

8.0 EXPERIMENTAL RESULTS

The computer-vision based facial expression recognition system described in the previous chapters has been trained and experimented with a large set of image sequences containing nine frequently occurred facial expressions of many subjects with various expression intensities. The experimental results are very exciting and have shown a great promise of our automatic recognition system.

8.1 Data Acquisition, Experimental Setup, and Digitizing

The database consists of 90 adult volunteers and 4 infants. The subjects included both male (35%) and female (65%). They ranged in both age (from 1 to 35 years of age) and ethnicity (81% Caucasian, 14% African-American, 4% Asian or Indian, and 1% Hispanic). The data acquisition was done in 8 sessions over a 2-month period¹. More than 400 image sequences and 8000 images were made available to this research.

Adult subjects were seated 2 meters directly in front of a standard VHS video camera, with a video rate of 30 frames per second, which was manually adjusted to capture a full-face frontal view. None of the subjects wore eyeglasses. Some of subjects had hair covering their foreheads, and several subjects wore caps, or had makeup on their brows, eyelids or lips. Overhead fluorescent and incandescent lights as well as two halogen lights attached to portable umbrellas were positioned to the front at 30 degrees left and right, and were adjusted to provide maximum illumination with a minimum of facial shadows (Figure 61). Although reflection and lighting may have varied across individuals because of different facial skin colors and different times, constant illumination was used for each subject. These constraints - constant illumination using

¹ This was done by Miss Adena J. Zlochower and Dr. Jeffrey F. Cohn, Department of Psychology, University of Pittsburgh.

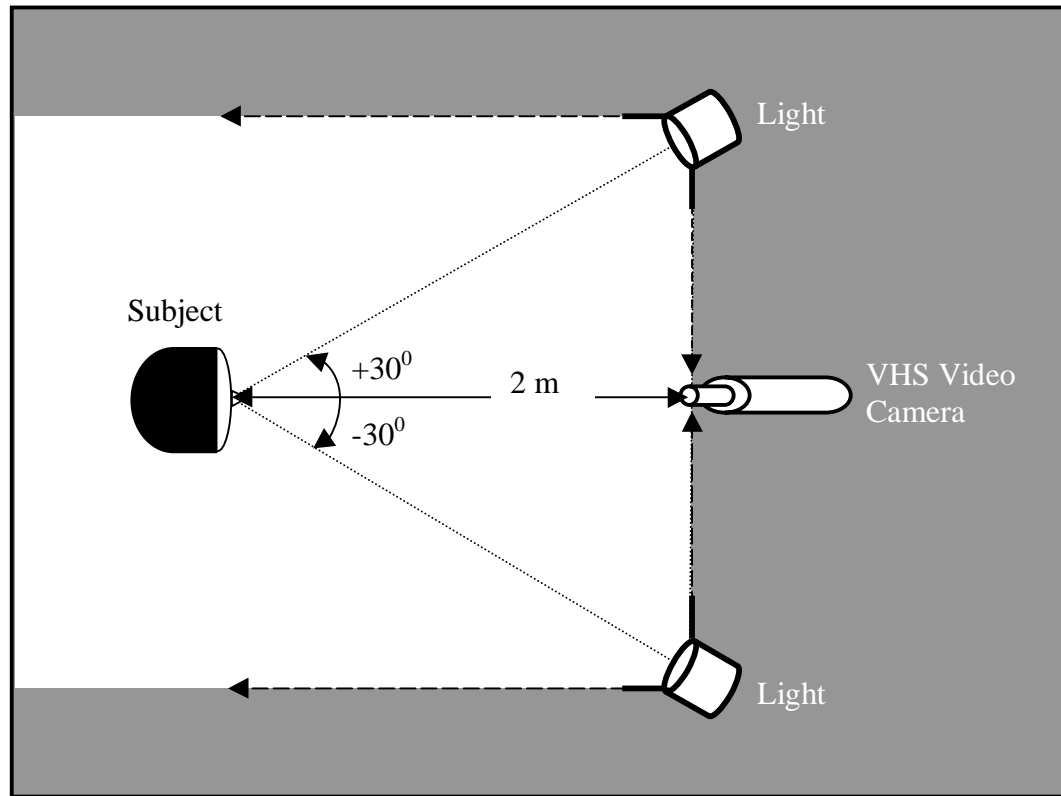


Figure 61 Experimental setup.

fixed light sources, and no eyeglasses - were imposed to minimize optical flow degradation.

None of the subjects were previously trained in displaying specific facial expressions. Prior to video recording, subjects practiced the expressions with FACS experts. During recording, subjects were free to look at the experts and copy their expressions. Subjects were asked to perform six basis expressions (joy, fear, anger, disgust, sadness, and surprise), and a series of “expression units” corresponding to individual AUs (*e.g.*, AU12) and AU combinations (*e.g.*, AU12+25) (see Table 3 for a complete list, and Figure 4 for “expression units”). Each expression was repeated 3 times, and the best expression was chosen. Each posed expression began from neutral, reached peak, and ended at neutral expressions again. There is at least a half second duration (15 frames) of

neutral expression between posed expressions. Even though these untraining subjects have seen the expressions demonstrated by experts, subjects still showed a range of posing ability. Not all of the expressions conformed to the “expression units,” such as the combination of “expression units”: AU1+2+4, AU12+20+25 and AU12+15. The spontaneous expressions showed more variability. In addition, facial expressions (non-rigid motion) with some out-of-plane head motion (rigid motion) such as yawing or pitch less than $\pm 10^0$ occurred concurrently, even though all subjects were viewed frontally.

Each frame of video sequence was automatically digitized into 490 x 640-pixel image on a Sun Sparc 20 workstation using the K2T digitizer. For feature point tracking and high gradient component analysis, the size of each frame was kept the same as the original 490 x 640-pixel image. To save the computing time when using dense flow tracking, the image size of each frame was automatically cropped to 417 x 385 pixels, which exactly covered the entire face and cut out the unnecessary background.

