Recognizing Facial Actions by Combining Geometric Features and Regional Appearance Patterns

Ying-li Tian   Takeo Kanade   Jeffrey F. Cohn

CMU-RI-TR-01-01

Robotics Institute, Carnegie Mellon University,
Pittsburgh, PA 15213

January, 2001

This research is sponsored by NIMH, under contract R01MH51435 "Facial Expression Analysis by Computer Image Processing".

Keywords: Facial expression analysis Action units Neural network Geometric features Regional Appearance Patterns Gabor wavelet

Abstract

In facial expression analysis, two principle approaches to extract facial features are geometric feature-based methods and appearance-based methods such as Gabor filters. In this paper, we combine these approaches in a feature-based system to recognize Facial Action Coding System (FACS) action units (AUs) in a complex database. The geometric facial features (including mouth, eyes, brows, and cheeks) are extracted using multi-state facial component models. After extraction, these features are represented parametricly. The regional facial appearance patterns are captured using a set of multi-scale and multi-orientation Gabor wavelet filters at specific locations. For the upper face, we recognize 8 AUs and neutral expression. The database consists of 606 image sequences from 107 adults of European, African, and Asian ancestry. AUs occur both alone and in combinations. Average recognition rate is 87.6% by using geometric facial features alone, 32% by using regional appearance patterns alone, 89.6% by combining both features, and 92.7% after refinement. For the lower face, we recognize 13 AUs and neutral expression. The database consists of 514 image sequences from 180 adults of European, African, and Asian ancestry. AUs occur both alone and in combinations. Average recognition rate is 84.7% by using geometric facial features alone, 82% by combining both features, and 87.4% after refinement.

1. Introduction

Facial expression is one of the most powerful, natural, and immediate means for human beings to communicate emotions and intentions. Often emotions are expressed through the face before they are verbalized. In the past decade, much progress has been made in building computer systems to understand and use this natural form of human communication [1, 3, 2, 5, 7, 8, 9, 11, 13, 15, 18, 22, 19, 23, 24, 25]. Two procedures are necessary for an automatic expression analysis system: facial feature extraction and facial expression recognition.

In facial feature extraction, there are mainly two types of approaches: geometric feature-based methods and appearance-based methods. In geometric feature-based methods, 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 whole-face or specific regions in a face image to extract a feature vector. Geometric feature extraction can be more computationally expensive, but is more robust to variation in face position, scale, size, and head orientation.

In facial expression recognition, most automatic expression analysis systems attempt to recognize a small set of prototypic expressions (i.e. joy, surprise, anger, sadness, fear, and disgust) [8, 13, 24, 26]. However, in everyday life, emotion is often communicated by changes in one or two discrete facial features, such as tightening the lips in anger or obliquely lowering the lip corners in sadness [4]. Some researchers [1, 5, 7, 11, 22, 19] have attempted to recognize the fine-grained changes in facial expression based on the Facial Action Coding System (FACS). We focus on AU recognition.

Zhang *et al.* [26] have compared two type of features, the geometric positions of 34 fiducial points on a face and 612 Gabor wavelet coefficients extracted from the face image at the fiducial points. The recognition rates for 6 emotion-specified expressions (e.g. joy and anger) were significantly higher for Gabor wavelet coefficients. Recognition of FACS AUs was not tested. The system of Lien *et al.* [11] used dense-flow, feature point tracking and edge extraction to recognize 3 upper face AUs (AU1+2, AU1+4, and AU4) and 6 lower face AUs. 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.* [7] 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, 7, 11] used a manual step to align the input images with a standard face image using the center of the eyes and mouth. In our previous system [21], multi-state face and facial component models are proposed for tracking and modeling the various facial features, including lips, eyes, brows, cheeks, and furrows. During tracking, detailed parametric descriptions of the facial features are extracted. With these parameters as the inputs, a group of action units(neutral expression, 6 upper face AUs, and 10 lower face AUs) are recognized whether they occur alone or in combinations. The system has achieved average recognition rates of 96.4% (95.4% if neutral expressions are excluded) for upper face AUs and 96.7% (95.6% with neutral expressions excluded) for lower face AUs.

In this report, we combine geometric facial features and regional facial appearance patterns in a feature-based system to recognize Facial Action Coding System (FACS) action units (AUs) in a complex database. We assume the first frame in the image sequence is neutral expression. The geometric facial features (including mouth, eyes, brows, and cheeks) are extracted using multi-state facial component models. The extracted features are represented as detailed parametric descriptions. The regional facial appearance patterns are captured using a set of multi-scale and multi-orientation Gabor wavelet coefficients at specific locations. For the upper face, we recognize 8 AUs and neutral expression. The database consists of 606 image sequences from 107 adults of European, African, and Asian ancestry. AUs occur both alone and in combinations. Average recognition rate is 87.6% by using geometric facial features alone, 32% by using regional appearance patterns alone, 89.6% by combining both features, and 92.7% after refinement. For the lower face, we recognize 13 AUs and neutral expression. The database consists of 514 image sequences from 180 adults of European, African, and Asian ancestry. AUs occur both alone and in combinations. Average recognition rate is 84.7% by using geometric facial features alone, 82% by combining both features, and 87.4% after refinement. Two AUs about head position also are recognized. Compared with our previous system [21], we recognize 8 more AUs (2 more in the upper face, 4 more in the lower face, and 2 for head position) in current system by combining the regional facial appearance patterns with the geometric facial features. Compared with the previous system, we recognize more AUs in more complex database by using more information.

2. Facial Feature Extraction

Contraction of 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 as a neutral expression. After initializing the templates of the permanent features in the first frame, both geometric facial features and regional appearance patterns are automatically extracted in the whole image sequence. No face crop or alignment is neccessary.

