Multimed Tools Appl (2018) 77:10063–10075
DOI 10.1007/s11042-017-4945-x

Human eye location algorithm based on multi-scale self-quotient image and morphological filtering for multimedia big data

Bin Song[1,2] **· Doo-Kwon Baik**[2] **· Shunxian Zhou**[3]

Received: 17 February 2017 / Revised: 22 May 2017 / Accepted: 13 June 2017 /
Published online: 24 July 2017

Abstract In order to reduce the effect on the eyes location caused by the variation of illumination and expression, this paper proposes a human eye location algorithm based on the multi-scale self-quotient image and morphological filtering. Firstly, the multi-scale self-quotient image is used to offset the lighting effects on the face, then the morphological open-close operation will be taken to enhance the local features around the eyes and relevant coefficient is used to roughly position the eyes. At last, the variance projection method will be used to analyze the roughly-positioned areas and binarize them to position accurately the central point of the eye. The experiments on the images from JAFFE Database, Yale B Database and AR database have shown that the proposed algorithm can well position the center of the eye, and it is robust to deal with the changes of illumination and expressions.

Keywords Multimedia Big Data · Multi-scale Self-quotient Image · Morphological Filtering · Human Eye Location · Variance Projection

✉ Doo-Kwon Baik
baikdk@korea.ac.kr

Bin Song
songbin@korea.ac.kr

Shunxian Zhou
zsx_hd@hnu.edu.cn

[1] School of Information Engineering, Henan University of Science and Technology, Luoyang 471023, China

[2] Department of Computer and Radio Communications Engineering, Korea University, Seoul 136-701, Republic of Korea

[3] College of Software and Communication Engineering, Xiangnan University, Chenzhou 423000, China

1 Introduction

With the popularity of cloud computing applications, the era of multimedia big data has arrived.Eyes location is a critically important step of the automatic recognition of the pose, face and expression [3]. This is because the space between eyes is least affected by the lighting or variation of facial expressions, it is often used as a standard of the rotation calibration, normalization and equalization of the image, and as the basic element for detecting and extracting other parts of the human face. How to carry out the correct and fast image recognition in the multimedia data is becoming more and more important.

At home and abroad, many experts have made more achievements in eyes location. Reinders et al. put forward a neuronic network-based eyes location method which can only identify the general location of the eyes, but not accurately position of the center of the eye [7]. Wang took advantage of the topographical features to position the eye [8]. Zhou improved the traditional integral mapping and put forward the generic projection function (GPF) [10]. Wu Yan et al. proposed to use gray-scale information and the pupil filters to position eyes [9]. Park made use of texture information composed of Gabor filter and round filter to position eyes [6]. This method is valid to some extent but too complex. Chen Huajie proposed a face feature location method based on Hierarchical Edge Orientation Field Matching (HEOFM) [1], but this method cannot position accurately the center of the eye.

The existing eye location methods do not take into account the problems caused by the illumination. For the face images which are taken in the uneven lighting conditions, the accuracy of the eyes location is generally not high. Moreover, the dramatic changes of the facial expressions will cause the missing of some eye features. For instance, when the eyes are closed, the white of the eye and the pupil will be hidden, which would lead to failure of some algorithms for eyes location based on the gray scale variation in the horizontal direction.

In order to reduce the effect on the eyes location caused by the variation of illumination and expression, this paper proposes a human eye location algorithm based on the multi-scale self-quotient image and morphological filtering (MSMF). Firstly, the multi-scale self-quotient image is used to offset the lighting effects on the face, then the morphological open-close operation will be taken to enhance the local features around the eyes and relevant coefficient is used to roughly position the eyes. At last, the variance projection method will be used to analyze the roughly-positioned areas and binarize them to position accurately the central point of the eye. The experiments on the images from JAFFE Database, Yale B Database and AR Database have shown that this algorithm can well position the center of the eye, and it is robust to deal with the changes of illumination and expressions.

The specific contributions of this paper include:

(1) An effective multi-scale self-quotient image optimization models for multimedia big data is proposed.
(2) An effective rough location of morphological filtering optimization models for multimedia big data is proposed.
(3) An effective precise eye location optimization models for multimedia big data is proposed.
(4) Performance analysis of the proposed algorithm and an evaluation of the algorithm with respect to other existing algorithms.

