

Eye-State Action Unit Detection by Gabor Wavelets

Ying-li Tian¹ Takeo Kanade¹ and Jeffrey F. Cohn^{1,2}

¹ Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213

² Department of Psychology, University of Pittsburgh, Pittsburgh, PA 15260

Email: {yltian, tk}@cs.cmu.edu jeffcohn@pitt.edu

<http://www.cs.cmu.edu/~face>

Abstract Eyes play important roles in emotion and paralinguistic communications. Detection of eye state is necessary for applications such as driver awareness systems. In this paper, we develop an automatic system to detect eye-state action units (AU) based on Facial Action Coding System (FACS) by use of Gabor wavelets in a nearly frontal-viewed image sequence. Three eye-state AU (AU 41, AU42, and AU43) are detected. After tracking the eye corners in the whole sequence, the eye appearance information is extracted at three points of each eye (i.e., inner corner, outer corner, and the point between the inner corner and the outer corner) as a set of multi-scale and multi-orientation Gabor coefficients. Then, the normalized Gabor coefficients are fed into a neural-network-based eye-state AU detector. An average recognition rate of 83% is obtained for 112 images from 17 image sequences of 12 subjects.

1. Introduction

Facial Action Coding System (FACS) action unit recognition attracts attention for facial expression analysis [1, 5, 6, 16, 14]. Eyes play important roles in emotion and paralinguistic communications. Detection of eye state (i.e. whether the eye is open or closed) is also necessary for applications such as driver awareness systems. Although many methods exist for eye feature extraction and eye tracking, detecting qualitative changes of eye states is relatively undeveloped [2, 4, 7, 9, 18, 19]. In our facial expression analysis system, we developed a dual-state eye model for eye tracking [15]. In that paper, two eye states are detected by geometry feature information of the iris. However, when the eye is narrowly-opened or the iris is difficult to detect, the eye state may be wrongly identified as closed. We believe that the eye appearance information will help to solve this difficulty and increase the number of AU that can be recognized in the eye region.

Recently, Gabor wavelet has been applied to image analysis, face recognition, facial expression analysis [3, 5, 10, 13, 17, 20]. This research suggests that Gabor wavelet is a promising tool to extract facial appearance information.

In this paper, we develop a facial appearance information based eye-state AU detection system to detect AU 41 (upper-lid droop), AU 42 (slit), and AU 43 (closed). Figure 1 depicts the overview of the eye-state AU detection system. First, the face position is detected and the initial positions of the eye corners are given in the first

frame of the nearly frontal face image sequence. The eye corners then are tracked in the image sequence. Next, a set of multi-scale and multi-orientation Gabor coefficients of three eye points are calculated for each eye. Finally, the normalized Gabor coefficients are fed into a neural-network-based detector to classify three states of the eye.

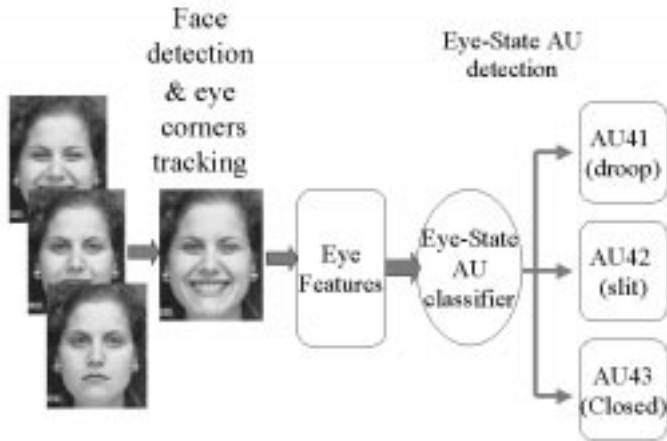


Fig. 1. Eye state detection system.

2. Eye-State AUs

In FACS, there are nine eye-state AUs i.e. AU5, AU6, AU7, AU41, AU42, AU43, AU44, AU45, and AU46. We recognized AU5 (eye wide), AU6 (infra-orbital raise), and AU7 (lower-lid raise) in previous work by feature-based information [11, 14]. In this paper, we recognize AU41, AU42, and AU43 by appearance information. The examples of these AUs are shown in Table 1. We classify these AU into three eye states: open (AU41), very narrow (AU42), and closed (AU43). The closed eye is defined as closure of the eyelid brought about by total relaxation of the levator palpebrae superioris muscle, which controls the motion of the upper eyelid. The closed eye may also involve weak contraction of the orbicularis oculi pars palpebralis muscle, a sphincter muscle that surrounds the eye orbit. The very narrow eye is defined as the eyelids appearing as narrowed as possible without being closed. Their appearance resembles a slit, the sclera is not visible, and the pupil may be difficult to distinguish. Relaxation of the levator palpebrae superioris is not quite complete. The open eye is defined as a barely detectable drooping of the upper eyelid or small to moderate drooping of the upper eyelid. See paper [8] for complete list of FACS action units.

