Evaluation of Gabor-Wavelet-Based Facial Action Unit Recognition in Image Sequences of Increasing Complexity

Ying-li Tian [1]   Takeo Kanade[2] and Jeffrey F. Cohn[2,3]

[1] IBM T. J. Watson Research Center, PO Box 704, Yorktown Heights, NY 10598
[2] Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213
[3] Department of Psychology, University of Pittsburgh, Pittsburgh, PA 15260
Email: yltian@us.ibm.com,   tk@cs.cmu.edu   jeffcohn@pitt.edu

Abstract

Previous work suggests that Gabor-wavelet-based methods can achieve high sensitivity and specificity for emotion-specified expressions (e.g., happy, sad) and single action units (AUs) of the Facial Action Coding System (FACS). This paper evaluates a Gabor-wavelet-based method to recognize AUs in image sequences of increasing complexity. A recognition rate of 83% is obtained for three single AUs when image sequences contain homogeneous subjects and are without observable head motion. The accuracy of AU recognition decreases to 32% when the number of AUs increases to nine and the image sequences consist of AU combinations, head motion, and non-homogeneous subjects. For comparison, an average recognition rate of 87.6% is achieved for the geometry-feature-based method. The best recognition is a rate of 92.7% obtained by combining Gabor wavelets and geometry features.

1. Introduction

In facial feature extraction of expression analysis, there are mainly two types of approaches: geometric feature-based methods and appearance-based methods [1, 2, 3, 5, 6, 7, 10, 11, 12, 13, 15, 17, 16, 18, 19]. The geometric facial features present the shape and locations of facial components (including mouth, eyes, brows, nose etc.). The facial components or facial feature points are extracted to form a feature vector that represents the face geometry. In appearance-based methods, image filters, such as Gabor wavelets, are applied to either the whole-face or specific regions in a face image to extract a feature vector.

Zhang *et al.* [20] have compared two type of features to recognize expressions, the geometric positions of 34 fiducial points on a face and 612 Gabor wavelet coefficients extracted from the face image at these 34 fiducial points. The recognition rates for six emotion-specified expressions (e.g. joy and anger) were significantly higher for Gabor wavelet coefficients. Recognition of FACS AUs was not tested. Bartlett *et al.* [1] compared optical flow, geometric features, and principle component analysis (PCA) to recognize 6 individual upper face AUs (AU1, AU2, AU4, AU5, AU6, and AU7) without combinations. The best performance was achieved by PCA. Donato *et al.* [5] compared several techniques for recognizing 6 single upper face AUs and 6 lower face AUs. These techniques include optical flow, principal component analysis, independent component analysis, local feature analysis, and Gabor wavelet representation. The best performances were obtained using a Gabor wavelet representation and independent component analysis. All of these systems [1, 5, 20] used a manual step to align each input image with a standard face image using the center of the eyes and mouth.

Previous work suggests that the appearance-based methods (specifically Gabor wavelets) can achieve high sensitivity and specificity for emotion-specified expressions (e.g., happy, sad) [11, 20] and single AUs [5] under four conditions. (1) Subjects were homogeneous either all Japanese or all Euro-American. (2) Head motion was excluded. (3) Face images were aligned and cropped to a standard size. (4) Specific-emotion expression or single AUs were recognized. In multi-culture society, expression recognition must be robust to variations of face shape, proportion, and skin color. Facial expression typically consists of AU combinations, that often occur together with head motion. AUs can occur either singly or in combination. When AUs occur in combination they may be *additive*, in which the combination does not change the appearance of the constituent AUs, or *non-additive*, in which the appearance of the constituents does change. The non-additive AU combinations make recognition more difficult.

In this paper, we investigate the AU recognition accuracy of Gabor wavelets for both single AUs and AU combinations. We also compare the Gabor-wavelet-based method and the geometry-feature-based method for AU recognition

in a more complex image database than have been used in previous studies of facial expression analysis using Gabor wavelets. The database consists of image sequences from subjects of European, African, and Asian ancestry. Small head motions and multiple AUs are included. For 3 single AUs without head motion, a recognition rate of 83% is obtained for the Gabor-wavelet-based method. When the number of recognized AUs increases to 9 and the image sequences consists of AU combinations, head motions, and non-homogeneous subjects, the accuracy of the Gabor-wavelet-based method decreases to 32%. In comparison, an average recognition rate of 87.6% is achieved for the geometry-feature-based method, and the best recognition rate of 92.7% obtained by combining the Gabor-wavelet-based method and the geometry-feature-based method.