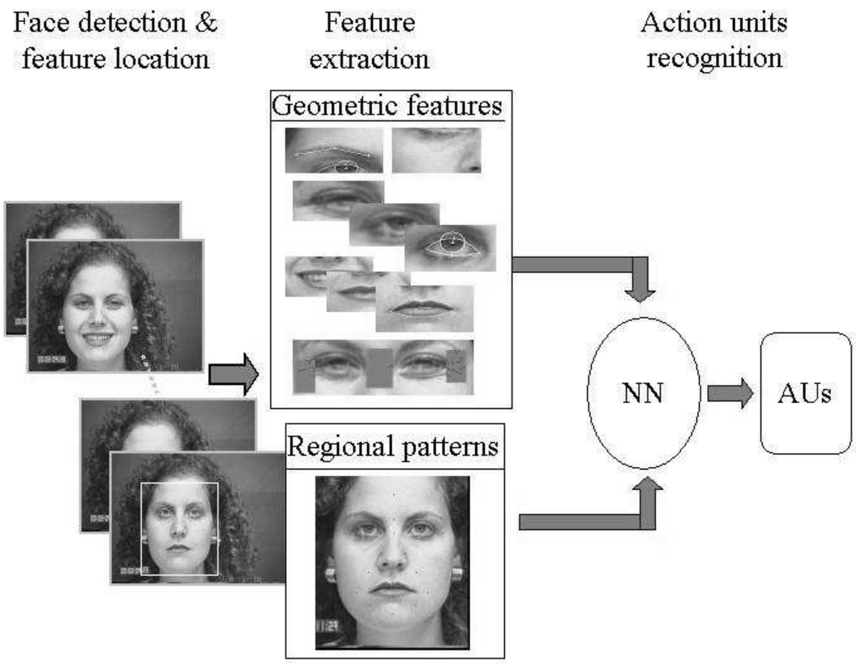


Figure 1. Feature-based Automatic Facial Action Analysis (AFA) System.

2.1. Geometric facial features

Lips: A three-state lip model represents open, closed and tightly closed lips. A different lip contour template is prepared for each lip state. The open and closed lip contours are modeled by two parabolic arcs, which are described by six parameters: the lip center position (xc, yc), the lip shape ($h1$, $h2$ and w), and the lip orientation (θ). For tightly closed lips, the dark mouth line connecting the lip corners represents the position, orientation, and shape.

Tracking of lip features uses color, shape, and motion. In the first frame, the approximate position of the lip template is detected automatically. It then is adjusted manually by moving four key points. A Gaussian mixture model represents the color distribution of the pixels inside of the lip template [14]. The details of our lip tracking algorithm have been presented in [20].

Eyes: In order to detect whether the eyes are open or closed, the degree of eye opening, and the location and radius of the iris. Two eye states are proposed: open and closed. For an open eye, the eye template is composed of a circle with three parameters (x_0, y_0, r) to model the iris and two parabolic arcs with six parameters $(x_c, y_c, h_1, h_2, w, \theta)$ to model the boundaries of the eye. For a closed eye, the template is reduced to 4 parameters: two for the position of each of the eye corners.

The open-eye template is adjusted manually in the first frame by moving 6 points for each eye. We found that the outer corners are more difficult to track than the inner corners; for this reason, the inner corners of the eyes are tracked first. The outer corners are then located on the line that connects the inner corners at a distance of the eye width as estimated in the first frame.

The iris provides important information about the eye states. Part of the iris is normally visible if the eye is open. Intensity and edge information are used to detect the iris. We have observed that the eyelid edge is noisy even in a good quality image. However, the lower part of the iris is almost always visible, and its edge is relatively clear for open eyes. Thus, we use a half circle mask to filter the iris edge. The radius of the iris circle template r_0 is determined in the first frame, since it is stable except for large out-of-plane head motion. The radius of the circle is increased or decreased slightly (δr) from r_0 so that it can vary between minimum radius $(r_0 - \delta r)$ and maximum radius $(r_0 + \delta r)$. The system determines that the iris is found when the following two conditions are satisfied: the edges in the mask are at their maximum; and the change in the inside circle average intensity is less than a threshold. Once the iris is located, the eye is determined to be open and the iris center is the iris mask center (x_0, y_0). The eyelid contours then are tracked. For a closed eye, a line connecting the inner and outer corners of the eye is used as the eye boundary.

Brow and cheek: Features in the brow and cheek areas are also important for expression analysis. Each left or right brow has one model – a triangular template with six parameters $(x1, y1)$, $(x2, y2)$, and $(x3, y3)$. Each cheek has also a similar six parameter down-ward triangular template model. Both brow and cheek templates are tracked using Lucas-Kanade algorithm [12].

Crows-feet wrinkles nasal root wrinkles: Crows-feet wrinkles appearing to the side of the outer eye corners are useful features for recognizing upper face AUs. For example, the lower eyelid is raised for both AU6 and AU7, but the crows-feet wrinkles appear for AU6 only. Nasal root wrinkles appearing to the nasal root are useful features for recognizing lower face AUs. For example, the upper lip is raised for both AU9 and AU10, but the nasal root wrinkles appear for AU9 only. Compared with the neutral frame, the wrinkle state is present if the wrinkles appear, deepen, or lengthen. Otherwise, it is absent.

After locating the outer corners of the eyes, edge detectors search locally in the lateral areas for crows-feet wrinkles. We compare the number of edge pixels E, of the current frame with the numbers of edge pixels E_0 of the first frame. If E/E_0 is larger than the threshold T, the crows-feet and nasal root wrinkles are present. Otherwise, they are absent.

2.2. Regional appearance patterns

(a) 20 locations in the upper face (b) 18 locations in the lower face

Figure 2. Locations to calculate Gabor coefficients.

We use Gabor wavelets to extract the regional appearance patterns 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 [7, 6, 10, 26]. Following Zhang *et al.* [26], we use the Gabor filter in a selective way, that is in particular facial locations instead of use 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 $p_{\mathbf{k}}(\mathbf{x})$

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

where the Gabor filter $p_{\mathbf{k}}(\mathbf{x})$ can be formulated [6]:

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

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

In our system, Gabor wavelet coefficients are calculated in 38 locations which are automatically defined based on the geometric features in the whole face. Of these 38 locations, there are 20 locations in the upper face (Figure 2(a)) and 18 locations in the lower face (Figure 2(b)). 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 in $\pi/8$. 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.

An example of the Gabor filtered images from one image sequence is shown in Figure 3. This figure shows the result of geometric feature extraction, the locations at which regional appearance patterns are calculated, and the corresponding Gabor filter responses for the second spatial frequency ($k_i = (\frac{\pi}{4})$) in horizontal and vertical orientations. More results for different subjects with out-of-plane head motion can be found at http://www.cs.cmu.edu//~face.

3. AU Recognition by Combining Geometric Features and Regional Appearance Patterns

3.1. Experimental Setup

Table 1. AUs to be recognized in the upper face

AU 1	AU 2	AU 4
		
Inner portion of the brows is raised.	Outer portion of the brows is raised.	Brows lowered and drawn together
AU 5	AU 6	AU 7
		
Upper eyelids are raised.	Cheeks are raised.	Lower eyelids are raised.
AU 41	AU 43/45/46	AU0(neutral)
		
Upper-lid is slightly lowered.	Eyes are completely closed.	Eyes, brow, and cheek are relaxed.