3. Localizing Eye Points

To extract information about change of eye appearance, the eye position first must be localized. Three points for each eye are used. As shown in Figure 2, these are the inner and outer corners, and the mid-point between them. At each point, multi-scale and multi-orientation Gabor wavelet coefficients are calculated.

Inner corner: We found that the inner corners of the eyes are the most stable features

Table 1. Eye states and corresponding FACS action units




Open	Very narrow	Closed
AU41	AU42	AU43/45/46
		
Upper-lid is slightly lowered.	Eyes are barely opened.	Eyes are completely closed.



Fig. 2. Three points for each eye are used to detect eye states.

in a face and are relatively insensitive to deformation by facial expression. We assume the initial location of the inner corner of the eye is given in the first frame. The inner corners of the eyes then are automatically tracked in the subsequent image sequence using a modified version of the Lucas-Kanade tracking algorithm [12], which estimates feature-point movement efficiently with sub-pixel accuracy.

We assume that intensity values of any given region (feature window size) do not change but merely shift from one position to another. Consider an intensity feature template $I_t(x)$ over a $n \times n$ region R in the reference image at time t . We wish to find the translation d of this region in the following frame $I_{t+1}(x + d)$ at time $t + 1$, by minimizing a cost function E defined as:

$$E = \sum_{x \in R} [I_{t+1}(x + d) - I_t(x)]^2. \quad (1)$$

The minimization for finding the translation d can be calculated in iterations (See paper[15] for details).

Outer corner and mid-point: Because the outer corners of the eyes are difficult to detect and less stable than the inner corners, we assume they are collinear with the inner corners. The width of the eye is obtained from the first frame. If there is not large head motion, the width of the eye will not change much. The approximate positions of the outer corners of eyes are calculated by the position of the inner corners and the eye widths.

After obtaining the inner and outer corners of the eyes in each frame, the middle points are easy to calculate from the position of the inner- and outer corners of the eyes.

4. Eye Appearance Information

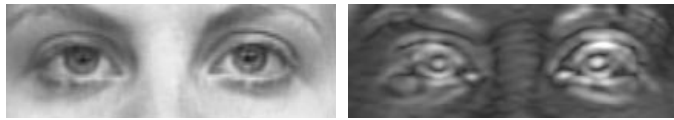
We use Gabor wavelet to extract the information about change of eye appearance as a set of multi-scale and multi-orientation coefficients. The response image can be written as a correlation of the input image $I(x)$, with

$$a_{\mathbf{k}}(\mathbf{x}_0) = \int I(x)p_{\mathbf{k}}(\mathbf{x} - \mathbf{x}_0)dx, \quad (2)$$

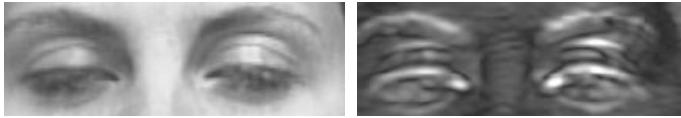
where the Gabor filter $p_{\mathbf{k}}(\mathbf{x})$ can be formulated [3]:

$$p_{\mathbf{k}}(\mathbf{x}) = \frac{k^2}{\sigma^2} \exp\left(-\frac{k^2}{2\sigma^2}x^2\right) \left(\exp(i\mathbf{k}\mathbf{x}) - \exp\left(-\frac{\sigma^2}{2}\right)\right) \quad (3)$$

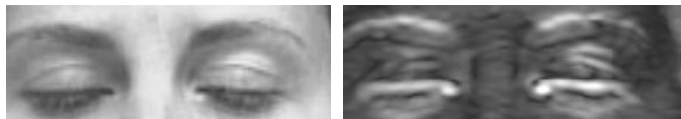
where \mathbf{k} is the characteristic wave vector.



(a) AU41 (open)



(b) AU42 (very narrow)



(c) AU43 (closed)

Fig. 3. Gabor images for different states of the eyes when the spatial frequency= $\frac{\pi}{4}$ in horizontal orientation.

In our system, we use $\sigma = \pi$ and three spatial frequencies with wavenumbers $k_i = (\frac{\pi}{2}, \frac{\pi}{4}, \frac{\pi}{8})$ and six orientations from 0 to π differing in $\pi/6$. Only the magnitudes are used because they vary slowly with the position while the phases are very sensitive. Therefore, for each point of the eye, we have 18 Gabor wavelet coefficients. Figure 3 shows the examples of different eye state and the corresponding Gabor filter responses for the second spatial frequency ($k_i = (\frac{\pi}{4})$) and horizontal orientation. The Gabor coefficients appear highly sensitive to eye states even when the images of eyes are very dark.