2. Facial Feature Extraction

Contracting the facial muscles produces changes in both the direction and magnitude of skin surface displacement, and in the appearance of permanent and transient facial features. Examples of permanent features are eyes, brow, and any furrows that have become permanent with age. Transient features include facial lines and furrows that are not present at rest. In order to analyze a sequence of images, we assume that the first frame is a neutral expression. After initializing the templates of the permanent features in the first frame, both geometric facial features and Gabor wavelets coefficients are automatically extracted the whole image sequence. No face crop or alignment is necessary.
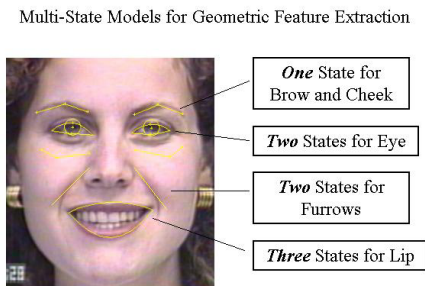
2.1. Geometric facial features

Figure 1. Multi-state models for geometric feature extraction.

To detect and track changes of facial components in near frontal face images, multi-state models are developed to extract the geometric facial features (Fig. 1). A three-state lip model describes lip state: open, closed, and tightly closed. A two-state model (open or closed) is used for each of the eyes. Each brow and cheek has a one-state model. Transient facial features, such as nasolabial furrows, have two states: present and absent. Given an image sequence, the region

of the face and approximate location of individual face features are detected automatically in the initial frame [14]. The contours of the face features and components then are adjusted manually in the initial frame. Both permanent (e.g., brows, eyes, lips) and transient (lines and furrows) face feature changes are automatically detected and tracked in the image sequence. We group 15 parameters which describe shape, motion, eye state, motion of brow and cheek, and furrows in the upper face. These parameters are geometrically normalized to compensate for image scale and in-plane head motion based two inner corners of the eyes. Details of geometric feature extraction can be found in paper [16].
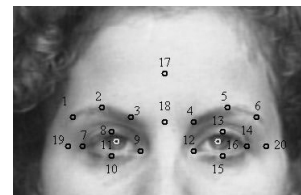
2.2. Gabor wavelets

Figure 2. Locations to calculate Gabor coefficients in upper face.

We use Gabor wavelets to extract the facial appearance changes as a set of multi-scale and multi-orientation coefficients. The Gabor filter may be applied to specific locations on a face or to the whole face image [4, 5, 9, 17, 20]. Following Zhang *et al.* [20], we use the Gabor filter in a selective way, for particular facial locations instead of the whole face image.

The response image of the Gabor filter can be written as a correlation of the input image $I(\mathbf{x})$, with the Gabor kernel $p_{\mathbf{k}}(\mathbf{x})$

$$a_{\mathbf{k}}(\mathbf{x_0}) = \int \int I(\mathbf{x})p_{\mathbf{k}}(\mathbf{x} - \mathbf{x_0})d\mathbf{x}, \qquad (1)$$

where the Gabor filter $p_{\mathbf{k}}(\mathbf{x})$ can be formulated [4]:

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

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

In our implementation, 800 Gabor wavelet coefficients are calculated in 20 locations which are automatically defined based on the geometric features in the upper face (Figure 2). We use $\sigma = \pi$, five spatial frequencies with wavenumbers $k_i = (\frac{\pi}{2}, \frac{\pi}{4}, \frac{\pi}{8}, \frac{\pi}{16}, \frac{\pi}{32})$, and 8 orientations from 0 to π differing by $\pi/8$. In general, $p_{\mathbf{k}}(\mathbf{x})$ is complex. In our approach, only the magnitudes are used because they vary slowly with the position while the phases are very sensitive. Therefore, for each location, we have 40 Gabor wavelet coefficients.

3. Evaluation of Gabor-Wavelet-Based AU Recognition in Image Sequences of Increasing Complexity

3.1. Experimental Setup

AUs to be Recognized: Figure 3 shows the AUs to be recognized and their Gabor images when the spatial frequency$=\frac{\pi}{4}$ in horizontal orientation. AU 43 (close) and AU 45 (blink) differ from each other in the duration of eye closure. Because AU duration is not considered and AU 46 (wink) is close eye only in left or right, we pool AU 43, AU 45, and AU 46 as one unit in this paper. AU 1 (inner brow raise), AU 2 (outer brow raise) and AU 4 (brows pull together and lower) describe actions of brows. Figure 3(h) shows an AU combination.