8.2 Segmentation and Coding by Human Observers (Ground Truth)

Before digitizing, the image sequences were segmented and coded by two certified FACS coders. Training a FACS coder is time consuming and takes approximately 100 hours to achieve acceptable levels of reliability, and coding criteria are subject to drift over the course of prolonged studies. It can take up to 10 hours of coding time per minute (30 frames/second) of taped facial behavior depending on the comprehensiveness of the system and the density of behavior changes.

Certified FACS coders segmented video tape from the beginning of the beginning duration, to the apex duration, and finally to the end of the ending duration to capture an expression sequence ⁽¹⁰³⁾. For each expression sequence, the beginning duration is defined as the last 2 ~ 4 frames of the neutral expression, which are prior to the facial movement, to the beginning of the apex duration. The ending duration is defined as from the end of the apex duration to the first 2 ~ 4 frames of the neutral expression, which are after facial movement. The apex duration is defined as the maximum movement of facial

motion, which is between the end of the beginning duration and the beginning of the ending duration. According to our experiments, generally there are at least 3 frame without any obvious movement at the apex duration (Figure 62). The velocity of facial motion for the beginning duration or the ending duration increases then decreases, like the oscillation of a spring between compression and release (Figure 62). Different expressions (from the beginning of the beginning duration to the end of the ending duration) correspond to different durations from 1/2 second (15 frames) to 3 1/3 seconds (100 frames).

All expression sequences were coded by two FACS experts at different dates. The overall agreement between the two FACS experts is 97%. Disagreement occurred due to fatigue during observation, which produced misclassification of subtly asymmetric expressions, eye blinking, or out-of-plane head motion. The agreement at eyebrow (upper face) expressions for AU1+4 was only 78%, because it is easy to confuse AU1+4 with either AU1 or AU4. The confusing between AU1+4 and AU1 occurs when very weak inner brows close together during inner brow raised, or the Ω shape of the furrow appears at the forehead during inner brow raised but without closing inner brows together (Figure 63.a). The confusion between AU1+4 and AU4 occurs when inner brows close together with eye blinking, head rotation pitching in the vertical direction, the Ω shape of furrow at the forehead, or asymmetric brow motion (Figure 63.b). In addition, the confusion among AU12+25, AU20+25 and AU12+20+25 also occurs sometimes, because these “expression units” have common muscle movement in the lip region (Figure 63.c). The final FACS AU coding for each expression sequence, which includes the subtle motion, instant motion (expression appears temporarily for only few frames but can not be seen at the peak expressions) and asymmetric motion, is set by the agreement of both FACS experts to be the ground true of our training and recognition processes.

In our experimental study, we used image sequences which start from the beginning of the beginning duration and end during the apex duration. These digitized image sequences are in arbitrary length varying from 9 to 47 frames. The average number of images per expression sequence is about 20.

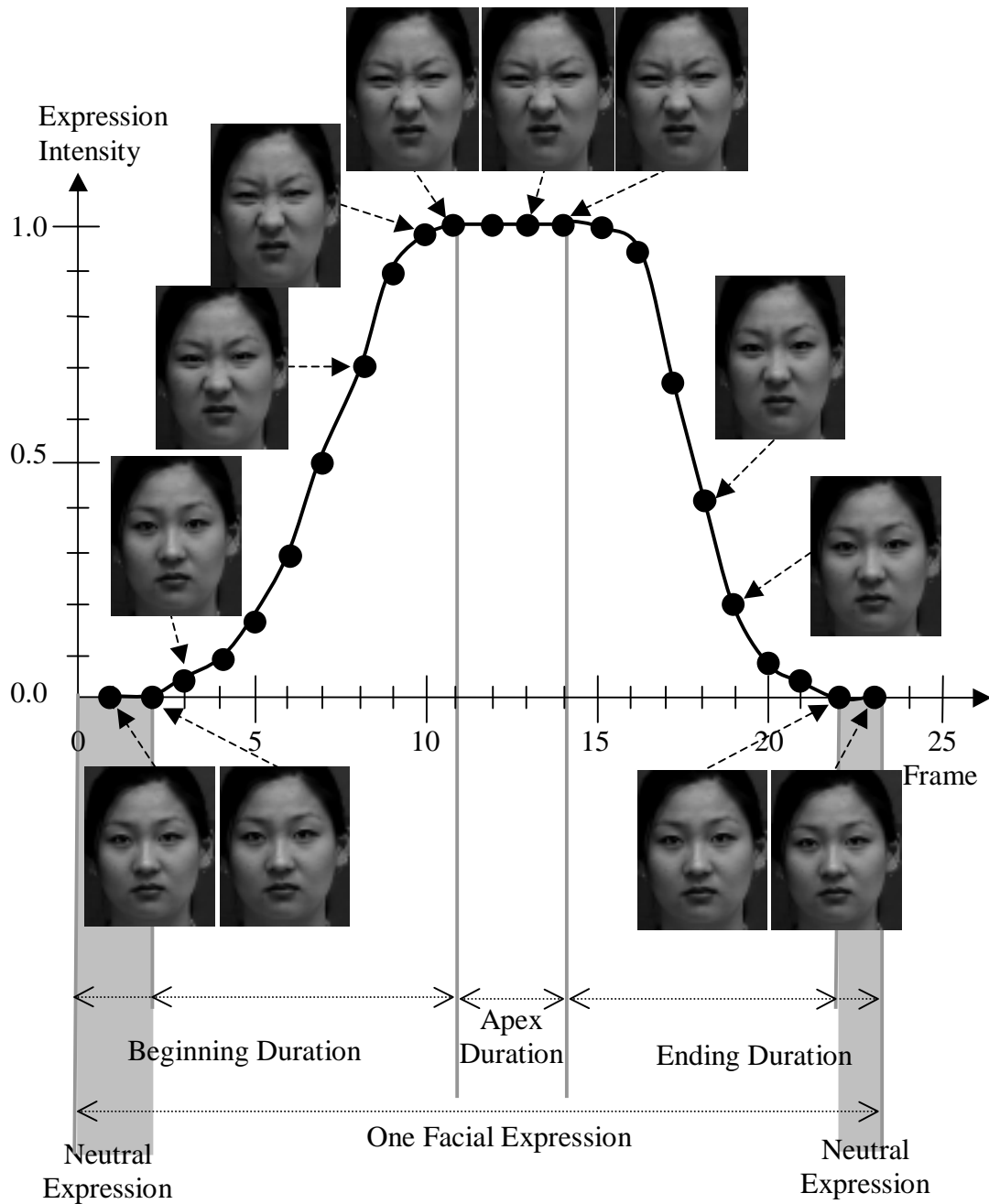
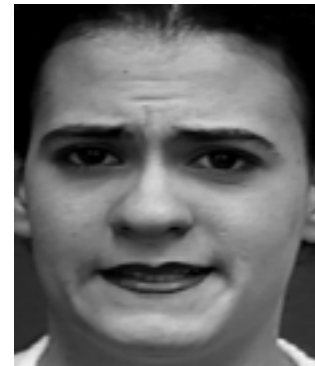


