Recognizing Action Units for Facial Expression Analysis

Ying-li Tian   Takeo Kanade   Jeffrey F. Cohn

CMU-RI-TR-99-40

Robotics Institute, Carnegie Mellon University,
Pittsburgh, PA 15213

December, 1999

This research is sponsored by NIMH, under contract 1R01MH51435 “Facial Expression Analysis by Computer Image Processing”.

Keywords: Facial expression analysis Action units Neural network

Abstract

Most automatic expression analysis systems attempt to recognize a small set of prototypic expressions (e.g. happiness and anger). Such prototypic expressions, however, occur infrequently. Human emotions and intentions are communicated more often by changes in one or two discrete facial features. We develop an automatic system to analyze subtle changes in facial expressions based on both permanent facial features (brows, eyes, mouth) and transient facial features (deepening of facial furrows) in a nearly frontal image sequence. Unlike most existing systems, our system attempts to recognize fine-grained changes in facial expression based on Facial Action Coding System (FACS) action units (AUs), instead of six basic expressions (e.g. happiness and anger). Multi-state face and facial component models are proposed for tracking and modeling different facial features, including lips, eyes, brows, cheeks, and their related wrinkles and facial furrows. Then we convert the results of tracking to detailed parametric descriptions of the facial features. With these features as the inputs, 11 lower face action units (AUs) and 7 upper face AUs are recognized by a neural network algorithm. A recognition rate of 96.7% for lower face AUs and 95% for upper face AUs is obtained respectively. The recognition results indicate that our system can identify action units regardless of whether they occurred singly or in combinations.

1. Introduction

Recently facial expression analysis has attracted attention in the computer vision literature [3, 5, 6, 9, 11, 13, 17, 19]. Most automatic expression analysis systems attempt to recognize a small set of prototypic expressions (i.e. joy, surprise, anger, sadness, fear, and disgust) [11, 17]. In everyday life, however, such prototypic expressions occur relatively infrequently. Instead, emotion is 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 [2]. Change in isolated features, especially in the area of the brows or eyelids, is typical of paralinguistic displays; for instance, raising the brows signals greeting. To capture the subtlety of human emotion and paralinguistic communication, automated recognition of fine-grained changes in facial expression is needed.

Ekman and Friesen [4] developed the Facial Action Coding System (FACS) for describing facial expressions. The FACS is a human-observer-based system designed to describe subtle changes in facial features. FACS consists of 44 action units, including those for head and eye positions. AUs are anatomically related to contraction of specific facial muscles. They can occur either singly or in combinations. AU combinations may be additive, in which case combination does not change the appearance of the constituents, or nonadditive, in which case the appearance of the constituents changes (analogous to co-articulation effects in speech). For action units that vary in intensity, a 5-point ordinal scale is used to measure the degree of muscle contraction. Although the number of atomic action units is small, more than 7,000 combinations of action units have been observed [12]. FACS provides the necessary detail with which to describe facial expression.

Automatic recognition of action units is a difficult problem. AUs have no quantitative definitions and as noted can appear in complex combinations. Several researchers have tried to recognize AUs [1, 3, 9]. The system of Lien *et al.* [9] used dense-flow, feature point tracking and edge extraction to recognize 6 upper face AUs or AU combinations (AU1+2, AU1+4, AU4, AU5, AU6, and AU7) and 9 lower face AUs and AU combinations (AU12, AU25, AU26, AU27, AU12+25, AU20+25, AU15+17, AU17+23+24, AU9+17). Bartlett *et al.* [1] recognized 6 individual upper face AUs (AU1, AU2, AU4, AU5, AU6, and AU7) but none occurred in combinations. The performance of their feature-based classifier on novel faces was 57%; on new images of faces used for training, the rate was 85.3%. By combining holistic spatial analysis and optical flow with local features in a hybrid system, Bartlett *et al.* increased accuracy to 90.9% correct. Donato *et al.* [3] compared several techniques for recognizing action units including optical flow, principal component analysis, independent component analysis, local feature analysis, and Gabor wavelet representation. Best performances were obtained by Gabor wavelet representation and independent component analysis which achieved a 95% average recognition rate for 6 upper face AUs and 6 lower face AUs.

In this report, we developed a feature-based AU recognition system. This system explicitly analyzes appearance changes in localized facial features. Since each AU is associated with a specific set of facial muscles, we believe that accurate geometrical modeling of facial features will lead to better recognition results. Furthermore, the knowledge of exact facial feature positions could benefit the area-based [17], holistic analysis [1], or optical flow based [9] classifiers. Figure 1 depicts the overview of the analysis system. First, the head orientation and face position are detected. Then, subtle changes in the facial components are measured. Motivated by FACS action units, these changes are represented as a collection of mid-level feature parameters. Finally, action units are classified by feeding these parameters to a neural network.

Because the appearance of facial features is dependent upon head orientation, we develop a multi-state model-based system for tracking facial features. Different head orientations and corresponding variation in the appearance of face components are defined as separate states. For each state, a corresponding description and one or more feature extraction methods are developed.

We separately represent all the facial features into two parameter groups for upper face and lower face because facial actions in the upper and lower face are relatively independent [4]. Fifteen parameters are used to describe eye shape, motion, and state, and brow and cheek motion, and upper face furrows for upper face. Nine parameters are used to describe the lip shape, lip motion, lip state, and lower face furrows for lower face.

After the facial features are correctly extracted and suitably represented, we employ a neural network to recognize the upper face AUs (Neutral, AU1, AU2, AU4, AU5, AU6, and AU7) and lower face AUs (Neutral, AU9, AU 10, AU 12, AU 15, AU 17, AU 20, AU 25, AU 26, AU 27, and AU23+24) respectively. Seven basic upper face AUs and eleven basic lower face AUs are identified regardless of whether they occurred singly or in combinations. For the upper face AU recognition, compared to Bartlett's [1] results by using the same database, our system achieves recognition accuracy with an average recognition rate of 95% with fewer parameters and in the more difficult case in which AUs may occur either individually or in additive and nonadditive combinations. For the lower face AU recognition, a previous attempt for a similar task [9] recognized 6 lower face AUs and combinations(AU 12, AU12+25, AU20+25, AU9+17, AU17+23+24, and AU15+17) with 88% average recognition rate by separate hidden Markov Models for each action unit or action unit combination. Compared to the previous results, our system achieves recognition accuracy with an average recognition rate of 96.71%. Difficult cases in which AUs occur either individually or in additive and nonadditive combinations are handled also.

