SDM-based Means of Gradient for Eye Center Localization

Yifan Xia[1], Jianwen Lou[1], Junyu Dong[2], Gongfa Li[3] and Hui Yu[1*]
[1]School of Creative Technologies, University of Portsmouth, Portsmouth, UK
[2]Department of Computer Science and Technology, Ocean University of China, Qingdao, China
[3]Key Laboratory of Metallurgical Equipment and Control Technology, Ministry of Education,
Wuhan University of Science and Technology, Wuhan 430081, China
Email: *hui.yu@port.ac.uk

***Abstract*—For eye gaze estimation and eye tracking, localizing eye center is a crucial requirement. This task is challenging work because of the significant variability of eye appearance in illumination, shape, color and viewing angles. In this paper, we improve the performance of means of gradient method in low resolution images, which could locate the eye center more accurately. The proposed method applies Supervised Descent Method (SDM), which has remarkable achievement in the field of face alignment, to improve the traditional means of gradient method in localizing eye center. We extensively evaluate our method on BioID database which is very challenging and realistic for eye center localization. Moreover, we have compared our method with existing state of the art methods and the results of the experiment confirm that the proposed method is an attractive alternative for eye center localization.**

Keywords—Eye Gaze Estimation; Eye Center Localization; Means of Gradient.

I. INTRODUCTION

For various computer vision areas, such as eye gaze estimation, eye tracking or other human-computer interaction applications, the localization of eye centers is one of the most important function. However, eye center localization is a tough challenge task in real life because there are some interference on visual images, such as variety illumination conditions and occlusion from other objects like hair and glasses.

Generally, according to the type of devices, two types of method can be utilized to locate the eye centers: infrared cameras and standard webcams. The first type of methods mainly apply specialized devices, such as some head-mounted devices which use infrared cameras, this method take advantage of infrared illumination to estimate the eye centers through corneal reflections. Although this technique can obtain a high accurate eye center location and is very popular in commercial areas, they are limited to some daylight applications and outdoor scenarios. Moreover, the uncomfortable and expensive devices could make this kind of method unattractive. Hence, it is very necessary to use standard webcams for eye center localization. In this paper, we mainly focus on methods about eye center localization for low resolution images by using standard webcam instead of specific hardware devices. The current approaches for eye center localization by using standard webcams could be divided into two categories: appearance-based and learning-based method.

The appearance, such as the color and circle shape of the pupil, some geometric characteristics of eye and surrounding structures, is available for appearance-based methods. Hamouz *et al*. [1] search the position of the eye center in the image by using Gabor-filter-based complex-valued statistical models. Valenti and Gevers [2] localize the eye center by using the isophote curvature. Based on [2], Valenti *et al*. [3] use a scale space pyramid and match the SIFT vector to improve isophote method. According to ace detection and CDF analysis, Asadifard and Shanbezadeh [4] propose a method which automatically detect the eye center. Means of gradient method proposed by Timm and Barth [5] is a very popular method in eye center localization. This method localizes eye centers by calculating the dot product of gradient vector and displacement vector. Based on image gradient, [6] also proposed improved method based on convolutio. The method of Skodras and Fakotakis [7] which distinguishes the eye and the skin by utilizing color is very similar to above method. Based on Hough transform, Soelistio *et al.* [8] localize the eye center points by using circle detection. Normally these methods are highly precise, however, they often lack robustness because of low resolution images, poor illumination and other challenging scenarios.

Another kind of method for localizing the eye center is learning-based methods which mainly rely on machine learning algorithms. Jesorsky *et al.* [9] use pupil center images to train a multi-layer perceptron to locate the pupil position. The method of Niu *et al.* [10] which apply 2D cascaded AdaBoost method to detect eye center. Turkan *et al*. [11] detect probable points of eye and then find the real location of eye using a high-pass filter and support vector machine classifier. Campadelli *et al*. [12] propose a method which uses properly selected Haar wavelet coefficients to train two Support Vector Machines for accurate eye detection. These above methods are more robust than appearance-based methods. However, this kind of method also have some drawbacks, like in some cases, the annotated training data is difficult to obtain.