The rest of this paper is organized as follows. Section 2 discusses the multi-scale self-quotient image optimization models for multimedia big data, followed by the effective rough location of morphological filtering optimization models is designed in Section 3. The effective precise eye location optimization models is discussed in Section 4. Section 5 shows the simulation experimental results, and Section 6 concludes the paper with summary and future research directions.

2 Multi-scale self-quotient image

The issue of variation of illumination is the most delicate one in the aspect of three-dimensional target identification (e.g. face recognition) based on the gray-scale image. This issue in recent years has aroused widespread concern, so has the issue of eye location which is confronted with the same problem. The elimination of the impact of illumination is of major significance for the accurate location of the eyes. But the self-quotient image presents a characteristic of stability of illumination which provides a solution to this problem. The self-quotient image of the image is defined as

$$R(x,y) = \frac{I(x,y)}{\hat{I}(x,y)} = \frac{I(x,y)}{F(x,y) \otimes I(x,y)} \tag{1}$$

where $I(x, y)$ is the smooth version of face recognition, $I(x,y)$ is the smooth filtering kernel, and $F(x, y)$ is the convolution operation, and the division operation is carried out point-by-point. The $R(x, y)$ is considered as the self-quotient image because it is a quotient image derived from itself.

But if the kernel is too small, and it is close to 1, the information of image reflectivity will lose greatly. If the kernel of is too large, the vignette effect will be found in the areas close to ladder-like boundary. It is therefore suggested the use of multi-scale technology to make the results more robust. The definition of the multi-scale self-quotient image is given as

$$R_M(x,y) = \frac{I(x,y)}{\sum\limits_{n=1}^{N} w_n \left[F(x,y,\sigma_n) \otimes I\left(x,y\right) \right]} \tag{2}$$

In which, $I(x, y)$ represents the function of the smooth filter kernel, i.e.

The scale parameter of the Gaussian function is given as

$$F(x,y,\sigma) = Ke^{-\left(x^2+y^2\right) \big/ \sigma^2} \tag{3}$$

In which, K is determined by the normalized formulation

$$\iint F(x,y,\sigma)dxdy = 1 \tag{4}$$

In the formula $w = (w_1 + w_2 + \cdots + w_n + \cdots w_N)$, w_n represents the weight of the nth smooth filtering kernel component, while $\sum\limits_{n=1}^{N} w_n = 1$. In general, $w_n = 1/N$. σ_n denotes the scale parameter of the n-th smooth filtering kernel function. If $N = 3$, the experiments show that the result would be better when σ is assigned with 5, 10 and 15.

Figure 1 shows the multi-scale self-quotient images of three images from Yale B Database which has been taken under the different illumination conditions and processed. It follows that self-quotient images being processed have eliminated the effect of the illumination. But at the same time the information of some important organs in the face has been saved, conducive to the eye location.

Fig. 1 Results of multi-scale self-quotient images. **a** Multi-scale self-quotient images, **b** Original images, **c** Intermediate results

3 Rough location of morphological filtering

The mathematical morphology is prepared by a group of algebraic operators including mainly the dilation and erosion as well as the open and close operations based on them. These operators present their own features in the binary images and grayscale images. So firstly let's take a look at gray morphology. Set $b(i,j)$ is the function of the input image, m is the structural element, then the Grayscale Dilation Function can be given as

$$(f \oplus b)(x,y) = \max_{0 \le i,j \le m-1} \left[f(x+i,y+j) + b\left(i,j\right) \right]. \tag{5}$$

The Grayscale Erosion Function can be given as

$$(f \Theta b)(x,y) = \min_{0 \le i,j \le m-1} \left[f(x+i,y+j) - b\left(i,j\right) \right]. \tag{6}$$

The operation of grayscale dilation is to determine the maximum value of in the area determined by structural elements. It can be used for the elimination of the dark details and the enhancement of the bright boundary. The operation of erosion is to determine the minimum value of in the same area. Then the output image would be like this: the gray scale value of brighter details on the boundary will be reduced, and the brighter boundary will shrink.

