PROJECTION PEAK ANALYSIS FOR RAPID EYE LOCALIZATION

Jingwen Dai, Dan Liu, Jianbo Su
Research Center of Intelligent Robotics, Shanghai Jiaotong University, Shanghai, 200240, China
{*daijingwen,ldan,jbsu*}@*sjtu.edu.cn*

Keywords: eye localization, threshold, segmentation, projection peak

Abstract: This paper presents a new method of projection peak analysis for rapid eye localization. First, the eye region is segmented from the face image by setting appropriate candidate window. Then, a threshold is obtained by histogram analysis of the eye region image to binarize and segment the eyes out of the eye region. Thus, a series of projection peak will be derived from vertical and horizontal gray projection curves on the binary image, which is used to confirm the positions of the eyes. The proposed eye-localization method does not need any a priori knowledge and training process. Experiments on three face databases show that this method is effective, accurate and rapid in eye localization, which is fit for real-time face recognition system.

1 INTRODUCTION

Recent years, building automatic face recognition system has become a hot topic in computer vision and pattern recognition areas. Some commercial systems have been developed and applied in public and individual security. Generally, an automatic face recognition system is composed of three steps, i.e. face detection, facial feature localization and face recognition. Face detection determines whether or not there are any faces in the image or video sequence and, if present, acquires the location and extent of each face. Facial feature localization obtains the location of salient feature points of face, i.e. eyes, nose, mouth etc. And face recognition identifies or verifies one or more persons in the scene using a stored database of faces.

Most researchers test their recognition algorithms under an assumption that the locations of facial feature points are given or obtained through some interactions between users, i.e. pointing out the positions of eyes manually. Roughly speaking, in the major face recognition algorithms, salient facial landmarks must be detected and faces must be correctly aligned before recognition. The performance of a face recognition algorithm is greatly affected by the accuracy of facial features alignment. The recognition rate of Fisherface, which is one of the most successful face recognition methods, is reduced by 10%, when the centers of the eyes have been inaccurately localized with a deviation of just one pixel from their true positions (S.G. Shan, 2004). What's more, the position of eyes is the precondition for the localization of other facial landmarks. Therefore, the localization of the eyes is essential to automatic face recognition systems.

Many representation approaches of facial feature localization have been proposed in the previous works, such as Hough transform (T.Kawaguchi, 2000), symmetry detector (D. Reisfeld, 1995), ASM (T. F. Cootes, 1998), AAM (T. F. Cootes, 2001), Adaboost (P. Viola, 2001), projection analysis (Kanade, 1973; Z.H. Zhou, 2004; G.H. Li, 2006). Among the approaches above, projection analysis is one of the most classical algorithms. In gray facial image, the gray value of facial features is lower than that of the skin. By utilizing this character, firstly, the projection analysis calculates the sum of gray value or gray function value along x-axis and y-axis respectively and find out the special change points, then aggregates the change points of different directions according to the prior knowledge, and finally obtains the location of facial landmarks. Compared with other methods, its computational complexity is very low which is important for real-time application. Furthermore, it does not need any training process. However, general projection analysis is not robust to the variation of face poses, illuminations, expressions, or additional accessories, such as glasses.

This paper is to propose a novel method of projection peak analysis for rapid and precise eye localization. First, we segment the eye region from the face image by setting appropriate candidate window. Second, by histogram analysis of the eye region image, we get a threshold and perform binary transformation to segment eye out of the eye region. Then, a series of projection peak will be derived from vertical and horizontal gray projection curves of the binary image, which is utilized to figure out the exact coordinates of the eye.

The rest of this paper is organized as follows. In Section 2, the process of eye localization with projection peak analysis is briefly illustrated. Section 3 shows the performance of our proposed method in three standard face databases and in real-time application, followed by conclusions in Section 4.

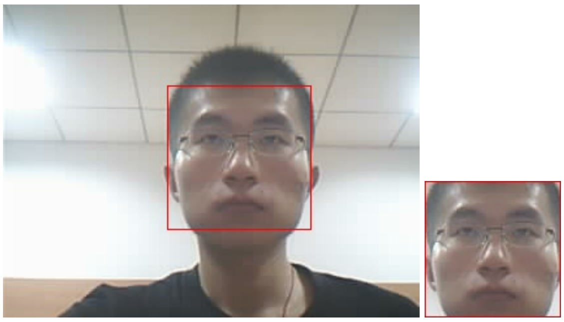
2 DESCRIPTION OF METHOD

The projection peak method we proposed is to be described in this section. This includes the selection of candidate window, the threshold for segmentation of eye region image, the gray projection and the analysis of projection peak.

