

Multimedia Tools and Applications (2020) 79:805–824
<https://doi.org/10.1007/s11042-019-08160-5>

Hybrid regression and isophote curvature for accurate eye center localization



Yifan Xia¹ · Jianwen Lou¹ · Junyu Dong² · Lin Qi² · Gongfa Li³ · Hui Yu¹ 

Received: 18 October 2018 / Revised: 15 July 2019 / Accepted: 2 September 2019 /

Published online: 16 September 2019

© The Author(s) 2019

Abstract

The eye center localization is a crucial requirement for various human-computer interaction applications such as eye gaze estimation and eye tracking. However, although significant progress has been made in the field of eye center localization in recent years, it is still very challenging for tasks under the significant variability situations caused by different illumination, shape, color and viewing angles. In this paper, we propose a hybrid regression and isophote curvature for accurate eye center localization under low resolution. The proposed method first applies the regression method, which is called Supervised Descent Method (SDM), to obtain the rough location of eye region and eye centers. SDM is robust against the appearance variations in the eye region. To make the center points more accurate, isophote curvature method is employed on the obtained eye region to obtain several candidate points of eye center. Finally, the proposed method selects several estimated eye center locations from the isophote curvature method and SDM as our candidates and a SDM-based means of gradient method further refine the candidate points. Therefore, we combine regression and isophote curvature method to achieve robustness and accuracy. In the experiment, we have extensively evaluated the proposed method on the two public databases which are very challenging and realistic for eye center localization and compared our method with existing state-of-the-art methods. The results of the experiment confirm that the proposed method outperforms the state-of-the-art methods with a significant improvement in accuracy and robustness and has less computational complexity.

Keywords Human-computer interaction · Eye tracking · Eye gaze estimation · Eye center localization

✉ Gongfa Li
ligongfa@wust.edu.cn

✉ Hui Yu
hui.yu@port.ac.uk

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, Wuhan University of Science and Technology, Wuhan, China

1 Introduction

Eye center localization is a fundamental step for various human-computer interaction tasks [4, 9, 13, 17, 24, 33, 36]. The accurate localization of eye centers has been proven to play an important role in many applications, such as eye gaze estimation, eye tracking, human attention control, driver monitoring system, sentiment analysis and other applications of computer vision. The accuracy and robustness of eye center localization significantly affects these applications' performance to some extent. Generally, according to the type of devices and sensors, the systems which locate the eye centers can be broadly classified into two categories: infrared camera-based system and standard webcam-based system. The infrared camera-based system mainly applies specialized devices, such as the infrared cameras or head-mounted devices, this kind of systems takes 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 sometimes limited to some daylight applications or outdoor scenarios. Moreover, the uncomfortable and expensive devices could make this kind of method unattractive.

Hence, there is a need for developing an alternative system for eye center localization. A standard webcam-based system for localizing the eye center can be easily implemented using the computer vision and image processing techniques. Unlike infrared camera-based system, standard webcam-based system only relies on a low-cost webcam and maintain the high accuracy and robustness meanwhile. Therefore, this paper mainly focuses on webcam-based system, which only uses low resolution images from a low-cost webcam to localize the eye center instead of specific hardware devices.

Standard webcam-based methods have achieved significant progress in the last few years, however, the localization of eye center is still a challenging task in real life due to interference, such as variety illumination conditions and occlusion from other objects like hair and glasses. The current approaches proposed in the literature for eye center localization using standard webcams could be divided into three categories: appearance-based methods, model-based methods, and hybrid methods.

Appearance-based methods use appearance information like the color and circle shape of the pupil, some geometric characteristics of eye and surrounding structures to localize eye center. Valenti and Gevers [29] proposed a method which used isophote curvature method to localize the eye center. An automatic eye center detection method based on ace detection and CDF analysis is proposed by Asadifard and Shanbezadeh [1]. Asteriadis et al. [2] proposed a method which only used appearance information like pixel and edge information to localize eyes on a face. Zhou and Geng [37] defined the generalized image projection function to detect the region and boundary of eyes. Leo et al. [22] proposed an approach to detect eye center which using analysis of image intensities and the local variability of the appearance. Bai et al. [3] proposed a localization algorithm which applied a low level, context free generalized symmetry transform. Timm and Barth [27] proposed a popular method for eye center localization called Means of gradient which calculated the dot product of gradient vector and displacement vector to localize the eye center. Based on [8, 27] proposed an improved method based on convolution. The method of Soelistio et al. [26] which localized the position of eye center utilizing Hough transform and circle detection. Although these methods are accurate, however, they are not robust enough under some situations like low resolution images, poor illumination and other challenging scenarios.

Model-based methods also called learning-based methods which mainly employ some machine learning algorithms. This kind of methods first train a model about eye appearance or structures using some images. And then use a set of features of images to fit the learned

model to estimate eye centers. A Gabor-filter-based complex-valued statistical model was utilized by Hamouz et al. [16] which used to predict the position of eye center. Kim et al. [19] localized eyes using a multi-scale approach based on Gabor vectors. Jesorsky et al. [18] trained a multi-layer perceptron to locate the position of eye center using pupil center images. A 2D cascaded AdaBoost method applied by Niu et al. [25] to detect eye centers. Kroon et al. [20] filtered the face image using a Fisher Linear Discriminant and then select the highest responses as eye center. Chen et al. [11] finished this task by using a hierarchical floatBoost and MLP classifier. Cristinacce et al. [12] used a coarse-to-fine method which use Active Appearance Model (AAM) to find the eye centers. Behnke [5] trained a hierarchical neural network with local recurrent connectivity to localize the eye center. Gou et al. [15] proposed a method which learn cascade regression models from synthetic photorealistic data to predict the eye center. The model-based methods are more robust than appearance-based methods. However, this kind of methods needs a large volume of annotated training data, which is difficult to obtain in many cases.