Database: The Cohn-Kanade expression database [8] is used in our experiments. The database contains image sequences from 210 subjects between the ages of 18 and 50 years. They were 69% female, 31% male, 81% Euro-American, 13% Afro-American, and 6% other groups. Over 90% of the subjects had no prior experience in FACS. Subjects were instructed by an experimenter to perform single AUs and AU combinations. Subjects sat directly in front of the camera and performed a series of facial behaviors which was recorded in an observation room. Image sequences with in-plane and limited out-of-plane motion are included. The image sequences began with a neutral face and were digitized into 640x480 pixel arrays with either 8-bit gray-scale or 24-bit color values. Face size varies between 90×80 and 220×200 pixels. No face alignment or cropping is performed.

AU Recognition NNs: We use a three-layer neural network with one hidden layer to recognize AUs by a standard back-propagation method. The network is shown in Figure 4, and could be divided into two components. The sub-network shown in Figure 4(a) is used for recognizing AU by using the geometric features alone. The inputs of the neural network are the 15 geometric feature parameters. The sub-network shown in Figure 4(b) is used for recognizing AUs by using Gabor wavelets. The inputs are the Gabor coefficients extracted based on 20 locations. For using both geometric features and regional appearance patterns, these two sub-networks are applied in concert. The outputs are the recognized AUs. Each output unit gives an estimate of the probability of the input image consisting of the associated AUs. The networks are trained to respond to the designated AUs whether they occur singly or in combination. When AUs occur in combination, multiple output nodes are excited.

3.2. Experimental Results

First, we report the recognition results of Gabor wavelets for single AUs (AU 41, AU 42, and AU 43). Then, AU Recognition of Gabor wavelets for image sequences of
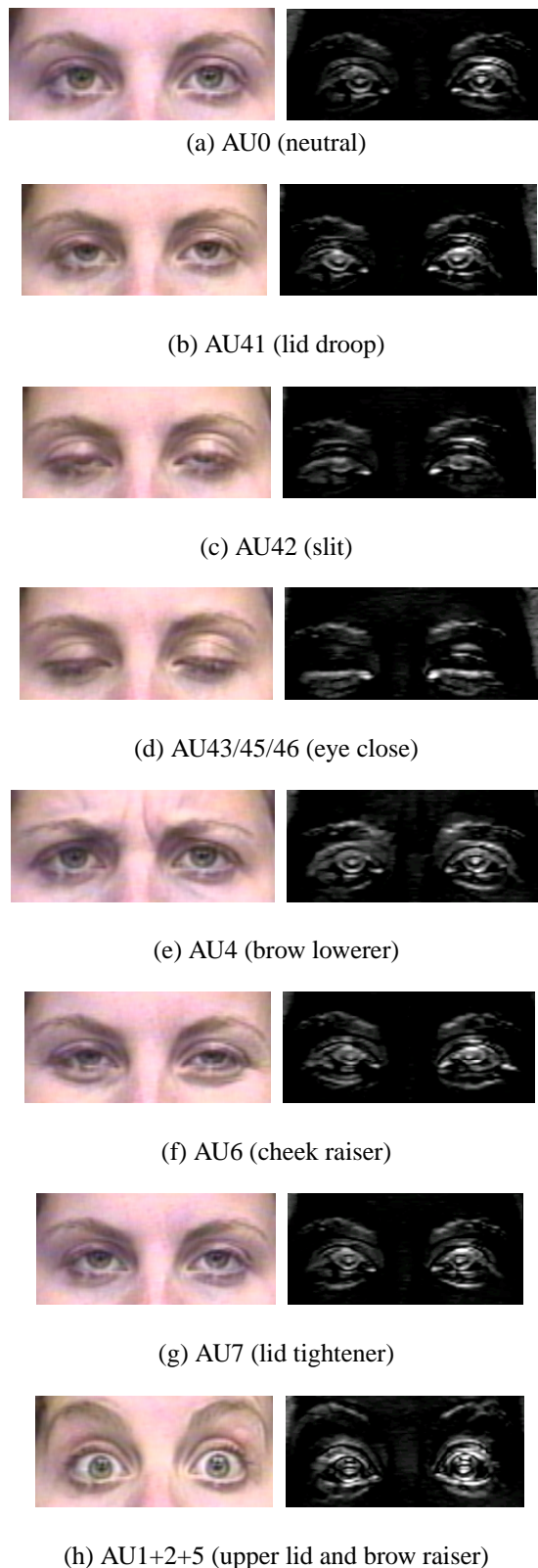