2.1 Selection of Candidate Window

To locate the position of eyes, first we need know the position of face in the image. There have been many face detection methods so far, such as Adaboost (P. Viola, 2001). After the face is detected, we segment the face region from image as shown in Fig.1(a), then eye localization can be carried out in the face region image as shown in Fig.1(b).

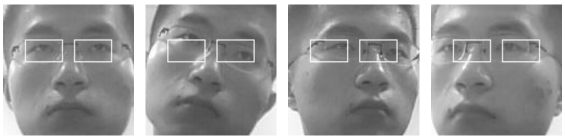
Generally, in a face, 1) two eyes must be in the upper part of human face; 2) The eyebrow must be above the eye; 3) Two eyes must be lie on both sides of centre-axis of frontal face image symmetrically, etc. These rules are of great help to reduce the searching area of eyes, which not only eliminates some interference, but also reduces the computation cost. We define the region of eye as eye candidate window, which must be robust to the variation of face poses. Fig. 2 shows the case that the eye candidate window are too small. On the contrary, if the eye candidate windows are too large, some interference including eyebrows, hair and the frame of glasses will be brought in, of



(a) (b)

Figure 1: (a) Face detection with Adaboost (b) Segmented face image

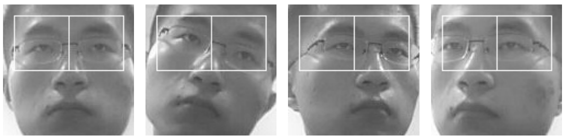
which the gray value is also low. So the interference has greatly influences on the eye localization through general projection analysis.



(a) (b) (c) (d)

Figure 2: (a) Candidate windows under ideal conditions (b - d) Eyes out of candidate windows when face poses changed

Fig. 3 shows the case that eyes are still in the candidate windows when face pose changed. However, we need to eliminate the interference (i.e. eyebrows, hair, the frame of glasses etc.) as much as possible to localize the positions of the eyes.



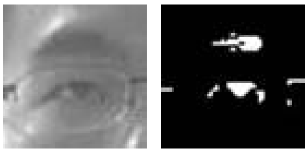
(a) (b) (c) (d)

Figure 3: (a) Candidate windows under ideal conditions (b - d) Eyes out of candidate windows when face poses changed

When the candidate windows are chosen, the region of eyes can be segmented as shown in Fig. 4(a).

2.2 Segmentation with a Threshold

Because of its intuitive properties and simplicity of implementation, image thresholding enjoys a central position in applications of image segmentation. And the crucial point of segmentation is to select a proper threshold. In the candidate window, it is obvious that the gray value of the pupil of the eye is lower than that in other regions. Hence, the threshold for segmentation is determined via histogram analysis. First we sort all the pixels in the candidate window ascendingly according to their gray value. Then the $p\%$ pixels with smaller gray value is set to be 255, while the rest is set with value 0. Thus the image of candidate window is binarized, as shown in Fig. 4(b).



(a) (b)

Figure 4: (a) Original image of eye region (b) Image after gray-histograms threshold segmentation

2.3 Gray Projection

Under ideal circumstances, the position of the pupil's center should be the position of the extreme point of the vertical projection curve (VPC) and the horizontal projection curve (HPC). However, there will be some interference (i.e. eyebrows, hair, frame of glasses etc.) owing to the larger candidate window, and their gray values are even lower than that of the pupil. What's more, because of the variation of illumination, there will be some shadows around eyeballs, which have great influence to the curve of projection. So it is unreasonable to take the extreme points of VPC and HPC as the position of pupil's center. To solve the problem mentioned above, we proposed a method of projection peak analysis(PPA),which will be described in the following section.

