

Robust Eye Center Localization Based on An Improved SVR Method

Zhiyong Wang¹, Haibin Cai², and Honghai Liu*¹

¹ State Key Laboratory of Mechanical System and Vibration,
Shanghai Jiao Tong University, Shanghai, China
{yzwang_sjtu, honghai.liu}@sjtu.edu.cn,
<http://bbl.sjtu.edu.cn/>

² Group of Intelligent System and Biomedical Robotics,
School of Creative Technologies, University of Portsmouth
Portsmouth, United Kingdom
{jlcaihaibin}@gmail.com

Abstract. Eye center localization is an important technique in gaze estimation, human computer interaction, virtual reality, etc., which attracts a lot of attention. Although a great deal of progress has been achieved over the past few years, the accuracy declines dramatically due to the low input image resolution, poor lighting conditions, side face, and eyes status such as closed or covered. To handle this issue, this paper proposes an improved support vector regression (SVR) method to detect the eye center based on the facial feature localization. Several image processing techniques were tried to improve the accuracy, and results showed that the SVR combining a Gaussian filter could get a better accuracy.

Keywords: Eye center localization, SVR, Gaussian filter

1 Introduction

Generally, the role of gaze in human communication is well known to be crucial due to its ability to convey intentions and describe attention. It gets many research focuses in many fields such as human computer interaction, virtual reality, etc. Accurate eye pupil center localization is the basis of the gaze estimation. Although promising progress has been achieved over the past few years, the accuracy of eye center localization declines dramatically due to the low resolution, poor lighting conditions of input image, side face, and complex eyes status such as closed or covered. Nowadays, face detection and facial feature points localization have been mature techniques. It is reasonable and implementable to extract and analyze the eye region separately, which makes the eye pupil center localization more accurate and faster.

Support vector regression (SVR), proposed by Drucker *et al.* [1] in 1997, was given a comprehensive introduction by Smola and Scholkopf [2]. Its basic principle is described as follows. For sample (\mathbf{x}, \mathbf{y}) , the traditional regression model is calculating the loss between the model output $\mathbf{f}(\mathbf{x})$ and actual output

\mathbf{y} . If and only if the output $\mathbf{f}(\mathbf{x})$ and actual output \mathbf{y} are identical, the loss is equals to zero. But in SVR, it is assumed that the deviation between $\mathbf{f}(\mathbf{x})$ and \mathbf{y} is acceptable within the threshold of ε . That is, the loss is calculated unless the absolute error between $\mathbf{f}(\mathbf{x})$ and \mathbf{y} is greater than ε .

This paper gave tentative work to improve the accuracy and robustness of the existing eye center localization methods. An improved SVR method combining with Gaussian filter and eye state judgment strategy was proposed in this paper. And the results showed that eye centers could be located accurately using our proposed method in several public data base. The basic principles of SVR and linear regression (LR) are introduced as follows.

SVR has been successfully applied in the detection of facial feature points [3], and plenty of source codes have been exposed on the OpenFace [4]. It is compute-efficient and robust against image resolution variation, illumination change, and face angle conversion.

LR is a statistical analysis method that uses regression analysis in mathematical statistics to determine the quantitative relationship between two or more variables. It is a commonly used regression method to eliminate linear error and improve the accuracy.

2 Related Works

The methods that can detect the eye center are generally divided into two categories: (1) model based method, and (2) appearance based method. In this part, some previous work are reviewed.

2.1 Model Based Methods

The model based methods, also known as feature based methods, detect the eye center using some priori knowledge of the eyes such as the dramatic change in the brightness at the edge of the iris and the geometric shape of iris. Daugman *et al.* [5] proposed an Integro-Differential Operator (IDO) method which utilized the drastic intensity change between the iris and sclera. Experimental results achieved good performance for high resolution eye images. In 2015, Cai *et al.* modified the IDO by optimizing the number of kernels and got a desirable accuracy in localizing the eye center in lower resolution image [6]. On the other hand, Hough transform, a classical method to find the circle, was also applied to find eye center as the visual part of the iris is around a circular arc [7]. Timm *et al.* [8] proposed a simple but efficient eye center localization method by calculating the means of gradient for each points in the image. Besides, the color and radial symmetry of the eyes were used to locate the eye center [9] [10]. Most of the these model-based methods performed well in dealing with the high resolution images, but the accuracy will drop when the image quality decreases such as in low resolution or poor illumination and even worse when the eyes are closed.