(a) AU0 (neutral)

(b) AU41 (lid droop)

(c) AU42 (slit)

(d) AU43/45/46 (eye close)

(e) AU4 (brow lowerer)

(f) AU6 (cheek raiser)

(g) AU7 (lid tightener)

(h) AU1+2+5 (upper lid and brow raiser)

Figure 3. AUs to be recognized and their Gabor images when the spatial frequency$=\frac{\pi}{4}$ in horizontal orientation.
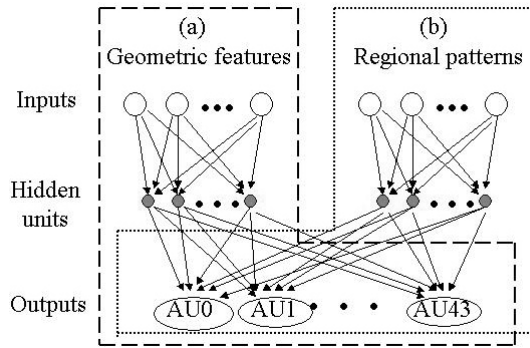
Figure 4. AU recognition neural networks.

increasing complexity is investigated. Because input sequences contain multiple AUs, several outcomes are possible. $Correct$ denotes that target AUs are recognized. $Missed$ denotes that some but not all of the target AUs are recognized. $False$ denotes that AUs that do not occur are falsely recognized. For comparison, the AU recognition results of the geometric-feature-based method are reported. Best AU recognition results are achieved by combining Gabor wavelets and geometric features.

AU Recognition of Gabor wavelets for single AUs: In this investigation, we focus on recognition of AU41, AU42, and AU43 by Gabor wavelets. We selected 33 sequences from 21 subjects for training and 17 sequences from 12 subjects for testing. All subjects are Euro-American without observable head motions. The data distribution of training and test data sets is shown in Table 1.

Table 1. Data distribution of training and test data sets for single AU recognition.

Data Set	AU 41	AU 42	AU 43	Total
$Train$	92	75	74	241
$Test$	56	40	16	112

Table 2 shows the recognition results for 3 single AUs (AU 41, AU 42, and AU 43) when we use three feature points of the eye and three spatial frequencies of Gabor wavelet $(\frac{\pi}{2}, \frac{\pi}{4}, \frac{\pi}{8})$. 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.

AU Recognition of Gabor Wavelets for AU Combinations in Image Sequences of Increasing Complexity:

In this evaluation, we test recognition accuracy of Gabor wavelets for AU combinations in a more complex database. The database consists of 606 image sequences from 107 subjects of European, African, and Asian ancestry. Most

Table 2. Recognition results of single AUs by using Gabor wavelets.

	AU 41	AU 42	AU 43
AU 41	52	4	0
AU 42	4	28	8
AU 43	0	3	13
Recognition rate: 83%			

image sequences contain AU combinations and some include small head motion. We split the image sequences into training (407 sequences from 59 subjects) and testing (199 sequences from 48 subjects) sets to ensure that the same subjects did not appear in both training and testing. Table 3 shows the AU distribution for training and test sets.

Table 3. AU distribution of training and test data sets in image sequences of increasing complexity.

Datasets	AU0	AU1	AU2	AU4	AU5	AU6	AU7	AU41	AU43
$Train$	407	163	124	157	80	98	36	74	94
$Test$	199	104	76	84	60	52	28	20	48

Table 4. AU Recognition of Gabor Wavelets for AU combinations in image sequences of increasing complexity.