2.4 Projection Peak Analysis

The main process of projection peak analysis(PPA) is described as follows. Firstly, we analyze the possibly existing interference. From top to bottom along the vertical direction, the possibly existing regions that have lower gray values are hair, eyebrow, the upper frame of glasses, pupil, and the lower frame of glasses successively. While from left to right along horizontal direction, the possibly existing regions that have lower gray values are hair, the outer frame of glasses, pupil, and the inner frame of glasses successively. After threshold segmentation, the interference mentioned above might exist simultaneously, separately or even be absent. Secondly, in the candidate window, for eyebrow and pupil, the gray value of them are lower, and compared with the interference described above, the areas occupied by them are larger. As a result, they could be segmented by means of selecting proper threshold (Sometimes the eyebrow is too sparse or too thin to be segmented out, but the pupil can be extracted as usual). Take the horizontal projection curve of binary image after threshold segmentation which is denoted by $P(y)$ for example. Through analysis, it is easily to find out there are some peaks in the curve, that can be represented as $[y_{11}, y_{12}], [y_{21}, y_{22}], \cdots, [y_{n1}, y_{n2}]$, in which $y_{11}, y_{21}, \cdots, y_{n1}$ are the rising edge and $y_{12}, y_{22}, \cdots, y_{n2}$ are the falling edge of the peaks. The peaks mentioned above are corresponding to pupil, eyebrow and other interference respectively. And the peak accompanied with pupil has broader breadth, larger area and little deviation from the center of image. For the vertical projection curve, we can get similar result. The result of projection is shown in Fig. 5. By this characteristic, we proposed a new method for rapid eye localization based on projection peak analysis. Regarding to every projection peak, we define an evaluation value U_n in Eq. (1),

$$U_n = \alpha W_n + \beta S_n + \gamma D_n, \quad (1)$$

$$W_n = y_{n2} - y_{n1}, \quad (2)$$

$$S_n = \sum_{y=y_{n1}}^{y_{n2}} P(y), \quad (3)$$

$$D_n = \left|\frac{y_{n2} - y_{n1}}{2} - Y_c\right|, \quad (4)$$

where W_n, S_n, D_n is the peak's breadth, peak's area and deviation between image's center and the peak's central axis respectively as defined in Eq. (2-4), and α, β, γ are weights. Y_c in Eq. (4) is the image center along y-axis.

According to Eq. (1), the U values of all projection peaks are calculated and then sorted. The peak having the maximum U value is considered as the one corresponding to the region of eyeball. By calculating the maximum point in this peak, the coordinates of the pupil's center is finally confirmed.

3 EXPERIMENTS

To validate the proposed method, we implement the method by C++, and evaluate the performance in three standard face databases and in real-time respectively. A widely accepted criterion (O. Jesorsky, 2001) is used to judge the quality of eye localization, which is a relative error measure based on the distances between the located and the accurate central points of the eyes. Let C_l and C_r be the manually extracted left and right eye positions, $C_l'$and C_r' be the located positions, $d_l$be the Euclidean distance between C_l and C_l', $d_r$be the Euclidean distance between C_r and C_r'. Then the relative error of this localization is defined as Eq. (5):

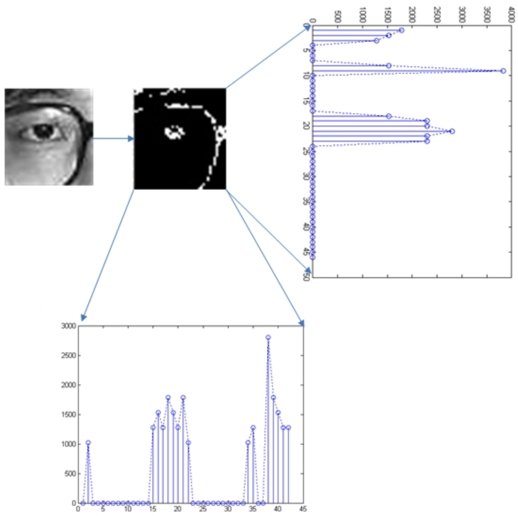


Figure 5: Horizontal and vertical projection peaks

$$err = \frac{\max(d_l, d_r)}{\|C_l - C_d\|}. \tag{5}$$

If $err < 0.25$, the localization is considered to be correct. Thus, for a face database comprising N images, the localization rate is defined as:

$$rate = \sum_{\substack{i=0 \\ err_i<0.25}}^{N} \frac{1}{N} \times 100\%, \tag{6}$$

where err_i is the err on the i-th image.

