Proceedings of the 7th
World Congress on Intelligent Control and Automation
June 25 - 27, 2008, Chongqing, China

A Novel Eye Localization Method Based on Spectral Residual Model*

Fei Yang, Jingwen Dai and Dan Liu

Department of Automation & Research Center of Intelligent Robotics
Shanghai Jiaotong University, Shanghai, 200240, China

{yangfei, daijingwen, ldan}@sjtu.edu.cn

Abstract **– The eyes are the most important facial landmarks on human face, for both human computer communication and face normalization. It is crucial to detect and locate the eyes accurately. This paper introduces a novel and simple method for eye localization. Based on the special gray lever distribution in the eye region, the spectral residual model is exploited to extract the rough eye region primarily. And then a coarse-to-fine searching strategy is adopted to locate the exact eye position. Compared with the Gabor transformation, the spectral residual transformation is lower in computational cost and more accurate in eye extraction. Some other experiments are carried out to investigate its robustness on variations of pose, orientation, scale, expression and illumination. More important, the proposed eye extraction method can be easily incorporated in other eye detection approaches as a combined processing module or employed in a real-time application.**

Index Terms - Eye localization, logarithmic spectrum, spectral residual model.

I. INTRODUCTION

The eyes are exceedingly salient on human face, which contain a great deal of information. They play important roles in understanding and interpreting a person's desires, needs and emotional states. These properties could be exploited to develop a large number of applications, e.g., multi-modal human-computer interface (HCI), facial expression analysis, driver awareness system [1]. Furthermore, over the last twenty years or so, face recognition has become an actively developing research field. The face in the input image must be normalized to a standard size, location and orientation before it is matched to database faces, and the accuracy of face normalization affects greatly the following analysis of face recognition. Currently such normalization is usually done by using some geometrical measurements among facial features. The eyes can be considered salient and relatively stable features in comparison with other facial features, and especially for their relatively constant interocular distance on the face. When we detect facial features, it is advantageous to detect eyes before the detection of other facial features [2, 3]. Therefore, the eye localization has a large impact on face normalization and the performance of a face recognition system. The following paragraphs summarize the various methods that have been utilized so far in the filed of eye detection.

Many eye detection methods have been developed during the last two decades. Generally speaking, eye detection can be divided into eye position detection and eye contour detection [2]. In this paper, eye detection means eye position detection or eye localization. The image based passive approaches are more widely used than the active infrared based approaches for no extra equipment is needed, thus this paper just concentrates on the former. Roughly speaking, there are three major approaches for eye detection: template based approaches, feature based approaches and appearance based approaches [4].

In the template based approaches, representative methods are deformable template based algorithms presented by Yuille et al. [5] and extended by Lam and Yan [6]. In this method, a generic eye model is designed first and the eye position can be obtained through a recursive searching process. However, the locating results are heavily affected by the eye model initialization and the image contrast. Moreover, deformable template suffers from two limitations. First, it is computationally expensive. Second, the weighting factors for energy terms have to be determined manually [4].

The feature based approaches explore the characteristics of the eyes to identify some distinctive features around the eyes. Projection function is widely used to locate the landmarks (i.e., corner points) of an eye for its low computational complexity and high effectiveness [7, 8]. It is observed that some eye landmarks are with relatively high contrast, such as the boundary points between eyelid and eyeball. Some other published papers employ the color distributions of the sclera and the flesh to extract the eye region [9]. These techniques are mostly fast and effective. However, these methods will fail if the eye is closed or partially occluded by hair. The eyebrow and face orientation may also degrade the performance of the discriminative function. Its performance depends on the accuracy of candidate eye window detection to a great extent.

The appearance based methods detect eyes based on face patterns. In these approaches, Jee et al. [10] detected face and eyes in such case that the eye objects were isolated from other face objects using classifier, which was trained using SVM (Support Vector Machine). Niu et al. [4] located eyes with classifier using 2D cascade AdaBoost algorithm. ASM (Active Shape Model) and AAM (Active Appearance Model) are also used for eye localization in some works are published for landmarks localization. However, to distinguish eyes from

* This work is partially supported by NSFC Grant No. 60428303.

978-1-4244-2114-5/08/$25.00 © 2008 IEEE.

other face objects, these methods usually need a large amount of training data under different face orientations and different illumination conditions before detection. And the geometric characteristics of people of different races are largely different that render this technique ineffective in practice.