Recent success of Supervised Descent Method (SDM) has prompted to apply to eye center localization. Therefore, in this paper, we propose a hybrid method which called SDM-based means of gradient method using both appearance-based and learning-based algorithms. This method uses SDM [13] to obtain rough eye center estimates and then optimize initial estimates by using means of gradient method. The structure of this paper is as follows. In Section II, we introduce the methodology including SDM, means of gradient and our proposed SDM-based means of gradient method. In Section III, we show experimental results on the open dataset to evaluate the performance of our proposed method and other existing methods for eye center localization. Finally, we conclude the paper in Section IV.

II. METHODOLOGY

In this section, we firstly review the basic SDM and means of gradient algorithm and then describe our improved algorithm in detail.

A. SDM Algorithm

Supervised Descent Method (SDM) [13] is designed for minimizing nonlinear least squares problems. It is widely used in face alignment. In the process of face alignment, given an image d, p landmarks are represented in this image using $d(x)$. h represents a feature extraction function which is non-linear such as SIFT features. In training stage, x_* is known which represents correct p landmarks. And (x_0) represents an average shape of lankmarks which is an initial configuration. Therefore, minimizing the following formula over Δx could represent face alignment:

$$f(x_0+\Delta x)=\left\|h\big(d(x_0+\Delta x)\big)-\varphi_*\right\|_2^2 \qquad (1)$$

where $\varphi_* = h(d(x_*))$ denotes the non-linear feature values of each manually labeled landmarks. φ_* and Δx are also known.

By using Newton's method, Eq. (1) could be minimized. The assumption of this method is that using a quadratic function in a neighborhood of the minimum can well approximate a smooth function $f(x)$. If the Hessian is positive definite, calculating a system of linear equations could obtain the minimum. The Newton updates to minimize Eq. (1) can be represented by using the following formula:

$$x_k = x_{k-1} - 2H^{-1}J_h^T(\varphi_{k-1}-\varphi_*) \qquad (2)$$

where $\varphi_{k-1} = h(d(x_{k-1}))$ denotes non-linear feature vector values of previous set of landmarks x_{k-1}. The Hessian and Jacobian at x_{k-1} are represented by using H and $\mathrm{J_h}$. However, in some cases, the non-linear feature like SIFT operator is not differentiable. It is very computationally expensive to minimize Eq. (1) using first or second order methods, which requires numerical approximations of the Jacobian and the Hessian.

In order to solve these issues, like the function of Hessian in the Newton's method, SDM is proposed which learns a sequence of descent directions and re-scaling factors. This method could produce a series of updates ($x_{k+1} = x_k + \Delta x_k$) starting from x_0 that converges to x_* in the training data. Therefore, SDM could learn a series of generic descent directions $\{R_k\}$ and bias terms $\{b_k\}$ from training data

$$x_k = x_{k-1} + R_{k-1}\varphi_{k-1} + b_{k-1} \qquad (3)$$

such that for all images of the training set the succession of x_k could converges to x_*. For more details on SDM, see [13].

B. Means of Gradient

The means of gradient method [5] localizes eye center in the image by calculating the dot product of gradient vector and displacement vector. In the circle (c represents center point and the point x_i lies on the boundary), the gradient vector $g_i$and the displacement vector d_i of the point x_i will have the same orientation and the dot product of two vectors will be the biggest (see Fig.1). Therefore, the eye center can be represented as the following formula:

$$c^* = \arg\max_c S(c) = \arg\max_c\{\tfrac{1}{N}\textstyle\sum_{i=1}^{N}(d_i^T\cdot g_i)^2\} \qquad (4)$$

$$d_i = \frac{x_i - c}{\|x_i - c\|_2} \quad \forall i: \|g_i\|_2 = 1 \qquad (5)$$

where c^* and c represent the final estimated eye center and the possible eye center respectively . d_i and g_i are scaled to unit length for an equal weight and this will improve robustness. N is the number of pixels of the image. This method localize eye center by calculating each pixels' dot products and the pixel is the final estimated eye center if the value is the biggest. For more details on Means of Gradient, see [5].