3.1 Performances on Standard Face Database

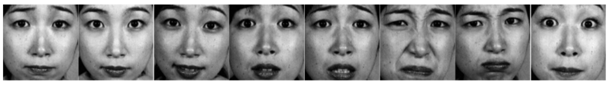
The FERET database (P. J. Phillips, 1998) is the most widely adopted benchmark for the evaluation of face recognition algorithm, in which there are 14051 human head-shoulder images, and among these images, there are 3816 images of which the position of two pupils have been labelled manually. The face regions in these images are taken out to form a test set. The BioID face database consists of 1521 frontal view gray level images with pupils' position manually marked. And this database features a large variety of illumination and face size, which is used to evaluate the performance of proposed method under different illuminations. The JAFFE face database (M. J. Lyons, 1998) is made up of 213 frontal view gray level images with a large variety of facial expressions posed by Japanese females. We use this database to test the method when facial expressions changed.

Let W and H be the width and height of facial image respectively. In our experiments, we set the parameters of eye candidate window as follow: Let $(W/12, H/12)$ and $(W/12, H/12)$ the origin of left and right eye candidate window, $(5 \times H/12, 5 \times H/12)$ be the size of two eyes candidate window.

Some test samples are shown in Fig. 6, including variations of face poses, illuminations, expressions, and accessories.

(a)

(b)



(c)

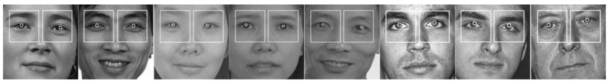
Figure 6: Some images from experimental database (a) Sample images from FERET (b) Sample images from BioID (c) Sample images from JAFFE

The results of experiment are shown in Table 1, where p is the threshold for threshold segmentation. It is shown that when p is set to 5, the best accurate rates of eye localization are obtained on all three databases.

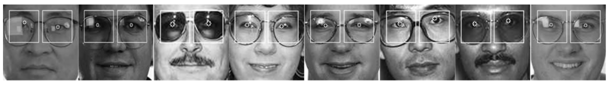
Table 1: Localization accuracy on standard face database

p(%)	FERET(%)	BioID(%)	JAFFE(%)
1	79.18	71.79	84.62
3	93.19	85.78	90.36
5	**98.92**	**95.87**	**99.26**
8	96.19	92.63	97.10
10	95.71	86.06	96.51

The eye localization by the proposed method are shown in Fig. 7 in which localization results of frontal facial images from FERET, BioID and JAFFE are shown in Fig.7 (a) (b) (f), respectively, with illumination and facial expression changes. The position of the eye is indexed by white "o"s in the images. Fig. 7(c) exhibits some localization results of facial images with multi-poses from FERET, from which we can confirm the robustness of the proposed method to the face pose changed. Some results of facial images with all kinds of glasses from FERET and BioID are shown in Fig. 7(b) and (e), it is evident that our method is able to eliminate the interference of glasses and achieve precise eye localization. Through the analysis of some mis-located samples (see Fig. 7(g)), it is found that the intense reflection from glasses leads to the pupils be invisible, which is the main reason for mis-localization.

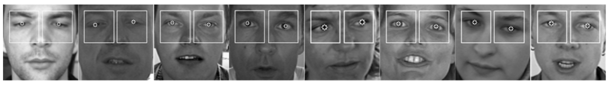


(a)

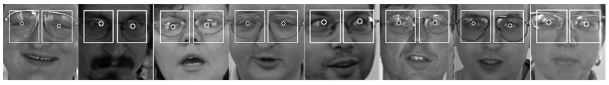


(b)

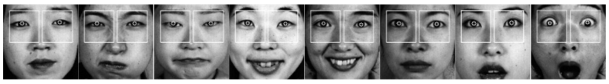
(c)



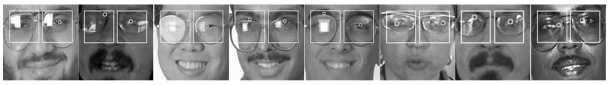
(d)



(e)



(f)



(g)

Figure 7: Some result images (a) Some frontal samples in FERET (b) Some samples with glasses in FERET (c) Some profile samples in FERET (d) Some frontal samples in BioID (e) Some samples with glasses in BioID (f) Some samples from JAFFE (g) Some samples that eyes are wrongly located