Figure 62 Each facial expression begins from the beginning duration, continues through the apex duration, and ends at the ending duration. In our current work, we segmented each facial expression to include only the beginning and apex durations.



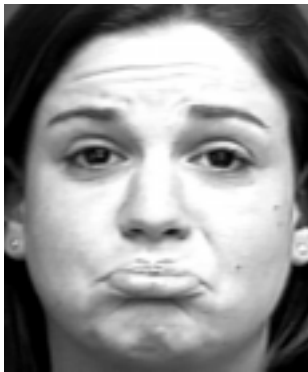
Standard AU1+4

AU1+4+20+25

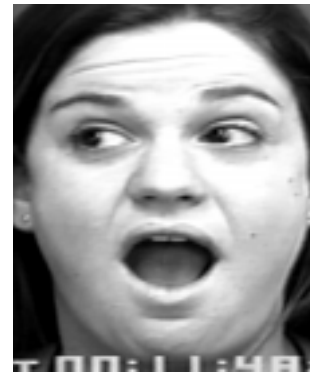


AU1+4+20+25

Medial portion of the eyebrows is raised (AU1) and pulled together (AU4).

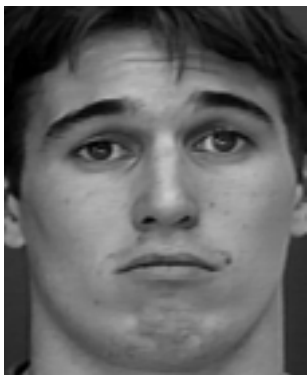


AU1+15+17



AU1+2+12+27

Both left (AU1) and right images (AU1+2) have the same permanent Ω shapes of furrows at her forehead (usually in AU1+4).



AU1+17



AU1+15+17

Ω shape of furrows (usually occurs in AU1+4) appears when the inner brows are raised (AU1).

Figure 63.a Standard AU1+4 expressions and manual misclassification of three AU1 expressions and one AU1+2 expression to AU1+4 expressions.

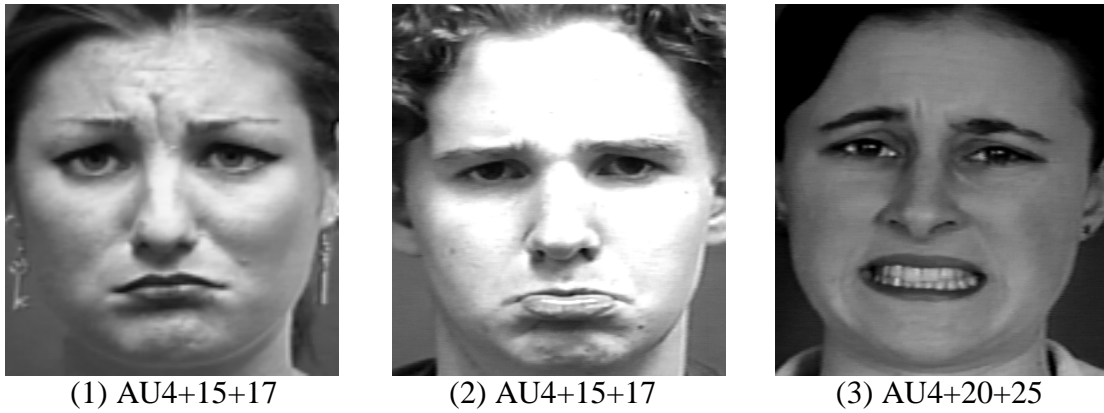


Figure 63.b Manual misclassification of three AU4 expressions to AU1+4 expressions. These mistakes are because of (1) Ω shape of furrows at the forehead, (2) confusing expression, and (3) asymmetric brow motion. The standard AU4 expression is shown in Figure 63.c (3).