(a) Tracked geometric features (frame 1, 10, and 19 from left to right)

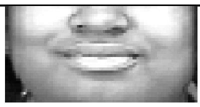
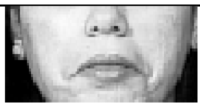
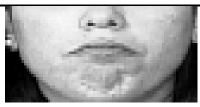
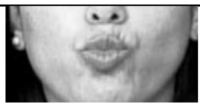
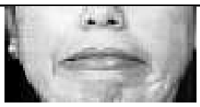
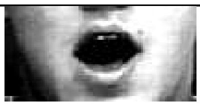
(b) Locations at images regional appearance patterns (frame 1, 10, and 19 from left to right) are calculated. Note that these positions correspond to the same physical locations in the face regardless of expression.

(c) Gabor filtered images in horizontal orientation (frame 1, 10, and 19 from left to right)

(d) Gabor filtered images in vertical orientation (frame 1, 10, and 19 from left to right)

Figure 3. Tracked geometric features, locations to calculate regional appearance patterns and Gabor filtered images of an image sequence. Note that the original image size is 640×480 in our experiments. For different subjects, face size varies between 90×80 and 220×200 pixels. For display purpose, images have been cropped to reduce space.

Table 2. AUs to be recognized in the lower face

AU 9	AU 10	AU 12	AU13
			
The infraorbital triangle and center of the upper lip are pulled upwards. Nasal root wrinkling is present.	The infraorbital triangle is pushed upwards. Upper lip is raised. Causes angular bend in shape of upper lip. Nasal root wrinkle is absent.	Lip corners are pulled obliquely.	Angle of lips up more sharply than AU 12. The red part of lips are not elongated.
AU 14	AU 15	AU17	AU 18
			
Lip corners are pulled inward and tightened.	The corners of the lips are pulled down.	The chin boss is pushed upwards.	The mouth is puckered (pushed forward and pulled medially). The red part of the lips appear taut.
AU 20	AU23	AU 24	AU 25
			
The lips and the lower portion of the nasolabial furrow are pulled pulled back laterally. The mouth is elongated.	Lips are tightened. Wrinkles above and below the lips, and muscle bulges below the lower lip.	Lips are pressed together, tightening and narrowing the lips.	Lips are relaxed and parted.
AU 27	$AU0(neutral)$		
			
Mouth stretched open and the mandible pulled downwards.	Lips relaxed and closed.		

AUs to be Recognized: We limit ourselves to 9 AUs (including AU0 which corresponds to neutral) in the upper face (Table 1). AU 43 (close) and AU 45 (blink) differ from each other in the duration of eye closure. Because AU duration is not considered, we pool AU43 and AU45 as one unit. Action units can occur either singly or in combinations. In the lower face, 14 AUs (including AU0 which corresponds to neutral) are recognized (Table 2).

The AU combinations may be additive, in which case the combination does not change the appearance of the constituents (e.g., AU1+5), or non-additive, in which case the appearance of the constituents does change (e.g., AU1+4). The non-additive AU combinations make recognition more difficult. In this report, we recognize the 9 AUs in the upper face and 14 AUs in the lower face, whether they occur singly or in combination with one exception. The combination AU6+7 is not included because its recognition requires sequential information.

Data set: The database we use in the experiments contains 582 image sequences from 180 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.

Table 3. AU distribution of training and test data sets for upper face.

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

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. We split the image sequences into training and testing sets to ensure that the same subjects did not appear in both training and testing.

For AU recognition in the upper face, 407 sequences from 59 subjects are used for training (some

Table 4. AU distribution of training and test data sets for lower face.

Datasets	AU0	AU9	AU10	AU12	AU13	AU14	AU15	AU17	AU18	AU20	AU23	AU24	AU25	AU27
$Train$	340	46	48	82	60	54	74	146	56	52	38	40	184	50
$Test$	174	28	12	68	14	12	36	68	6	24	18	20	106	42

sequences are multiple used for different frames with different AUs) and 199 sequences from 48 subjects are used for test. Table 3 shows the AU distribution for training and test sets in the upper face.

For AU recognition in the lower face, 340 sequences from 110 subjects are used for training and 174 sequences from 70 subjects are used for test. Please note that some sequences are multiple used based on two reasons. The first is there are different AUs for different frames in the same sequence. The second is that we have not enough sequences for some AUs (for example, AU10 and AU18). Table 4 shows the AU distribution for training and test sets in the lower face.

Feature Representation: In our experiments, two types of features are used: geometric features and regional appearance patterns. In the upper face, the geometric features are represented as 16 parameters. Of these, 12 parameters describe the motion and shape of eyes, brows, cheeks, 1 describes the motion of forehead, 2 parameters describe the state of crows-feet wrinkles, and 1 parameter describes the distance between brows. In the lower face, the geometric features are represented as 8 parameters. Of these, 6 parameters describe lip shape, state and motion, 1 describes the chin motion, and 1 describes the furrows in the nasal root regions. These parameters are normalized by using the ratios of the current feature values to that of the neutral frame.

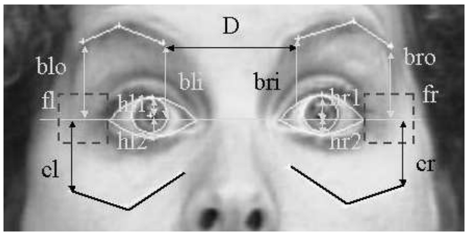


Figure 4. Geometric facial features are represented based on eye inner corners.