2.2 Appearance Based Methods

The appearance based methods focus on the eye holistic appearance and surrounding structures, which needs a lot of data to train models to detect the eye center. Thus, many machine learning algorithms have been applied to locate the eye center. Markuvs *et al.* [11] trained an ensemble of randomized regression trees for the eye center localization. Bayesian models was employed in [12] to locate the eye center. Considering the state of the eyes, Gou *et al.* [13] proposed a cascaded regression framework to detect the eye center. Although these methods need a large amount of training data for model training, a good result can be achieved even the images are in low resolution.

3 Method

This paper proposed a novel method which combined classical image processing method and machine learning algorithms to detect the eye pupil center. Two datasets, LFPW [14] and HELEN [15], were used to train the model. This model is based on the characteristics of the intensity distribution around the eyes. In order to enhance the original image quality, some strategies were used to image preprocessing methods such as Gaussian filter, histogram equalization, gradient features, etc. before training the support vector regression model and linear regression model. The specific method is shown in Fig.1.

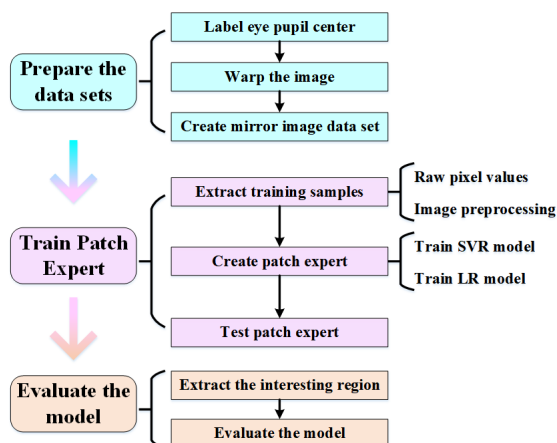


Fig. 1. The framework of the improved SVR eye center detect method.

3.1 Data Preprocessing

To train the eye localization model, we used two large image datasets LFPW and HELEN which were used to train the models detecting facial landmarks. The

LFPW training set contains 811 PNG images and test set has 224 PNG images. There are 2000 JPG images and 330 images correspondingly in the HELEN dataset. These two datasets are regarded as the most comprehensive datasets because they collect samples of different skin color, race, age, and situation in various resolutions. Besides, the original images, the 68 landmarks of face are also provided in the datasets. The distribution of landmarks are showed in Fig.2.

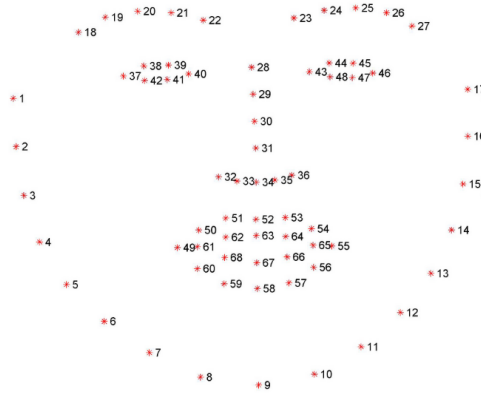


Fig. 2. Distribution of the face landmarks.

To collect the training data, the eye pupil centers of each picture in the two large datasets were labeled manually. In order to facilitate the unified processing of subsequent data, we warped all the images into 400*400 pixels, and transferred them into gray scale images. Besides, image mirroring was used to increase the training data.

3.2 Train The SVR Model

Extract The Training Samples. The samples were extracted from the input images. Both the gray scale image and the landmark data containing 68 points were read. It is easy to obtain the eye region from the face feature points distribution matrix. The feature points around the right eye sclera are related to point 37,38,39,40,41,42 (37 as the outer corner feature point, 40 as to the inner eye corner, 38 and 39 for the upper eyelid edge points, 41 and 42 of lower eyelid edge points, the specific distribution is shown in Fig3.

The hexagonal region with isoclines is our interesting region. Likewise, point 43, 44, 45, 46, 47, 48, are the related points around the sclera of the left eye. Following we will take the right eye as an example to explain this method.

