

# CONVOLUTION-BASED MEANS OF GRADIENT FOR FAST EYE CENTER LOCALIZATION

HAI-BIN CAI<sup>1,3</sup>, HUI YU<sup>1</sup>, CHUN-YAN YAO<sup>4</sup>, SHEN-YONG CHEN<sup>3</sup>, HONG-HAI LIU<sup>1,2</sup>

<sup>1</sup>School of Computing, University of Portsmouth, UK

<sup>2</sup>State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University, China.

<sup>3</sup>College of Computer Science, Zhejiang University of Technology, China

<sup>4</sup>College of Mechanical Engineering, Zhejiang University of Technology, China

E-MAIL: haibin.cai@port.ac.uk, hui.yu@port.ac.uk, ycy@zjut.edu.cn, sy@ieee.org, honghai.liu@icloud.com

## Abstract:

Localizing eye center is primary challenge for applications involving gaze estimation, face recognition and human machine interaction. The challenge is caused by significant variability of eye appearance in illumination, shape, color, viewing angle and dynamics, and computation related issues. In this paper, we propose a convolution-based means of gradient method to efficiently and accurately locate the eye center in low resolution images. Priority of enhancing its computation is achieved by the use of FFT transform and fewer identified pixels of circular boundary of potential eye centres. The proposed algorithm is validated in the research database platform of BioID face database. The experimental results confirm that the proposed outperforms the-state-of-art methods and its potential in real-time eye gaze tracking related applications.

## Keywords:

Eye center location; Iris center location; Convolution; Gradients

## 1. Introduction

As one of the most silent facial features, eyes can reflect human emotion and focus attention. Efficient and accurate eye centers localization has become an active research topic and opens a wide range of applications including face alignment, gaze estimation, visual focus of attention, expression analysis, human machine interaction, psychology and assistive devices for disable people. Many eye center localization methods have been proposed in the past three decades, however due to the individuality of eyes, occlusion, dynamics and illumination, fast and accurate eye center localization has been a challenging problem [1]. Although high accurate eye center location

can be obtained through high quality eye tracking systems, the intrusive or expensive devices make these existing methods unattractive [2]. In this paper we focus on non-intrusive fast eye localization methods for low resolution images without the need of specific hardware devices. The current eye center detection methods can be roughly classified into three categories, namely appearance-based, shape-based and hybrid methods.

Appearance-based methods take the whole eye image as input, regarding the eye localization problem as mapping high dimensional features to low dimensional eye positions, detect eyes using the colour distribution or filter response. Jesorsky et al. [3] propose a shape comparison approach using Hausdorff Distance to achieve robust face detection to deal with changes in illumination and background and then locate the pupil by a multi-layer perceptron trained with pupil center images. Hamouz et al. [4] represent facial features by Gabor-filter-based complex-valued statistical models, which searches the best eye center location candidate in whole given image. Niu et al. [5] bootstrap positive samples and negative samples to design two cascade classifiers respectively, and then present a 2D cascaded AdaBoost method for eye localization. Markus et al. [6] describe a non-intrusive system which only needs a camera and makes use of an ensemble of randomized regression trees for eye pupil localization.

Unlike appearance-based methods which rely on models built directly on the appearance of the eye region, shape-based methods make use of the prior geometric characteristics of the eye or surrounding structures such as limbus and eye corner, eyelids etc. Valenti et al. [7] take advantage of the isophote curvature where curves connecting points are of equal intensity to design a voting scheme for pupil localization. Asadifard et al. [8] calculate the histogram cumulative density function of the

eye region which is determined by applying the boosted cascade face detector proposed by Viola and Jones [9], followed by a minimum intensity pixels filter to locate the eye center. Timm et al. [10] propose a gradient based method utilizing the dot product of image gradient vector and displacement vector for eye center localization .

Hybrid methods try to combine the advantages of appearance-based methods and shape-based methods to achieve better performance. Turkan et al. [11] firstly highlight edges of the face region and detect candidate points for each eye through a high-pass filter, and then classify all the candidate points using a support vector machine based classifier, finally the location of each eye is estimated according to the most probable ones among the candidate points. Campadelli et al. [12] make use of properly selected Haar wavelet coefficients to train two Support Vector Machines for accurate eye detection. Valenti et al. [2] add scale invariance to their former isophote based method using a scale space pyramid and finally determine the eye location by matching the SIFT vector of each candidate within a database.