Hybrid methods integrate the advantages of appearance-based and model-based method in one system. Turkan et al. [28] first obtained some probable positions of eye center and then use a high-pass filter and support vector machine classifier for these probable positions to determine the final position of eye center. The method of Campadelli et al. [10] trained two Support Vector Machines to locate the eye center by utilizing properly selected Haar wavelet coefficients. Based on [29], Valenti et al. [30] used a hybrid method to improve isophote method which used mean shift and machine learning algorithm to overcome occlusions problems from the eyelids under certain lighting conditions. Although this method is accurate and efficient, it may predict the position which is far from the real eye center like eyebrow or eye eyelids. In order to further improve accuracy and robustness, in this paper, we propose a hybrid method based on regression and isophote curvature method. The proposed method first applies the regression method called Supervised Descent Method (SDM) [34] to obtain the rough location of eye region and eye centers. After that the isophote curvature method [29] is employed on the obtained eye region to obtain several candidate points of eye center. Finally, the proposed method selects several estimated eye center locations from the isophote curvature method and SDM as our candidates and a SDM-based means of gradient method, which employs the means of gradient method [27] further calculates the refined location of the eye centers from these candidates under the constraint of the SDM rough estimation. This kind of methods have two advantages: (1) based on regression method, estimated eye centers and eye region are used to make the results more robust; (2) the final refined eye center is determined by the results of regression and isophote curvature method, which means we take the results of regression as a constraint and retain the accuracy achieved by isophote curvature method. The contributions of this paper are the following:

- A novel hybrid method based on regression and isophote curvature method regression is proposed for eye center localization, which well balances the performance and the computational costs.
- To improve of accuracy and robustness in eye center localization, the proposed algorithm employs a two stage coarse to fine strategy which first roughly select some candidates of eye center location and then filter them to find real eye center points.

The structure of this paper is as follows. Section 2 describes the methodology about our proposed hybrid regression and isophote curvature method. Section 3 shows results of experiment on the open dataset to evaluate the performance of our proposed method and other existing methods for eye center localization. Finally, the general discussion and conclusion are given in Sections 4 and 5 respectively.

2 Methodology

This section mainly describes the proposed method for eye center localization in detail. Figure 1 shows the brief flowchart of our proposed method. The proposed method firstly detects the face bounding box of test images for SDM initialization using the human face detection method. And then use a regression method proposed called supervised descent method (SDM) [34] to estimate the rough location of eye region and eye centers. The result of estimation can be used as a constraint to improve the accuracy and robustness of whole hybrid method. This method has low computation and memory cost but can achieve high accuracy and robustness in the field of face alignment. Since the estimation result of eye center from the regression method is not accurate enough in some cases, the proposed method also integrates isophote curvature [29], the accuracy and robustness of them has been proven in previous research. The proposed method selects several estimated eye center locations from the isophote curvature method and SDM as our candidates. And finally uses a SDM-based means of gradient method which integrates means of gradient and the constraint of the SDM rough estimation to decide the final eye center location from these candidates. The following part will explain the details of each part of the hybrid method involved in the whole process.

2.1 Regression method

The first step of the proposed method is to use a regression method called supervised descent method (SDM) [34] to estimate the rough location of eye region and eye centers, which improve the robustness of the proposed hybrid method. The success of the regression method in face alignment has prompted to apply to eye center localization. In the process of face alignment, the proposed method extends the number of facial landmarks of training

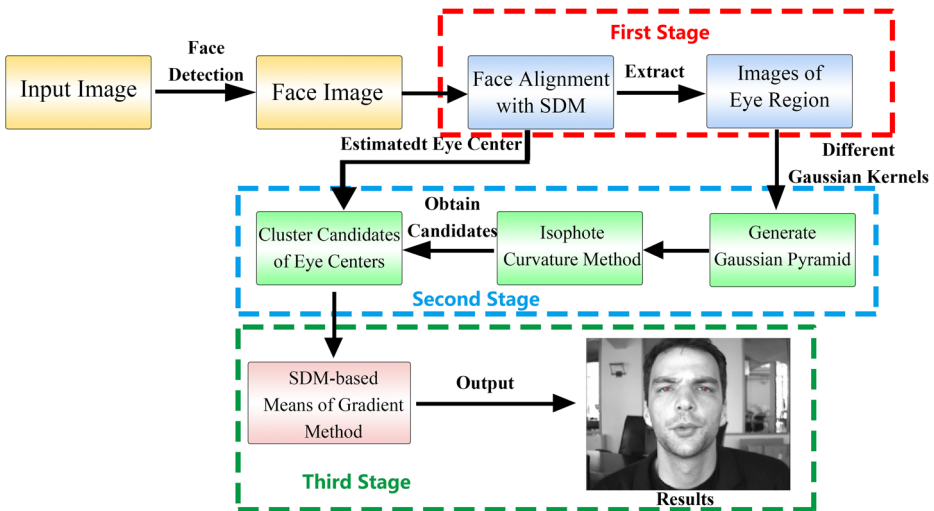


Fig. 1 Flowchart of our proposed method for eye center localization. First, we use face detection algorithm to obtain the face region, and then use SDM on face region to obtain the eye region and estimated eye center point. After that using isophote curvature method to obtain several eye center candidates from eye region images of Gaussian pyramid which generated using different Gaussian kernels. Finally, we use a SDM-based means of gradient to determine the final eye center position

data which includes the eye center and landmarks around the eye center. Therefore, the proposed method not only obtains facial landmarks including eye center but also obtains the eye region according to the landmarks around eyes. The details of SDM are as follows.

In the process of face alignment, $d(x)$ indexes p landmarks of the image d and h is a non-linear feature extraction function (such as SIFT features [23]). During training, x_* represents known and correct p landmarks of the image. And (x_0) is an initial configuration of the landmarks which corresponds to an average shape. Therefore, a formula can be used to indicate the face alignment process which minimizes the following function over Δx :

$$f(x_0 + \Delta x) = \|h(d(x_0 + \Delta x)) - \varphi_*\|_2^2 \tag{1}$$

where $\varphi_* = h(d(x_*))$ represents the SIFT feature values in the manually labeled landmarks. And φ_* and Δx are known during training.

Equation (1) could be minimized by using Newton’s method. Newton’s method assumes 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, the minimum can be found by solving a system of linear equations. The Newton updates to minimize (1) would be:

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

where $\varphi_{k-1} = h(d(x_{k-1}))$ is the feature vector extracted at the previous set of landmark locations, x_{k-1} , H and J_h are the Hessian and Jacobian evaluated at x_{k-1} . However, the SIFT operator is not differentiable and minimizing (1) using first or second order methods requires numerical approximations of the Jacobian and the Hessian, which are very computationally expensive.