The distance between the eye corners can be calculated with some horizontal coordinates of point 37 and point 40. For the convenience of processing, we

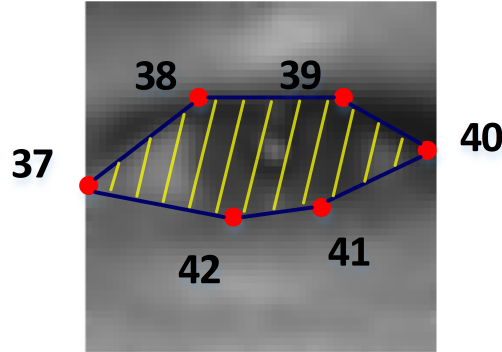


Fig. 3. Distribution of feature points around the corner of the eye.

extract a square area with a center that is the horizontal midpoint of point 37 and point 40. The width of the square equals the distance of eye corners. This eye area is usually 27×27 pixels. In order to better extract the characteristics of the iris of human eyes in different states, bilinear interpolation was used for the interest area, which was enlarged to 51×51 pixels. The enlarged iris radius is about 10 to 12 pixels. To make the template contain the iris completely, 25×25 pixels is taken as the template size, that is, the samples dimension is 625. In this way, the area of interest of 51×51 pixels can be sampled according to the template size. Total 27×27 samples, also calling patches, can be extracted. As for the labels, if we take this problem as a classification problem, the true eye pupil center will be the only one positive sample. In this case, the training will fail. Thus the Gaussian distribution was used to generate the labels.

The center of the Gaussian distribution, $(0, 0)$, is the actual center of the iris localization coordinates, and the variance equals 1. The responses of the Gaussian distribution were set to be the labels of patches correspondingly. The patch center is farther away from the center of the iris, the corresponding response is close to 0. Inversely, the patch center is closer to the iris center, the label will be closer to 1. Thus a binary classification problem was turned into a fuzzy problem. In this way, the problem of sample imbalance between positive and negative samples can be solved.

Because the actual coordinates of the six feature points around the eye have been obtained, the iris center can only be located within the hexagon. The hexagon taking the six feature points as the vertex were used to narrow the sampling range. Only the labels of which coordinates were within the hexagon were retained. Meanwhile, all the outer ones were set to 0. Thus, each patch got a label, and the patches with a label of 0 will be discarded to further improve the pertinence of the sample. After completing this step, we got sample data for training the model.

There were two kinds of input feature used to train the model: the raw pixel values and the gradient intensity values calculated by convolution with

Soble operator. In order to simplify the calculation and improve the speed, the absolute value addition strategy is used to approximate the gradient. The specific calculation method is showed as followings:

$$\left\{ \begin{array}{l} G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * A \\ G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} * A \\ G = \sqrt{G_x^2 + G_y^2} \\ G \approx |G_x + G_y| \end{array} \right. \quad (1)$$

Meanwhile, this paper tried two classical image processing methods, Gaussian blur and histogram balance, to improve the accuracy. Gaussian filter is used to smooth the interest area, eliminate Gaussian noise and ignore details. Histogram equalization is beneficial to improve the contrast of the images, especially for the images with a concentrated pixel distribution. There are four manners tried totally as shown in Fig.4. Correspondingly, the processed the input eye regions are showed as Fig.5.

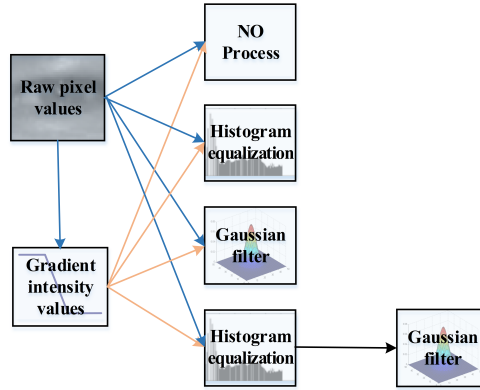


Fig. 4. The four manners used in the image preprocessing.