Other methods that have been adopted for eye detection include wavelets, principal component analysis, fuzzy logic, neural networks, evolutionary computation and hidden Markov models. To obtain a higher performance in various applied environment, combination methods are usually adopted in eye localization applications. The coarse-to-fine searching strategy is a promising way in this field. It first detects the eye window and then searches and locates the precise eye position. The Gabor transformation is a common used approach to extract eye region in the first stage, and some encouraging works using this technique have been finished by far [11]. However, the drawback of this approach is that the algorithms based on the Gabor wavelets need heavy computation and which limits its use in real-time application.

In this paper, we propose a fast and robust method for eye localization. The method analyzes face image in statistical spectral domain and then extracts features on the face. The effect of the method is similar to the Gabor transformation in a way. However, it is lower in computational cost and more accurate.

The remainder of this paper is organized as follows. Section II describes the spectral residual model in detail. In section III, the process of eye localization based on spectral residual model is briefly illustrated. Experimental results are given and analyzed in Section IV and conclusions are made in Section V.

II. Spectral Residual Model

Hou and Zhang presented the spectral residual method for salient object detection in [12]. In the paper, they applied the model to statistical scene analysis and tried to draw the biological relevance to human vision through multiple experimental studies. Their work laid emphasis on this attempt. We do not continue their efforts on simulating human preattentive vision functionally. However, we are just inspired by the spectral residual model of images. The model is described as follows.

First, the spectral residual model adopts the logarithmic spectrum representation of an image. The logarithmic spectrum $L(F)$ can be obtained by

$$L(F) = \log(M(F)), \qquad (1)$$

where $M(F)$ denotes the Fourier spectrum of the image.

To our knowledge, the logarithmic spectral amplitude method can be traced back to [13] developed by Ephraim and Malah, was proved very efficient in reducing musical residual noise phenomena for speech enhancement. It was then used for extracting the information on the echo of organ tissues from a preprocessed input image in the field of ultrasonic imaging [14]. In these cases, the power spectrum is, moreover, almost always represented on a logarithmic scale. When the gain applied to a signal varies, the shape of the logarithmic power spectrum is preserved; the spectrum is simply shifted up or down. Because of this translation invariant property, an "innovation component" which appears as convolutional effect on the waveform and as multiplicative effect on the linear power spectrum, becomes simply additive constant on the logarithmic power spectrum [15]. Now the logarithmic spectral amplitude method is very useful in a variety of signal and image processing applications [16-20].

Given an image of size $M \times N$, let $I(x,y)$ be the gray level distribution of the image. As we are only concerned with digital images, we restrict this discussion to the Discrete Fourier Transform (DFT). The definition of the Discrete Fourier Transform is given by the equation

$$F(u,v) = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} I(x,y)e^{-j2\pi(ux/M+vy/N)}, \qquad (2)$$

where u = 0, 1, 2, …, M-1 and v = 0, 1, 2, …, N-1.

The input image $I(x,y)$ is the spatial domain equivalent. The output $F(u,v)$ of the transform represents the image in the frequency domain. The image in the spatial and frequency domain is of the same size. In order to decrease the number of required computations, we employ the Fast Fourier Transform (FFT) to compute the DFT.

Note that $I(x,y)$ is the image and is real, but $F(u,v)$ (abbreviate as F) is the DFT and is complex. Generally, F is represented by its magnitude $M(F)$ and phase $P(F)$ rather that its *REAL* and *IMAGINARY* parts, where:

$$M(F) = \sqrt{REAL(F)^2 + IMANIARY(F)^2}, \qquad (3)$$

$$P(F) = \arctan(REAL(F)/IMANIARY(F)). \qquad (4)$$

The averaged curve of logarithmic spectra indicates a local linearity and it is approximated by convolution operation

$$A(F) = h_n * L(F), \qquad (5)$$

where h_n is an $n \times n$ matrix defined by

$$h_n = \frac{1}{n^2}\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}.$$

The spectral residual R(F) is defined as follows:

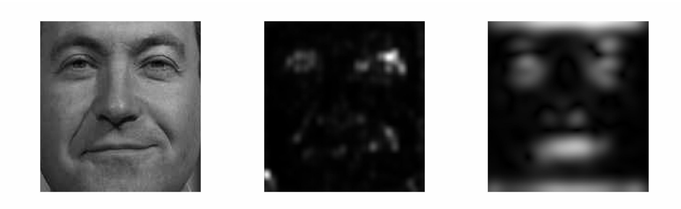
$$R(F) = L(F) - A(F). \qquad (6)$$

Therefore the spectral residual $R(F)$ can be obtained by (1), (2), (3), (5) and (6). Finally, $R(F)$ is transformed back to the spatial domain by

$$S(x,y) = abs(\mathbb{F}^{-1}\{\exp(R(F) + i * P(F))\}), \qquad (7)$$

where $\mathbb{F}^{-1}\{\bullet\}$ denotes the inverse Fourier transform.