### 1.1. Our contributions

We propose an convolution-based means of gradient method for eye center localization. The method has less computation complexity than the original means of gradient method [10] and can run in real-time. Furthermore, the proposed method compares well with the current state of the art in terms of accuracy. And the experiments on the public available BioID database shows that the algorithm can also deal with different illumination changing, glass wearing situations. The rest of this paper is organized as follows: a brief introduction of the means of gradient method is presented in Section II. Section III describes the detail of the proposed convolution-based means of gradient method. Section IV presents the methodology along with the data set used in this study and the experimental results. Finally the paper is concluded with discussions in Section V.

## 2. Method

Many eye center localization methods have been proposed in the recent years. And the means of gradient method [10] has attracted much attention due to its easy implementation and high accuracy which is reported to have the best average performance for eye center localization. However the speed of the eye localization method obstructs its application in real time systems. In this paper, we propose a convolution-based means of gradient method to improve both the speed and the accuracy.

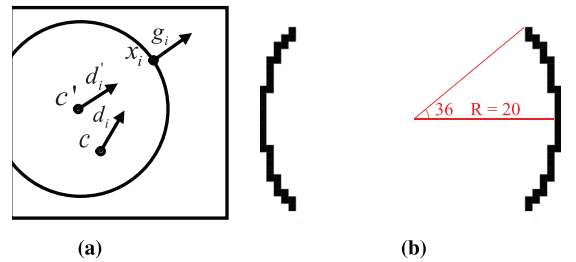
### 2.1. Means of Gradients

Timms et al. [10] propose an accurate eye center localization method by using means of gradients. The means of gradient method makes use of the relationship between a possible eye center and the vector field of all the image gradients. The eye center is found through the following equation:

$$c' = \arg \max_c \left\{ \frac{1}{N} \sum_{i=1}^N (\mathbf{d}_i^T \cdot \mathbf{g}_i)^2 \right\} \quad (1)$$

$$\mathbf{d}_i = \frac{x_i - c}{\|x_i - c\|_2} \quad \forall i : \|\mathbf{g}_i\|_2 = 1$$

where  $c'$  denote the located eye center position and  $c$  is the possible eye center. The dot product will reach the biggest if the displacement vector  $d_i$  and the gradient vector  $g_i$  have the same orientation which will happen if the point  $x_i$  lies on the boundary of the circle whose center point is  $c$  as is shown in Fig. 1a. The displacement vectors  $d_i$  and gradient vectors  $g_i$  are scaled to unit length to obtain an equal weight for all pixel positions. And  $N$  represents for the number of pixels of the image. The algorithm calculates dot products of the normalised displacement vectors and the gradient vectors for every possible eye center. And each pixel in the image is a potential eye center. The pixel that has the maximum value of mean of dot products is regarded as the final eye center.



**FIGURE 1.** (a) The example of dot products, where  $c$  is the center of the circle,  $x_i$  is one of the points on the circular boundary.  $d_i$  is the displacement vector and  $g_i$  is the gradient vector (b) The designed mask with a radius of 20 pixels

### 2.2. Convolution-Based Means of Gradients

Although the means of gradient eye localization method can locate eye center accurately, the heavy computation cost makes the algorithm less attractive for real time applications. The

computation complexity of this method is  $O(N^2)$  where  $N$  stands for the number of pixels of the eye area. The algorithm calculate the dot product of all the displacement vectors  $d_i$  and the gradient vectors  $g_i$ . Thus for a possible eye center, all the pixels in the eye image is used for the dot product. Although the computational complexity can be decreased by considering only the same orientation displacement vectors and the gradient vectors that have a significant magnitude, the accuracy will drop dramatically.