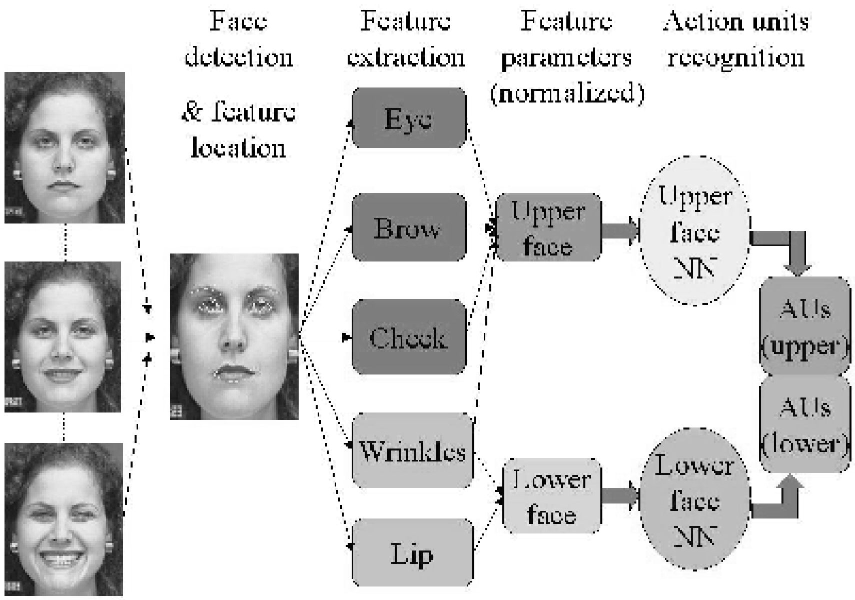


Figure 1. Feature based action unit recognition system.

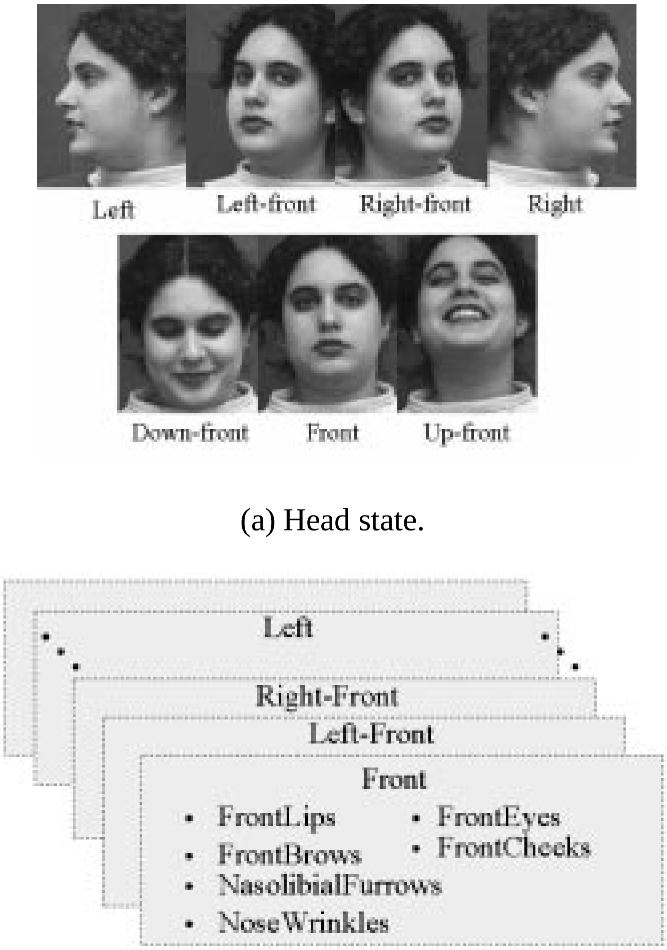
2. Multi-State Models for Face and Facial Components

2.1. Multi-state face model

Head orientation is a significant factor that affects the appearance of a face. Based on the head orientation, seven head states are defined in Figure 2. To develop more robust facial expression recognition system, head state will be considered. For the different head states, facial components, such as lips, appear very differently, requiring specific facial component models. For example, the facial component models for a front face include $FrontLips$, $FrontEyes$ (left and right), $FrontCheeks$(left and right), $NasolabialFurrows$, and $Nosewrinkles$. The right face includes only the component models $SideLips$, $Righteye$, $Rightbrow$, and $Rightcheek$. In our current system, we assume the face images are nearly front view with possible in-plane head rotations.

2.2. Multi-state face component models

Different face component models must be used for different states. For example, a lip model of the front face doesn't work for a profile face. Here, we give the detailed facial component models for the nearly front-view face. Both the permanent components such as lips, eyes, brows, cheeks and the transient components such as furrows are considered. Based on the different appearances of different components, different geometric models are used to model the component's location, shape, and appearance. Each



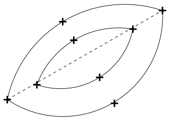
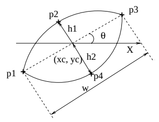
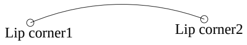
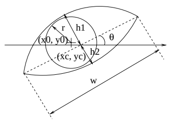
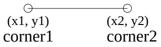
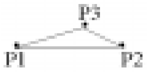
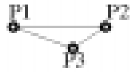
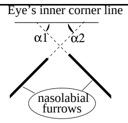
(a) Head state.

(b) Different facial components used for each head state.

Figure 2. Multiple state face model. (a) The head state can be left, left-front, front, right-front, right, down, and up. (b) Different facial component models are used for different head states.

component employs a multi-state model corresponding to different component states. For example, a three-state lip model is defined to describe the lip states: open, closed, and tightly closed. A two-state eye model is used to model open and closed eye. There is one state for brow and cheek. Present and absent are use to model states of the transient facial features. The multi-state component models for different components are described in Table 1.