Because the inner corners of the eyes are most reliably detected and their relative position is unaf-

Table 5. Geometric feature representation for upper face

Permanent features (Left and right)		
Inner brow motion (r_{binner})	Outer brow motion (r_{bouter})	Eye height ($r_{eheight}$)
$r_{binner} = \frac{bi - bi_0}{bi_0}$. If $r_{binner} > 0$, Inner brow move up.	$r_{bouter} = \frac{bo - bo_0}{bo_0}$. If $r_{bouter} > 0$, Outer brow move up.	$r_{eheight} = \frac{(h1+h2)-(h1_0+h2_0)}{(h1_0+h2_0)}$. If $r_{eheight} > 0$, Eye height increases.
Eye top lid motion (r_{top})	Eye bottom lid motion (r_{btm})	Cheek motion (r_{cheek})
$r_{top} = \frac{h1 - h1_0}{h1_0}$. If $r_{top} > 0$, Eye top lid move up.	$r_{btm} = -\frac{h2 - h2_0}{h2_0}$. If $r_{btm} > 0$, Eye bottom lid move up.	$r_{cheek} = -\frac{c - c_0}{c_0}$. If $r_{cheek} > 0$, Cheek move up.
Other features		
Distance of brows (D_{brow})	Left crows-feet wrinkles (W_{left})	Right crows-feet wrinkles (W_{right})
$D_{brow} = \frac{D - D_0}{D_0}$.	If $W_{left} = 1$, Left crows-feet wrinkle present.	If $W_{right} = 1$, Right crows-feet wrinkle present.
Forehead Motion ($r_{forehead}$)		
$r_{forehead} = \frac{fh - fh_0}{fh_0}$. If $r_{forehead} > 0$, Forehead move up.		

Table 6. Lower face feature representation for AUs recognition

Permanent features		
Lip height (r_{height})	Lip width (r_{width})	Left lip corner motion (r_{left})
r_{height} $=\frac{(h1+h2)-(h1_0+h2_0)}{(h1_0+h2_0)}$. If $r_{height} > 0$, lip height increases.	r_{width} $=\frac{w-w_0}{w_0}$. If $r_{width} > 0$, lip width increases.	r_{left} $=-\frac{D_{left}-D_{left0}}{D_{left0}}$. If $r_{left} > 0$, left lip corner moves up.
Right lip corner (r_{right})	Top lip motion (r_{top})	Bottom lip motion(r_{btm})
r_{right} $=-\frac{D_{right}-D_{right0}}{D_{right0}}$. If $r_{right} > 0$, right lip corner moves up.	r_{top} $=-\frac{D_{top}-D_{top0}}{D_{top0}}$. If $r_{top} > 0$, top lip moves up.	r_{btm} $=-\frac{D_{btm}-D_{btm0}}{D_{btm0}}$. If $r_{btm} > 0$, bottom lip moves up.
Transient features		
State of nasal root wrinkles (S_{nosew})	Motion of chin (r_{chin})	
If $S_{nosew} = 1$, nasal root wrinkles present.	(r_{chin}) $==-\frac{D_{chin}-D_{chin0}}{D_{chin0}}$. If $r_{chin} > 0$, Chin moves up.	

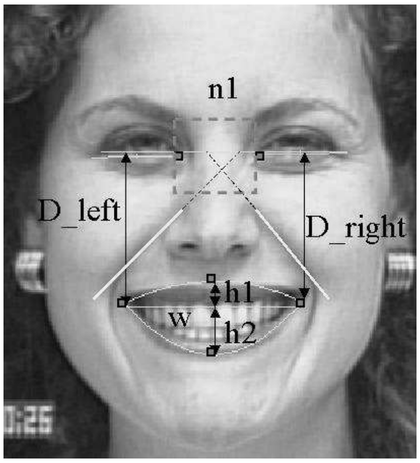


Figure 5. Lower face features. $h1$ **and** $h2$ **are the top and bottom lip heights;** w **is the lip width;** D_{left} **is the distance between the left lip corner and eye inner corners line;** D_{right} **is the distance between the right lip corner and eye inner corners line;** $n1$ **is the nasal root area.**

fected by muscle contraction, we define the x-axis as the line connecting the inner corners of eyes. The geometric feature parameters are represented in this coordinate system as shown in Figure 4 and Figure 5. The definitions of the geometric feature parameters are listed in Table 5 and Table 6. In order to remove the effects of the different size of face images in different image sequences, all the parameters (except two parameters of crows-feet wrinkles) are normalized by dividing by the distances between each feature and the line connecting two inner corners of eyes in the first frame (neutral expression).

The regional appearance patterns are represented by 40 Gabor wavelet coefficients corresponding 5-scale and 8-orientation. In the upper face, these coefficients are calculated at 20 specific locations. Therefore, there are 800 ($5\times8\times20$) Gabor coefficients in upper face. In our experiments, we have found that 480 Gabor coefficients of three middle scales perform better than use all 5 scales. In the lower face, these coefficients are calculated at 18 specific locations. Therefore, there are 720 ($5\times8\times20$) Gabor coefficients in upper face. In our experiments, we have found that 432 Gabor coefficients of last three scales perform better than use all 5 scales.

AU Recognition NN: 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 6, and could be divided into two components. The sunnetwork shown in Figure 6(a) is used for recognizing AU by using the geometric features alone. The inputs of the neural network are the 16 geometric feature parameters in the upper face and 8 parameters in the lower face. The subnetwork shown in Figure 6(b) is used for recognizing AUs by using the regional appearance patterns alone. The inputs are 480 Gabor coefficients extracted based on 20 locations in the upper face and 432 Gabor coefficients extracted based on 18 locations in the lower face. For using both geometric features and regional appearance patterns, these two subnetworks are applied in concert. The outputs are the 9 AUs shown in Table 1 for the upper face and the 14 AUs shown in Table 2 for the lower face. 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.

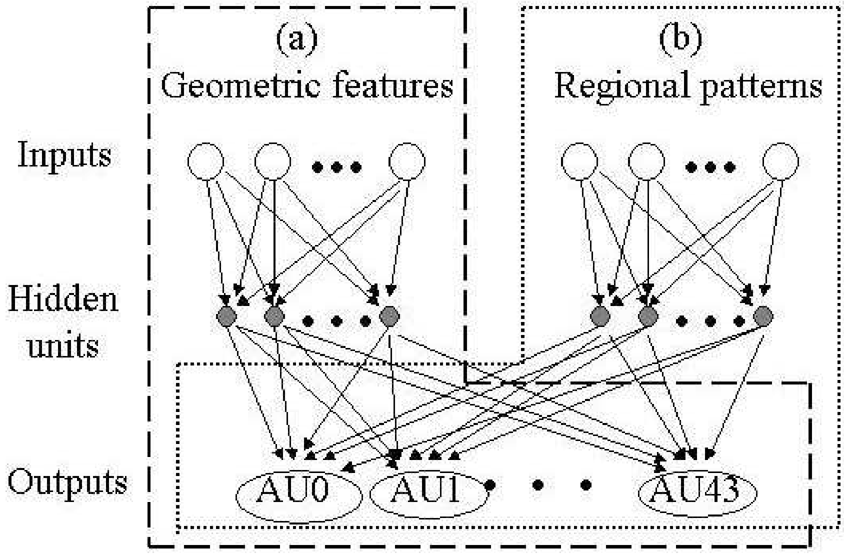


Figure 6. AU recognition neural networks.