5. Eye State Detection

5.1. Image Databases

We have been developing a large-scale database for promoting quantitative study of facial expression analysis [8]. The database currently contains a recording of the facial behavior of 210 adults who are 18 to 50 years old; 69% female and 31% male; and 81% Euro-American, 13% Africa-American, and 6% other groups. Subjects sat directly in front of the camera and performed a series of facial expressions that included single AUs and AU combinations. To date, 1,917 image sequences of 182 subjects have been FACS coded for either the entire sequence or target action units. Approximately fifteen percent of the 1,917 sequences were coded by a second certified FACS coder to validate the accuracy of the coding.

In this investigation, we focus on AU41, AU42, and AU43. We selected 33 sequences from 21 subjects for training and 17 sequences from 12 subjects for testing. The data distribution of training and test data sets for eye states is shown in Table 2.

Table 2. Data distribution of training and test data sets.

Data Set	Eye states included			
	Open	Narrow	Closed	Total
<i>Train</i>	92	75	74	241
<i>Test</i>	56	40	16	112

To assess how reliably trained observers could make these distinctions, two research assistants with expertise in FACS independently coded image sequences totaling 139 frames. Inter-observer agreement between them averaged 89%. More specifically, inter-observer agreement was 94% for AU41, 84% for AU42, and 77% for AU43. For FACS coders, the distinction between very narrow (AU 42) and closed (AU 43) was more difficult.

5.2. Neural network-based eye state detector

As shown in Figure 4, we use a three-layer neural network with one hidden layer to detect eye states. The inputs to the network are the Gabor coefficients of the eye feature points. The outputs are the three states of the eyes.

In our system, the inputs of the neural network are normalized to have approximately zero mean and equal variance.

5.3. Experimental Evaluations

We conducted three experiments to evaluate the performance of our system. The first is detection of three states of the eye by using three feature points of the eye. The second is the investigation of the importance of each feature points to eye state detection. Finally, we study the significance of image scales.

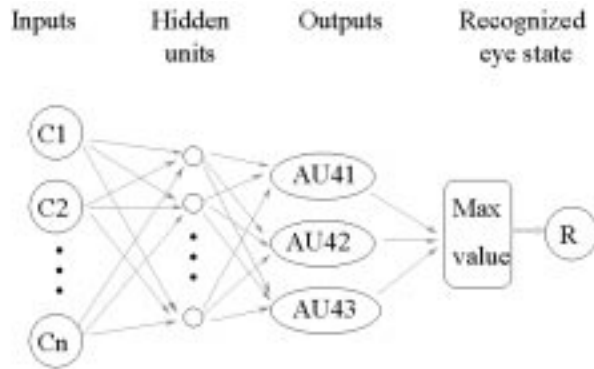


Fig. 4. Neural network-based detector for three states of the eye. The inputs are the Gabor coefficients of the eye feature points, and the output is one label out of the three states of the eyes.

Results of eye state detection: Table 3 shows the detection results for three eye states when we use three feature points of the eye and three different spatial frequencies of Gabor wavelet. The average recognition rate is 83%. More specifically, 93% for AU41, 70% for AU42, and 81% for AU43. These are comparable to the reliability of different human coders.

Compared to expression analysis, three eye states are unnecessary for driver awareness systems. Very narrow eye and closed eye can be combined into one class in driver awareness systems. In that case, the accuracy of detection increases to 93%.

Table 3. Detection results by using three feature points of the eye. The numbers in bold means can be combined into one class in driver awareness systems.

	Recognized eye states		
	Open	Narrow	Closed
Open	52	4	0
Narrow	4	28	8
Closed	0	3	13
Recognition rate of three states: 83%			
Recognition rate of two states: 93%			

Importance of eye feature points: We also carried out experiments on detection of the three eye states by using one point (the inner corner) of the eye and two points (the inner corner and the middle point) of the eye. The recognition rates for using different points of the eye are list in Table 4. The recognition rate of 81.3% for two points is close to that (83%) for three points. When only the inner corner of the eye is used, the recognition rate decreases to 66%. When only the outer corner of the eye is used, the recognition rate decreases to 38%. The inner corner and middle point carry more useful information than the outer corner for eye state detection.

Table 4. Detection results for three eye states by using different feature points of the eye. We found that the inner corner and middle point carry more useful information than the outer corner for eye state detection.

Used eye feature points				
1 point		2 points		3 points
Inner corner	Outer corner	Outer & middle	Inner & middle	Inner, outer, & middle
66%	38%	61.2%	81.3%	83%

Significance of different image scales: To investigate the effects of the different spatial frequencies, we evaluated the experiments by using two of the spatial frequencies (i.e., wavenumber $k_i = (\frac{\pi}{2}, \frac{\pi}{4}, \frac{\pi}{8})$). Table 5 shows the resulting comparisons. An 80% recognition rate is achieved when we use $k_i = (\frac{\pi}{4}, \frac{\pi}{8})$. It is higher than the recognition rate 74% when we use the higher spatial frequencies (i.e., $k_i = (\frac{\pi}{2}, \frac{\pi}{4})$).