The Open is defined as follows:

$$Open(f) = f \circ b = (f \Theta b) \oplus b \tag{7}$$

The Close is defined as follows:

$$Close(f) = f \bullet b = (f \oplus b) \Theta b \tag{8}$$

The Open is to carry out dilation after erosion, which can diminish local small blocks of bright areas. The Close is to carry out erosion after dilation, which can make up for the local small blocks of dark areas.

While the eye areas contain both the bright blockettes such as the white of eye (crest value) and dark blockettes such as the pupil, a small block of ingestion and dark areas of the valley (bottom value). In order to efficiently extract the area information of eyes, the filtering function of eyes is introduced as follows:

$$G(x,y) = R_t(x,y) \times h(x,y) \tag{9}$$

where $R_t(x, y)$ is a low-pass smooth filter. The parameter of the smooth filtering template adopted by this paper is 7 × 7. While

$$R_t(x,y) = \left| R_M(x,y) - Open\left(R_M\left(x,y\right)\right) \right| + \left| R_M(x,y) - Close\left(R_M\left(x,y\right)\right) \right| \tag{10}$$

In which, $|R_M(x, y) - Open(R_M(x, y))|$can enhance the peak value of the image that is shown in Fig. 2(b), $|R_M(x, y) - Close(R_M(x, y))|$can enhance the bottom value that is shown in Fig. 2(c),$R_t(x, y)$ can enhance both the peak and the bottom values that is shown in Fig. 2(d). $G(x, y)$ is the result of $R_t(x, y)$ after being fuzzy that is shown in Fig. 2(e). The three-dimensional representation is shown in picture (f). As it can be seen that highlighted area of $G(x, y)$corresponds the local change of the areas of eyes and mouth. Therefore, the eye feature has been effectively enlarged and enhanced.

This method can effectively highlight the information of eyes, and reduce the effect of such things as glasses and hair. Moreover, it is not sensitive to the open and close of the eyes. In order to extract the effective eye area, there is a need for binaryzation. Set the threshold of the adaptive binaryzation as follows:

$$thr = mean(G(x,y)) + 0.1 \times \left[\max(G(x,y)) - mean\left(G\left(x,y\right)\right) \right] \tag{11}$$

Use thr to carry out boundary binaryzation of the eye filtering diagram $G(x, y)$ and obtain the candidate eye areas $M(x, y)$.

$$M(x,y) = \begin{cases} 1, & G(x,y) > thre \\ 0, & otherwise \end{cases} \tag{12}$$

And then use $M(x, y)$ to extract the left and right eyes from $R_M(x, y)$, and then evaluate them. If there is a N candidate eyes, it needs to carry out (N× (N-1))/2 matches, which will lead to a more complicated operations. In order to reduce search areas, it can add priori knowledge to set some constraints. For example, eye areas should be located on the top of the face, and the areas a certain distance blow eye areas are flat cheek areas; the distance range between the centers of eyes is