3.2. Experimental Results

We report the recognition results for geometric features alone, regional appearance patterns alone, and both of them together. Because input sequences contain one or more 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.

3.2.1 AU Recognition Using Geometric Features Alone

(1) Upper face: Using the 16 parameters for geometric features, we achieved average recognition- and false alarm rates of 87.6% and 6.4% respectively (Table 7). Recognition of individual AUs is good with the exception of AU7. Most instances of AU7 are of low intensity, which may have been a factor.

Table 7. Upper face AU Recognition only 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%				

(2) Lower face: Using the 8 parameters for geometric features, we achieved average recognition- and false alarm rates of 84.7% and 7.3% respectively (Table 8). Recognition of individual AUs is good with the exception of AU10 and AU14.

3.2.2 AU Recognition Using Regional Appearance Patterns Alone

In this experiment, only the regional appearance patterns are used for AU recognition. The inputs are 480 Gabor coefficients (3 spatial frequencies in 8 orientations, applied at 20 locations). The recognition results are summarized in Table 9. 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.

For AU recognition in the lower face, the neural network does not converge by using regional appearance patterns alone.

Table 8. Lower AU Recognition only Using Geometric Features.

AUs	Total	Correct	Missed	False
AU9	28	18	10	0
AU10	12	0	12	1
AU12	68	62	6	2
AU13	14	13	1	3
AU14	12	5	7	1
AU15	36	25	11	6
AU17	68	46	22	4
AU18	6	6	0	0
AU20	24	22	2	1
AU23	18	13	5	0
AU24	20	13	7	0
AU25	106	96	10	10
AU27	42	39	3	5
AU0	174	174	0	13
Total	628	532	96	46
Average Recognition Rate: 84.7%				
False Alarm Rate: 7.3%				

Table 9. Upper face AU Recognition Using Regional Appearance Patterns.

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%				

3.2.3 AU Recognition Combining Geometric features and Regional Appearance Patterns

(1) Upper face: In this experiment, both geometric features and regional appearance patterns are fed to the network. The inputs are 16 geometric feature and 480 Gabor coefficients (3 spatial frequencies in 8 orientations applied at 20 locations). The recognition results are shown in Table 10. We achieved average recognition- and false alarm rates of 89.6% and 7.6% respectively.

Table 10. Upper face AU Recognition Combining Geometric Features and Regional Appearance Patterns.

AUs	Total	Correct	Missed	False
AU1	104	101	3	1
AU2	76	76	0	3
AU4	84	75	9	8
AU5	60	47	13	6
AU6	52	42	10	6
AU7	28	3	25	0
AU41	20	12	8	6
AU43	48	46	2	13
AU0	199	199	0	8
Total	671	601	70	51
Average Recognition Rate: 89.6%				
False Alarm Rate: 7.6%				

(2) Lower face: In this experiment, both geometric features and regional appearance patterns are fed to the network. The inputs are 8 geometric feature and 432 Gabor coefficients (3 last spatial frequencies in 8 orientations applied at 18 locations). The recognition results are shown in Table 11. We achieved average recognition- and false alarm rates of 82% and 14.2% respectively.

(3) Significance of Image Scales: From our experiments, we also found that the Gabor wavelet coefficients at each image scale do not play the same role. In regional patterns extraction, we calculate 5 spatial frequencies ($k = (\frac{\pi}{2}), (\frac{\pi}{4}),(\frac{\pi}{8}),(\frac{\pi}{16}),(\frac{\pi}{32})$). To test the significance of image scales, we perform the experiments for AU recognition in upper and lower face by combining the geometric features with the first 3 spatial frequencies ($k = (\frac{\pi}{2}), (\frac{\pi}{4}),(\frac{\pi}{8})$) , 3 middle spatial frequencies ($k = (\frac{\pi}{4}), (\frac{\pi}{8}),(\frac{\pi}{16})$) , and 3 last spatial frequencies ($k = (\frac{\pi}{8}), (\frac{\pi}{16}),(\frac{\pi}{32})$) respectively. We found that the best results were obtained by using 3 middle frequencies for the upper face and 3 last frequencies for the lower face (Figure 7).

3.2.4 Refinement of AU Recognition results

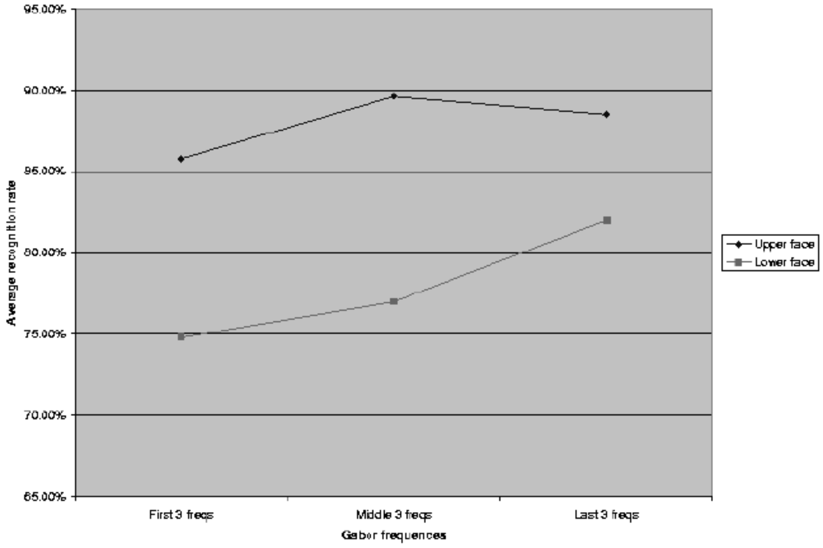


Figure 7. Significance of Image Scales. The best results were achieved by using 3 middle frequencies for the upper face and 3 last frequencies for the lower face.

Table 11. Lower face AU Recognition Combining Geometric Features and Regional Appearance Patterns.