SDM addresses the problems of the SIFT operator is not differentiable and reduces computational complexity by learning a series of descent directions and re-scaling factors (done by the Hessian in the case of Newton’s method) such that it produces a sequence of updates ($x_{k+1} = x_k + \Delta x_k$) starting from x_0 that converges to x_* in the training data. That is, SDM learns from training data a sequence of generic descent directions $\{R_k\}$ and bias terms $\{b_k\}$

$$x_k = x_{k-1} + R_{k-1}\varphi_{k-1} + b_{k-1} \tag{3}$$

such that the succession of x_k converges to x_* for all images in the training set. For more details on SDM, see [34]. Therefore, by using SDM, the position of eye center c_e and other landmarks around the eye centers can be obtained. And the eye regions R_e of the image can also be determined. Figure 2 shows the examples of the rough estimated eye centers and eye regions using SDM.

2.2 Isophote curvature method

The second step of the proposed method is to get some estimated candidates on the estimated eye region R_e by using an appearance-based method. According to the prior knowledge about eye, we find that it has a lot of difference with other part of the face because of its special appearance like the color and shape. Therefore, the appearance-based methods are very popular for eye center localization, which could maintain high accuracy. Among these methods, the feature of the shape of pupil, which is considered to be circular, is often as the basis of these methods. Therefore, in our work, the proposed method also uses this feature to realize appearance-based method. And chooses isophote curvature [29] to get some estimated candidates, since it maintains the high accuracy and robustness in previous research and use circular feature of eye to estimate eye center. The following part will explain the details of this method and how to obtain several candidate points of eye center.



Fig. 2 The examples of the rough estimated eye centers and eye

For the isophote curvature method, the location of eye center is at the center which should be surrounded by these isophote curves. Therefore, in order to estimate the eye center, the proposed method first calculates the isophote curves k . The isophote curves can be represented as the following formula:

$$k = - \frac{L_y^2 L_{xx} - 2L_x L_{xy} L_y + L_x^2 L_{yy}}{(L_y^2 + L_x^2)^{\frac{3}{2}}} \quad (4)$$

where L_x , L_y , and L_{xx} , L_{xy} , L_{yy} are the first and second order derivatives of the luminance function $L(x, y)$ in the x and y dimensions respectively. And then the displacement vectors $\{D_x, D_y\}$ to the estimated position of the centers can also be represented as the following formula:

$$\{D_x, D_y\} = - \frac{\{L_x, L_y\} (L_y^2 + L_x^2)}{L_y^2 L_{xx} - 2L_x L_{xy} L_y + L_x^2 L_{yy}} \quad (5)$$

According to displacement vectors, each pixel of the image can be mapped into an accumulator which also called center map. And a weighted curvedness is given to represent the possibility of each pixel to be the eye center. Therefore, the following formula can be obtained:

$$curvedness = \sqrt{L_{xx}^2 + 2L_{xy}^2 L_{yy}^2} \quad (6)$$

where the estimated location of eye center is the highest response on the center map (shown in Fig. 3). More details about isophote curvature method can be found in [29].

Although the isophote curvature method could detect eye center accurately, they often suffer from low robustness and accuracy under some non-ideal cases like other appearance-based methods. For example, the eye center is estimated at the position on the eyebrow or eye eyelids which is far from the real eye center in some cases. This is because the isophote curvature method may lead to local maxima. Moreover, the method could obtain accurate results if only contains the eye. If the image of eye regions has other components such as hair, eyebrows, or glasses, it can lead to inaccurate results. There are two cues to improve isophote curvature method: (1) before using the isophote curvature method to localize the eye center position, we always smooth images of eye region using Gaussian filtering to eliminate the effect of noise, and the different choice of a Gaussian kernel can lead to different results of eye center position and accuracy of eye center estimation; (2)

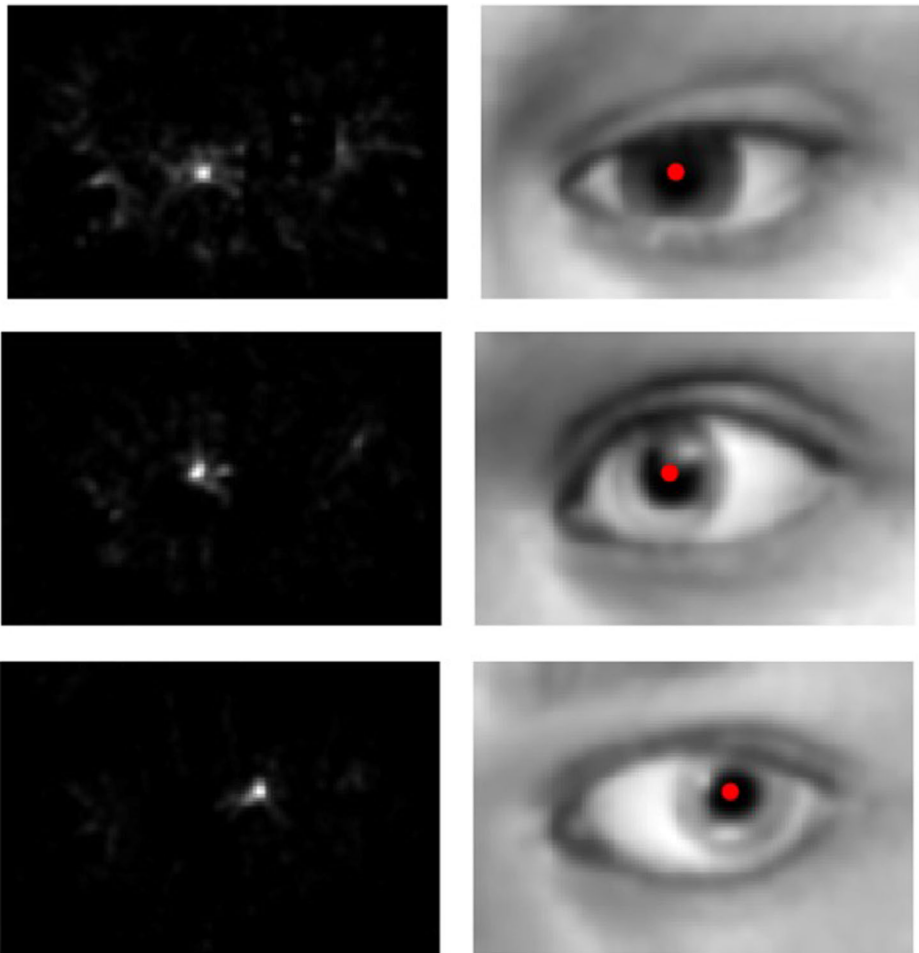


Fig. 3 The examples of isophote curvature method. The left column shows the center map, and the right is the eye center estimation

Valenti et al. [30] use multi-scale image information to improve the method. Inspired by these two cues, the proposed method in this paper applies a strategy which could obtain some candidate points to improve the possibility of obtaining the real eye center position from isophote curvature method.