Table 1. Multi-state facial component models of a front face

Component	State	Description/Feature
Lip	Opened	
	Closed	
	Tightly closed	
Eye	Open	
	Closed	
Brow	Present	
Cheek	Present	
Furrow	Present	
	Absent	

3. Facial Feature Extraction

Contraction of the facial muscles produces changes in both the direction and magnitude of the motion on the skin surface and in the appearance of permanent and transient facial features. Examples of permanent features are the lips, eyes, and any furrows that have become permanent with age. Transient features include any facial lines and furrows that are not present at rest. We assume that the first frame is in a neutral expression. After initializing the templates of the permanent features in the first frame, both permanent and transient features can be tracked and detected in the whole image sequence regardless of the states of facial components. The tracking results show that our method is robust for tracking facial features even when there is large out of plane head rotation.

3.1. Permanent features

Lip features: A three-state lip model is used for tracking and modeling lip features. As shown in Table 1, we classify the mouth states into open, closed, and tightly closed. Different lip templates are used to obtain the lip contours. Currently, we use the same template for open and closed mouth. Two parabolic arcs are used to model the position, orientation, and shape of the lips. The template of open and closed lips has six parameters: lip center (xc, yc), lip shape ($h1$, $h2$ and w), and lip orientation (θ). For a tightly closed mouth, the dark mouth line connecting lip corners is detected from the image to model the position, orientation, and shape of the tightly closed lips.

After the lip template is manually located for the neutral expression in the first frame, the lip color is obtained by modeling as a Gaussian mixture. The shape and location of the lip template for the image sequence is automatically tracked by feature point tracking. Then, the lip shape and color information are used to determine the lip state and state transitions. The detailed lip tracking method can be found in paper [15].

Eye features: Most eye trackers developed so far are for open eyes and simply track the eye locations. However, for recognizing facial action units, we need to recognize the state of eyes, whether they are open or closed, and the parameters of an eye model, the location and radius of the iris, and the corners and height of the open eye. As shown in Table 1, the eye model consists of "open" and "closed".

The iris provides important information about the eye state. If the eye is open, part of the iris normally will be visible. Otherwise, the eye is closed. For the different states, specific eye templates and different algorithms are used to obtain eye features.

For an open eye, we assume the outer contour of the eye is symmetrical about the perpendicular bisector to the line connecting two eye corners. The template, illustrated in Table 1, is composed of a circle with three parameters (x_0, y_0, r) and two parabolic arcs with six parameters $(x_c, y_c, h_1, h_2, w, \theta)$. This is the same eye template as Yuille's except for two points located at the center of the whites [18]. For a closed eye, the template is reduced to 4 parameters for each of the eye corners.

The default eye state is open. Locating the open eye template in the first frame, the eye's inner corner is tracked accurately by feature point tracking. We found that the outer corners are hard to track and less stable than the inner corners, so we assume the outer corners are on the line that connects the inner corners. Then, the outer corners can be obtained by the eye width, which is calculated from the first frame.

Intensity and edge information are used to detect an iris because the iris provides important information about the eye state. A half-circle iris mask is used to obtain correct iris edges. If the iris is detected, the eye is open and the iris center is the iris mask center (x_0, y_0). In an image sequence, the eyelid contours are tracked for open eyes by feature point tracking. For a closed eye, we do not need to track the eyelid contours. A line connects the inner and outer corners of the eye is used as the eye boundary. The detailed eye feature tracking techniques can be found in paper [14].

Brow and cheek features: Features in the brow and cheek areas are also important to facial expression analysis. For the brow and cheek, one state is used respectively, a triangular template with six parameters $(x1, y1)$, $(x2, y2)$, and $(x3, y3)$ is used to model the position of brow or cheek. Both brow and cheek are tracked by feature point tracking. A modified version of the gradient tracking algorithm [10] is used to track these points for the whole image sequence. Some permanent facial feature tracking results for different expressions are shown in Figure 3. More facial feature tracking results can be found in http://www.cs.cmu.edu/∼face.

3.2. Transient features

Facial motion produces transient features. Wrinkles and furrows appear perpendicular to the motion direction of the activated muscle. These transient features provide crucial information for the recognition of action units. Contraction of the corrugator muscle, for instance, produces vertical furrows between the brows, which is coded in FACS as AU 4, while contraction of the medial portion of the frontalis muscle (AU 1) causes horizontal wrinkling in the center of the forehead.

Some of these lines and furrows may become permanent with age. Permanent crows-feet wrinkles around the outside corners of the eyes, which is characteristic of AU 6 when transient, are common in adults but not in infants. When lines and furrows become permanent facial features, contraction of the corresponding muscles produces changes in their appearance, such as deepening or lengthening. The presence or absence of the furrows in a face image can be determined by geometric feature analysis [9, 8], or by eigen-analysis [7, 16]. Kwon and Lobo [8] detect furrows by snake to classify pictures of people into different age groups. Lien [9] detected whole face horizontal, vertical and diagonal edges for face expression recognition.

In our system, we currently detect nasolabial furrows, nose wrinkles, and crows feet wrinkles. We define them in two states: present and absent. Compared to the neutral frame, the wrinkle state is present if the wrinkles appear, deepen, or lengthen. Otherwise, it is absent. After obtaining the permanent facial features, the areas with furrows related to different AUs can be decided by the permanent facial feature locations. We define the nasolabial furrow area as the area between eye's inner corners line and lip corners line. The nose wrinkle area is a square between two eye inner corners. The crows feet wrinkle areas are beside the eye outer corners.

We use canny edge detector to detect the edge information in these areas. For nose wrinkles and crows feet wrinkles, we compare the edge pixel numbers E of current frame with the edge pixel numbers E_0 of the first frame in the wrinkle areas. If E/E_0 large than the threshold T, the furrows are present. Otherwise, the furrows are absent. For the nasolabial furrows, we detect the continued diagonal edges. The nasolabial furrow detection results are shown in Fig. 4.