AUs	Total	Correct	Missed	False
AU9	28	18	10	0
AU10	12	7	5	6
AU12	68	61	7	2
AU13	14	6	8	5
AU14	12	0	12	15
AU15	36	21	15	2
AU17	68	52	16	5
AU18	6	0	6	2
AU20	24	19	5	1
AU23	18	14	4	0
AU24	20	13	7	0
AU25	106	100	6	11
AU27	42	40	2	10
AU0	174	164	10	30
Total	628	515	113	89
Average Recognition Rate: 82%				
False Alarm Rate: 14.2%				

Sung [17] provided some formalization of how a set of identically trained detection can be used together to improve accuracy. He argued that if the errors made by a detector are independent, then by having a set of networks vote on the result, the number of overall errors will be reduced [16]. Rowley [16] presented two strategies to improve the reliability of the face detector: clean-up the outputs from an individual network, and arbitrating among multiple networks. Here, we use arbitration among multiple networks to improve the AU recognition results.

(1) Upper face: For refining the AU recognition, we train a new neural network. The inputs of the network are the output values of the networks by using geometric features and by combining both regional patterns and tracking features. The outputs are the AUs. In comparison to the results of using either the geometric features or the regional appearance patterns alone, combining these features increases accuracy, recognition performance has been improved to 92.7% as shown in Table 12.

(2) Lower face: For refining the AU recognition result in the lower face, Oring the outputs of the two networks is used for AU10, AU17, and AU25. For other AUs, the outputs of the network by using

Table 12. Refined AU Recognition Results for the Upper face.

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%				

geometric features are used. The average recognition rate is improved to 87.4

3.2.5 Lower face AU Recognition Using the Same Datasets as PAMI paper

For compare the AU recognition result with our PAMI paper [21], we use the same datasets to train and test the networks. In the training and test data, several sequences with AU combinations are recoded as different with the PAMI paper. For example, the sequence AU10+15+17 is recoded as AU10+17 and the sequences AU12+26 is recoded as AU12+25. The AU recognition results by using tracking features alone, combination of both tracking features and regional patterns, and after refinement are shown in Table 14, Table 15, and Table 16 respectively. The average recognition rates are 94.5%, 94.5%, and 96.8% with very low false alarm rates. Although the average recognition rates are same by using tracking features alone and both tracking features and regional patterns, AU10 and AU17 are improved when we add the regional patterns. Compare with the PAMI paper, the recognition rates are a little lower but with very low false alarm rates. It is normal to have 1% to 2% differences for the average recognition rate when we stop the training at different time for the different networks. Higher recognition rates can be achieved with higher false alarm.

4. Conclusion and Discussion

We summarize the AU recognition results in Figure 8 and Figure 9. In Figure 8, four recognition rates for each AU are described by histograms. The gray histogram shows recognition results based on only

Table 13. Refined AU Recognition Results for the Lower Face.

AUs	Total	Correct	Missed	False
AU9	28	18	10	0
AU10	12	7	5	7
AU12	68	62	6	2
AU13	14	13	1	3
AU14	12	5	7	1
AU15	36	25	11	6
AU17	68	52	16	7
AU18	6	6	0	0
AU20	24	22	2	1
AU23	18	13	5	0
AU24	20	13	7	0
AU25	106	100	6	11
AU27	42	39	3	5
AU0	174	174	0	13
Total	628	549	79	56
Average Recognition Rate: 87.4%				
False Alarm Rate: 8.8%				

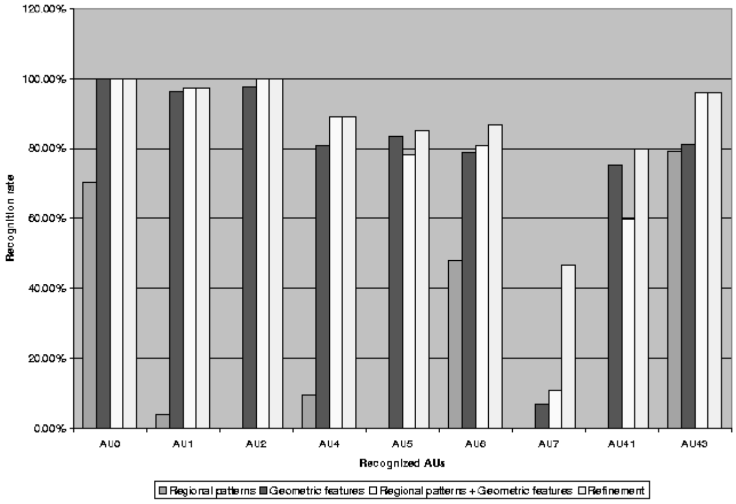


Figure 8. AU recognition results for the upper face. The gray histogram shows recognition results based on only geometric features. The dark gray histogram shows recognition results based only on regional appearance patterns, the white histogram shows results obtained using both types of features, and the bright gray histogram shows the refinement of the recognition results.

Table 14. Lower face AU recognition results by using tracking features alone

<table>
<tr><td colspan="2" rowspan="3">Actual AUs</td><td rowspan="3">Samples</td><td colspan="4">Recognized AUs</td></tr>
<tr><td rowspan="2">$Correct$</td><td colspan="2">$Partially\ correct$</td><td rowspan="2">$Incorrect$</td></tr>
<tr><td>$Missing\ AUs$</td><td>$Extra\ AUs$</td></tr>
<tr><td colspan="2">AU 9</td><td>2</td><td>2</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 10</td><td>4</td><td>0</td><td>4</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 12</td><td>4</td><td>4</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 15</td><td>2</td><td>2</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 17</td><td>6</td><td>4</td><td>2</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 20</td><td>2</td><td>2</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 25</td><td>30</td><td>30</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 26</td><td>12</td><td>10</td><td>-</td><td>-</td><td>2(AU 25)</td></tr>
<tr><td colspan="2">AU 27</td><td>8</td><td>8</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 23+24</td><td>0</td><td>-</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 9+17</td><td>12</td><td>12</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 9+17+23+24</td><td>2</td><td>2</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 9+25</td><td>2</td><td>2</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 10+17</td><td>4</td><td>0</td><td>2(AU 17), 2(AU 10)</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 10+25</td><td>2</td><td>0</td><td>2(AU 25)</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 12+25</td><td>10</td><td>10</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 15+17</td><td>10</td><td>10</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 17+23+24</td><td>4</td><td>4</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 20+25</td><td>10</td><td>10</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">$NEUTRAL$</td><td>63</td><td>63</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td rowspan="6">With respect to samples</td><td rowspan="2">Total No. of input samples</td><td>126</td><td>112</td><td colspan="3" rowspan="2">14</td></tr>
<tr><td>189</td><td>175</td></tr>
<tr><td rowspan="2">Recognition rate of samples</td><td colspan="5">88.9% (excluding $NEUTRAL$)</td></tr>
<tr><td colspan="5">94.5% (including $NEUTRAL$)</td></tr>
<tr><td rowspan="2">False alarm of samples</td><td colspan="5">1.6% (excluding $NEUTRAL$)</td></tr>
<tr><td colspan="5">1.1% (including $NEUTRAL$)</td></tr>
<tr><td rowspan="6">With respect to AU components</td><td rowspan="2">Total No. of AUs</td><td>190</td><td>176</td><td rowspan="2">12</td><td rowspan="2">0</td><td rowspan="2">2</td></tr>
<tr><td>253</td><td>239</td></tr>
<tr><td rowspan="2">Recognition rate of AUs</td><td colspan="5">92.6% (excluding $NEUTRAL$)</td></tr>
<tr><td colspan="5">94.5% (including $NEUTRAL$)</td></tr>
<tr><td rowspan="2">False alarm of AUs</td><td colspan="5">1.1% (excluding $NEUTRAL$)</td></tr>
<tr><td colspan="5">0.8% (including $NEUTRAL$)</td></tr>
</table>