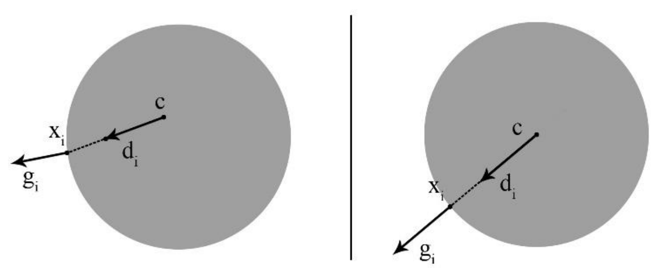


Fig. 1. The exmaple of means of gradient method

C. SDM-based Means of Gradient

Although the above Means of Gradient method could detect eye center accurately, but often suffers from the lack robustness problem in low resolution images or poor illumination. Under some conditions, the maximum value of means of dot products will lead to wrong center estimation because of local maxima. For example, the means of gradient method could obtain accurate results if only contains the eye. However, the rough eye regions which has other structures, such as hair, eyebrows, or glasses, can lead to bad results. Therefore, we propose an SDM-based means of gradient method to increase robustness and accuracy.

The success of SDM has prompted to apply to eye center localization. The originally proposed SDM algorithm is used to detect the facial landmarks. In this paper, we extend facial landmarks including eye center position and other eye region landmarks to obtain an accurate eye region and rough eye center estimate in the image. After obtaining an initial eye center estimate, we use means of gradient in the eye region of the image to optimize this initial eye center estimate to get an accurate eye center.

Our method combines the means of gradient method, but the original method is not robust enough nor are any priors for rough eye center estimate from SDM algorithm being considered. Therefore, the original formula is modified to the following one:

$$c^* = argmaxS(c) = \arg\max_c\{\frac{1}{N}\sum_{i=1}^{N}\left(d_i^T \cdot g_i\right)^2 - w_c(c_x - a_0)^2 - w_c\left(c_y - b_0\right)^2\} \qquad (6)$$

where (a_0, b_0) is the center estimate from SDM and w_c is the weight which represents grey value at possible eye center c. We regard maximum value of this formula as the final eye center, because it should have the maximum value of means of dot products and at the same time it should be close to initial center estimate to eye center. On the other hand, we apply a weight w_c for each possible eye center c, which is more easier to distinguish dark centers and bright centers because the eye center is usually more darker compared to other regions.

III. Experiments

In this section, the dataset and the evaluation metric we used in our experiment are firstly introduced; and then we display our experimental results of our proposed method; finally, we compare our proposed method with other existing methods on an open dataset.

A. Datasets And Evaluation Metric

In our experiment, we train the SDM using 1852 images selected from COFW [14] dataset. COFW dataset contain many faces which are from real-world conditions. Because differences of pose, expression, accessories and objects (e.g. sunglasses, hats, microphones, etc.), the shape and occlusions of these faces have large variations. All images have 29 manually labeled landmarks which include the eye center position.

We have used the BioID dataset [15] to evaluate the performance of our proposed method since this dataset has the most challenging images for eye center localization and many recent results are available. The dataset contains 1521 grey level images from 23 different subjects under various illumination conditions, pose and locations. The image size is 286 × 384 and we can approximately consider these images are from a low-resolution webcam because of image quality. The centers of eft and right eye are labeled and could be obtained for all images in this dataset. In some images the eyes are closed and occluded with glasses or hair. Moreover, the head of subjects sometimes affected by shadows or turn away from the camera. The BioID database is considered as one of the most challenging and realistic database because of these issues.