4. Facial Feature Representation

Each action unit of FACS is anatomically related to contraction of a specific facial muscle. For instance, AU 12 (oblique raising of the lip corners) results from contraction of the zygomaticus major muscle, AU 20 (lip stretch) from contraction of the risorius muscle, and AU 15 (oblique lowering of the lip corners) from contraction of the depressor anguli muscle. Such muscle contractions produce motion in the overlying skin and deform shape or location of the facial components. In order to recognize the

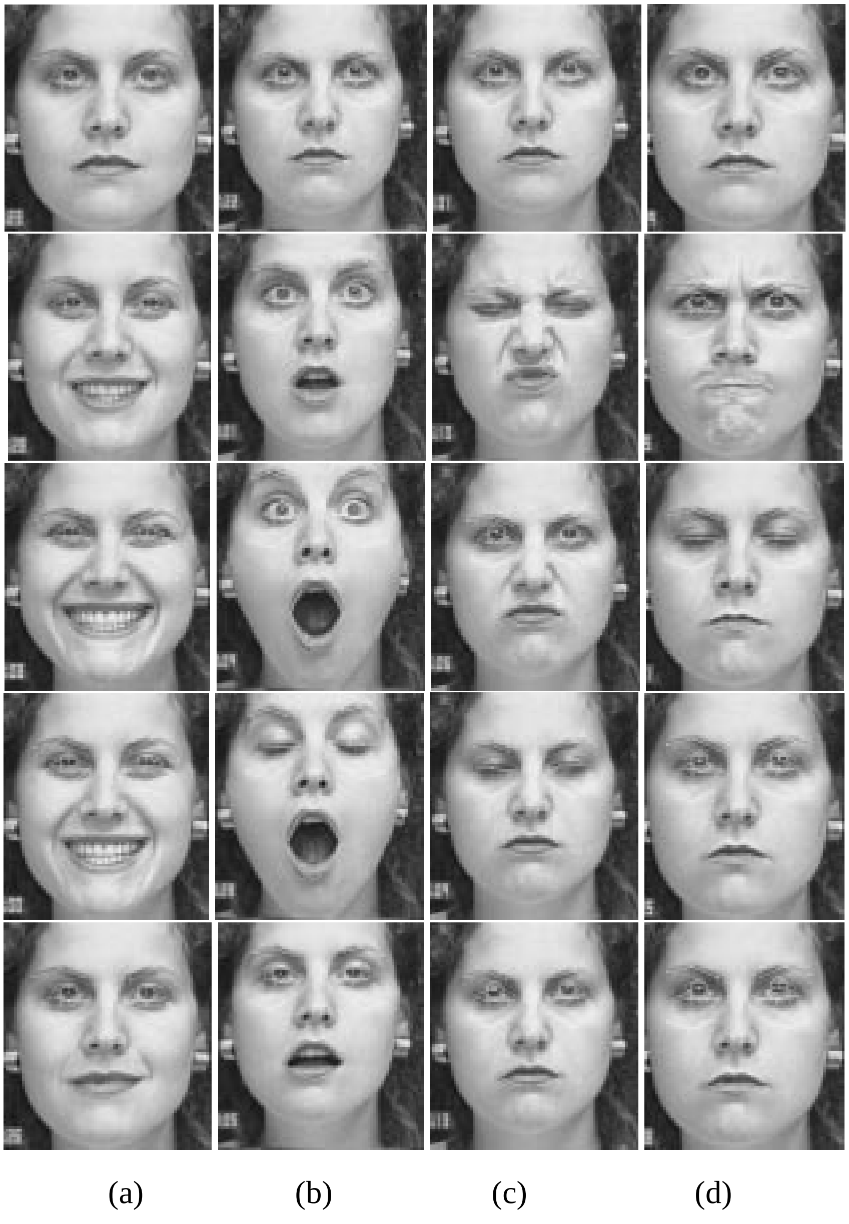


Figure 3. Permanent feature tracking results for different expressions. (a) Narrowing eyes and opened smiled mouth. (b) Large open eye, blinking and large opened mouth. (c) Tight closed eye and eye blinking. (4) Tightly closed mouth and blinking.

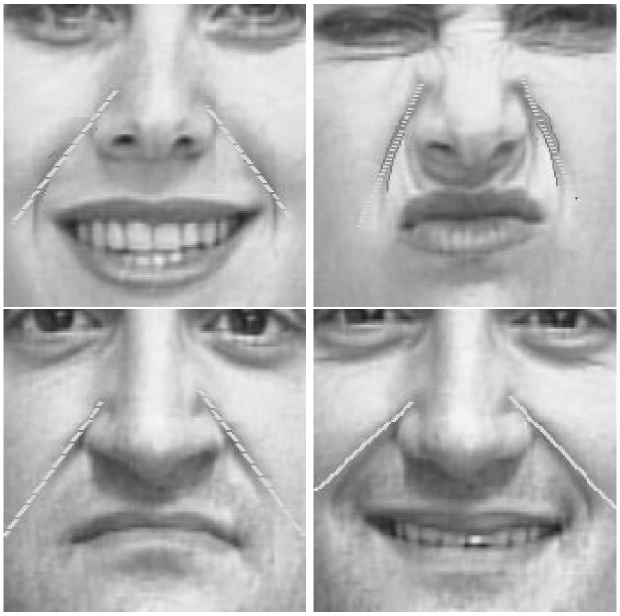


Figure 4. Nasolabial furrow detection results. For the same subject, the nasolabial furrow angle(between the nasolabial furrow and the line connected eye inner corners) is different for different expressions.

subtle changes of face expression, we represent the upper face features and lower face features into a group of suitable parameters respectively because facial actions in the upper face have little influence on facial motion in lower face, and vice versa [4].

For defining these parameters, we first define the basic coordinate system. Because the eye's inner corners are the most stable features in the face and are relatively insensitive to deformation by facial expressions, we define the x-axis as the line connecting two inner corners of eyes and the y-axis as perpendicular to x-axis. In order to remove the effects of the different size of face images in different image sequences, all the parameters except those about wrinkles' states are calculated in ratio scores by comparison to the neutral frame.

4.1. Upper Face Feature Representation

We represent the upper face features as 15 parameters. Of these, 12 parameters describe the motion and shape of eyes, brows, and cheeks. 2 parameters describe the state of crows feet wrinkles, and 1 parameter describes the distance between brows. Figure 5 shows the coordinate system and the parameter meanings. The definitions of upper face parameters are listed in Table 2.

Table 2. Upper face feature representation for AU recognition

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.

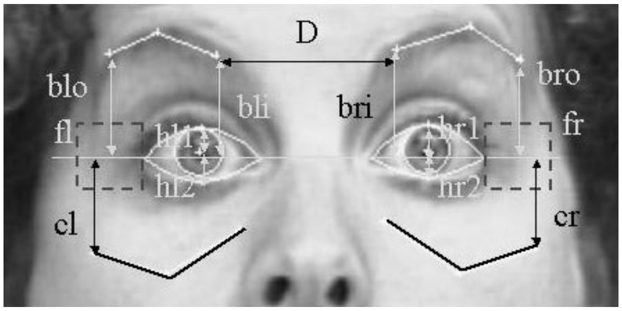


Figure 5. Upper face features. $hl(hl1 + hl2)$ and $hr(hr1 + hr2)$ are the height of left eye and right eye; D is the distance between brows; cl and cr are the motion of left cheek and right cheek. bli and bri are the motion of the inner part of left brow and right brow. blo and bro are the motion of the outer part of left brow and right brow. fl and fr are the left and right crows feet wrinkle areas.