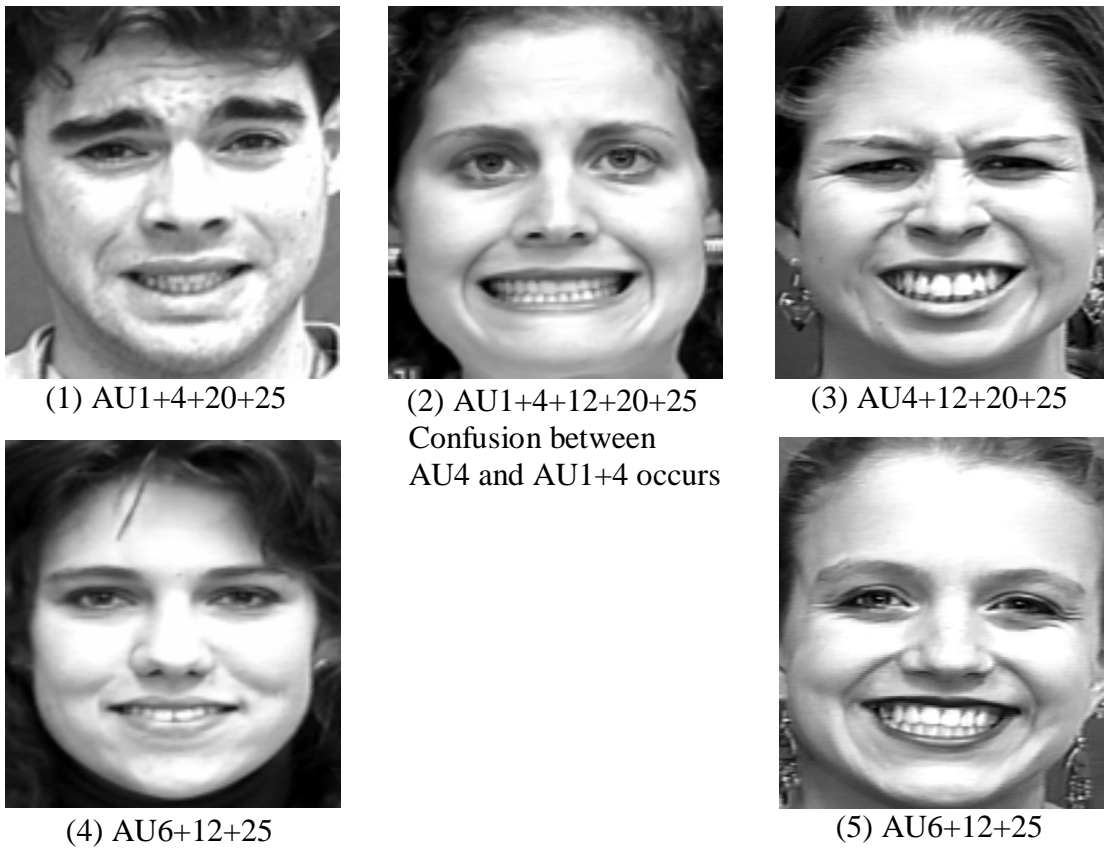


Figure 63.c Confusions among AU12+25, AU20+25 (also in Figure 63.a and 63.b) and AU12+20+25.

8.3 Automatic Expression Recognition

Our goal is to discriminate subtle differences in facial expressions for the upper face region: AU4, AU1+4, and AU1+2, and for the lower face region: AU12, AU6+12+25, AU20+25, AU9+17, AU17+23+24, and AU15+17. The experimental image sequences were processed for extraction of expression information and coding. About one half of them were used in training, and the other half in testing. From these two sets, subsets were processed by three methods (facial feature point tracking, dense flow tracking with PCA, and high gradient component analysis). The extracted expression information is normalized using affine transformation, converted to displacement vector sequences, weight vector sequences, and mean-variance vector sequences, and then vector quantized into symbol sequences for use in training and recognition processes (Figure 64).

8.3.1 Training Process

In reality, the same facial expression may appear different among individuals because of different motion intensities. To design a robust recognition system, the training data were selected to cover all possible facial actions and expression intensities for each facial expression (Figure 65). Motions in upper facial expressions and in lower facial expressions were separately extracted. For upper facial expressions, the training data consist of 100 image sequences for high gradient component analysis, 60 image sequences for facial feature point tracking, and of which a subset of 45 sequences for dense flow tracking. For lower facial expressions, the training data consist of 120, 120, and 60 image sequences, respectively (Table 9). We used a smaller subset of data for dense flow tracking because of its requirement of excessive processing time.

Before using HMMs for training or recognition process, any motion vector sequence is preprocessed by vector quantization to an observable symbol sequence O . The codebooks are created based on their corresponding training data. The codebook size M , which is power of 2, is chosen to be less than or equal to the total number of training