Up till now, the spectral residual representation of the input image $I(x,y)$ is derived as $S(x,y)$. The output image $S(x,y)$ should be smoothed with a Gaussian filter in order to achieve better effect of features extraction. We name the above whole transformation process as Spectral Residual Transformation (SRT). The output image of SRT is named Spectral Residual Image (SRI) in the rest of this paper. Fig. 1 shows original face image, transformed results by SRT and by the Gabor transformation, respectively. The sample face image is selected from FERET database.



(a) (b) (c)

Fig. 1 Original face image and its transformed images. (a) Original image. (b) Corresponding spectral residual image. (c) The original image convoluted with a horizontal-orientation and fifth-scale Gabor kernel.

III. Eye Localization Method

Recently, face detection technique has become mature, so we prefer to implement eye localization after obtaining the face image rather than detect eye directly. We first detect face based on Viola's boosted cascade algorithm [21] and then start eye detection on the assumption that the face has already been accurately detected. The flowchart of our method is shown in Fig. 2.

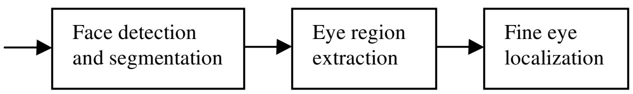


Fig. 2 Flowchart of the proposed eye localization method based on spectral residual model.

A. *Eye region extraction*

Based on the spectral residual model, the eye region can generally be segmented by following steps:

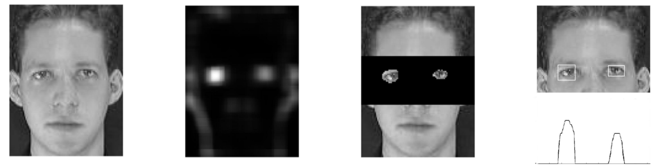
1) SRT: input an image, implement spectral residual transformation and output the SRI.

2) Binarization: convert SRI to a binary image by thresholding. We use single threshold to segment the eye region roughly. The threshold lever can be decided by empirical way. The selection of threshold is a trade-off problem between false alarm and neglect of eye regions.

3) Morphological processes: in order to detect the eye region precisely, we employ binary dilation and erosion to expand the eye region. Then we remove all connected components (regions) that have fewer pixels from a binary image.

4) Integral projection: according to the prior structure knowledge of a face image, we assume that the eyes region locates in the rectangle area in the relative position of a face image. For example, in Fig. 3, we assume that the eyes locate in the rectangle area whose top is H/3 from the top of the image and its bottom is H/3 from the bottom of the image (H is the height of the image). In the binary image, we perform integral projection within the rectangle area in vertical direction. We can separate the left half face region from the right half face region roughly due to analyzing the vertical projection curve. And then we do integral projection in horizontal direction within the left and the right half face region, respectively.

5) Segment the eye region: search maximum connected regions in the left half face region and the right half face region respectively. We refer to the integral projection curve and usually consider the maximum connected region as the potential left or right eye region. So far, we extract the relatively precise eye regions using the method based on the residual spectral model, which is called SReye method in the rest of this paper.



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

Fig. 3 Eye region extraction. (a) Original image. (b) SRI. (c) Binarization in the limited region. (d) Vertical projection curve and eye region extraction.

B. *Fine eye localization*

Generally, we can consider the center of the rectangles derived above as the location of eyes, for the exciting eye region extraction performance of proposed spectral residual model. But if we want to achieve more accurate effect of eye localization, we should employ other techniques, e.g., a simple integral projection function is introduced in this paper. In Fig. 4, we illustrate this coarse-to-fine strategy by applying integral projection function to a previously extracted eye region. We can obtain two integral projection curves in vertical and horizontal directions, for the pupil and iris are darker than the sclera, eyelid or skin in gray scale distribution image. We search the peak of the integral projection curve from the two ends to the middle to determine the edge of iris region, and the corresponding position is considered as a more accurate location of eye.

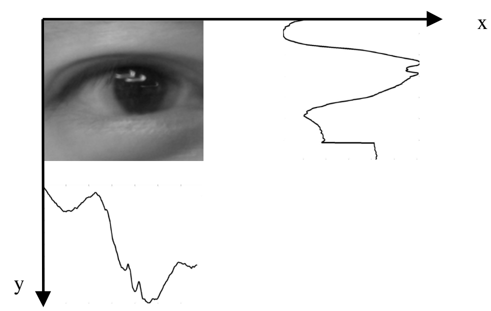


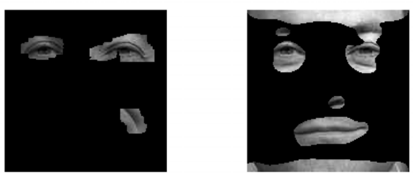
Fig. 4 Using integral projection function to locate the x-coordinate and y-coordinate of the eye.