Create The Patch Expert. The sample data was divided into three parts 80% of them was used to train SVR model, and the inputs were the patches and labels. Thus, a 625 dimension SVR model, ω , could be obtained. In order to further improve the accuracy of the model, 10% of the sample data was used to train the linear fitting model, of which the input is trained SVR model ω ,

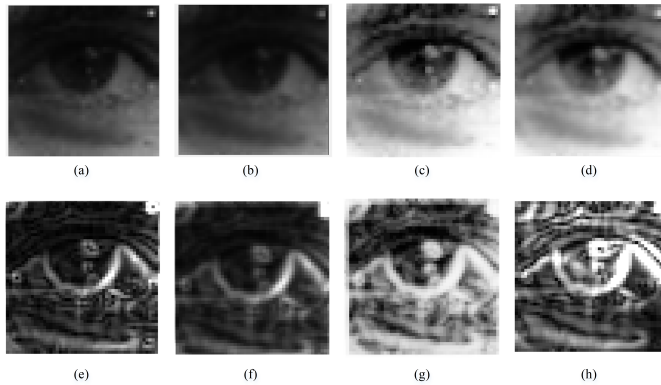


Fig. 5. Examples after the different process. Where (a)~(d) is the original images, processed by Gaussian filter, histogram equalization, and both of these two processing;(e)~(h) is the corresponding gradient images.

patches, and labels. The output is parameter a and b . The specific calculation method is as follows:

$$\begin{cases} y = ax + b \\ x = \omega \cdot patch \\ y = label \end{cases} ; \quad (2)$$

Combining the SVR and LR, a 627 dimensional patch expert was trained. The remaining 10% of the sample dataset was used to calculate the error and correlation of the training model.

$$response = -\frac{1}{e^{a \cdot \omega x + b}}. \quad (3)$$

Where, x is the pixels of the patch in 625 dimension. The response is belong to 0 to 1. And the RMSE (root mean squared error) and correlation were calculated.

3.3 Evaluate The SVR Model

To evaluate the model gathered from the last step, two public available dataset including GI4E[16], and BioID[17], were used to detect the eye pupil center.

Datasets. The BioID database has been regarded to be a challenging dataset for its images have a relative low resolution of 384*288 pixels. There are 1521 grayscale images which have various eye conditions such as poor illumination, wearing glasses, closed eyes, etc. The GI4E database contains 1236 images of 103 subjects with 12 gaze directions. The resolution of the images are 800*600 pixels, so it is considered to be a normal web camera setup.

Detect The Eye Pupil Center. The square areas of interest should be extracted and scaled up to 51*51 pixels. All of the response of the 729 test patches can be calculated through the patch expert. If the biggest response is larger than 0.15, the localization with the biggest response will be selected to be the detected eye pupil center. On the contrary, if the biggest response is less than 0.15, a preset value will set to be the eye center location which is related to the six feature points. The following Fig.6 illustrate a couple of distributions of the responses.

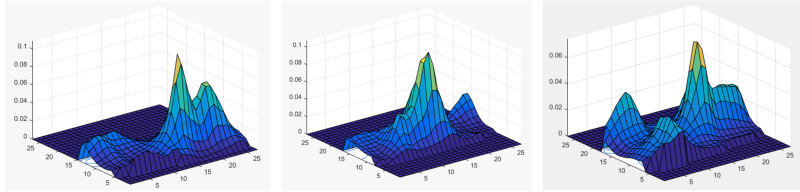


Fig. 6. The distributions of responses for some test images.

4 Experimental Results

4.1 Accuracy Measurement

The method calculating the accuracy of detect eye pupil center was first introduced by Jesorsky *et al.* [17]. Considering the large difference between two eyes and distances of two pupil centers (among different individuals), the normalized error was regarded as the evaluation criterion, which can be computed as:

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

where \mathbf{d}_l and \mathbf{d}_r are the Euclidean distances between the detected pupil center and actual pupil center for the left eye and right eye, respectively. The labeled eye centers are regarded as the ground truth. \mathbf{d} is the actual Euclidean distance between two eye pupil centers, correspondingly.

4.2 Results

It is acknowledged that the Gaussian filter can eliminate Gaussian noise, and histogram equalization can improve image contrast. In this paper, we tried these two methods to improve the quality of the input images. Besides, the gradient features are commonly used to find the object boundary, which may be a reasonable method to detect the boundary of the pupil due to the obvious difference between inside and outside in the pupil region. Table 1 shows the test results of

the database GI4e. It can be discovered that no matter which image processing method was used, the results of raw pixels is better than the gradient. Meanwhile, only the method using the Gaussian filter alone gathered a better accuracy. Although the histogram equalization can improve the image contrast, the result reveals that the histogram equalization does not make any improvement and even makes result worse. The results show that which method is able to detect the eye pupil center accurately if the image resolution is not too low. And the raw pixel values with the Gaussian filter performs best, similar conclusion can be achieved as well in the BioID dataset. The accuracy to detect eye center is as high as 96.84%, which is enough for the applications such as gaze estimation, and eye tracking. The Fig.7 shows some results of the eye localization.