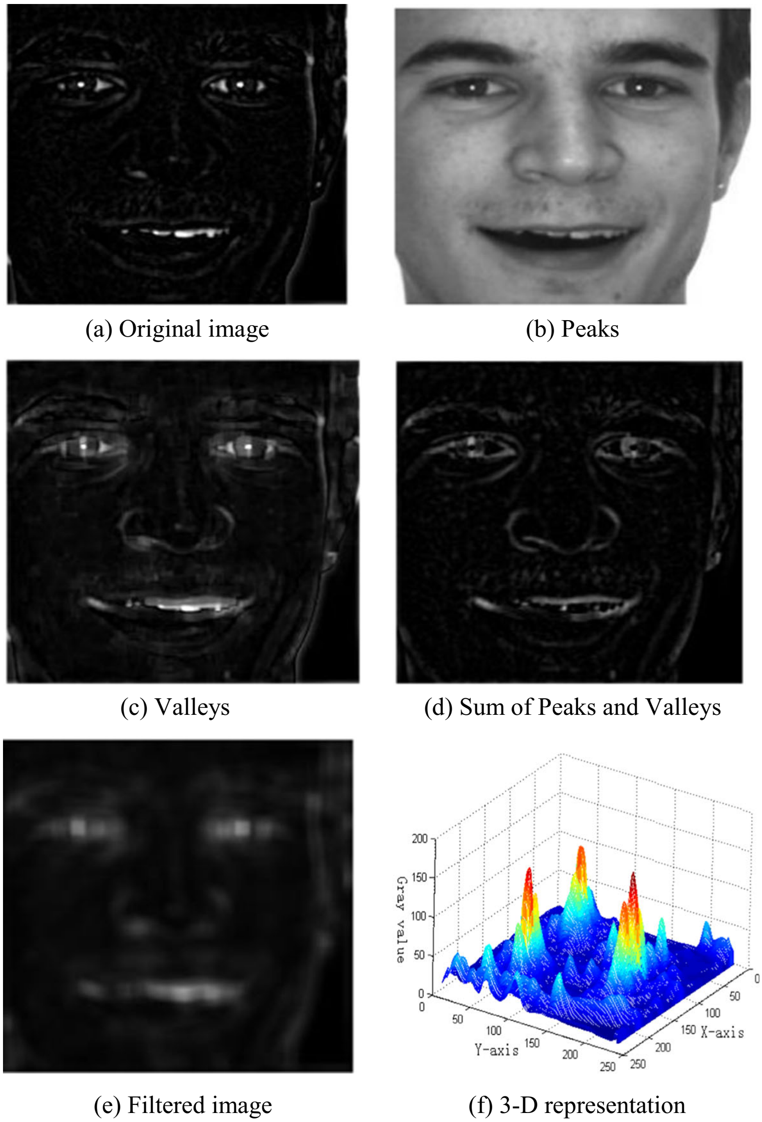


Fig. 2 Result of eye image after morphological filtering **a** Original image **b** Peaks **c** Valleys **d** Sum of peaks and valleys **e** Filtered image **f** 3-D representation

established. Once the distance is longer than 2/3 or shorter than 1/3 of the width of the image, it can safely determine it is not the eye pairs. Set the distance between the centers of two candidate areas as l, then from the multi-scale self-quotient image $R_M(x,y)$ remove two candidate eye window areas with the size of $0.4\,l \times 0.8\,l$ and calculate the correlation coefficient of the candidate eye window areas. The computation of the correlation coefficient is:

$$S = \frac{1}{|La_l - La_r| + |M_l - M_r|} \tag{13}$$

where La_l and La_r are the gray values of the left and right candidate eye areas, M_l and M_r are the average gray values.

Use the computation given above to determine the correlation of all candidate eye areas. The eyes with highest coefficient can be determined as the right eye pairs.

4 Precise eye location

In order to obtain the accurate location of the eyes, it needs to position for the second time the eye of each pair. Because eye areas contain eyebrow areas, the improved IPF (Integral Projection Function), that is, VPF (Variance Projection Function) proposed in [2] could be adopted to make it more robust. Variance projection function is:

$$\begin{cases} \delta_v^2 = \dfrac{1}{y_2 - y_1} \sum_{y=y_1}^{y_2} [I(x, y_i) - V_{mean}(x)]^2 \\ \delta_h^2 = \dfrac{1}{x_2 - x_1} \sum_{x=x_1}^{x_2} [I(x_i, y) - H_{mean}(y)]^2 \end{cases} \tag{14}$$