AUs	Total	Correct	Missed	False
AU1	104	4	100	8
AU2	76	0	76	0
AU4	84	8	76	3
AU5	60	0	60	0
AU6	52	25	27	0
AU7	28	0	28	0
AU41	20	0	20	0
AU43	48	38	10	0
AU0	199	140	59	208
Total	671	215	456	219
Average Recognition Rate: 32%				
False Alarm Rate: 32.6%				

In the experiment, total 800 Gabor wavelet coefficients corresponding 5-scale and 8-orientation are calculated at 20 specific locations. We have found that 480 Gabor coefficients of three middle scales perform better than use all 5 scales. The inputs are 480 Gabor coefficients (3 spatial frequencies in 8 orientations, applied at 20 locations). The

recognition results are summarized in Table 4. We have achieved average recognition- and false alarm rates of 32% and 32.6% respectively. Recognition is adequate only for AU6, AU43, and AU0. The appearance changes associate with these AUs are detected often occurred in specific regions for AU6 and AU43 comparing with AU0. For example, crows-feet wrinkles often appear for AU6 and the eyes look qualitatively different when they are open and closed (AU43). Use of PCA to reduce the dimensionality of the Gabor wavelet coefficients failed to increase recognition accuracy.

AU Recognition of Geometric Features for AU Combinations in Image Sequences of Increasing Complexity: For comparison, using the 15 parameters of geometric features, we achieved average recognition- and false alarm rates of 87.6% and 6.4% respectively (Table 5). Recognition of individual AUs is good with the exception of AU7. Most instances of AU7 are of low intensity, which change only 1 or 2 pixels in face image and cannot be extracted by geometry-feature-based method.

Table 5. AU Recognition Using Geometric Features.

AUs	Total	Correct	Missed	False
AU1	104	100	4	0
AU2	76	74	2	4
AU4	84	68	16	5
AU5	60	50	10	8
AU6	52	41	11	5
AU7	28	2	26	0
AU41	20	15	5	7
AU43	48	39	9	10
AU0	199	199	0	4
Total	671	588	83	43
Average Recognition Rate: 87.6%				
False Alarm Rate: 6.4%				

AU Recognition of Combining Geometric features and Gabor Wavelets for AU Combinations in Image Sequences of Increasing Complexity: In this experiment, both geometric features and Gabor wavelets are fed to the network. The inputs are 15 geometric feature and 480 Gabor coefficients (3 spatial frequencies in 8 orientations applied at 20 locations). The recognition results are shown in Table 6. In comparison to the results of using either the geometric features or the Gabor wavelets alone, combining these features increases the accuracy of AU recognition, recognition performance has been improved to 92.7% from 87.6% and 32% respectively.

4. Conclusion and Discussion

We summarize the AU recognition results by using Gabor wavelets alone, geometric features alone, and both of them in Figure 5. Three recognition rates for each AU are described by histograms. The gray histogram shows recognition results based on Gabor wavelets. The dark gray histogram shows recognition results based on geometric features, and the white histogram shows results obtained using both types of features. Using Gabor wavelets alone, recognition is adequate only for AU6, AU43, and AU0. Using geometric features, recognition is consistently good with the exception of AU7. The results using geometric features alone are consistent with previous research that shows high AU recognition rates for this approach. Combining both types of features, the recognition performance increased for all AUs.

Table 6. AU recognition results by combining Gabor wavelets and geometric features.

AUs	Total	Correct	Missed	False
AU1	104	101	3	4
AU2	76	76	0	6
AU4	84	75	9	11
AU5	60	51	9	8
AU6	52	45	7	7
AU7	28	13	15	0
AU41	20	16	4	3
AU43	48	46	2	11
AU0	199	199	0	1
Total	671	622	49	51
Average Recognition Rate: 92.7%				
False Alarm Rate: 7.6%				

Consistent with previous studies, we found that Gabor wavelets work well for single AU recognition for homogeneous subjects without head motion. However, for recognition of AU combinations when image sequences include non-homogeneous subjects with small head motions, we are surprised to find relatively poor recognition using this approach. In summary, several factors may account for the difference. First, the previous studies used homogeneous subjects. For instance, Zhang et al. included only Japanese and Donato et al. included only Euro-American. We use diverse subjects of European, African, and Asian ancestry. Second, the previous studies recognized emotion-specified expressions or only single AUs. We tested the Gabor-wavelet-based method on both single AUs and AU combinations, including non-additive combinations in which the occurrence of one AU modifies another. Third, the previous studies manually aligned and cropped face images. We omitted this preprocessing step. Our geometric features and the locations to calculate Gabor coefficients were robust to head motion. These differences suggest that any advantage of Gabor wavelets in facial expression recognition may depend on manual preprocessing and may fail to generalize to
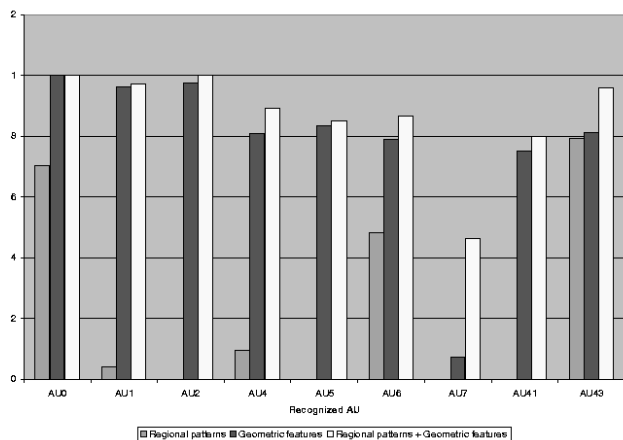
Figure 5. Comparison of AU recognition results by using different type of features in image sequences of increasing complexity. The gray histogram shows recognition results using Gabor wavelets. The dark gray histogram shows recognition results using geometric facial features, and the white histogram shows results obtained using both types of features.