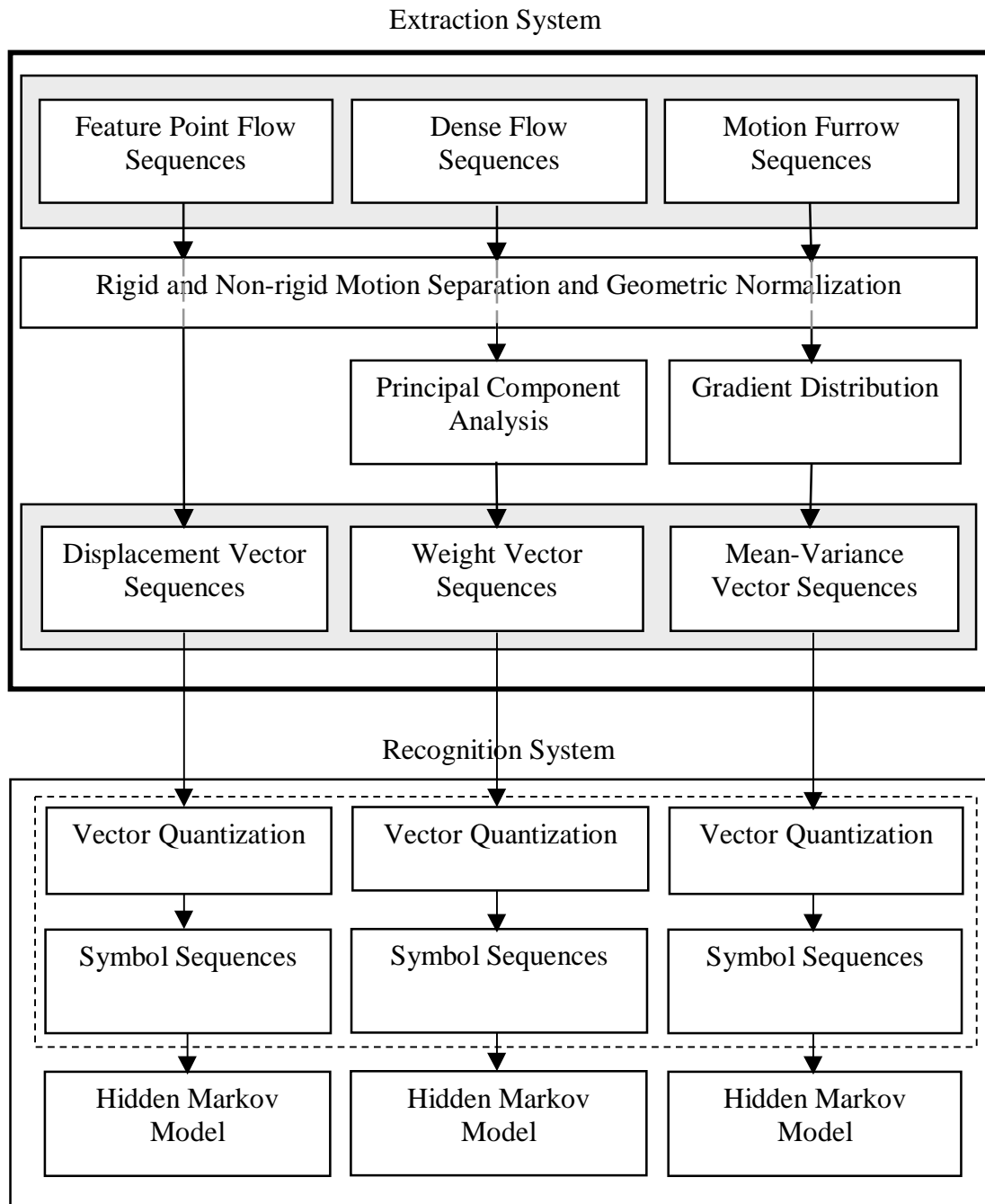


Figure 64 Three sets of extracted information as inputs to the recognition system using Hidden Markov Models.

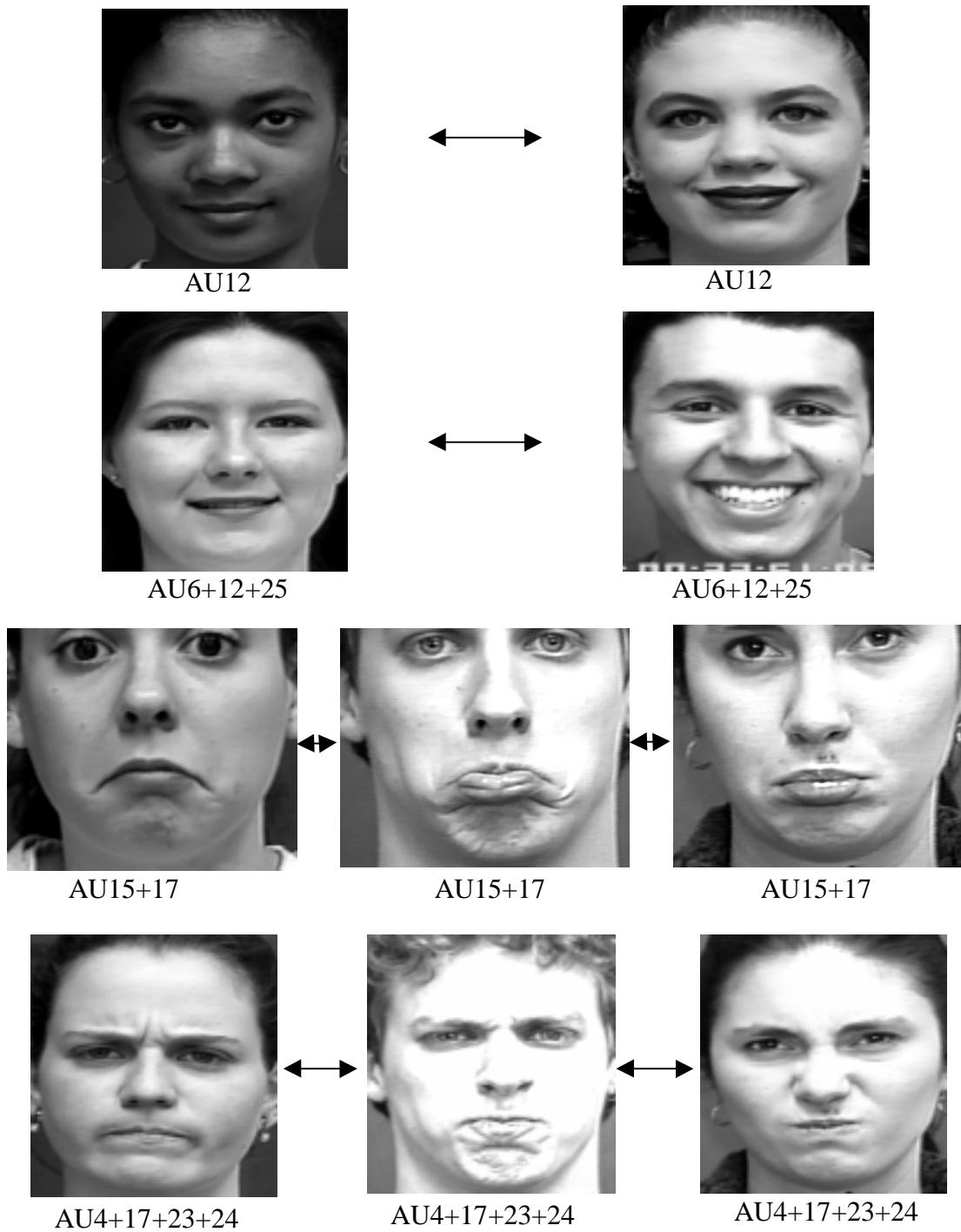
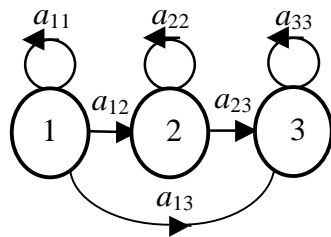


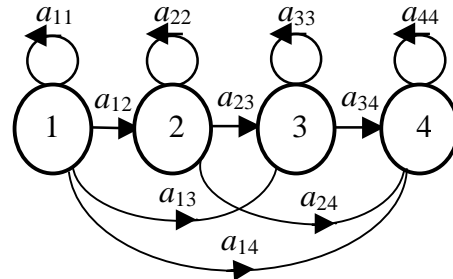
Figure 65 The images at the same row have the same facial expressions, but different facial actions or expression intensities.