Table 15. Lower face AU recognition results by using both tracking features and regional patterns

<table>
<tr><th colspan="2" rowspan="3">Actual AUs</th><th rowspan="3">Samples</th><th colspan="4">Recognized AUs</th></tr>
<tr><th rowspan="2">Correct</th><th colspan="2">Partially correct</th><th rowspan="2">Incorrect</th></tr>
<tr><th>Missing AUs</th><th>Extra AUs</th></tr>
<tr><td colspan="2">AU 9</td><td>2</td><td>2</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 10</td><td>4</td><td>2</td><td>2</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 12</td><td>4</td><td>4</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 15</td><td>2</td><td>2</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 17</td><td>6</td><td>6</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 20</td><td>2</td><td>-</td><td>2</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 25</td><td>30</td><td>28</td><td>2</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 26</td><td>12</td><td>10</td><td>-</td><td>-</td><td>2(AU 25)</td></tr>
<tr><td colspan="2">AU 27</td><td>8</td><td>6</td><td>-</td><td>2(AU 10+27)</td><td>-</td></tr>
<tr><td colspan="2">AU 23+24</td><td>0</td><td>-</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 9+17</td><td>12</td><td>11</td><td>1(AU 17)</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 9+17+23+24</td><td>2</td><td>2</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 9+25</td><td>2</td><td>1</td><td>1(AU 9)</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 10+17</td><td>4</td><td>2</td><td>2</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 10+25</td><td>2</td><td>0</td><td>2(AU 25)</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 12+25</td><td>10</td><td>10</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 15+17</td><td>10</td><td>10</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 17+23+24</td><td>4</td><td>4</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 20+25</td><td>10</td><td>10</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">$NEUTRAL$</td><td>63</td><td>63</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td rowspan="6">With respect to samples</td><td rowspan="2">Total No. of input samples</td><td>126</td><td>110</td><td colspan="3" rowspan="2">16</td></tr>
<tr><td>189</td><td>173</td></tr>
<tr><td rowspan="2">Recognition rate of samples</td><td colspan="5">87.3% (excluding $NEUTRAL$)</td></tr>
<tr><td colspan="5">91.5% (including $NEUTRAL$)</td></tr>
<tr><td rowspan="2">False alarm of samples</td><td colspan="5">3.2% (excluding $NEUTRAL$)</td></tr>
<tr><td colspan="5">2.2% (including $NEUTRAL$)</td></tr>
<tr><td rowspan="6">With respect to AU components</td><td rowspan="2">Total No. of AUs</td><td>190</td><td>176</td><td rowspan="2">12</td><td rowspan="2">2</td><td rowspan="2">2</td></tr>
<tr><td>253</td><td>239</td></tr>
<tr><td rowspan="2">Recognition rate of AUs</td><td colspan="5">92.6% (excluding $NEUTRAL$)</td></tr>
<tr><td colspan="5">94.5% (including $NEUTRAL$)</td></tr>
<tr><td rowspan="2">False alarm of AUs</td><td colspan="5">2.2% (excluding $NEUTRAL$)</td></tr>
<tr><td colspan="5">1.6% (including $NEUTRAL$)</td></tr>
</table>

Table 16. Lower face AU recognition results after refinement

<table>
<tr><td colspan="2" rowspan="3">Actual AUs</td><td rowspan="3">Samples</td><td colspan="4">Recognized AUs</td></tr>
<tr><td rowspan="2">$Correct$</td><td colspan="2">$Partially\ correct$</td><td rowspan="2">$Incorrect$</td></tr>
<tr><td>$Missing\ AUs$</td><td>$Extra\ AUs$</td></tr>
<tr><td colspan="2">AU 9</td><td>2</td><td>2</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 10</td><td>4</td><td>4</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 12</td><td>4</td><td>4</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 15</td><td>2</td><td>2</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 17</td><td>6</td><td>6</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 20</td><td>2</td><td>2</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 25</td><td>30</td><td>30</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 26</td><td>12</td><td>10</td><td>-</td><td>-</td><td>2(AU 25)</td></tr>
<tr><td colspan="2">AU 27</td><td>8</td><td>8</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 23+24</td><td>0</td><td>-</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 9+17</td><td>12</td><td>12</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 9+17+23+24</td><td>2</td><td>2</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 9+25</td><td>2</td><td>2</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 10+17</td><td>4</td><td>0</td><td>2(AU 17), 2(AU 10)</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 10+25</td><td>2</td><td>0</td><td>2(AU 25)</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 12+25</td><td>10</td><td>10</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 15+17</td><td>10</td><td>10</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 17+23+24</td><td>4</td><td>4</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">AU 20+25</td><td>10</td><td>10</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">$NEUTRAL$</td><td>63</td><td>63</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td rowspan="6">With respect to samples</td><td rowspan="2">Total No. of input samples</td><td>126</td><td>118</td><td colspan="3" rowspan="2">8</td></tr>
<tr><td>189</td><td>181</td></tr>
<tr><td rowspan="2">Recognition rate of samples</td><td colspan="5">93.7% (excluding $NEUTRAL$)</td></tr>
<tr><td colspan="5">95.8% (including $NEUTRAL$)</td></tr>
<tr><td rowspan="2">False alarm of samples</td><td colspan="5">1.6% (excluding $NEUTRAL$)</td></tr>
<tr><td colspan="5">1.1% (including $NEUTRAL$)</td></tr>
<tr><td rowspan="6">With respect to AU components</td><td rowspan="2">Total No. of AUs</td><td>190</td><td>182</td><td rowspan="2">6</td><td rowspan="2">0</td><td rowspan="2">2</td></tr>
<tr><td>253</td><td>245</td></tr>
<tr><td rowspan="2">Recognition rate of AUs</td><td colspan="5">95.8% (excluding $NEUTRAL$)</td></tr>
<tr><td colspan="5">96.8% (including $NEUTRAL$)</td></tr>
<tr><td rowspan="2">False alarm of AUs</td><td colspan="5">1.1% (excluding $NEUTRAL$)</td></tr>
<tr><td colspan="5">0.8% (including $NEUTRAL$)</td></tr>
</table>