heterogeneous subjects and more varied facial expression. Combining Gabor wavelet coefficients and geometric features resulted in the best performance.

Acknowledgements

References

[1] M. Bartlett, J. Hager, P.Ekman, and T. Sejnowski. Measuring facial expressions by computer image analysis. *Psychophysiology*, 36:253–264, 1999.

[2] M. J. Black and Y. Yacoob. Trcking and recognizing rigid and non-rigid facial motions using local parametric models of image motion. In *Proc. Of International conference on Computer Vision*, pages 374–381, 1995.

[3] M. J. Black and Y. Yacoob. Recognizing facial expressions in image sequences using local parameterized models of image motion. *International Journal of Computer Vision*, 25(1):23–48, October 1997.

[4] J. Daugmen. Complete discrete 2-d gabor transforms by neutral networks for image analysis and compression. *IEEE Transaction on Acoustic, Speech and Signal Processing*, 36(7):1169–1179, July 1988.

[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] I. A. Essa and A. P. Pentland. Coding, analysis, interpretation, and recognition of facial expressions. *IEEE Transc. On Pattern Analysis and Machine Intelligence*, 19(7):757–763, JULY 1997.

[7] K. Fukui and O. Yamaguchi. Facial feature point extraction method based on combination of shape extraction and pattern matching. *Systems and Computers in Japan*, 29(6):49–58, 1998.

[8] T. Kanade, J. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. In *Proceedings of International Conference on Face and Gesture Recognition*, pages 46–53, March, 2000.

[9] T. Lee. Image representation using 2d gabor wavelets. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 18(10):959–971, Octobor 1996.

[10] J.-J. J. Lien, T. Kanade, J. F. Cohn, and C. C. Li. Detection, tracking, and classification of action units in facial expression. *Journal of Robotics and Autonomous System*, 31:131–146, 2000.

[11] 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.

[12] K. Mase. Recognition of facial expression from optical flow. *IEICE Transactions*, E. 74(10):3474–3483, October 1991.

[13] M. Rosenblum, Y. Yacoob, and L. S. Davis. Human expression recognition from motion using a radial basis function network archtecture. *IEEE Transactions On Neural Network*, 7(5):1121–1138, 1996.

[14] H. A. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *IEEE Transactions On Pattern Analysis and Machine intelligence*, 20(1):23–38, January 1998.

[15] D. Terzopoulos and K. Waters. Analysis of facial images using physical and anatomical models. In *IEEE International Conference on Computer Vision*, pages 727–732, 1990.

[16] Y. Tian, T. Kanade, and J. Cohn. Recognizing action units for facial expression analysis. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 23(2):1–19, February 2001.

[17] Y. Tian, T. Kanade, and J. Cohn. Eye-state action unit detection by gabor wavelets. In *Proceedings of International Conference on Multi-modal Interfaces (ICMI 2000)*, pages 143–150, Sept, 2000.

[18] Y. Yacoob and M. J. Black. Parameterized modeling and recognition of activities. In *Proc of the 6th International Conference on Computer Vision, Bombay, India*, pages 120–127, 1998.

[19] Y. Yacoob and L. S. Davis. Recognizing human facial expression from long image sequences using optical flow. *IEEE Transactions On Pattern Analysis and machine Intelligence*, 18(6):636–642, June 1996.

[20] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu. Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron. In *International Workshop on Automatic Face and Gesture Recognition*, pages 454–459, 1998.