The evaluation metric for eye center localization is called maximum normalized error. It calculates the maximum error from the worst estimations of both eye. Jesorsky *et al.*[9] propose this kind of evaluation metric and the definition is shown as follows:

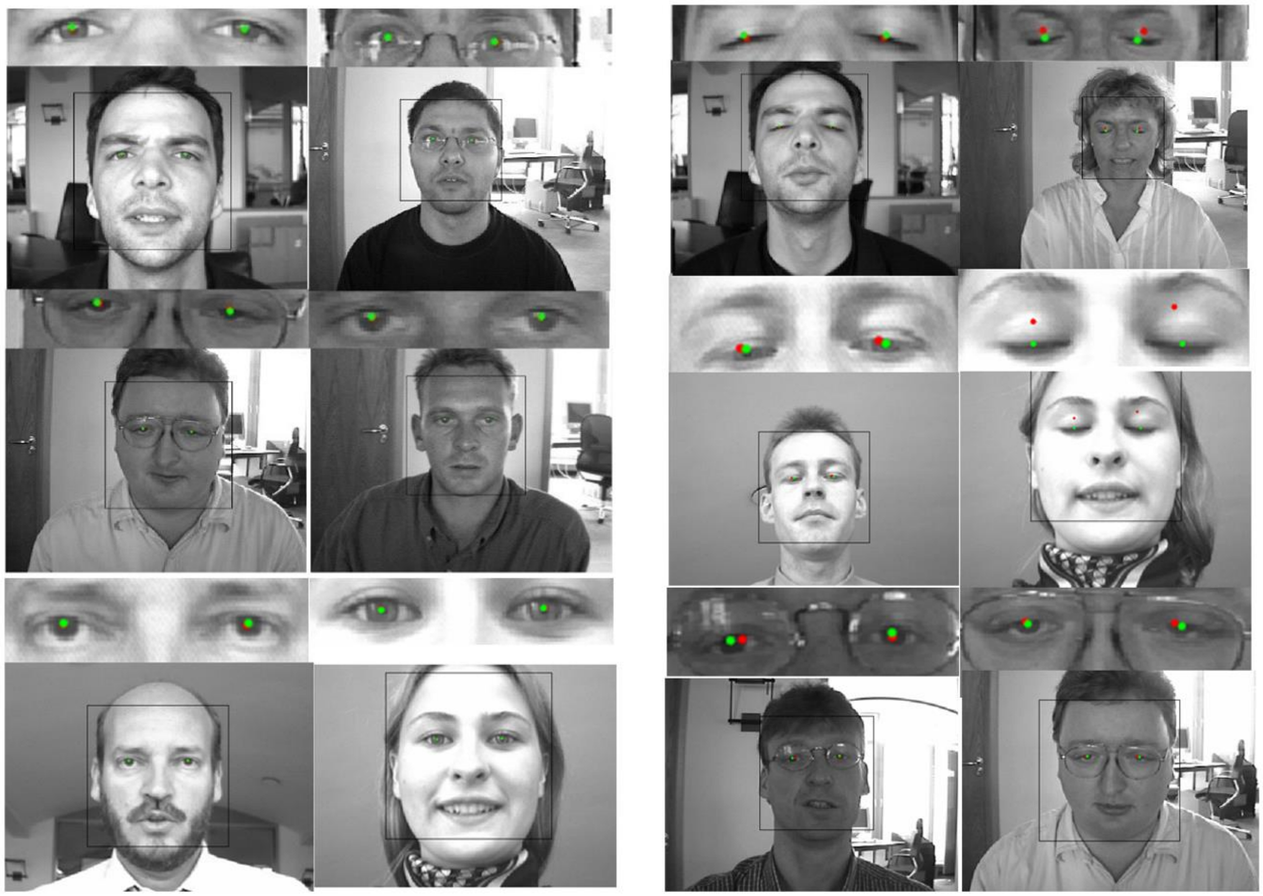
$$e \leq \frac{\max(d_l, d_r)}{d} \qquad (7)$$

where d_l and d_r represent the Euclidean distances between the estimated position and the ground truth of left and right eye centers respectively. And d is the Euclidean distance between the left and right eyes in ground truth. This meanings of evaluation metric are shown as follows: $e \leq 0.05$ equals the diameter of pupil; $e \leq 0.1$ corresponds to the diameter of iris; $e \leq 0.25$ means distance between the eye center and the eye corners. When estimate eye center, the normalized error less than or equal to 0.25 indicates that the detected eye center point might be located within the eye region. Therefore, for evaluating methods of eye center localization, we mainly focus on the performance which maximum normalized error is $e \leq 0.25$.