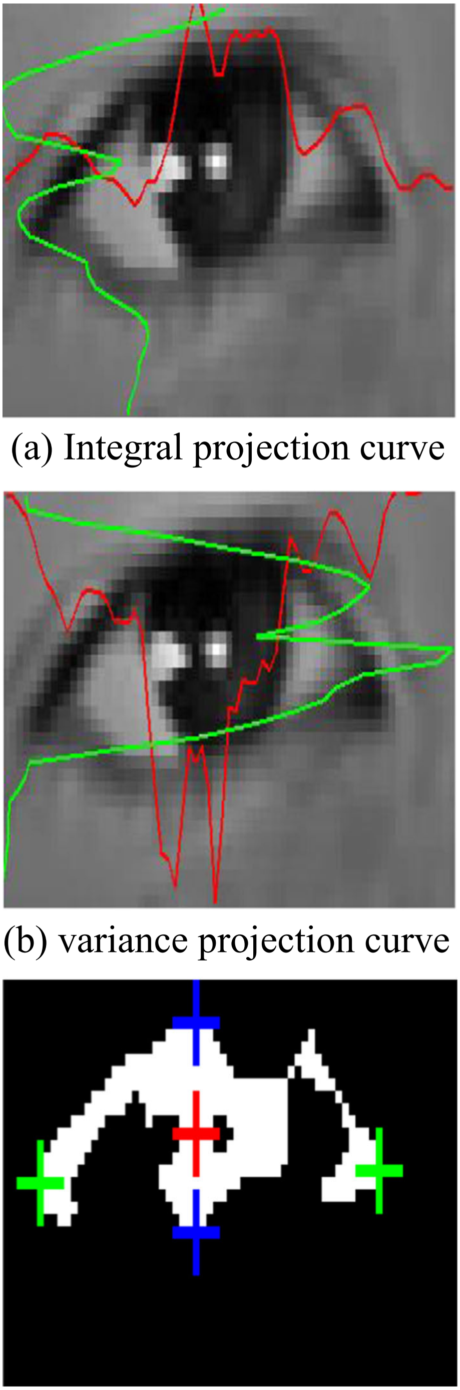
In which, V_{mean} and H_{mean} are the traditional average values of IPF in vertical and horizontal direction.

To project the candidate eye areas with VPF, which is shown in Fig. 3(b), the red curve represents the vertical projection, while the green curve the horizontal projection). Compared with IPF that is shown in Fig. 3(a), this method can more accurately determine the position of the center of the pupil. The pupil area can be obtained by VPF curve, but the center of the eye is always not the center of the pupil. This is because the location of the pupil will change with the variation of the direction of the eye sight. In order to obtain precise central point of the eye, we can set the diameter of the pupil as d, and then extract an area around the pupil with the size of 2d × 4d, and carry out binaryzation. After the analysis of the area extracted, four coordinates of the eye area can be acquired. They are coordinate of the highest point of the center T, the coordinate of the lowest point of the center B, the leftmost coordinate L, and the rightmost coordinate R. Therefore, the coordinate of the center of eye C can be expressed as C = (T + B + L + R)/4. The central point got with this formula not only is closer to the real center of the eye, but also presents better anti-interference ability. Fig. 3(c) shows the result of location. The up and down points of the eye are represented as blue "+", the left and right points as green "+", and the central point of the eye as red "+".

5 Experimental results and analysis

The images are taken respectively from JAFFE Database, Yale B Database and AR Database. We take Rainer Lienhart's improved AdaBoost method [5] to detect the face. All faces of the images can be precisely positioned. The experiment uses the measurement standard put forward by Jesorsky et al. [4]. Set the manually labeled central points of the left and right eyes as C_l and C_r. The location of the left and right eyes detected is C_l and C_r . d_l is the Euclidean distance between C_l and C_l, while dr is the Euclidean distance between Cr and C_r . d_{lr} is the Euclidean distance between C_l and C_r. Therefore, the definition of relative error of location is:

$$err = \frac{\max(d_l, d_r)}{d_{lr}} \tag{15}$$



(a) Integral projection curve

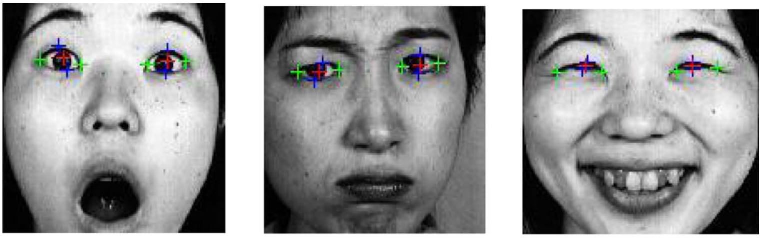
(b) variance projection curve

(c) Eye Location result

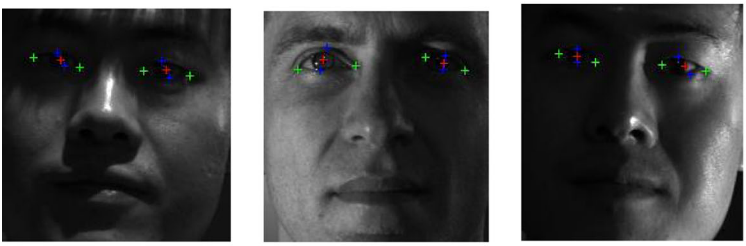
Fig. 3 Precise eye location **a** Integral projection curve, **b** variance projection curve **c** Eye location result