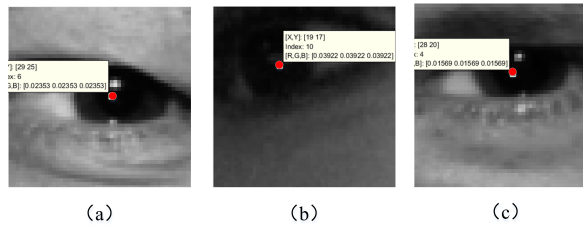


Fig. 7. The results of the eye localization

Table 1. Comparison of maximum normalized error in the GI4E database

Patch type	Process	$e < 0.05$	$e < 0.1$	$e < 0.25$
raw pixel	none	96.32%	99.65%	100.00%
gradient	gradient	83.79%	98.86%	100.00%
raw pixel	Gaussian filter	96.84%	99.74%	100.00%
gradient	Gaussian filter	82.47%	99.04%	100.00%
raw pixel	histogram	95.71%	99.21%	100.00%
gradient	histogram	80.98%	99.47%	100.00%
raw pixel	histogram+	96.67%	99.65%	100.00%
gradient	Gaussian filter	84.14%	99.74%	100.00%
gradient	histogram+	84.14%	99.74%	100.00%
gradient	Gaussian filter	84.14%	99.74%	100.00%

Table 2 presents the comparison of the proposed method with some state-of-the-art methods using the same database BioID and the same accuracy evaluation criterion. Although the accuracy based on the proposed method was not the

highest when the error was less than 0.05, the result was better than the state-of-the-art methods as error is less than 0.1.

Table 2. Comparison of maximum normalized error in the LFPW database

Patch type	e<0.05	e<0.1	e<0.25
Campadelli2009[18]	80.70%	93.20%	99.30%
Asadifard2010[19]	47.00%	86.00%	96.00%
Timm2011[20]	82.5%	93.40%	98.00%
Valenti2012[21]	86.10%	91.70%	97.90%
Leo2014[22]	80.70%	87.30%	94.00%
Anjith2016[23]	85.00%	94.30%	-
Poulopoulos2017[10]	87.10%	98.00%	100.00%
Proposed	82.15%	98.70%	100.00%

To verify the robustness of the proposed algorithm, we also tested the dataset LFPW. Table 3 shows the results for the dataset LFPE. Although the results are not as high as the method mentioned in the previous related works, it is robust to locate the eye center in the rough condition with an acceptable accuracy.

Table 3. Comparison of maximum normalized error in the LFPW database

Process	e<0.05	e<0.1	e<0.25
none	81.42%	97.34%	100.00%
Gauss filter	84.07%	98.23%	100.00%
histogram	83.19%	97.34%	100.00%
histogram+ Gauss filter	83.78%	98.23%	100.00%

5 Conclusion

This paper proposed an appearance-based method to detect the eye pupil center with an improved SVR. Though this method cannot get the highest accuracy among the current research, it can be robust against the variations of face state and image quality with an acceptable estimation accuracy. On the other hand, Gaussian filter and histogram equalization were tried to improve the accuracy in this paper. Results show that the gradient features and histogram process for every pixel make no improvement for the SVR-based method. Meanwhile,

the Gaussian filter can perform well to improve the accuracy of SVR-based eye center localization.

Acknowledgments. This work is supported by the National Natural Science Foundation of China (No. 61733011, 51575338).