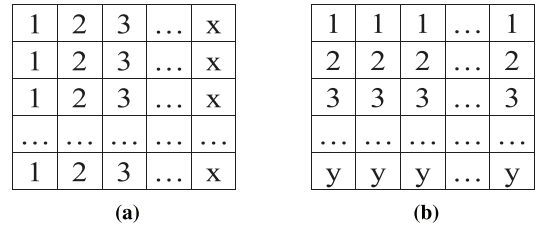
We propose a convolution-based means of gradient method which reduces the computation cost while at the same time improve the eye localization accuray. In the proposed method only the pixels on the circular boundary of a possible eye center are used to calculate the dot product. So the computation complexity is greatly reduced. And the influence of other points such as eyelids, eye corners in the dot product are also eliminated. Different sizes of masks are built to convolute the eye images. Each mask contains a circle whose center point is at the center of the masks and the pixels value on the boundary of the circle are normalised. Assuming the radius of the circle is  $r$ , then both the width and height of the built mask will be  $2r + 1$ . Fig. 1b shows a mask with a radius of 20 pixels, the black squares represent pixels with a weight and the white part of the pixels have a weight of zero. In order to deal with the obscure of the eye lids, the upper and lower part of the pixels around the circle is set to be zero and the angel of the arc is set to be 36.

We propose to apply convolution to the dot product and its equation can be further extend to the following:

$$\begin{aligned}
& \sum_{i=1}^n \mathbf{d}_i^T \cdot \mathbf{g}_i \\
&= \sum_{i=1}^n (x_{di}x_{gi} + y_{di}y_{gi}) \\
&= \sum_{i=1}^n ((x_i - x_c)x_{gi} + (y_i - y_c)y_{gi}) \\
&= \sum_{i=1}^n (x_i x_{gi}) - x_c \sum_{i=1}^n x_{gi} + \sum_i (y_i y_{gi}) - y_c \sum_i y_{gi}
\end{aligned} \tag{2}$$

where the  $d_i$  is the displacement vector and  $g_i$  is the gradient vector. And the coordinate of the  $d_i$  is  $(x_{di}, y_{di})$  which can be calculated by the difference between the circular boundary point  $x_i$  and the possible eye center  $c$ . The coordinate of the  $g_i$  is  $(x_{gi}, y_{gi})$  which can be calculated through partial derivatives or other methods for computing image gradients. In order to represent the  $x_i, x_c, y_i, y_c$ , we built two position image which is the same size as the eye region image as is shown in Fig. 3.

So the  $\sum_{i=1}^n (x_i x_{gi})$  can be calculated by firstly multiply each pixels of the position  $x$  image with the  $x$  direction gradient image and then using the former designed mask to convolute the result image. And the  $x_c \sum_{i=1}^n x_{gi}$  can be calculated by firstly convolute the designed mask with the  $x$  direction gradient image and then multiply each pixels of the result image with the position  $y$  image.



**FIGURE 2.** (a) The position image for  $x$  direction (b) The position image for  $y$  direction

Since only the pixels on the boundary of the circles are used to calculate the dot products, the other pixels of the eye image won't effect the performance. Thus we propose to directly use the sum of the dot products rather than the square of the dot products. The final position of eye center is determined by searching the maximum response of the following equation:

$$\max_{(r, x_0, y_0)} \left( \frac{1}{r} \sum_{i=1}^n \mathbf{d}_i^T \cdot \mathbf{g}_i \right) \tag{3}$$

where  $r$  is the radius of the circle within the masks.  $d_i$  and  $g_i$  are the displacement vector and the gradient vector respectively.  $(x_0, y_0)$  is the eye center coordinate. In order to locate the eye center, the proposed method searches the maximum value of equation 3 by changing radius and center points. The computation complexity is reduced because the FFT is employed in the realisation of convolution where only 2 cycles DFT and 1 cycle IDFT is performed. So the computational complexity of the convolution-based means of gradient is  $O(P \log_2(P)N)$ , where  $N$  is the number of the pixels of the eye image and  $P$  is the counts that needs to satisfy  $P \leq X + Y + C$ ,  $X, Y$  are the number of rows and columns of the center coordinate and  $C$  is a constant number.

### 3. Evaluation

In our experiments, the boosted cascade face detector proposed by Viola and Jones [9] with default parameters is used to located the face position in the image. Then the rough

eye regions are extracted through the anthropometric relations between the face and eye as stated in [2],[6],[10]. And the eye center is located in the rough eye area by applying the convolution-based means of gradient method.