Table 9 Different Hidden Markov Models for 3 upper facial expressions and 6 lower facial expressions.

| Methods for the Extraction of Expression Information | Feature Point Tracking | Dense Flow Tracking with Principal Component Analysis | High Gradient Component Analysis in the Spatio-Temporal Domain |
|--|---|---|--|
| Codebook Size (M) for the Upper Facial Expressions | $M = 16$ (60 training symbol sequences) | $M = 16$ (45 training symbol sequences) | $M = 32$ (100 training symbol sequences) |
| Hidden Markov Model (HMM) | 2nd-order 3-state left-right HMM | 2nd-order 3-state left-right HMM | 2nd-order 3-state left-right HMM |
| Codebook Size (M) for the Lower Facial Expressions | $M = 32$ (120 training symbol sequences) | $M = 16$ (60 training symbol sequences) | $M = 32$ (120 training symbol sequences) |
| Hidden Markov Model (HMM) | 3rd-order 4-state left-right HMM | 3rd-order 4-state left-right HMM | 3rd-order 4-state left-right HMM |



The 2nd-order 3-state left-right Hidden Markov Model for AU4, AU1+4, and AU1+2.



The 3rd-order 4-state left-right Hidden Markov Model for AU12, AU6+12+25, AU20+25, AU9+17, AU17+23+24, and AU15+17.

frames divided by 50. So, the codebook size for the training data having 45 and 60 image sequences (around $45 \times 20 = 900$ and $60 \times 20 = 1200$ frames) is $M = 2^4$ ($2^4 \leq 900/50 = 18$, and $1200/50 = 24 < 2^5$) (Table 9). The codebook size for the training data having 100 and 120 image sequences (around $100 \times 20 = 2000$, and $120 \times 20 = 2400$ frames) is $M = 2^5$ ($2^5 \leq 2000/50 = 40$, and $2400/50 = 48 < 2^6$) (Table 9). The 12- and 20-dimensional displacement vectors from feature point tracking (for upper and lower facial expressions, respectively) (Figure 19), the 20- and 30-dimensional weight vectors from the dense flow tracking with principal component analysis (PCA) (Figures 29 and 30), and the 32- and 32-dimensional mean-variance vectors from the high gradient component analysis (Figure 41) are each vector quantized to one codeword (or observable symbol) o_t , $0 \leq o_t \leq M-1$, according to its respective codebook, where the subscript t denotes the frame t .

Based on these training symbol sequences, we determine the HMM topology using the method that we have developed. Thus, a 2nd-order 3-state left-right HMM and a 3rd-order 4-state left-right HMM are used for modeling the three upper facial expressions and six lower facial expressions, respectively (Table 9).

There are three sets of information extracted from the upper and lower facial expressions data by three methods. For each set of data, two sets of HMM parameters $\lambda = (\pi, A, B)$ are trained. Parameter sets λ_{AU4} , λ_{AU1+4} , and λ_{AU1+2} characterize the most likely occurrences of the three upper facial “expression units” (individual AUs or AU combinations), and sets λ_{AU12} , $\lambda_{AU6+12+25}$, $\lambda_{AU20+25}$, λ_{AU9+17} , $\lambda_{AU17+23+24}$, and $\lambda_{AU15+17}$ characterize the six lower facial “expression units.” These trained HMM parameter sets serve to evaluate any observable symbol sequence of facial expression to give the most likely classification (Figure 66).

To initialize the training of each model parameter set, the initial state probability at the first state π_1 is set to 1, and the rest states are set to 0. Each element of the state-transition probability matrix $A_{N \times N}$ and the output observable symbol probability matrix $B_{M \times N}$ is initialized to a very small value (say, 10^{-6}) of a uniformly distributed random variable. The Baum-Welch method is then applied to estimate the parameters $\lambda = (\pi, A, B)$ in iterations based on the Forward procedure (variable α) and the Backward procedure

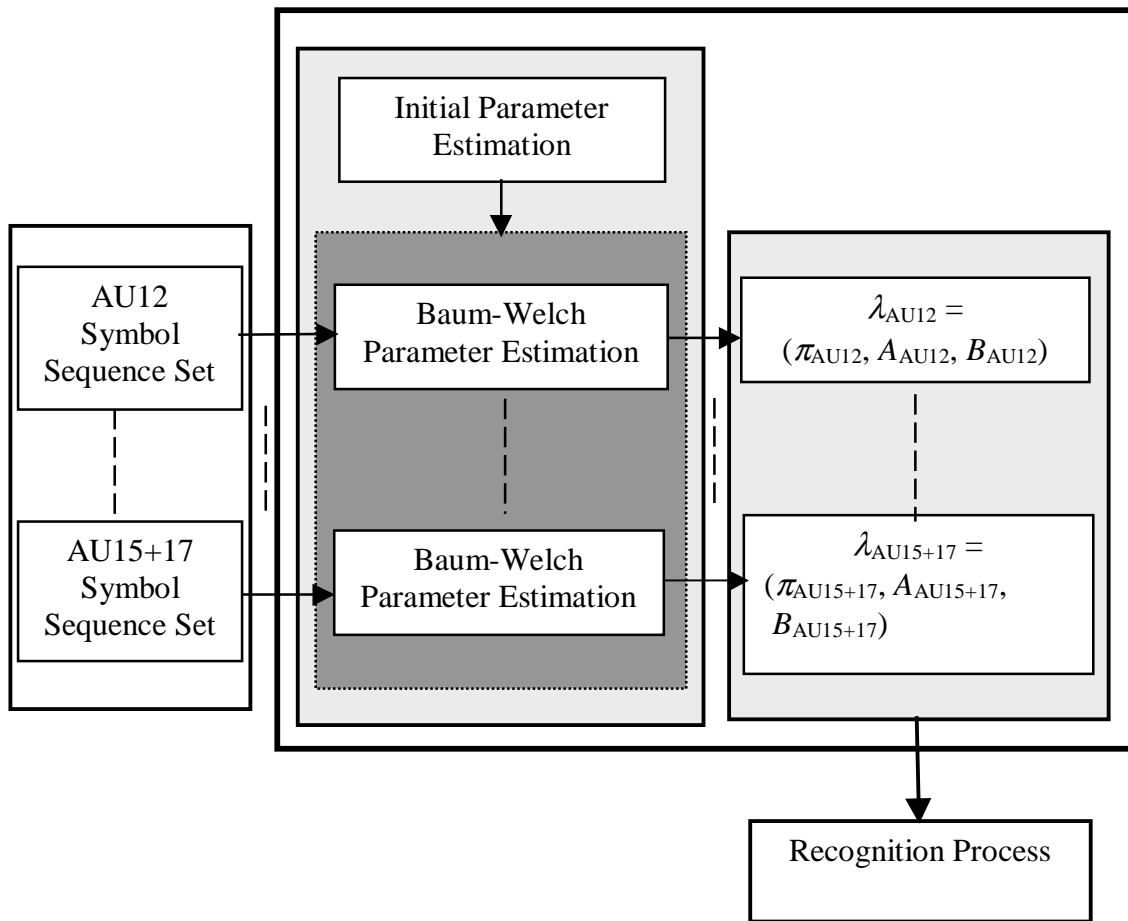


Figure 66 The training process for the Hidden Markov Model (an example for the lower facial expressions: AU12, AU6+12+25, AU20+25, AU9+17, AU17+23+24 and AU15+17).