4.2. Lower Face Feature Representation

We define nine parameters to represent the lower face features from the tracked facial features. Of these, 6 parameters describe the permanent features of lip shape, lip state and lip motion, and 3 parameters describe the transient features of the nasolabial furrows and nose wrinkles.

We notice that if the nasolabial furrow is present, there are different angles between the nasolabial furrow and x-axis for different action units. For example, the nasolanial furrow angle of AU9 or AU10 is larger than that of AU12. So we use the angle to represent its orientation if it is present. Although the nose wrinkles are located in the upper face, but we classify the parameter of them in the lower face feature because it is related to the lower face AUs.

The definitions of lower face parameters are listed in Table 3. These feature data are affine aligned by calculating them based on the line connected two inner corners of eyes and normalized for individual differences in facial conformation by converting to ratio scores. The parameter meanings are shown in Figure 6.

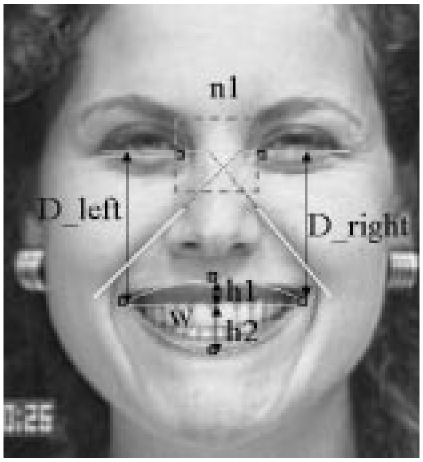


Figure 6. Lower face features. $h1$ **and** $h2$ **are the top and bottom lip heights;** w **is the lip width;** D_{left} **is the distance between the left lip corner and eye inner corners line;** D_{right} **is the distance between the right lip corner and eye inner corners line;** $n1$ **is the nose wrinkle area.**

5. Facial Action Unit Definitions

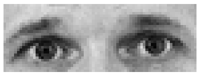
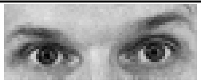
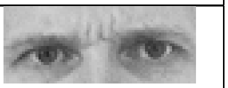
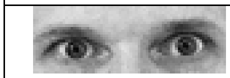
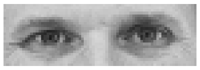
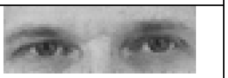
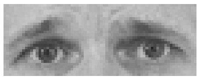
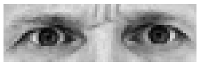
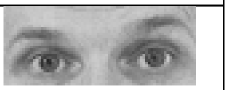
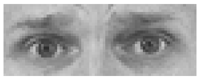
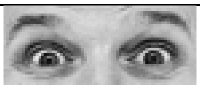
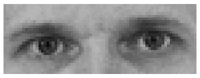
Ekman and Friesen [4] developed the Facial Action Coding System (FACS) for describing facial expressions by action units (AUs) or AU combinations. 30 FACS AUs are anatomically related to

Table 3. Representation of lower face features 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 move 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 move up.	$r_{top} = -\frac{D_{top}-D_{top0}}{D_{top0}}$. If r_{top}>0, top lip move up.	$r_{btm} = -\frac{D_{btm}-D_{btm0}}{D_{btm0}}$. If r_{btm}>0, bottom lip move up.
Transient features		
Left nasolibial furrow angle (Ang_{left})	Right nasolibial furrow angle (Ang_{right})	State of nose wrinkles (S_{nosew})
Left nasolibial furrow present with angle Ang_{left}.	Left nasolibial furrow present with angle Ang_{right}.	If $S_{nosew} = 1$, nose wrinkles present.

contraction of a specific set of facial muscles. Of thses, 12 are for upper face, and 18 are for lower face. Action units can occur either singly or in combinations. The action unit combinations may be additive such as AU1+5, in which case combination does not change the appearance of the constituents, or nonadditive, in which case the appearance of the constituents does change such as AU1+4. Although the number of atomic action units is small, more than 7,000 combinations of action units have been observed [12]. FACS provides the necessary detail with which to describe facial expression.

Table 4. Basic upper face action units or AU combinations

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 1+4	AU 4+5	AU 1+2
		
Medial portion of the brows is raised and pulled together.	Brows lowered and drawn together and upper eyelids are raised.	Inner and outer portions of the brows are raised.
AU 1+2+4	AU1+2+5+6+7	AU0(neutral)
		
Brows are pulled together and upward.	Brow, eyelids, and cheek are raised.	Eyes, brow, and cheek are relaxed.

5.1. Upper Face Action Units

Table 4 shows the definitions of 7 individual upper face AUs and 5 non-additive combinations involving these action units. As an example of a non-additive effect, AU4 appears differently depending on whether it occurs alone or in combination with AU1, as in AU1+4. When AU1 occurs alone, the brows are drawn together and lowered. In AU1+4, the brows are drawn together but are raised by the action of AU 1. As another example, it is difficult to notice any difference between the static images of AU2 and AU1+2 because the action of AU2 pulls the inner brow up, which results in a very similar appearance to AU1+2. In contrast, the action of AU1 alone has little effect on the outer brow.

5.2. Lower face action units

Table 5 shows the definitions of 11 lower face AUs or AU combinations.

6. Image Database

6.1. Image Database for Upper Face AU Recognition

We use the database of Bartlett *et al.* [1] for upper face AUs recognition. This image database was obtained from 24 Caucasian subjects, consisting of 12 males and 12 females. Each image sequence consists of 6-8 frames, beginning with a neutral or with very low magnitude facial actions and ending with a high magnitude facial actions. For each sequence, action units were coded by a certified FACS coder.

For this investigation, 236 image sequences from 24 subjects were processed. Of these, 99 image sequences contain only individual upper face AUs, and 137 image sequences contain upper-face AU combinations. Training and testing are performed on the initial and final two frames in each image sequence. For some of the image sequences, lighting normalizations were performed.