### 3.1. Accuracy measurement

The accuracy measure of eye localization is calculated in normalized error which records the maximum error of both eye points. The measure was introduced by Jesorsky et al. [3] and is defined as follows:

$$e = \frac{\max(d_l, d_r)}{d} \quad (4)$$

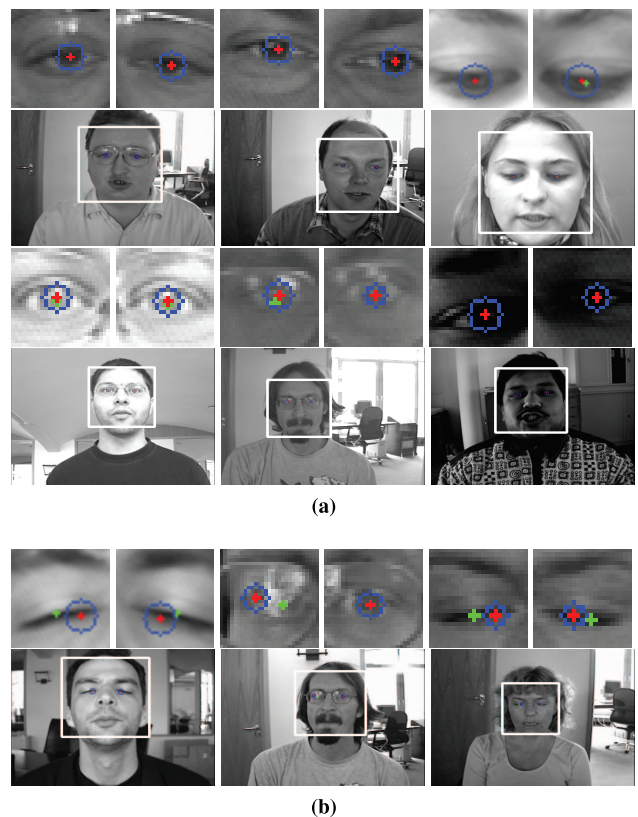
where  $d_l$  and  $d_r$  are the Euclidean distances between the detected left and right eye centers and the ones in the ground truth and  $d$  is the Euclidean distance between the left and right eyes in the ground truth. Herein, a relative error of  $e \leq 0.25$  equals a distance of half an eye width,  $e \leq 0.1$  means the diameter of iris and  $e \leq 0.05$  corresponds to the length of pupil. In order to compare the proposed method with other published methods which also employ normalized error measure, the minimum normalized error which is calculate by the minimum of  $d_l$  and  $d_r$  and the average normalized error which is obtained by the average of  $d_l$  and  $d_r$  are also calculated.

### 3.2. Results

The proposed method is validated on the BioID face database [13] which consists of 1521 grayscale images of 23 different people with a resolution of  $384 \times 288$  pixels [14]. The images in the database are head and shoulder frontal view images with a large variety of illumination, background, scale and pose. Some people in the database are wearing glasses. In some images the eyes are closed or completely hidden by reflections on the glasses. Because of these conditions, the BioID database is considered to be challenging and realistic. We discard those images where the face detector failed to find the face as in [2],[6],[10]. The number of the image with detected face is 1470.

Figs. 3a and 3b present some snapshots with accurate eye centre locations and inaccurate eye locations respectively using the convolution-based means of gradient method. The proposed method can locate the eye centre coloured by red along with the radius of iris. And the green point in the eye image is the ground truth of eye center. The proposed method is able to deal with changing illumination, shadows, squint, reflect of the light from iris and the presence of glasses. On the other hand, the proposed method failed to localize the eye center where the

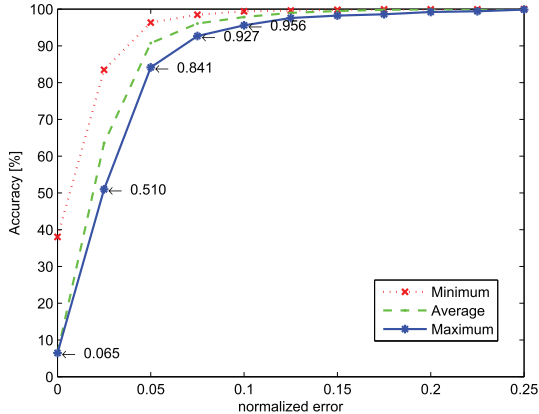
eyes are closed or strong highlights exist because of the glasses. In order to fully test the performance of the proposed method and provide comparison with other methods, we calculate the maximum normalized error, minimum normalised error and average normalized error. The minimum normalized error select the best performance of the right and left eyes, so its accuracy is higher then the maximum normalized error while the accuracy of average normalized error is between them. The accuracy performances of the proposed method is shown in Fig. 4 where the minimum, average and maximum normalised error are drawn.