In this section, a general eye localization method is introduced by using a face image chosen from the ORL face Database. For better effect of visual illustration, this face image is not cropped to a local face image such as the Fig. 1(a), which contains no ear and few hairs. In fact, it is easy to get more precise local face image by means of current face detection algorithms. In this case, our eye localization method would perform better and need less other additional processes. And if we combine the SReye method with another more complex method in the second stage, the overall performance of eye localization could be improved.

IV. Experiments And Analyses

The SReye algorithm is proved very simple and effective by the above illustration. Note that the SReye algorithm can extract eye region precisely without the interference of eyebrow, which appears as a common problem when using complexion model or the Gabor transformation model to extract eye regions. With these algorithms, separating eye from eyebrow is often an intractable task. In order to compare the performance of the SReye algorithm and that of the Gabor transformation, we take Fig. 1 for example, and the results are shown in Fig. 5. Fig. 5(a) shows that the eye regions extracted by the SReye algorithm exclude eyebrow. Fig. 5(b) shows that the eye regions extracted by the Gabor transformation include parts of eyebrow.

Furthermore, because of the lower computational cost, the SReye algorithm is faster than the Gabor transformation. In order to compare the computational cost of the two algorithms, we test their time performances on a PC equipped with a Intel Pentium IV 1.8G CPU and a 256MB RAM. We program the two algorithms and run them using Microsoft Visual C++ 6.0 development environment which runs on a Window desktop operating system. The size of the sample face image is 92×112. The SRT processing takes about 12ms. On the contrary, the same sample image is convoluted with one Gabor kernel and this operation takes about 55ms, practically more than one Gabor kernels are needed in the literature.



(a) (b)

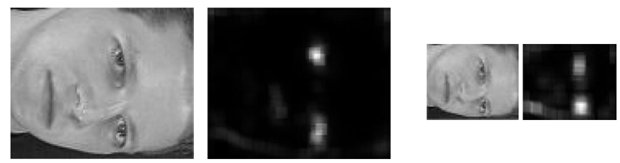
Fig. 5 Results of the two algorithms applied to the sample face image (a) Face under mask of binary SRI. (b) Face under mask of binary Gabor transformation image.

Other experiments are carried out to assess how robust the SReye algorithm is by various face images. In order to investigate the performance of the SReye model across variation of face pose, Fig. 6 is given to show the results of the SRT. It shows that eye region can be highlighted regardless of the changes of the face pose.

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

Fig. 6 Four face poses and their SRI. (a) Turn right. (b) Turn left. (c) Tilt up. (d) Tilt down.

As is well known that, the Fourier domain is rotated according to rotation of the input image and then the Fourier image can be re-transformed to the spatial domain, so the Fourier transform allows for rotation and scale invariance. Combined with the logarithmic transform, also rotation and scale invariance can be achieved. Therefore, we can expect that our SReye algorithm is rotation and scale invariant too. In order to test this assumption, the face image on the upper left of Fig. 6 is rotated by 90° and reduced to 1/4 of the original size. Then the SReye algorithm is applied to test them respectively. The results are shown in Fig. 7, which prove the orientation and scale invariance of our SReye algorithm. In contrast, the convolution kernel should be adjusted to the orientation variance of face image when the Gabor transformation is employed, and the adjustment usually couldn't be fulfilled automatically.



(a) (b)

Fig. 7 Robustness of the SReye algorithm on orientation and scale variance.

In order to evaluate the performance of the SReye algorithm, we choose 1500 images from FERET database [23] to build the experimental test set, which consists of face images with different illumination, pose and facial expression. The experimental result achieves high accuracy over 95%. Some correctly located samples are given in Fig. 8. Experimental results show that our method can accurately locate the eyes, and is robust to variations of face pose, expression and illumination.

Fig. 8 Eyes are correctly located.

V. Conclusions

This paper proposes a novel eye localization method based on spectral residual model. First, the gray scale face image is transformed by the spectral residual model. Second, the eye region is extracted from the face image. Subsequently, a simple integral projection approach is employed to locate the exact eye position. Even though the SRT algorithm is simple, the results are promising and could be further improved by combination with more powerful techniques. Some sample face images are also given to illustrate its robustness of eye extraction across rotation, orientation and scale variance. Our method makes no use of memory, training or a database. More important, it can be easily incorporated in other approaches as a front-end filter or exploited in a time-constrained application of eye localization.