To test our algorithm on the individual AUs, we randomly generate training and testing sets from the 99 image sequences, as shown in Table 6. In $TrainS3$ and $TestS3$, we ensure that the subjects do not appear in both training and testing sets.

To test our algorithm on the both individuall AUs and AU combinations, we generate a training set ($TrainC1$) and a testing set ($TestC1$) as shown in Table 6.

Table 5. Basic lower face action units or AU combination

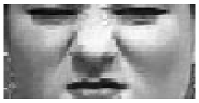
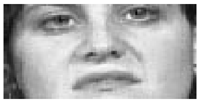
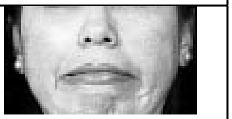
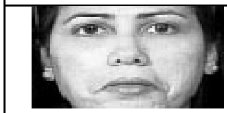
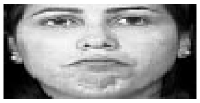
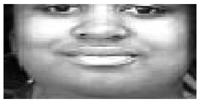
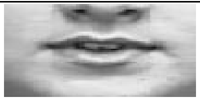
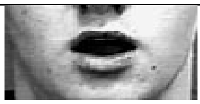
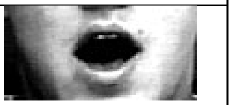
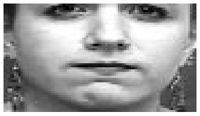
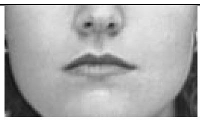
AU 9	AU 10	AU20
		
The infraorbital triangle and center of the upper lip are pulled upwards. Nose wrinkling is present.	The infraorbital triangle is pushed upwards. Upper lip is raised. Nose wrinkle is absent.	The lips and the lower portion of the nasolabial furrow are pulled pulled back laterally. The mouth is elongated.
AU 15	AU 17	AU12
		
The corner of the lips are pulled down.	The chin boss is pushed upwards.	Lip corners are pulled obliquely.
AU 25	AU 26	AU27
		
Lips are relaxed and parted.	Lips are relaxed and parted; mandible is lowered.	Mouth stretched, open and the mandible pulled downwards.
AU 23+24	neutral	
		
Lips tightened, narrowed, and pressed together.	Lips relaxed and closed.	

Table 6. Data distribution of each data set for upper face AU recognition.

Single AU Data Sets								
AUs	AU0	AU1	AU2	AU4	AU5	AU6	AU7	Total
$TrainS1$	47	14	12	16	22	12	8	141
$TestS1$	52	14	12	20	24	14	20	156
$TrainS2$	76	20	18	32	34	20	28	228
$TestS2$	23	8	6	4	12	6	10	69
$TrainS3$	52	18	14	14	18	24	16	156
$TestS3$	47	10	10	22	28	2	22	141
AU Combination Data Sets								
$TrainC1$	214	148	90	116	110	90	150	918
$TestC1$	22	18	12	10	10	14	8	94

6.2. Image Database for Lower Face AU Recognition

We use the data of *Pitt-CMU AU-Coded Face Expression Image Database* for lower face AU recognition. The database currently includes 1917 image sequences from 182 adult subjects of varying ethnicity, performing multiple tokens of 29 of 30 primary FACS action units. Subjects sat directly in front of the camera and performed a series of facial expressions that included single action units (e.g., AU 12, or smile) and combinations of action units (e.g., AU 6+12+25). Each expression sequence began from a neutral face. For each sequence, action units were coded by a certified FACS coder.

Total 463 image sequences from 122 adults (65% female, 35% male, 85% European-American, 15% African-American or Asian, ages 18 to 35 years) are processed for lower face action unit recognition. Some of the image sequences are with more action unit combinations such as AU9+17, AU10+17, AU12+25, AU15+17+23, AU9+17+23+24, and AU17+20+26. For each image sequence, we use the neutral frame and two peak frames. 400 image sequences are used as training data and 63 different image sequences are used as test data. The training and testing data sets are shown in Table 7.

Table 7. Training data set for lower face AU recognition

	neutral	AU9	AU10	AU12	AU15	AU17	AU20	AU25	AU26	AU27	AU23+24	Total
Train Set	400	38	30	160	136	204	72	212	94	96	48	1220
Test Set	63	16	12	14	12	36	12	50	14	8	6	243

7. Face Action Units Recognition

7.1. Upper Face Action Units Recognition

We used three-layer neural networks with one hidden layer. The inputs of the neural networks are the 15 parameters shown in Table 2. Three separate neural networks were evaluated. For comparison with Bartlett's results, the first NN is for recognizing individual AUs only. The second NN is for recognizing AU combinations when only modeling 7 individual upper face AUs. The third NN is for recognizing AU combinations when separately modeling nonadditive AU combinations. The desired number of hidden units to achieve a good recognition was also investigated.

7.1.1 Upper Face Individual AU Recognition

The NN outputs are 7 individual upper face AUs. Each output unit gives an estimate of the probability of the input image consisting of the associated action units. From experiments, we have found 6 hidden units are sufficient.

In order to recognize individual action units, we used the training and testing data that include individual AUs only. Table 8 shows results of our NN on the $TrainS1$, $TestS1$ training and testing sets. A 92.3% recognition rate was obtained. When we increase the training data by using $TrainS2$ and test by using $TestS2$, a 92% recognition rate was obtained.

For detecting the system's robustness to new faces, we tested our algorithm on the $TrainS3$/$TestS3$ training/testing sets. The recognition results are shown in Table 9. The average recognition rate is 92.9% with zero false alarms. For the misidentifications between AUs, although the probability of the output units of the labeled AU is very close to the highest probability, it was treated as an incorrect result. For

Table 8. AU recognition for single AUs on $TrainS1$ and $TestS1$. The rows correspond to NN outputs, and columns correspond to human labels.

	AU0	AU1	AU2	AU4	AU5	AU6	AU7
AU0	52	0	0	0	0	0	0
AU1	0	12	2	0	0	0	0
AU2	0	3	9	0	0	0	0
AU4	0	0	0	20	0	0	0
AU5	2	0	0	0	22	0	0
AU6	0	0	0	0	0	12	2
AU7	1	0	0	0	0	2	17
Average Recognition rate: 92.3%							

example, if we obtain the probability of AU1 and AU2 with AU1=0.59 and AU2=0.55 for a labeled AU2, it means that AU2 was misidentified as AU1. When we tested the NN trained on single AU image sequences on data set containing AU combinations, we found the recognition rate decreases to 78.7%.

