke, A.H., et al., The Chronos eye tracker description and verification study, CHRONOS VISION GmbH, Wiesenweg 9, 12247 Berlin. www.chronos-vision.de/eyetracking.
- [8] Fitzgibbon, A., M. Pilu, and R.B. Fisher, Direct least square fitting of ellipses. *Pattern Analysis and Machine Intelligence, IEEE Transactions on.* **21**(5): p. 476-480. (1999).
- [9] Donna, J.W. and S. Mubarak, A fast algorithm for active contours and curvature estimation. *CVGIP: Image Underst.* **55**(1): p. 14-26. (1992).
- [10] Trucco, E. and A. Verri, *Introductory Techniques for 3-D Computer Vision* Prentice Hall, pp 108-113. (1998).
- [11] Mulligan, J.B. *Image Processing for Improved Eye Tracking Accuracy.* in *Behavior Research Methods, Instrumentation and Computers.* (1997).

Author Biography

Feng Li received a B.E. degree in Information Engineering and M.E. degree in Optical Engineering from Zhejiang University, China, in 2000 and 2003, respectively.

He is currently a PhD candidate in Imaging Science in the Visual Perception Laboratory, Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology, Rochester, NY, USA. His PhD thesis entails building a new image-based eye tracking system.