References

1. Drucker H, Burges C J C, Kaufman L.: Support Vector Regression Machines.[J]. *Advances in Neural Information Processing Systems*, 1996, 28(7):779-784.
2. Smola, A.J. and B. Schölkopf.: A tutorial on support vector regression.[J]. *Statistics and Computing*, 2004, 14(3):199-222.
3. Baltrusaitis T, Robinson P, Morency L P.: Constrained Local Neural Fields for Robust Facial Landmark Detection in the Wild[C]. *IEEE International Conference on Computer Vision Workshops*. IEEE Computer Society, 2013:354-361.
4. Baltrusaitis T, Robinson P, Morency L P.: OpenFace: An open source facial behavior analysis toolkit[C]. *IEEE Winter Conference on Applications of Computer Vision*. IEEE, 2016:1-10.
5. Daugman J G.: High Confidence Visual Recognition of Persons by a Test of Statistical Independence[M]. IEEE Computer Society, 1993.
6. Cai, H., Liu, B., Zhang, J., Chen, S., Liu, H.: Visual Focus of Attention Estimation Using Eye Center Localization[J]. *IEEE Systems Journal*, 2017, PP(99):1-6.
7. Huan N V, Kim H.: A Novel Circle Detection Method for Iris Segmentation[C]. *Congress on Image and Signal Processing*, Vol. IEEE Computer Society, 2008:620-624.
8. Timm F, Barth E.: Accurate Eye Centre Localisation by Means of Gradients[C]. *Visapp 2011 - Proceedings of the Sixth International Conference on Computer Vision Theory and Applications*, Vilamoura, Algarve, Portugal, 5-7 March. DBLP, 2011:125-130.
9. Skodras E, Fakotakis N.: Precise localization of eye centers in low resolution color images[M]. *Butterworth-Heinemann*, 2015.
10. Pouloupoulos N, Psarakis E Z.: A new high precision eye center localization technique[C]. *IEEE International Conference on Image Processing*. IEEE, 2017:2806-2810.
11. Markuš, N., Frljak, M., Pandžić, I. S., Ahlberg, J., Forchheimer, R.: Eye pupil localization with an ensemble of randomized trees[J]. *Pattern Recognition*, 2014, 47(2):578-587.
12. Everingham M, Zisserman A.: Regression and classification approaches to eye localization in face images[C]. *International Conference on Automatic Face and Gesture Recognition*. IEEE Computer Society, 2006:441-448.
13. Gou, C., Wu, Y., Wang, K., Wang, K., Wang, F. Y., Ji, Q.: A joint cascaded framework for simultaneous eye detection and eye state estimation[J]. *Pattern Recognition*, 2017, 67(1):23-31.
14. Belhumeur, P. N., Jacobs, D. W., Kriegman, D. J., Kumar, N.: Localizing parts of faces using a consensus of exemplars[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2013, 35(12):2930-2940.
15. Le, V., Brandt, J., Bourdev, L., Bourdev, L., Huang, T. S.: Interactive facial feature localization[C]. *European Conference on Computer Vision*. Springer-Verlag, 2012:679-692.

16. Villanueva, A., Ponz, V., Sesma-Sanchez, L., Ariz, M., Porta, S., Cabeza, R.: Hybrid method based on topography for robust detection of iris center and eye corners[J]. *Acm Transactions on Multimedia Computing Communications & Applications*, 2013, 9(4):1-20.
17. Jesorsky O, Kirchberg K J, Frischholz R.: Robust Face Detection Using the Hausdorff Distance[C]. *International Conference on Audio- and Video-Based Biometric Person Authentication*. Springer-Verlag, 2001:90-95.
18. Campadelli P, Lanzarotti R, Lipori G.: Precise Eye Localization through a General-to-specific Model Definition[C]. *British Machine Vision Conference 2006*, Edinburgh, Uk, September. DBLP, 2006:187-196.
19. Asadifard M, Shanbezadeh J.: Automatic Adaptive Center of Pupil Detection Using Face Detection and CDF Analysis[J]. *Lecture Notes in Engineering & Computer Science*, 2010, 2180(1):13-14.
20. Timm F, Barth E.: Accurate Eye Centre Localisation by Means of Gradients[C]. *Visapp 2011 - Proceedings of the Sixth International Conference on Computer Vision Theory and Applications*, Vilamoura, Algarve, Portugal, 5-7 March. DBLP, 2011:125-130.
21. Valenti R, Gevers T.: Accurate eye center location through invariant isocentric patterns.[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2012, 34(9):1785-1798.
22. Leo, M., Cazzato, D., De, M. T., Distante, C.: Unsupervised eye pupil localization through differential geometry and local self-similarity matching.[J]. *PLoS ONE*,9,8(2014-8-14), 2014, 9(8):e102829.
23. George A, Routray A.: Fast and accurate algorithm for eye localisation for gaze tracking in low-resolution images[J]. *Iet Computer Vision*, 2017, 10(7):660-669.