(variable β). After each iteration, the estimated probability for each element of these parameters is smoothed by setting a numeric floor 0.0001 to avoid zeroing the parameter element and producing an unreliable result. They are then renormalized to meet the required statistical constraint, and go on for further iteration. The trained parameter values for $\lambda_{AU6+12+25}$, for example, are shown in Table 10.

Table 10 The trained parameter set $\lambda = (\pi, A, B)$ of the 3rd-order 4-state Hidden Markov Model, whose topology is determined in Figures 58 and 60, for the lower facial expression AU6+12+25 using dense flow tracking method (codebook size $M=16$).

| π | | | | |
|-------|--------------------|--------------------|--------------------|--------------------|
| | State 1 | State 2 | State 3 | State 4 |
| | 1.0000000000000000 | 0.0000000000000000 | 0.0000000000000000 | 0.0000000000000000 |

| A | | | | |
|-------|--------------------|--------------------|--------------------|--------------------|
| State | State 1 | State 2 | State 3 | State 4 |
| 1 | 0.852713734696394 | 0.147286265303606 | 0.0000000000000000 | 0.0000000000000000 |
| 2 | 0.0000000000000000 | 0.530337222246666 | 0.106073169816378 | 0.363589607936957 |
| 3 | 0.0000000000000000 | 0.0000000000000000 | 0.861441737692501 | 0.138558262307499 |
| 4 | 0.0000000000000000 | 0.0000000000000000 | 0.0000000000000000 | 1.0000000000000000 |

| B | | | | |
|--------|--------------------|--------------------|--------------------|--------------------|
| Symbol | State 1 | State 2 | State 3 | State 4 |
| 0 | 0.0000000000000000 | 0.0000000000000000 | 0.0000000000000000 | 0.0000000000000000 |
| 1 | 0.0000000000000000 | 0.000006813844404 | 0.028598228023638 | 0.0000000000000000 |
| 2 | 0.999999999231619 | 0.133964815503377 | 0.0000000000000000 | 0.0000000000000000 |
| 3 | 0.0000000000000000 | 0.0000000000000000 | 0.113344687439651 | 0.534742185973886 |
| 4 | 0.0000000000000000 | 0.0000000000000000 | 0.0000000000000000 | 0.0000000000000000 |
| 5 | 0.0000000000000000 | 0.017482535428337 | 0.858057084052733 | 0.000000182200935 |
| 6 | 0.0000000000000000 | 0.0000000000000000 | 0.0000000000000000 | 0.0000000000000000 |
| 7 | 0.0000000000000000 | 0.0000000000000000 | 0.0000000000000000 | 0.0000000000000000 |
| 8 | 0.0000000000000000 | 0.0000000000000000 | 0.0000000000000000 | 0.0000000000000000 |
| 9 | 0.000000000587439 | 0.424273591136245 | 0.000000000483978 | 0.0000000000000000 |
| 10 | 0.000000000180942 | 0.424272244087637 | 0.0000000000000000 | 0.0000000000000000 |
| 11 | 0.0000000000000000 | 0.0000000000000000 | 0.0000000000000000 | 0.465257631825179 |
| 12 | 0.0000000000000000 | 0.0000000000000000 | 0.0000000000000000 | 0.0000000000000000 |
| 13 | 0.0000000000000000 | 0.0000000000000000 | 0.0000000000000000 | 0.0000000000000000 |
| 14 | 0.0000000000000000 | 0.0000000000000000 | 0.0000000000000000 | 0.0000000000000000 |
| 15 | 0.0000000000000000 | 0.0000000000000000 | 0.0000000000000000 | 0.0000000000000000 |