In our work, the details of candidate strategy are shown as follows. As shown in Fig. 1, before using the isophote curvature method to estimate the eye center, Gaussian kernels is first employed to obtain the different images in different scales which can be regard as Gaussian pyramid. For each image, the hybrid appearance-based method is used to find the candidate points. Not only select the point which lead to maximum value in (6) as the eye center position candidate, but also the second and the third, and so on. Therefore, several candidate points for eye center localization can be obtained (shown in Fig. 4). The advantage of this strategy is that with the increase of the number of the candidate points, the possibility of including the real eye center position also increase. The expectation is that among those candidates, at least one point is the right position. Considering the computational efficiency, balancing the number of candidate points is necessary. We thus, use different Gaussian kernels to generate 5 images and select the biggest points as candidate points $\{c_1, c_2, c_3, c_4, c_5\}$. And we also regard the estimated eye center point c_e from SDM as a candidate point. Therefore, C_c can be used to represent these candidate points:

$$C_c = \{c_1, c_2, c_3, c_4, c_5, c_e\} \quad (7)$$

2.3 SDM-based means of gradient method

After obtaining six candidates of eye center C_c , a SDM-based means of gradient method is used to further refine the eye center position. This method uses the means of gradient method [27] to select these candidate points to get an accurate eye center under the constraint of the rough estimation from SDM.

The means of gradient method is similar to the isophote curvature method which is also on the basis of the circular shape of eye. It localizes eye center in the image by calculating the dot product of gradient vector and displacement vector. The theory of this method is shown in Fig. 5. If point x_i lies on the boundary of the circle, the orientation of the gradient

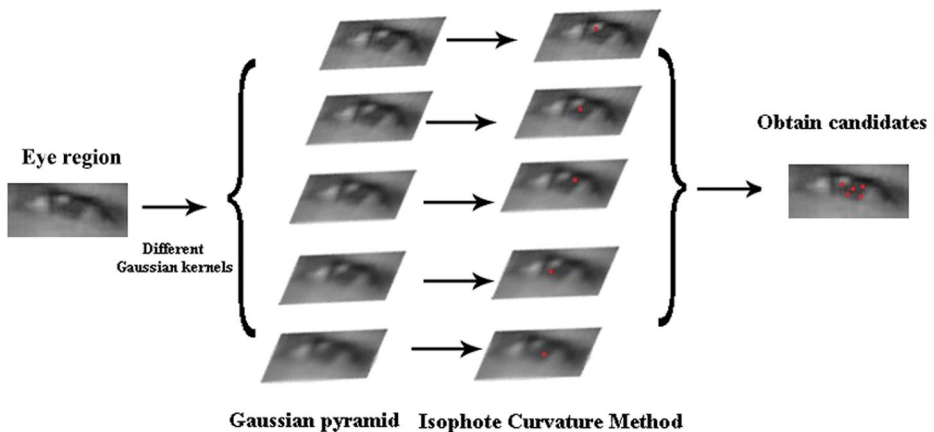


Fig. 4 The candidate strategy for the isophote curvature method

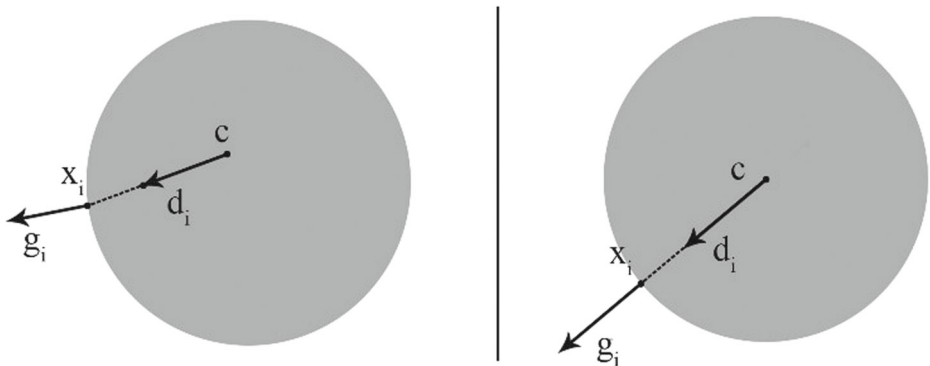


Fig. 5 The example of means of gradient method (c is center point)

vector g_i and the displacement vector d_i are regarded the same and their dot product will be the biggest. Therefore, the following formulas are used to represent the eye center:

$$S(c) = \frac{1}{N} \sum_{i=1}^N \left(d_i^T \cdot g_i \right)^2 \tag{8}$$

$$C_e = \arg \min_c S(c) \tag{9}$$

$$d_i = \frac{x_i - c}{\|x_i - c\|} \forall i : \|g_i\|_2 = 1 \tag{10}$$

where C_e and c represent the final estimated eye center and the possible eye center respectively. And d_i and g_i are scaled to the unit length for an equal weight to improve robustness. N is the number of pixels of the image. The pixel with the biggest dot product value is regarded as the final estimated eye center. For more details on Means of Gradient, see [27].

Like the isophote curvature method and other appearance-based method, the original method is not robust enough nor are any priors for rough eye center estimate from SDM algorithm being considered. Therefore, in our work, we improve the means of gradient method which integrates constraint of the rough estimation from SDM. By integrating SDM, the robustness and accuracy of the means of gradient method can be improved. The original formula of means of gradient is modified to the following one:

$$C_e = \arg \min_{c \in C_c} \left\{ S(c) - w_c (c - c_e)^2 \right\} \tag{11}$$

where c_e is the estimated eye center from SDM, w_c is the weight which represents grey value at possible eye center c which belongs to six candidates C_c . This method localizes eye center by calculating this formula and the pixel is the final eye center C_e when the value is the biggest. This is because the means of gradient method has the maximum value and at the same time it is close to estimated eye center from SDM. Moreover, because the eye center is usually darker compared to other regions, a weight w_c , namely the grey value, is utilized for each possible eye center c to determine the eye center.