This paper describes work in progress and experimental results are limited but encouraging. Further experiments attempt to quantify the limits of the technique. Also several technical details must be resolved to provide improved results. It is also a promising way to combine this method with other sophisticated methods in order to improve eye localization jointly.

ACKNOWLEDGMENT

The first author would like to thank Xiaodi Hou, Xi Chen and Zhiqiang Li for their valuable discussions.

REFERENCES

[1] H. Tan and Y. Zhang, "Detecting eye blink states by tracking iris and eyelids," *Pattern Recognition Letters*, vol. 27, no. 6, pp. 667-675, 2006.

[2] T. Kawaguchi and M. Rizon, "Iris detection using intensity and edge information," *Pattern Recognition*, vol. 36, no. 2, pp. 549-562, 2003.

[3] W. Zhang, et al, "Precise eye localization with AdaBoost and fast radial symmetry," *Computational Intelligence and Security 2006*, pp. 1068-1077, 2006.

[4] Z. Niu, et al, "2D cascaded AdaBoost for eye localization," *Proceeding of International Conference on Pattern Recognition ICPR2006*, vol. 2, pp. 1216-1219, August 2006.

[5] A. Yuille, P. Hallinan, and D. Cohen, "Feature extraction from faces using deformable templates," *International Journal of Computer Vision*, vol. 8, no. 2, pp. 99-111, 1992.

[6] K. M. Lam and H. Yan, "Locating and extracting the eye in human face images," *Pattern Recognition*, vol. 29, no. 5, pp. 771-779, 1996.

[7] Z. Zhou and X. Geng, "Projection functions for eye detection", *Pattern Recognition*, vol. 37, no. 5, pp. 1049-1056, 2004.

[8] G. Feng and P. Yuen, "Variance projection function and its application to eye detection for human face recognition," *Pattern Recognition Letters*, vol. 19, no. 9, pp. 899-906, 1998.

[9] H. K. Jee, K. H. Lee, and S. B. Pan, "Eye and face detection using SVM," *Proc. of Int. Conf. on Intelligent Sensors, Sensor Networks and Information Processing*, pp. 577-580, 2004.

[10]G. Feng and P. Yuen, "Multi-cues eye detection on gray intensity image," *Pattern Recognition*, vol. 34, no. 5, pp. 1033-1046, 2001.

[11]P. Yang, et al, "A novel pupil localization method based on Gaboreye model and radial symmetry operator," *International Conference on Image Processing 2004*, vol. 1, pp. 67 – 70, October 2004.

[12]X. Hou and L. Zhang, "Saliency detection: a spectral residual approach," *IEEE Conference on Computer Vision and Pattern Recognition 2007*, pp. 1-8, June 2007.

[13]Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 443–445, April 1985.

[14]S. J. Park, J. G. Son, and N. C. Kim, "Organ recognition in ultrasound images using log power spectrum," *Medical Imaging 2002: Ultrasonic Imaging and Signal Processing, Proc. SPIE*, Vol. 4687, pp. 387-394, April, 2002.

[15]M. J. Hunt, "Signal representation," *Survey of the State of the Art in Human Language Technology*, chapter 1.3, p.10, Cambridge University Press, March 1998.

[16]I. Cohen, "Optimal speech enhancement under signal presence uncertainty using log spectral amplitude estimator," *IEEE Signal Processing Letters*, vol. 9, no.4, pp. 113-116, April 2002.

[17]A. Oliva and A. Torralba, "Modeling the shape of the scene: a holistic representation of the spatial envelope," *International Journal of Computer Vision*, vol. 42, no. 3, pp. 145–175, February 2001.

[18]A. Torralba and A. Oliva, "Depth estimation from image structure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1226–1238, February 2002.

[19]A. Torralba, "Modeling global scene factors in attention," *Journal of Optical Society of America A*, vol. 20, no. 7, pp. 1407-1418, February 2003.

[20]Y. Wang, J. Wu, and Z. Wang, "A two-dimensional robust algorithm based on Mel-bank log-spectrum," *IEEE Symposium International Communications and Information Technology*, vol. 1, pp. 759-763, October 2005.

[21]P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Proc IEEE Conference on Computer Vision and Pattern Recognition 2001*, pp. 511-518, 2001.

[22]AT&T Laboratories Cambridge, *ORL Database*, 1992-1994.

[23]P. Phillips, et al, "The FERET evaluation methodology for face-recognition algorithm," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090-1034, 2000.