Fig. 2. The examples of the rough eye regions

B. Results

We evaluate our proposed method on the BioID dataset [15]. Therefore, we firstly detect the face bounding box of images for SDM initialization using the method of Viola and Jones [16] with default parameters. If the face detector fail to find the face, we discard those images. The number of the



(a) (b)

Fig. 3. (a) The example of accurate eye center estimation. (b) The example of inaccurate eye center estimation

image with detected face is 1471. And then use the facial landmarks from SDM algorithm to extract the rough eye regions (see Fig. 2). And applying the SDM-based means of gradient method to detect the eye center in these rough eye areas.

The qualitative results of the our proposed method are shown in Fig. 3. Fig. 3a and 3b display some examples of accurate and inaccurate eye locations respectively by using the SDM-based means of gradient method. The red point in the image is the eye center which we estimate, and the green point is the ground truth of eye center. According to the Fig. 3a, our approach could estimate the eye center accurately no matter the presence of glasses, shadows or hair. But according to the Fig. 3b, it indicated that the proposed method might fail to find the accurate eye center if the eyes are closed.

Fig. 4 shows the quantitative results of the our proposed method. According to the definition of the normalized error, we calculate the maximum normalized error. Our approach can reach an accuracy of 87.1%($e \leq 0.05$), 98.7%($e \leq 0.1$), and 99.9%($e \leq 0.25$)for localizing eye center. In order to further evaluate our method, we also calculate minimum normalized error using the minimum of d_l and d_r and the average normalized error using the average of d_l and d_r. The minimum normalized error is the best performance of the eye center estimation, so its accuracy is higher than the maximum normalized error while the accuracy of average normalized error is between them. All the quantitative results are shown in Fig. 4.

C. Comparison with existing approaches

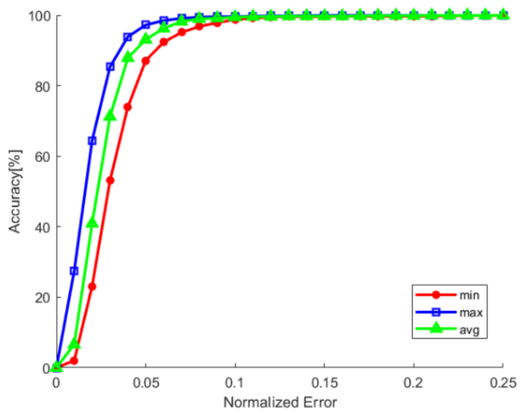


Fig. 4. Quantitative analysis of the proposed approach for the BioID database.

We also compare the proposed method with other existing methods on BioID database by using the maximum normalized error. Table. 1 shows the comparison result. According to the Table. 1, we display 11 kinds of methods about localizing eye center in order to do some comparisons with our method. Among all of the listed methods, Table 1 indicates that our method outperforms all other 11 kinds of methods when $e \leq 0.05$, $e \leq 0.10$, and $e \leq 0.25$. Therefore, our proposed method achieves a higher performance. We can conclude that our method is accurate and robust in some images of challenging scenarios which have low resolution, occlusion and poor illumination.

Table 1. Comparison of our method with other 11 kinds of methods (**Bold** value indicates best accuracy).