Table 5. Detection results for three eye states by using different spatial frequencies.

Spatial frequencies		
$k = (\frac{\pi}{2}, \frac{\pi}{4})$	$k = (\frac{\pi}{4}, \frac{\pi}{8})$	$k = (\frac{\pi}{2}, \frac{\pi}{4}, \frac{\pi}{8})$
74%	80%	83%

6. Conclusion

In this paper, we developed an appearance-based system to detect eye-state AUs: AU41, AU42, and AU43. After localizing three feature points for each eye, a set of multi-scale and multi-orientation Gabor coefficients is extracted. The Gabor coefficients are fed to a neural-network-based detector to learn the correlations between the Gabor coefficient patterns and specific eye states. A recognition rate of 83% was obtained for 112 images from 17 image sequences of 12 subjects. This is comparable to the agreement between different human coders. We have found that the inner corner of the eye contains more useful information than the outer corner of the eye and the lower spatial frequencies contribute more than the higher spatial frequencies.

Acknowledgements

The authors would like to thank Bethany Peters for processing the images, Lala Ambadar and Karen Schmidt for coding the eye states and calculating the reliability between different coders. This work is supported by NIMH grant R01 MH51435.

References

- [1] M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski. Measuring facial expressions by computer image analysis. *Psychophysiology*, 36:253–264, 1999.
- [2] G. Chow and X. Li. Towards a system for automatic facial feature detection. *Pattern Recognition*, 26(12):1739–1755, 1993.
- [3] J. Daugmen. Complete discrete 2-d gabor transforms by neural networks for image analysis and compression. *IEEE Transaction on Acoustic, Speech and Signal Processing*, 36(7):1169–1179, July 1988.
- [4] J. Deng and F. Lai. Region-based template deformation and masking for eye-feature extraction and description. *Pattern Recognition*, 30(3):403–419, 1997.
- [5] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski. Classifying facial actions. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 21(10):974–989, October 1999.
- [6] B. Fasel and J. Luttin. Recognition of asymmetric facial action unit activities and intensities. In *Proceedings of International Conference of Pattern Recognition*, 2000.
- [7] L. Huang and C. W. Chen. Human facial feature extraction for face interpretation and recognition. *Pattern Recognition*, 25(12):1435–1444, 1992.
- [8] T. Kanade, J. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. In *Proceedings of International Conference on Face and Gesture Recognition*, March, 2000.
- [9] K. Lam and H. Yan. Locating and extracting the eye in human face images. *Pattern Recognition*, 29(5):771–779, 1996.
- [10] T. Lee. Image representation using 2d gabor wavelets. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 18(10):959–971, October 1996.
- [11] J.-J. J. Lien, T. Kanade, J. F. Chon, and C. C. Li. Detection, tracking, and classification of action units in facial expression. *Journal of Robotics and Autonomous System*, in press.
- [12] B. Lucas and T. Kanade. An interactive image registration technique with an application in stereo vision. In *The 7th International Joint Conference on Artificial Intelligence*, pages 674–679, 1981.
- [13] M. Lyons, S. Akamasku, M. Kamachi, and J. Gyoba. Coding facial expressions with gabor wavelets. In *Proceedings of International Conference on Face and Gesture Recognition*, 1998.
- [14] Y. Tian, T. Kanade, and J. Cohn. Recognizing upper face actions for facial expression analysis. In *Proc. Of CVPR'2000*, 2000.
- [15] Y. Tian, T. Kanade, and J. Cohn. Dual-state parametric eye tracking. In *Proceedings of International Conference on Face and Gesture Recognition*, March, 2000.
- [16] Y. Tian, T. Kanade, and J. Cohn. Recognizing lower face actions for facial expression analysis. In *Proceedings of International Conference on Face and Gesture Recognition*, March, 2000.
- [17] L. Wiskott, J. M. Fellous, N. Kruger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 19(7):775–779, July 1997.
- [18] X. Xie, R. Sudhakar, and H. Zhuang. On improving eye feature extraction using deformable templates. *Pattern Recognition*, 27(6):791–799, 1994.
- [19] A. Yuille, P. Haallinan, and D. S. Cohen. Feature extraction from faces using deformable templates. *International Journal of Computer Vision*, 8(2):99–111, 1992.
- [20] Z. Zhang. Feature-based facial expression recognition: Sensitivity analysis and experiments with a multi-layer perceptron. *International Journal of Pattern Recognition and Artificial Intelligence*, 13(6):893–911, 1999.