Table 1 The division experimental results of three databases

	JAFFE database	Yale B database	AR database(without glasses)	AR database
Total	213	800	637	882
Detected	213	759	635	729
Accuracy	100.00%	94.88%	99.69%	82.65%

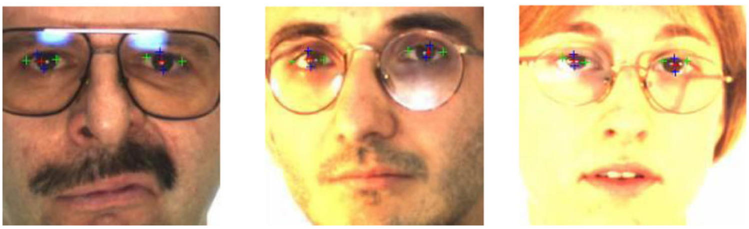


(a) Examples of Eye Location results on JAFFE database



(b) Examples of Eye Location results on Yale B database

(c) Examples of Eye Location results on AR database(faces without glasses)



(d) Examples of Eye Location results on AR database (faces with glasses)

Fig. 4 Examples of Eye Location results on three databases **a** Examples of eye location results on JAFFE database **b** Examples of eye location results on Yale B database **c** Examples of eye location results on AR database (faces without glasses) **d** Examples of eye location results on AR database (faces with glasses)

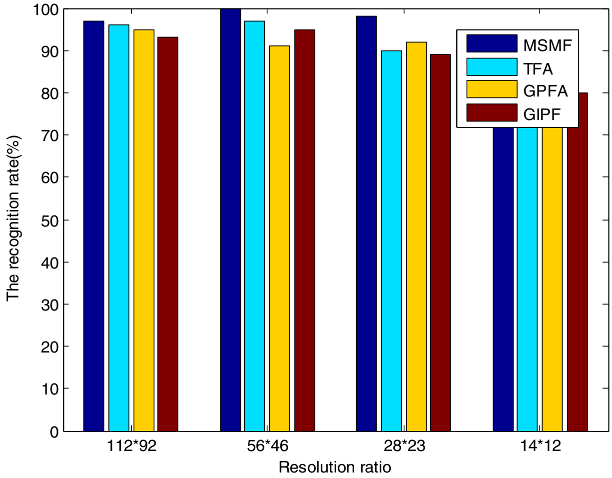


Fig. 5 The recognition rate of three algorithms with different resolution ratio for JAFFE samples

When err <0.25, the location is correct. In general, d_{lr} is approximately twice the width of the eye. So err = 0.25 means that the distance of the larger one (either d_l or d_r) is approximately one-half the width of the eye, that is, the error below 0.25 is acceptable. But when err ≥0.25, then it is assumed that the eye is not correctly positioned.

As can be seen from the Table 1 the proposed algorithm has achieved high accuracy rate in the experiments on expression variation, illumination variation and illumination combined with expression variation. It indicates that this algorithm is much robust for the variation of illumination and expression. Figure 4 shows part of results of the experiments mentioned above. The experiment on fourth group demonstrates that, due to the interference caused by the glasses, especially the reflection of light brought about by the spectacle-frames and lenses, the accuracy rate drops sharply.

We compare the performance of our MSMF algorithm with Terrain Feature Algorithm (TFA) [8], General Projection Function Algorithm (GPFA) [10] and the Gray-scale

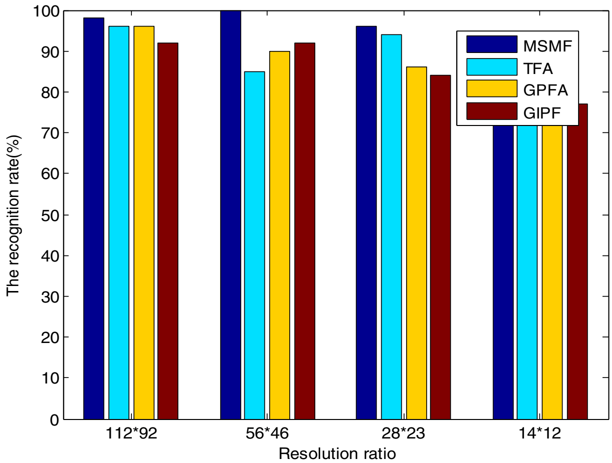


Fig. 6 The recognition rate of three algorithms with different resolution ratio for Yale B samples