**FIGURE 3.** (a) Snapshots with accurate eye center estimation. (b) Snapshots with inaccurate eye center estimation.

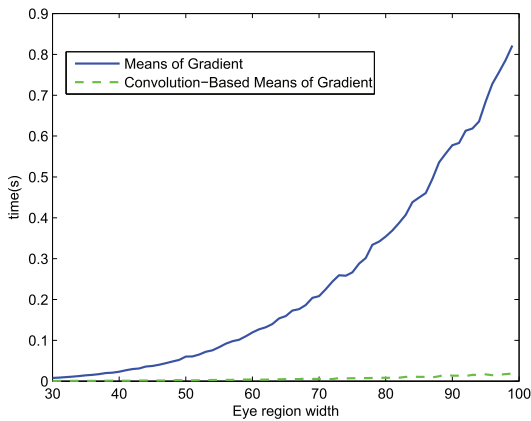
### 3.3. Comparison with the State of the Art

The proposed method has a low computational cost in comparison with the means of gradient method. In order to compare the time performance, both methods have been implemented by C++ in Mac. Both of the methods deal with the same rectangle



**FIGURE 4.** The accuracy of the convolution-based means of gradient on the BioID database. The three lines correspond to the minimum, average and maximum normalised error from the top to the bottom respectively.

area of the one eye region of the BioID database. The image is then resized to test the time performance of the two methods. Fig. 5 shows the the processing time of the two methods under different size of images. It is obvious that the proposed method cost less time in processing one image than that of the means of gradient method.



**FIGURE 5.** The processing time of the means of gradient method and convolution-based method on different size of eye images

We also compare the proposed method with other state of the art methods in terms of accuracy. It is reported that many eye center detection methods have employed the BioID database and the maximum normalised error measurement to evaluate

their performances. Table 1 presents the comparison of the proposed method with the other state of art methods using the same database and same accuracy measure. Although the highest accuracy  $e \leq 0.05$  is obtained by Markus et al. [6], their method needs to train plenty of images and the images need to be annotated which require extra works. Meanwhile, the proposed method get a relatively higher performance than most of the compared methods in terms of  $e \leq 0.10$  and  $e \leq 0.25$ . On the other hand, the proposed method does not need to be trained and can be easily combined into several existing applications without significant changes. It is evident that the proposed method outperforms most of the current eye localization methods that do not require any training process.

**TABLE 1.** Comparison of maximum normalized error in the BioID database

Method	$e \leq 0.05$	$e \leq 0.10$	$e \leq 0.25$
Bai[15]	37.0%	64.0%	96.0%
Turkan[11]	19.0%	73.7%	99.5%
Valenti[7]	84.1%	90.9%	98.5%
Valenti[2]	86.1%	91.7%	97.9%
Markus[6]	89.9%	97.1%	99.7%
Hamouz[4]	59.0%	77.0%	93.0%
Jesorsky[3]	40.0%	79.0%	91.8%
Cristinacce[16]	56.0%	96.0%	98.0%
Niu[5]	75.0%	93.0%	98.0%
Kroon[17]	65.0%	87.0%	98.8%
Asteriadis[18]	74.0%	81.7%	97.4%
Asadifard[8]	47.0%	86.0%	96.0%
Campadelli[12]	62.0%	85.2%	96.1%
Timm[10]	82.5%	93.4%	98.0%
Proposed CMoG	84.1%	95.6%	99.8%

## 4. Conclusion

We propose a convolution based means of gradient method for fast and accurate eye center location. Compared to the original means of gradient eye location method, the proposed method only calculates the gradient of the pixels on a circle which is the most important feature between the iris and sclera. Thus the impact of other less important pixels can be eliminated and the accuracy is improved. And the speed has been greatly improved because of the use of fewer pixels and FFT transform. Different kinds of masks are designed to convolute the eye image and the final position of eye center and its radius are determined by searching the maximum of the result images. Further more, the accuracy of the proposed method outperforms

most of the state of the art methods and can be easily combined into real-time eye gaze tracking related assistive applications.