Table 9. AU recognition for single AUs when all test data come from new subjects who were not used for training.

	AU0	AU1	AU2	AU4	AU5	AU6	AU7
AU0	47	0	0	0	0	0	0
AU1	0	10	0	0	0	0	0
AU2	1	2	7	0	0	0	0
AU4	2	0	0	20	0	0	0
AU5	2	0	0	0	26	0	0
AU6	2	0	0	0	0	0	0
AU7	1	0	0	0	0	0	21
Average Recognition rate: 92.9%							

7.1.2 Upper Face AU Combination Recognition When Modeling 7 Individual AUs

This NN is similar to the one used in the previous section, except that more than one output units could fire. We also restrict the output to be the first 7 individual AUs. For the additive and nonadditive AU combinations, the same value is given for each corresponding individual AUs in training data set. For example, for AU1+2+4, the outputs are AU1=1.0, AU2=1.0, and AU4=1.0. From experiments, we found we need to increase the number of hidden units from 6 to 12.

Table 10 shows the results of our NN on the ($TrainC1$)/($TestC1$) training/testing set. A 95% average recognition rate is achieved, with a false alarm rate of 6.4%. The higher false alarm rate comes from the AU combination. For example, if we obtained the recognition results with AU1=0.59 and AU2=0.55 for a labeled AU2, it was treated as AU1+AU2. This means AU2 is recognized but with AU1 as a false alarm.

Table 10. AU recognition for AU combinations when modeling 7 single AUs only.

AU	No.	Correct	false	Missed	Confused	Recognition rate
0	22	22	-	-	-	100%
1	18	18	-	-	-	100%
2	12	12	2	-	-	100%
4	10	10	4	-	-	100%
5	10	7	-	1	2	70%
6	14	12	-	2	-	85.7%
7	8	8	-	-	-	100%
Total	94	89	6	3	2	95%
False alarm: 6.4%						

7.1.3 Upper Face AU Combination Recognition When Modeling Nonadditive Combinations

For this NN, we separately model the nonadditive AU combinations. The 11 outputs consist of 7 individual upper face AUs and 4 non-additive AU combinations (AU1+2, AU1+4, AU4+5, and AU1+2+4). The non-additive AU combinations and the corresponding individual AUs strongly depend on each other. Table 11 shows the correlations between AU1, AU2, AU4, AU5, AU1+2, AU1+2+4, AU1+4, and AU4+5 used in the training set. We set the values based on the appearances of these AUs or combinations.

Table 12 shows the results of our NN on the ($TrainC1$)/($TestC1$) training/testing set. An average recognition rate of 93.7% is achieved, with a slightly lower false alarm rate of 4.5%. In this case, modeling separately the nonadditive combinations does not improve recognition rate due to the fact that

Table 11. The correlation of AU1, AU2, AU4, AU5, AU1+2, AU1+2+4, AU1+4, and AU4+5.

AU1+2	AU1+2 (1.0)	AU1 (1.0)	AU2 (0.5)
AU1+2+4	AU1+2+4 (1.0)	AU1+2 (0.5)	AU1+4 (0.5)
AU1+4	AU1+4 (1.0)	AU1 (0.5)	AU4 (0.5)
AU4+5	AU4+5 (1.0)	AU4 (0.9)	AU5 (0.5)

the AUs in these combinations strongly depend on each other.

Table 12. AU recognition for AU combinations by modeling the non-additive AU combinations as separate AUs.

AU	No.	Correct	false	Missed	Confused	Recognition rate
0	25	25	-	-	-	100%
1	22	20	-	2	-	91%
2	16	14	2	2	-	87.5%
4	14	14	-	-	-	100%
5	10	8	2	2	-	80%
6	14	13	-	1	-	93%
7	10	10	1	-	-	100%
Total	111	104	5	7	-	93.7%
False alarm: 4.5%						

7.2. Lower Face Action Units Recognition

We used a three-layer neural network with one hidden layer to recognize the lower face action units. The inputs of the neural network are the lower face feature parameters shown in Table 3. 7 parameters are used except two parameters of the nasolabial furrows. We don't use the angles of the nasolabial furrows because they are varied much for the different subjects. Generally, we use them to analyze the different expressions of same subject.

Two separate neural networks are trained for lower face AU recognition. The outputs of the first NN ignore the nonadditive combinations and only models 11 basic single action units which are shown in Table 5. We use AU 23+24 instead of AU23 and AU24 because they almost occur together. The outputs of the second one separately models some nonadditive combinations such as AU9+17 and AU10+17 besides the basic single action units.

The recognition results for modeling basic lower face AUs only are shown in Table 13 with recognition rate of 96.3%. The recognition results for modeling non-additive AU combinations are shown in Table 14 with average recognition rate of 96.71%. We found that separately model the nonadditive combinations slightly increase lower action unit recognition accuracy.

All the misidentifications come from AU10, AU17, and AU26. All the mistakes of AU26 are confused by AU25. It is reasonable because both AU25 and AU26 are with parted lips. But for AU26, the mandible is lowered. We did not use the jaw motion information in current system. All the mistakes of AU10 and AU17 are caused by the image sequences with AU combination AU10+17. Two combinations AU10+17 are classified to AU10+12. One combination of AU10+17 is classified as AU10 (missing AU17). The combination AU 10+17 modified the single AU's appearance. The neural network needs to learn the modification by more training data of AU 10+17. There are only ten examples of AU10+17 in 1220 training data in our current system. More data about AU10+17 is collecting for future training. Our system is able to identify action units regardless of whether they occurred singly or in combinations. Our system is trained with the large number of subjects, which included African-Americans and Asians in addition to European-Americans, thus providing a sufficient test of how well the initial training analyses generalized to new image sequences.

For evaluating the necessity of including the nonadditive combinations, we also train a neural network using 11 basic lower face action units as the outputs. For the same test data set, the average recognition rate is 96.3%.

8. Conclusion and Discussion

We developed a feature-based facial expression recognition system to recognize both individual AUs and AU combinations. To localize the subtle changes in the appearance of facial features, we developed a multi-state method of tracking facial features that uses convergent methods of feature analysis. It has

Table 13. Lower face action unit recognition results for modeling basic lower face AUs only.