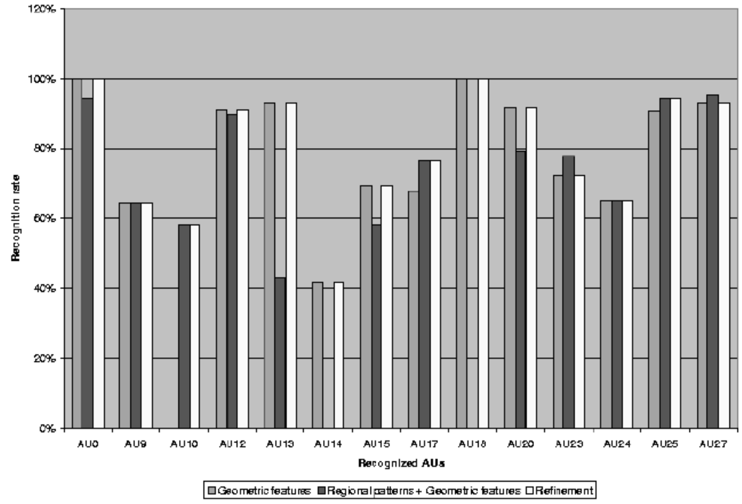


Figure 9. AU recognition results for the lower face. The gray histogram shows recognition results based on only geometric features. The dark gray histogram shows recognition results based on both geometric features and regional appearance patterns, the white histogram shows the refinement of the recognition results.

geometric features. The dark gray histogram shows recognition results based only on regional appearance patterns, the white histogram shows results obtained using both types of features, and the bright gray histogram shows the refinement of the recognition results. Using regional appearance patterns 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.

In Figure 9, three recognition rates for each AU are described by histograms. The gray histogram shows recognition results based on only geometric features. The dark gray histogram shows recognition results based on both geometric features and regional appearance patterns, the white histogram shows the refinement of the recognition results. Using regional appearance patterns alone, the neural network did not converge. Using geometric features, recognition is consistently good with the exception of AU10. 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 AU10, AU17, and AU25.

While previous studies have achieved high accuracy using Gabor wavelet coefficients [1, 26], we are surprised to find relatively poor recognition using this approach. Several factors may account for the difference. First, the previous studies manually aligned and cropped face images. We omitted this preprocessing step. Our geometric feature-based approach requires no image alignment and cropping, and we wished to retain this advantage. Second, the previous studies used very homogeneous subjects. Zhang et al., for instance, included only Japanese. We use diverse subjects of European, African, and Asian ancestry. Third, the previous studies recognized emotion-specified expressions or upper-face expressions in which only a single AU was present. We focus on both single AUs and AU combinations, including non-additive combinations in which the occurrence of one AU modifies another. We also recognized more upper-face AUs than in previous work. These differences suggest that any advantage of Gabor wavelets in facial expression recognition may depend on manual preprocessing and may fail to generalize to heterogeneous subjects and more varied facial expression. Combining Gabor wavelet coefficients and geometric features resulted in the best performance. Using both types of features, we achieved 92.7% accuracy in recognizing 9 AUs for the upper face and 87.4% accuracy in recognizing 14 AUs for the lower whether they occurred alone or in complex combinations.

5. Future work

From the experiment results, we have found that adding the regional patterns which extracted by multi-scale and multi-orientation Gabor wavelets can increase the recognition rate for some specific AUs. I think that there are several things we should do in the future research. (1) Use the sequential information to recognized more complex AU combinations (for example, AU10+15+17) by TDNN or HMM. (2) Do more analysis about the regional pattern (can use feature selection) to analyze which scale or orientation of Gabor coefficients are most useful for recognizing specific AUs. (3) Do more analysis about the regional pattern (can use feature selection) to analyze which positions on the face are most useful for specific recognizing AUs.

Acknowledgements

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] 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. M. Carroll and J. Russell. Facial expression in hollywood's portrayal of emotion. *Journal of Personality and Social Psychology.*, 72:164–176, 1997.

[5] J. F. Cohn, A. J. Zlochower, J. Lien, and T. Kanade. Automated face analysis by feature point tracking has high concurrent validity with manual facs coding. *Psychophysiology*, 36:35–43, 1999.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

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

[11] 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.

[12] B. Lucas and T. Kanade. An interative image registration technique with an application in stereo vision. In *The 7th International Joint Conference on Artificial Intelligence*, pages 674–679, 1981.

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

[14] R. R. Rao. *Audio-Visal Interaction in Multimedia*. PHD Thesis, Electrical Engineering, Georgia Institute of Technology, 1998.

[15] 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.

[16] H. A. Rowley. *Neural Network-Based Face Detection*. PhD Thesis, Carnegie Mellon University, May, 1999. Available as CMU Technical Report CMU-CS-99-117.

[17] K.-K. Sung. *Learning and Example Selection for Object and Pattern Detection*. PhD Thesis, MIT AI Lab, January, 1996. Available as AI Technical Report 1572.

[18] 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.

[19] Y. Tian, T. Kanade, and J. Cohn. Recognizing upper face actions for facial expression analysis. In *Proc. Of CVPR'2000*, pages I294–301, 2000.

[20] Y. Tian, T. Kanade, and J. Cohn. Robust lip tracking by combining shape, color and motion. In *Proc. Of ACCV'2000*, pages 1040–1045, 2000.

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

[22] 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*, pages 484–490, March, 2000.

[23] 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.

[24] 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.

[25] Y. Yacoob, H. Lam, and L. Davis. Recognizing face showing expressions. In *Proc. Of the international workshop on Automatic Face and Gesture Recognition, Zurich*, 1995.

[26] 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.