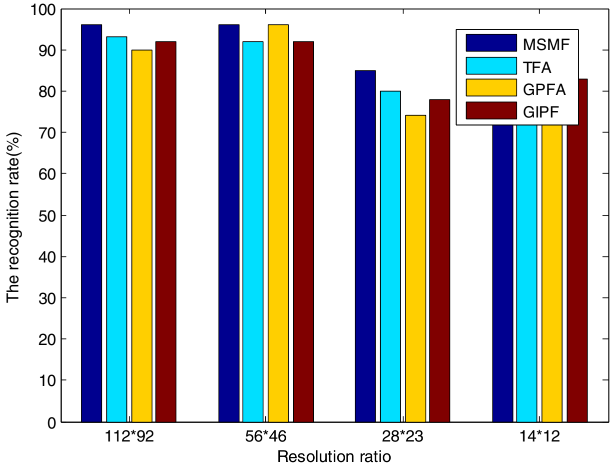


Fig. 7 The recognition rate of three algorithms with different resolution ratio for AR samples

Information and Pupil Filter (GIPF) [9]. All the four algorithms have the same performance which makes use of texture information that is constituted of Gabor Filter and round filter. Fig. 5 provides different recognition rate of eye location with different algorithms for the images from JAFFE database under the conditions of err <0.25. Fig 6. shows the recognition rate of four algorithms with different resolution ratio for Yale B samples. Fig. 7. shows the recognition rate of four algorithms with different resolution ratio for AR samples.

Figure. 8 shows the recognition rate of the four algorithms with different degrees, Fig. 9. Shows the recognition rate of three algorithms on different expressions, and Fig. 10. shows the recognition time of the four algorithms with different degrees. Through the above experimental results, it is not difficult to see that the recognition rate of our proposed MSMF is the highest in most cases, and its recognition time is the least. The proposed algorithm can be applied to the multimedia bid data.