3.2 Comparison with Other Methods

Adaboost (P. Viola, 2001) is a general and effective method for object detection. In order to obtain the Adaboost eye detector, we choose 4532 images of eye with a resolution of 20×20 as positive examples. The negative examples are obtained by a bootstrap process (P. Viola, 2001), and during every strong classifier training process, 2236 images without eye are employed as negative examples. Finally we obtained an eye detector with 16 strong classifiers. Experiments of eye localization with trained Adaboost detector are carried out in three standard face database previously mentioned and the results are shown in Table 2. The results of HPF (Z.H. Zhou, 2004) and our method are also enumerated in Table II. Through comparison among the experiment results of three different methods, it is evidently illuminated that our method can obtain great accuracy on the basis of little time consuming. In addition, without any training process, our method is convenience to implement.

Table 2: Comparison with other methods on localization accuracy and average time consumed

	Our Method	Adaboost	HPF
FERET(%)	98.92	98.93	—-
BioID(%)	95.87	96.03	94.81
JAFFE(%)	99.26	99.47	97.18
Time(ms)	0.56	10.36	0.49

3.3 Performance on Real-Time Face Recognition System

For an image of the size 128×128, it approximately takes 0.56ms to locate eyes position by a PC of 1.8G CPU and 256M memory. Moreover, the proposed method has been added into the face recognition system developed by Research Center of Intelligent Robotics (RCIR), Shanghai Jiaotong University (See Fig. 8). The rapid and accurate eye localization ensure the system to achieve better precision rate.

Figure 8: Real-time eye localization

4 CONCLUSIONS

In this paper, a novel method is proposed, which achieves rapid and accurate eye localization by making use of the static rules of human face on the basis of uncomplicated computation. In order to eliminate the interferences (i.e. hair, eyebrow, glasses) around eye region, we improve the general projection method by projection peak analysis. Experimental results show that our method is effective, accurate and rapid in eye localization, especially when the face poses, illuminations, expressions, and accessories varied. Owing to the lower computation cost, our method can satisfy the requirement of real-time face recognition system well. Although high performance is achieved by the proposed method, we should also mention that one drawback of our method lies in the situation when the pupils are occluded for some reasons, i.e. the intense light reflection from glasses. Further efforts will be focused on how to solve this drawback.

ACKNOWLEDGMENT

The work reported was partially sponsored by National Natural Science Foundation of China (No. 60675041) and Program for New Century Excellent Talents of Ministry of Education, China (NCET-06-0398).

REFERENCES

D. Reisfeld, H. Wolfson, Y. Y. (1995). Context free attentional operators: the generalized symmetry transform. *International Journal of Computer Vision*, 14(2):119–130.

G.H. Li, X.P. Cai, X. L. (2006). An efficient face normalization algorithm based on eyes detection. In *Proceedings of International Conference on Intelligent Robots and Systems*.

Kanade, T. (1973). *Picture Processing System by Computer Complex and Recognition of Human Faces*. PhD thesis, Kyoto University.

M. J. Lyons, S. Akamatsu, M. K. (1998). Coding facial expressions with gabor wavelets. In *Proceeding of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition*.

O. Jesorsky, K. Kirchberg, R. F. (2001). Robust face detection using the hausdorff distance. In *Proceeding of the 3rd International Conference on Audio- and Video-based Biometric Person Authentication*.

P. J. Phillips, H. Wechsler, J. H. (1998). The feret database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing*, 16(5):295–306.

P. Viola, M. J. (2001). Rapid object detection using a boosted cascade of simple features. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*.

S.G. Shan, Y.Z. Chang, W. G. (2004). Curse of misalignment in face recognition: Problem and a novel mis-alignment learning solution. In *Proceeding of the 6th IEEE International Conference on Automatic Face and Gesture Recognition*.

T. F. Cootes, G. J. Edwards, C. J. T. (1998). Active appearance models. In *Europeon Conference on Computer Vision*.

T. F. Cootes, G. J. Edwards, C. J. T. (2001). Active appearance models. *IEEE Trans. Pattern Anal. Machine Intell.*, 23(6):681–685.

T.Kawaguchi, D. Hikada, M. R. (2000). Detection of the eyes from human faces by hough transform and separability filter. In *Proceedings of International Conference of Image Processing*.

Z.H. Zhou, X. G. (2004). Projection functions for eye detection. *Pattern Recognition*, 37(5):1049–1056.