3 Experimental results

This section first introduces the database and parameter settings used in our experiment. And then describe the evaluation metric of our experiment. Finally, this section displays our

experimental results of our proposed method and compares our proposed method with other existing methods on two open databases.

3.1 Database

The COFW [7] database is used as our training data for training SDM. The COFW database consists of 1852 images which mainly includes human faces under real-world conditions. The human faces of this database have major variations in terms of shape and occlusions. It is because that the pose, expression and accessories (e.g. sunglasses, hats, microphones, etc.) have big differences. The database also provides the 29 manually labeled landmarks of each image including the position of eye center.

We evaluate the performance of our proposed method and other existing methods on BioID database [6], which is the most challenging databases for eye center localization publicly available. There are many research results are available since this database is widely used in previous research for eye center localization. The database consists of 1521 grey level images, which are captured from 23 different subjects under different conditions (e.g. illumination, pose and locations, etc.). And we can approximately consider the images of this database are captured from a low-resolution webcam because image size is 286×384 . The database provides labeled left eye center and right eye center of each image. Moreover, some images of this database are captured under some challenging situations like closed eyes, occlusion from glasses or hair, affection from shadows and far away from the camera. Therefore, the BioID database is considered as one of the most challenging and realistic databases for eye center localization that are currently publicly available. Another test set is GI4E [31] contains 1236 high quality RGB images of 103 subjects with 12 different gaze directions. These images have a resolution of 800×600 px, which are similar to images acquired by a normal camera. The eye centers are also labeled in this database.

3.2 Evaluation metric

We use the evaluation metric called maximum normalized error which is proposed by Jesorsky et al. [18] to evaluate the performance of each method for eye center localization. This kind of metric calculates the maximum error from the worst estimations of both eyes and the definition is shown as follows:

$$e \leq \frac{\max(d_l, d_r)}{d} \quad (12)$$

where d_l and d_r are the Euclidean distances between the estimated left and right eye center positions and the ground truth position of left and right eye center respectively. And d represents the Euclidean distance between the ground truth of left and right eye center. According to previous research, there are some thresholds which have special meanings for eye center localization: $e \leq 0.05$ means the diameter of pupil; $e \leq 0.1$ equals the diameter of iris; $e \leq 0.25$ is distance between the eye center and the eye corners. In order to make the estimated eye center point locate in the eye region, the normalized error should be less than or equal to 0.25. Therefore, for evaluating methods of eye center localization, it mainly focuses on the performance with the maximum normalized error less than or equal to 0.25.

3.3 Results

In our experiment, the proposed method is implemented using Matlab on a laptop with an Intel Core i5 at 2.50GHz processor and 16GB of RAM memory. We use the human face detection method proposed by Viola and Jones [32] with default parameters, which is a very popular method. For SDM, we implement the method according to [34]. Training and test sets are from COFW database which is labeled using 29 landmarks including eye center position to train the model. And the eye region is determined by using 4 landmarks around the eye center (left, right, top and bottom). And for the isophote curvature method, 5 different Gaussian kernels which are (4,4), (5,5), (6,6), (7,7) and (8,8) are used to obtain 5 candidates. And we also blur center map before the biggest value using a Gaussian filter is selected. We find that the Gaussian filter with size (9,9) could lead to better results.

We evaluate our proposed method on the BioID and GI4E database. Therefore, our method firstly detects the face bounding box of images for SDM initialization using the method of Viola and Jones [32] with default parameters. If the face detector fails to find the face, those images are discard and then use our proposed hybrid regression and isophote curvature method to detect the eye center. Figure 6 shows example results of the proposed method on BioID database and GI4E database. Red points are used to represent the eye center positions and the green points indicate the ground truth positions of eye center labeled in the database. The normalized error of each eye center is also labeled in this figure. In terms of accuracy, Fig. 6 shows that the proposed method can work well and localize the eye center correctly when the the eyes are visible or no occlusion from other objects. Our method could maintain high accuracy and low normalized error. And in the robustness of our method, we mainly display the performance on the objects with glasses of the BioID and GI4E database which is shown in Fig. 7 since the challenges for eye center localization in realistic scene mainly come from the strong reflection or occlusion from glasses. It is clear that the performance of our proposed method is robust and has low normalized error even though the persons wearing glasses.

Apart from the qualitative results shown in Figs. 6 and 7, we have obtained the overall performance of the proposed approach on BioID and GI4E database using three metrics: the maximum normalized error, the minimum normalized error and the average normalized error. The minimum normalized error and the average normalized error are used to evaluate



Fig. 6 The qualitative results of our proposed approach on the BioID and GI4E database. (The red points represent the estimated eye center positions by our proposed approach. And the green points represent the ground truth. The normalized error of each eye center is also labeled in red.)



Fig. 7 The qualitative results on the objects with glasses of the BioID and GI4E database. (The red points represent the estimated eye center positions by our proposed approach. And the green points represent the ground truth. The normalized error of each eye center is also labeled in red.)

the performance to give an upper bound and an average error. The minimum normalized error and the average normalized error replace maximum function in (12) with the minimum and average function respectively. Their corresponding normalized error curves are shown in Fig. 8. For quantitative results, we mainly focus on metric of the maximum normalized error in this work.

For the whole BioID database, the maximum normalized error in Fig. 8a shows that our approach can reach an accuracy of 88.1% ($e \leq 0.05$), 98.8% ($e \leq 0.10$) and 100% ($e \leq 0.25$) for localizing eye center. And in Fig. 8b, the accuracy is 80.7% ($e \leq 0.05$), 97.7% ($e \leq 0.10$) and 100% ($e \leq 0.25$) on the objects with glasses of the BioID database. For the GI4E database, Fig. 8 also shows that our method has good performance with the accuracy of 98.5% ($e \leq 0.05$), 99.9% ($e \leq 0.10$) and 100% ($e \leq 0.25$) on the whole GI4E database and 89.6% ($e \leq 0.05$), 99.0% ($e \leq 0.10$) and 100% ($e \leq 0.25$) on the objects with glasses of the GI4E database respectively.