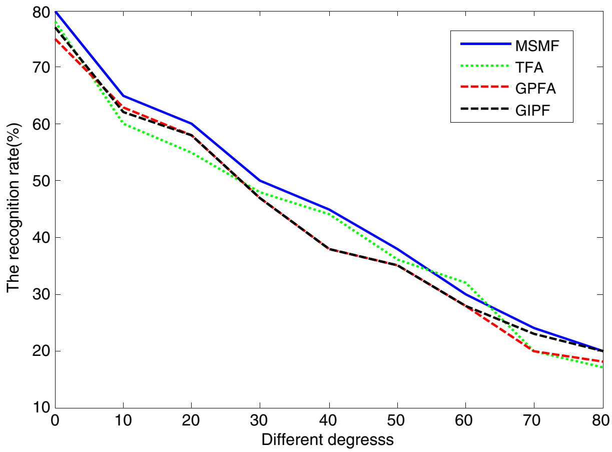


Fig. 8 The recognition rate of the four algorithms with different degrees

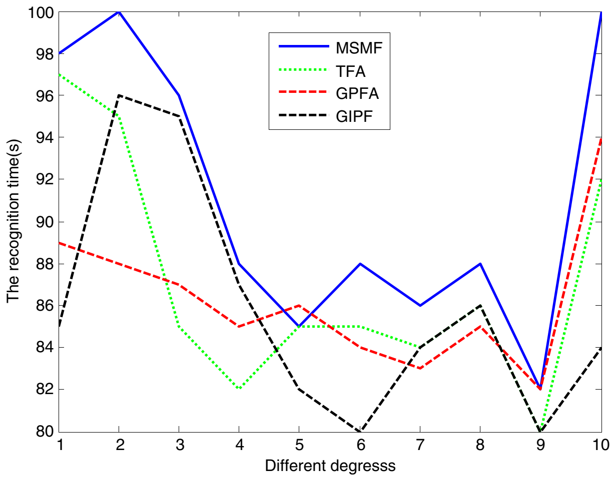


Fig. 9 The recognition rate of three algorithms on different expressions

6 Conclusions

In order to reduce the effect on the eyes location caused by the variation of illumination and expression, this paper proposes a human eye location algorithm based on the multi-scale self-quotient image and morphological filtering. Firstly, the multi-scale self-quotient image is used to offset the lighting effects on the face, then the morphological open-close operation will be taken to enhance the local features around the eyes and relevant coefficient is used to roughly position the eyes. At last, the variance projection method will be used to analyze the roughly-positioned areas and binarize them to position accurately the central point of the eye. The experiments on the images from JAFFE Database, Yale B Database and AR Database have shown that this algorithm can well position the center of the eye, and it is robust to deal with the changes of illumination and expressions. The further research is how to reduce or eliminate the influence of the glasses frame side and the reflection of the lens on the eye location.

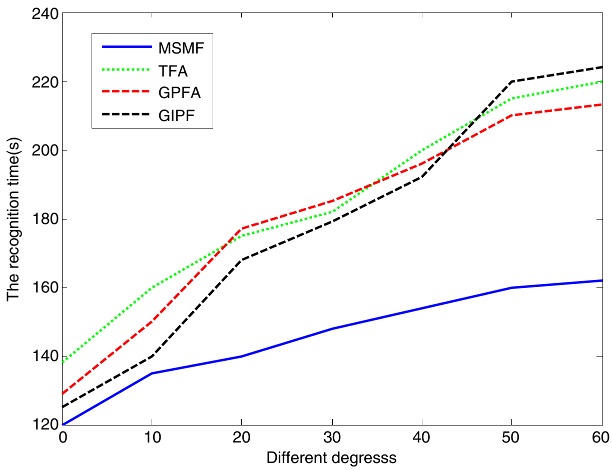


Fig. 10 The recognition time of the four algorithms with different degrees

Acknowledgements This work was supported by the National Natural Science Foundation of China under Grant No. 61402165, and the Scientific Research Fund of Hunan Provincial Education Department Project under Grant No. 15C1288 and No. 16B244. The authors would like to thank the anonymous reviewers and the editor for the very instructive suggestions that led to the much improved quality of this paper.

References

1. Chen H-j, Wei W (2007) Hierarchical edge orientation field matching based facial feature localization[J]. J Optoelectron Laser 18(2):241–244
2. Feng GC, Yuen PC (1998) Variance projection function and its application to eye detection for human face recognition[J]. Pattern Recogn Lett 19(9):899–906
3. Fu X-f, Wei W (2008) Local binary patterns-based facial expression recognition using supervised laplacianfaces[J]. J Optoelectron Laser 19(6):823–827
4. Jesorsky O, Kirchberg KJ, Robert W (2001) Robust face detection using the Hausdorff distance. Proceedings of the Third International Conference on Audio- and Video-Based Biometric Person Authentication [C]. Halmstad, Sweden: Springer, 90–95
5. Lienhart R, Maydt J (2002) An extended set of haar-like features for rapid object detection. Proceedings of IEEE International Conference on Image Processing[C]. Rochester, New York, USA: IEEE Press, 900–903
6. Park CW, Kwak JM, Park H, et al. (2007) An effective method for eye detection based on texture information[A]. International Conference on Convergence Information Technology[C]. Gyeongju, Korea: IEEE Computer Society, 586–589
7. Reinders MJT, Koch RWC, Gerbrands J J (1996) Locating facial features in image sequences using neural networks. Proceedings of the 2nd International Conference on Automatic Face and Gesture Recognition[C]. Washington, DC, USA: IEEE Computer Society, 230–235
8. Wang J, Yin L (2005) Eye detection under unconstrained background by the terrain feature[A]. Proceedings of IEEE Conference on Multimedia and Expo[C]. Amsterdam, Netherlands: IEEE Computer Society, 1528–1531
9. Yan W, Yang Y, Liping W (2005) An eye location algorithm based on the gray information and the pupil filter[J]. Comput Eng Appl 33:45–47
10. Zhou Z-H, Geng X (2004) Projection functions for eye detection[J]. Pattern Recogn 37(5):1049–1056