Method	$e \leq 0.05$	$e \leq 0.10$	$e \leq 0.25$
Asadifard and Shanbezadeh [4]	47.0%	86.0%	96.0%
Timm and Barth [5]	82.5%	93.4%	98.0%
Valenti and Gevers [2]	84.1%	90.9%	98.5%
Turkan *et al.* [11]	18.6%	73.7%	99.6%
Campadelli *et al.* [12]	62.0%	85.2%	96.1%
Niu *et al.* [10]	75.0%	93.0%	97.0%
Hamouz *et al.* [1]	58.6%	75.0%	91.0%
Jesorsky *et al.* [9]	38.0%	78.8%	91.8%
Soelistio *et al.* [8]	80.8%	95.2%	99.4%
Valenti and Gevers [3]	86.1%	91.7%	97.9%
Cai *et al.* [6]	84.1%	95.6%	99.8%
Our Method	**87.1%**	**98.7%**	**99.9%**

The proposed method outperforms a series of existing methods on localizing eye center. The performance of our proposed method is better because of its combination of SDM. SDM-based means of gradient method could simply improve the performance. The use of SDM method to obtain the estimated eye region and rough eye center point and remove most of the possible obstructions, which overcome the weakness of means of gradient method, such as some interferences from eyebrow, eyelid, and hair.

IV. Conclusion

We have proposed a SDM-based means of gradient method to localize eye center. It is a hybrid method based on both appearance-based and learning-based methods. This method uses SDM to get rough eye center estimates and then optimizes initial estimates by using means of gradient method. Compared to the original means of gradient method, the proposed method calculates the gradient of the pixels which is close to the SDM initial estimate. Thus, the impact of other pixels can be eliminated and the accuracy and robustness has been be improved. Moreover, the accuracy of the proposed method outperforms a series of the existing methods on localizing eye center and could be very easily combined into eye gaze estimation or other real-time applications.

However, the performance of our proposed method depends on rough estimated eye center point from SDM method and the presence of a visible pupil. For SDM, the performance mainly relies on the labeled training dataset. For the presence of a visible pupil, the errors have various sources: poor illumination conditions, closed eyes, strong reflection from glasses and possible obstructions, inaccurate face detection. Future work will address these limitations of our proposed method and then further improve the performance for eye center localization.

Acknowledgment

This work was supported by the EPSRC through project 4D Facial Sensing and Modelling (EP/N025849/1). This work was supported by the Royal Academy of Engineering through the project Multimodal Data-based Mental Workload and Stress Assessment for Assistive Brain-Computer Interface (NRCP1516/1/74). This work was supported in part by the Open Fund of the Key Laboratory for Metallurgical Equipment and Control of Ministry of Education in Wuhan University of Science and Technology [grant number 2017B05].

References

[1] M. Hamouz, J. Kittler, J.-K. Kamarainen, P. Paala-nen, 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

[2] 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

[3] 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

[4] M. Asadifard and J. Shanbezadeh "Automatic adaptive center of pupil detection using face detection and cdf analysis" Proc. Int. Multitconf. of Engineers and Computer Scientists, vol. 1, pp. 130-133, 2010

[5] 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

[6] Cai, Hai-Bin, et al. "Convolution-based means of gradient for fast eye center localization." Machine Learning and Cybernetics (ICMLC), 2015 International Conference on. Vol. 2. IEEE.

[7] E. Skodras, N. Fakotakis, Precise localization of eye centers in low resolution color images, Image Vis. Comput. 36 (2015) 51–60.

[8] Soelistio Y E, Postma E, Maes A. Circle-based eye center localization (CECL). Machine Vision Applications (MVA), 14th IAPR International Conference on. IEEE, 2015: 349-352.

[9] 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

[10] 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

[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] Xiong X, De la Torre F. Supervised descent method and its applications to face alignment. Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on. IEEE, 2013: 532-539.

[14] Burgos-Artizzu X P, Perona P, Dollár P. Robust face landmark estimation under occlusion. Computer Vision (ICCV), 2013 IEEE International Conference on. IEEE, 2013: 1513-1520.

[15] Bioid dataset, https://www.bioid.com/About/BioID-Face-Database.

[16] P. Viola and M. J. Jones "Robust real-time face detection" Int. J. Computer Vision, vol. 57, no. 2, pp. 137-154, 2004.