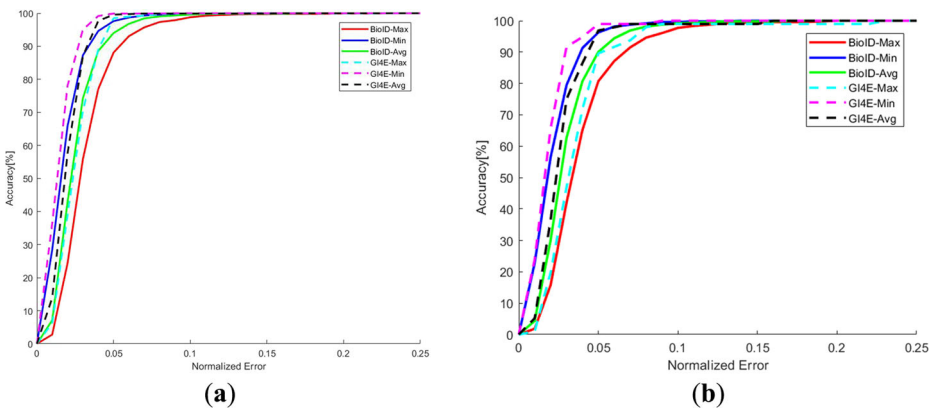


Fig. 8 **a** Quantitative analysis of the proposed approach on the whole BioID and GI4E database. **b** Quantitative analysis of the proposed approach on the objects with glasses of the BioID and GI4E database

3.4 Comparison with existing approaches

In order to further investigate the performance of our proposed method, we also extensively compare the proposed method with other existing state-of-the-art methods on BioID and GI4E database using the maximum normalized error evolution metric. The comparison results between our approach and existing state-of-the-art methods tested on the BioID and GI4E database are shown in Tables 1 and 2 respectively. From the Table 1, it shows results of 23 comparisons for eye center localization including appearance-based method, model-based method and hybrid method, which contains most of the eye center localization methods in recent years. Moreover, five kinds of thresholds e are used as the evolution metric which are widely used in previous research: $e \leq 0.05$, $e \leq 0.10$, $e \leq 0.15$, $e \leq 0.20$ and $e \leq 0.25$. From the Table 1, it shows that the proposed method does not have the best performance at $e \leq 0.05$, which ranks the second in all methods. The method of Gou et al. [15] which ranks the first at $e \leq 0.05$, but it requires a bigger dataset which is about 10730 images for training. However, our proposed method only use a smaller dataset which includes 1852 images for training. In addition, when the maximum normalized error is increased, the performance of our method will be better. And our method has more

Table 1 Comparison of our method with other methods on BioID database (bold value indicates best accuracy, * indicates second best accuracy and () indicates third best accuracy)

Method	$e \leq 0.05$	$e \leq 0.10$	$e \leq 0.15$	$e \leq 0.20$	$e \leq 0.25$
Asadifard and Shanbezadeh [1]	47.0%	86.0%	89.0%	93.0%	96.0%
Kim et al. [19]	–	(96.4%)	–	–	98.8%
Zhou and Geng [37]	–	–	–	–	94.8%
Timm and Barth [27]	82.5%	93.4%	95.2%	96.4%	98.0%
Leo et al. [22]	80.7%	87.3%	88.8%	90.9%	–
Leo et al. [21]	78.0%	86.0%	–	–	90.0%
Cai et al. [8]	84.1%	95.6%	–	–	99.8%*
Kroon et al. [20]	65.0%	87.0%	–	–	98.8%
Valenti and Gevers [29]	84.1%	90.9%	93.8%	97.0%	98.5%
Turkan et al. [28]	18.6%	73.7%	94.2%	(98.7%)	(99.6%)
Campadelli et al. [10]	62.0%	85.2%	87.6%	91.6%	96.1%
Niu et al. [25]	75.0%	93.0%	95.8%	96.4%	97.0%
Chen et al. [11]	–	89.7%	–	–	95.7%
Hamouz et al. [16]	58.6%	75.0%	80.8%	87.6%	91.0%
Cristinacce et al. [12]	57.0%	96.0%	96.5%	97.0%	97.1%
Behnke [5]	37.0%	86.0%	95.0%	97.5%	98.0%
Jesorsky et al. [18]	38.0%	78.8%	84.7%	87.2%	91.8%
Soelistio et al. [26]	80.8%	95.2%	97.8%*	98.9%*	99.4%
Bai et al. [3]	37.0%	64.0%	–	–	96.0%
Valenti and Gevers [30]	(86.1%)	91.7%	–	–	97.9%
Zhang et al. [35]	85.7%	93.7%	–	–	99.2%
George et al. [14]	85.1%	94.3%	(96.7%)	98.1%	–
Gou et al. [15]	89.2%	98.0%*	–	–	99.8%*
Our Method	88.1%*	98.8%	99.7%	99.9%	100%

Table 2 Comparison of our method with other methods on GI4E database (bold value indicates best accuracy, * indicates second best accuracy and () indicates third best accuracy)

Method	$e \leq 0.05$	$e \leq 0.10$	$e \leq 0.15$	$e \leq 0.20$	$e \leq 0.25$
Timm and Barth [27]	92.4%	96%	(96.9%)	–	98.0%
Villanueva et al. [31]	93.9%	97.3%	98%*	–	98.5%
Zhang et al. [35]	(97.9%)	(99.6%)	–	–	99.9%*
George et al. [14]	89.3%	92.3%	93.6%	94.2%*	–
Gou et al. [15]	98.2%*	99.8%*	–	–	(99.8%)
Our Method	98.5%	99.9%	99.9%	99.9%	100%

significant advantages. When it comes to $e \leq 0.10$, $e \leq 0.15$, $e \leq 0.20$ and $e \leq 0.25$, our method outperforms all other methods and achieves the most accurate results. It can conclude that under some challenging scenarios like low resolution, occlusion and poor illumination, our method can localize the eye center position more accurately and robustly.