## Acknowledgment

This work was supported by EU seventh framework programme under grant agreement No. 611391, Development of Robot-Enhanced Therapy for Children with Autism Spectrum Disorders (DREAM) and National Natural Science Foundation of China (No. 61325019).

## References

- [1] D. W. Hansen and Q. Ji, "In the eye of the beholder: A survey of models for eyes and gaze," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 478–500, 2010.
- [2] R. Valenti and T. Gevers, "Accurate eye center location through invariant isocentric patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 9, pp. 1785–1798, 2012.
- [3] O. Jesorsky, K. J. Kirchberg, and R. W. Frischholz, "Robust face detection using the hausdorff distance," *Proc. 3rd Int. Conf. Audio- and Video-based Biometric Person Authentication*, pp. 90–95, 2001.
- [4] M. Hamouz, J. Kittler, J.-K. Kamarainen, P. Paalanen, H. Kalviainen, and J. Matas, "Feature-based affine-invariant localization of faces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 9, pp. 1490–1495, 2005.
- [5] Z. Niu, S. Shan, S. Yan, X. Chen, and W. Gao, "2d cascaded adaboost for eye localization," *Proc. 18th Int. Conf. Pattern Recognition*, vol. 2, pp. 1216–1219, 2006.
- [6] N. Markuš, M. Frljak, I. S. Pandžić, J. Ahlberg, and R. Forchheimer, "Eye pupil localization with an ensemble of randomized trees," *Pattern Recognition*, vol. 47, no. 2, pp. 578–587, 2014.
- [7] R. Valenti and T. Gevers, "Accurate eye center location and tracking using isophote curvature," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1–8, 2008.
- [8] M. Asadifard and J. Shanbezadeh, "Automatic adaptive center of pupil detection using face detection and cdf analysis," *Proc. Int. MultiConf. of Engineers and Computer Scientists*, vol. 1, pp. 130–133, 2010.
- [9] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.
- [10] F. Timm and E. Barth, "Accurate eye centre localisation by means of gradients," *Proc. 6th Int. Conf. Computer Vision Theory and Applications*, pp. 125–130, 2011.
- [11] M. Türkan, M. Pardas, and A. E. Cetin, "Human eye localization using edge projections," *Proc. Conf. Computer Vision Theory and Applications*, pp. 410–415, 2007.
- [12] P. Campadelli, R. Lanzarotti, and G. Lipori, "Precise eye localization through a general-to-specific model definition," *Proc. British Machine Vision Conf.*, pp. 187–196, 2006.
- [13] The BioID Face Database. [Online]. Available: <http://www.bioid.com/downloads/facedb/index.php>
- [14] F. Jan, I. Usman, and S. Agha, "A non-circular iris localization algorithm using image projection function and gray level statistics," *Optik-Int. J. Light Electron Opt.*, vol. 124, no. 18, pp. 3187–3193, 2013.
- [15] L. Bai, L. Shen, and Y. Wang, "A novel eye location algorithm based on radial symmetry transform," *Proc. 18th Int. Conf. Pattern Recognition*, vol. 3, pp. 511–514, 2006.
- [16] D. Cristinacce, T. F. Cootes, and I. M. Scott, "A multi-stage approach to facial feature detection," *Proc. British Machine Vision Conf.*, pp. 1–10, 2004.
- [17] B. Kroon, A. Hanjalic, and S. M. Maas, "Eye localization for face matching: is it always useful and under what conditions?" *Proc. Int. Conf. Content-based Image and Video Retrieval*, pp. 379–388, 2008.
- [18] S. Asteriadis, N. Nikolaidis, A. Hajdu, and I. Pitas, "An eye detection algorithm using pixel to edge information," *Proc. Int. Symp. on Control, Comm. and Signal Processing*, 2006.