AU	No.	Correct	false	Missed	Confused	Recognition rate
0	63	63	-	-	-	100%
9	16	16	-	-	-	100%
10	12	9	-	3	-	91.67%
12	14	14	1	-	-	100%
15	12	10	-	-	2	100%
17	36	36	-	-	- (AU12)	94.44%
20	12	12	2	-	-	100%
25	50	50	4	-	-	100%
26	14	10	-	-	4 (AU25)	64.29%
27	8	8	-	-	-	100%
23+24	6	6	-	-	-	100%
Total	243	235	7	3	6	96.3%

Table 14. Lower face action unit recognition results for modeling non-additive AU combinations.

AU	No.	Correct	false	Missed	Confused	Recognition rate
0	63	63	-	-	-	100%
9	16	16	-	-	-	100%
10	12	11	-	1	-	91.67%
12	14	14	2	-	-	100%
15	12	12	-	-	-	100%
17	36	34	-	-	2 (AU12)	94.44%
20	12	12	-	-	-	100%
25	50	50	5	-	-	100%
26	14	9	-	-	5 (AU25)	64.29%
27	8	8	-	-	-	100%
23+24	6	6	-	-	-	100%
Total	243	235	7	1	7	96.71%

high sensitivity and specificity for subtle differences in facial expressions. All the facial features are represented in a group of feature parameters.

The network was able to learn the correlations between facial feature parameter patterns and specific action units. Although often correlated, these effects of muscle contraction potentially provide unique information about facial expression. Action units 9 and 10 in FACS, for instance, are closely related expressions of disgust that are produced by variant regions of the same muscle. The shape of the nasolabial furrow and the state of nose wrinkles distinguishe between them. Changes in the appearance of facial features also can affect the reliability of measurements of pixel motion in the face image. Closing of the lips or blinking of the eyes produces occlusion, which can confound optical flow estimation. Unless information about both motion and feature appearance are considered, accuracy of facial expression analysis and, in particular, sensitivity to subtle differences in expression may be impaired. A recognition rate of 95% was achieved for seven basic upper face AUs. Eleven basic lower face action units are recognized and 96.71% of action units were correctly classified.

Unlike previous methods [9] which build a separate model for each AU and AU combination, we build a single model that recognizes AUs whether they occur singly or in combinations. This is an important capability since the number of possible AU combinations is too large (over 7000) for each combination to be modeled separately.

Using the same database, Bartlett *et al.* [1] recognized only 6 single upper face action units but no combinations. The performance of their feature-based classifier on novel faces was 57%; on new images of a face used for training, the rate was 85.3%. After they combined holistic spatial analysis, feature measures and optical flow, they obtained their best performance at 90.9% correct. Compared to their system, our feature-based classifier obtained a higher performance rate about 92.5% on both novel faces and new images of a face used for training for individual AU recognition. Moreover, our system works well for a more difficult case in which AUs occur either individually or in additive and nonadditive AU combinations. 95% of upper face AUs or AU combinations are correctly classified regardless of whether these action units occur singly or in combination. Those disagreements that did occur were from nonadditive AU combinations such as AU1+2, AU1+4, AU1+2+4, AU4+5, and AU6+7. As a result, more analysis of the nonadditive AU combinations should be done in future.

From the experimental results, we have the following observations:

1. The recognition performance from facial feature measurements is comparable to holistic analysis and Gabor wavelet representation for AU recognition.
2. 5 to 7 hidden units are sufficient to code 7 individual upper face AUs. 10 to 16 hidden units are needed when AUs may occur either singly or in complex combinations.
3. For upper face AU recognition, separately modeling nonadditive AU combinations affords no increase in the recognition accuracy. In contrast, separately modeling nonadditive AU combinations affords slightly increase in the recognition accuracy for lower face AU recognition.
4. After using sufficient data to train the NN, recognition accuracy is stable for recognizing AUs of new faces.

In summary, the face image analysis system demonstrated concurrent validity with manual FACS coding. The multi-state model based convergent-measures approach was proved to capture the subtle changes of facial features. In the test set, which included subjects of mixed ethnicity, average recognition accuracy for 11 basic action units in the lower face was 96.71%, for 7 basic action units in the upper face was 95%, regardless of these action units occur singly or in combinations. This is comparable to the level of inter-observer agreement achieved in manual FACS coding and represents advancement over the existing computer-vision systems that can recognize only a small set of prototypic expressions that vary in many facial regions.

Acknowledgements

The authors would like to thank Paul Ekman, Human Interaction Laboratory, University of California, San Francisco, for providing the database. The authors also thank Zara Ambadar, Bethany Peters, and Michelle Lemenager for processing the images. 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] J. M. Carroll and J. Russell. Facial expression in hollywood's portrayal of emotion. *Journal of Personality and Social Psychology.*, 72:164–176, 1997.

[3] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski. Classifying facial actions. *International Journal of Pattern Analysis and Machine Intelligence*, 21(10):974–989, October 1999.

[4] P. Ekman and W. V. Friesen. *The Facial Action Coding System: A Technique For The Measurement of Facial Movement*. Consulting Psychologists Press Inc., San Francisco, CA, 1978.

[5] 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.

[6] 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.

[7] M. Kirby and L. Sirovich. Application of the k-l procedure for the characterization of human faces. *IEEE Transc. On Pattern Analysis and Machine Intelligence*, 12(1):103–108, Jan. 1990.

[8] Y. Kwon and N. Lobo. Age classification from facial images. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 762–767, 1994.

[9] 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.

[10] 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.

[11] K. Mase. Recognition of facial expression from optical flow. *IEICE Transc.*, E. 74(10):3474–3483, 0ctober 1991.

[12] K. Scherer and P. Ekman. *Handbook of methods in nonverbal behavior research*. Cambridge University Press, Cambridge, UK, 1982.

[13] 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.

[14] Y. Tian, T. Kanade, and J. Cohn. Dual-state parametric eye tracking. In *Submit to International Conference on Face and Gesture Recognition*, 1999.

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

[16] M. Turk and A. Pentland. face recognition using eigenfaces. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 586–591, 1991.

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

[18] 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.

[19] Z. Zhang. Feature-based facial expression recognition: Sensitivity analysis and experiments with a multilayer perceptron. *International Journal of Pattern Recognition and Artificial Intelligence*, 13(6):893–911, 1999.