GI4E is another evaluation database for eye center localization, which contains images with high quality taken by normal cameras. The results on GI4E are listed in Table 2. We have compared the proposed method with 5 state-of-art methods. As shown in Table 2, the performance on GI4E database is better than that on BioID in general. The proposed approach achieves the best performance on GI4E database with the accuracy of 98.5% ($e \leq 0.05$), 99.9% ($e \leq 0.10$), 99.9% ($e \leq 0.15$), 99.9% ($e \leq 0.20$) and 100% ($e \leq 0.25$) respectively.

Another important consideration in evaluating the algorithm for eye center localization is its computational complexity. The computational complexity is measured by average processing time for each input image. We have conducted a comparison in the processing time of locating the eye centers BioID database. In this case, the proposed algorithm is implemented on Matlab with a standard laptop with an Intel Core i5 at 2.50GHz processor and 16GB of RAM memory. We choose three methods from [21, 22] and [15] to compare with our method since they provide available experimental results about average processing time and we can easily compare with them. The comparison of our method and other methods in average processing time is shown in Table 3. In our experiment, the proposed method is more efficient and faster than all other state-of-the-art methods, taking 45ms per image on average. This shows that our proposed method is suitable for real time applications.

4 Discussion

In the experiment, the proposed method outperforms a series of existing state-of-the-art methods for eye center localization since it can combine the advantages of regression and isophote curvature method with a coarse to fine strategy (1) the regression method can improve the robustness and accuracy. Integrating regression method can obtain the estimated eye region and rough eye center point, which can remove most of the possible obstructions such as some interferences from eyebrow, eyelid, and hair and overcome the low robustness

Table 3 Comparison of average processing time of per image on the BioID database

Method	Leo et al. [21]	Leo et al. [22]	Gou et al. [15]	Our Method
Time(ms)	333	333	67	45

problems of isophote curvature method; (2) the whole algorithm is a coarse to fine process, the points are filtered twice by using isophote curvature method and SDM-based means of gradient method to find the real eye center point, which the impact of other points can be eliminated and the accuracy will increase.

Although the proposed hybrid regression and isophote curvature method is robust and accurate when evaluated on the testing databases, it still has some issues. First, the number of candidates is very important for the whole method. In our method, the isophote curvature method is used to generate several candidates, but the parameters are needed to tune to obtain the best performance. For each image, different Gaussian kernels are used to generate different processed images and then five candidates of eye center which is the biggest value from these processed images using isophote curvature method are selected in our experiment. Not only select the maximum value points as candidates, but also the second and the third, and so on. More candidates may lead to better performance because it covers a larger possibility. But at the same time the computational efficiency will decline with the increase of the number of candidates. It is necessary to find a balance between performance and computational efficiency. Second, the proposed method can work well when the pupils are visible, such as the open eyes, no strong reflection from glasses and no occlusion from other objects. Therefore, in some special cases, the method cannot obtain the right position of eye center. Finally, another factor that affects performance is the choice of the regression method. The performance of SDM mainly relies on the number and quality of labeled training data. The false results of face detection also lead to bad performance in SDM.

5 Conclusion

In this paper, we propose a method which is a hybrid regression and isophote curvature method for eye center localization. The proposed hybrid regression and isophote curvature method applied a coarse to fine strategy to obtain the candidates of the eye center based on the possibility.

A SDM-based means of gradient method is finally used to find the accurate eye center position from these candidates. We extensively evaluate our method and compare with the existing state-of-the-art methods on two public databases. The performance of the proposed method outperforms a series of the existing state-of-the-art methods for eye center localization and has less computational complexity. It is robust and accurate for eye center localization and can be very easily combined into eye gaze estimation, eye tracking or other real-time applications.

Future work will address these limitations of our proposed method and then further improve the performance. For example, if more constraints between two eye centers and other landmarks around the eye center are applied, we can locate the center based on another eye center point or other landmarks which may be a good solution to the problem of closed eye, strong reflection and occlusion. At the same time, augmenting more labeled training data may increase the performance of regression method. Moreover, the deep learning method is also an effective approach given sufficient training data.

Acknowledgements 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]. And this work was supported by grants of National Natural Science Foundation of China (Grant No. 51575407).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Asadifard M, Shanbezadeh J (2010) Automatic adaptive center of pupil detection using face detection and cdf analysis. In: Proceedings of the international multicongference of engineers and computer scientists, vol 1, p 3
2. Asteriadis S, Nikolaidis N, Hajdu A, Pitas I (2006) An eye detection algorithm using pixel to edge information. In: Int. symp. on control, commun. and sign. proc
3. Bai L, Shen L, Wang Y (2006) A novel eye location algorithm based on radial symmetry transform. In: 18th international conference on pattern recognition, 2006. ICPR 2006, vol 3, pp 511–514. IEEE
4. Bao H, Fang W, Guo B, Wang J (2019) Real-time wide-view eye tracking based on resolving the spatial depth. *Multimed Tools Appl* 78(11):14,633–14,655
5. Behnke S (2002) Learning face localization using hierarchical recurrent networks. In: International conference on artificial neural networks, pp 1319–1324. Springer
6. Bioid dataset. <https://www.bioid.com/About/BioID-Face-Database>
7. Burgos-Artizzu XP, Perona P, Dollár P (2013) Robust face landmark estimation under occlusion. In: Proceedings of the IEEE international conference on computer vision, pp 1513–1520
8. Cai HB, Yu H, Yao CY, Chen SY, Liu HH (2015) Convolution-based means of gradient for fast eye center localization. In: 2015 International conference on machine learning and cybernetics (ICMLC), vol 2, pp 759–764. IEEE
9. Cai H, Yu H, Zhou X, Liu H (2016) Robust gaze estimation via normalized iris center-eye corner vector. In: International conference on intelligent robotics and applications, pp 300–309. Springer
10. Campadelli P, Lanzarotti R, Lipori G (2006) Precise eye localization through a general-to-specific model definition. In: BMVC, vol 1, pp 187–196. Citeseer
11. Chen D, Tang X, Ou Z, Xi N (2006) A hierarchical floatboost and mlp classifier for mobile phone embedded eye location system. In: International symposium on neural networks, pp 20–25. Springer
12. Cristinacce D, Cootes TF, Scott IM (2004) A multi-stage approach to facial feature detection. In: BMVC, vol 1, pp 277–286
13. Cyganek B, Gruszczynski S (2014) Hybrid computer vision system for drivers' eye recognition and fatigue monitoring. *Neurocomputing* 126:78–94
14. George A, Routray A (2016) Fast and accurate algorithm for eye localisation for gaze tracking in low-resolution images. *IET Comput Vis* 10(7):660–669
15. Gou C, Wu Y, Wang K, Wang FY, Ji Q (2016) Learning-by-synthesis for accurate eye detection. In: 2016 23rd international conference on pattern recognition (ICPR), pp 3362–3367. IEEE
16. Hamouz M, Kittler J, Kamarainen JK, Paalanen P, Kalviainen H, Matas J (2005) Feature-based affine-invariant localization of faces. *IEEE Trans Pattern Anal Mach Intell* 27(9):1490–1495
17. Jang YM, Mallipeddi R, Lee S, Kwak HW, Lee M (2014) Human intention recognition based on eyeball movement pattern and pupil size variation. *Neurocomputing* 128:421–432
18. Jesorsky O, Kirchberg KJ, Frischholz RW (2001) Robust face detection using the hausdorff distance. In: International conference on audio-and video-based biometric person authentication, pp 90–95. Springer
19. Kim S, Chung ST, Jung S, Oh D, Kim J, Cho S (2007) Multi-scale gabor feature based eye localization. *World Acad Sci Eng Technol* 21:483–487
20. Kroon B, Hanjalic A, Maas SM (2008) Eye localization for face matching: is it always useful and under what conditions? In: Proceedings of the 2008 international conference on Content-based image and video retrieval, pp 379–388. ACM
21. Leo M, Cazzato D, De Marco T, Distanto C (2013) Unsupervised approach for the accurate localization of the pupils in near-frontal facial images. *J Electron Imaging* 22(3):033033
22. Leo M, Cazzato D, De Marco T, Distanto C (2014) Unsupervised eye pupil localization through differential geometry and local self-similarity matching. *PLoS one* 9(8):e102829

23. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. *Int J Comput Vis* 60(2):91–110
24. Narayanan R, Rangan VP, Gopalakrishnan U, Hariharan B (2019) Multiparty gaze preservation through perspective switching for interactive elearning environments. *Multimedia Tools and Applications* pp 1–34
25. Niu Z, Shan S, Yan S, Chen X, Gao W (2006) 2d cascaded adaboost for eye localization. In: 18th International conference on pattern recognition, 2006. ICPR 2006, vol 2, pp 1216–1219. IEEE
26. Soelistic YE, Postma E, Maes A (2015) Circle-based eye center localization (cecl). In: 2015 14th IAPR international conference on machine vision applications (MVA), pp 349–352. IEEE
27. Timm F, Barth E (2011) Accurate eye centre localisation by means of gradients. *Visapp* 11:125–130
28. Türkan M, Pardas M, Cetin AE (2007) Human eye localization using edge projections. In: *VISAPP* (1), pp 410–415
29. Valenti R, Gevers T (2008) Accurate eye center location and tracking using isophote curvature. In: IEEE Conference on computer vision and pattern recognition, 2008. CVPR 2008, pp 1–8
30. Valenti R, Gevers T (2012) Accurate eye center location through invariant isocentric patterns. *IEEE Trans Pattern Anal Mach Intell* 34(9):1785–1798
31. Villanueva A, Ponz V, Sesma-Sanchez L, Ariz M, Porta S, Cabeza R (2013) Hybrid method based on topography for robust detection of iris center and eye corners. *ACM Trans Multimed Comput Commun Appl (TOMM)* 9(4):25
32. Viola P, Jones MJ (2004) Robust real-time face detection. *Int J Comput Vis* 57(2):137–154
33. Wang W, Huang XD (2019) Virtual plate based controlling strategy of toy play for robot's communication development in ja space. *Int J Autom Comput* 16(1):93–101
34. Xiong X, De la Torre F (2013) Supervised descent method and its applications to face alignment. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 532–539
35. Zhang W, Smith ML, Smith LN, Farooq A (2016) Eye center localization and gaze gesture recognition for human–computer interaction. *JOSA A* 33(3):314–325
36. Zhang C, Yao R, Cai J (2018) Efficient eye typing with 9-direction gaze estimation. *Multimed Tools Appl* 77(15):19,679–19,696
37. Zhou ZH, Geng X (2004) Projection functions for eye detection. *Pattern Recogn* 37(5):1049–1056

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Yifan Xia received his M.Sc. degrees from Ocean University of China in 2017. He is currently pursuing the Ph.D. degree in the School of Creative Technologies in University of Portsmouth. His research interests include computer vision and eye gaze estimation.



Jianwen Lou received his M.Sc. degrees from Ocean University of China in 2016. He is currently pursuing the Ph.D. degree in the School of Creative Technologies in University of Portsmouth. His research interests include 2D facial tracking, facial animation and machine learning.



Junyu Dong received his B.Sc. and M.Sc. from the department of Applied Mathematics at Ocean University of China in 1993 and 1999 respectively and received his Ph.D. in November 2003 in Heriot-Watt University, UK. He is currently a professor and the head of the Department of Computer Science and Technology in Ocean University of China. His research interests include computer vision, underwater image processing and machine learning, with more than 10 research projects supported by NSFC, MOST and other funding agencies. He has published more than 100 journal and conference papers.



Lin Qi received his B.Sc. and M.Sc. degrees from Ocean University of China in 2005 and 2008 respectively and received his Ph.D. in computer science from Heriot-Watt University in 2012. He is currently an associate professor in the Department of Computer Science and Technology in Ocean University of China. His research interests include computer vision and visual perception.



Gongfa Li received the Ph.D. degree in Wuhan University of Science and Technology, Wuhan, China. He is currently a professor in Wuhan University of Science and Technology. His major research interests are computer aided engineering, mechanical CAD/CAE, modelling and optimal control of complex industrial process.



Hui Yu is a Professor with the University of Portsmouth, UK. His research interests include vision, computer graphics and application of machine learning to above areas, particularly in human machine interaction, image processing and recognition, Virtual/Augmented reality, 3D reconstruction, robotics and geometric processing of human/facial performances. He is Associate Editor of IEEE Transactions on Human-Machine Systems and the Neurocomputing journal.