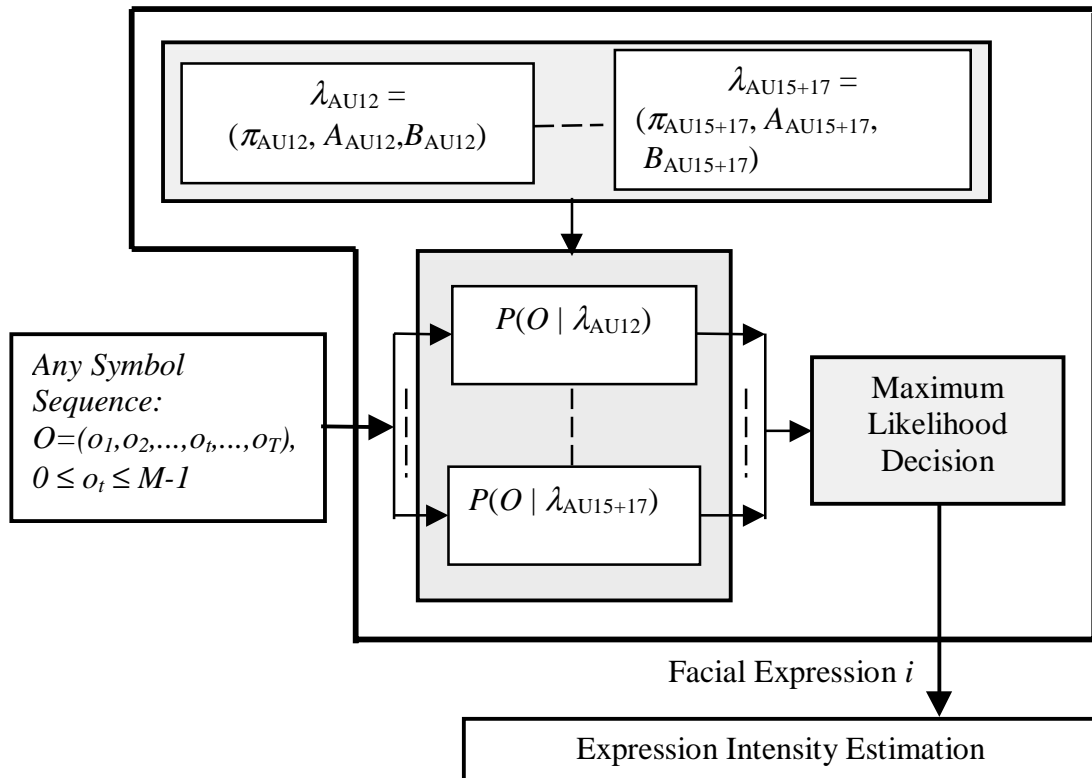


Figure 67 The recognition process for the Hidden Markov Model (an example for the lower facial expressions: AU12, AU6+12+25, AU20+25, AU9+17, AU17+23+24 and AU15+17).

8.3.2 Recognition Results

From the testing data, any observable symbol sequence O is evaluated or recognized by selecting the maximum output probability $P(O | \lambda_i)$ from the HMM parameter set λ_i , where λ_i is one of the HMM parameter set such as λ_{AU12} , $\lambda_{AU6+12+25}$, $\lambda_{AU20+25}$, λ_{AU9+17} , $\lambda_{AU17+23+24}$ and $\lambda_{AU15+17}$ for the lower facial expressions (Figure 67). If the output probability $P(O | \lambda_i)$ (usually it is close to 1) is greater than other output probability $P(O | \lambda_k)$, then the symbol sequence O is recognized as the facial expression represented by the model λ_i .

The recognition result using the HMM classifiers was evaluated by comparison with the coding of human observers coding taken as the ground truth. Two FACS experts

agreed on 97% of the collected facial expressions. The agreement for the AU1+4 was only 78%, since AU1+4 was easily confused with AU1 or AU4. The other point of disagreement was AU12+25, AU20+25 and AU12+20+25.

Because the computation time of different extraction methods is very different (the dense flow tracking is very time consuming when compared with the other two methods), and these methods were developed at different time (about 6 months apart), the number of image sequences used in experiments are not the same as indicated in Table 11. Those used in the dense flow study is a subset of the image sequences used in the feature point tracking study, which is a subset of those used in the high gradient components study. This situation is true for both training and testing. The test results of each study are given in Table 12, 13 and 14, respectively. The average recognition rate of the three upper facial expressions is 85% by feature point tracking, 92% by dense flow tracking with PCA, and 85% by high gradient component analysis. These results are based on 60, 45, and 100 training image sequences and 75, 60, and 160 testing image sequences. The average recognition rate of the six lower face expressions is 88% by feature point tracking, 92% by dense flow tracking with PCA, and 81% by high gradient component detection, based on 120, 60, and 120 training image sequences and 150, 90, and 150 testing image sequences (Table 11).

Comparing the recognition results of three different extraction methods, it is obvious that the dense flow tracking with PCA has the best performance for all facial expressions tested except AU17+23+24, for which the feature point tracking method is better, 92% against 87% (Table 12 and 13). The high gradient component analysis has the worst performance (Table 14). This is because dense flow tracking includes the entire motion information of a facial expression, such as allowing tracking a textureless region, which provides more complete information for the recognition process; but it requires substantially more computation time. It is subject to error due to occlusion (hair covering the forehead) or large discontinuities appearance of tongue or teeth when the mouth opens). The latter occurred in the case of 2-level dense flow estimation used for saving processing time.

Table 11 The number of the training and testing image sequences (the average number of frames per image sequence is 20) and their corresponding recognition rates.

| | Training Image Sequences | | |
|--|---------------------------------|------------------------------|----------------------------------|
| Three Methods | Feature Point Tracking | Dense Flow Tracking with PCA | High Gradient Component Analysis |
| No. of Sequences for the Upper Facial Expressions | 60 | 45 | 100 |
| No. of Sequences for the Lower Facial Expressions | 120 | 60 | 120 |
| | Testing Image Sequences | | |
| Three Methods | Feature Point Tracking | Dense Flow Tracking with PCA | High Gradient Component Analysis |
| No. of Sequences for the Upper Facial Expressions | 75 | 60 | 160 |
| Recognition Results | 85% | 92% | 85% |
| No. of Sequences for the Lower Facial Expressions | 150 | 90 | 150 |
| Recognition Results | 88% | 92% | 81% |

Table 12 Recognition results of the feature point tracking method. (The number given in each block is the number of testing image sequences.)

The average recognition rate for three upper facial expressions is 85% based on 75 testing image sequences.

| HMM \ Human | AU4 | AU1+4 | AU1+2 | Recognition Rate |
|-------------|-----|-------|-------|------------------|
| AU4 | 22 | 3 | 0 | 88% |
| AU1+4 | 4 | 19 | 2 | 76% |
| AU1+2 | 0 | 2 | 23 | 92% |

The average recognition rate for six lower facial expressions is 88% based on 150 testing image sequences.

| HMM \ Human | AU12 | AU6+12+25 | AU20+25 | AU9+17 | AU17+23+24 | AU15+17 | Recognition Rate |
|-------------|------|-----------|---------|--------|------------|---------|------------------|
| AU12 | 25 | 0 | 0 | 0 | 0 | 0 | 100% |
| AU6+12+25 | 0 | 21 | 4 | 0 | 0 | 0 | 84% |
| AU20+25 | 0 | 5 | 20 | 0 | 0 | 0 | 80% |
| AU9+17 | 0 | 0 | 0 | 22 | 3 | 0 | 88% |
| AU17+23+24 | 0 | 0 | 0 | 0 | 23 | 2 | 92% |
| AU15+17 | 0 | 0 | 0 | 1 | 3 | 21 | 84% |

Table 13 Recognition results of the dense flow tracking method. (The number given in each block is the number of testing image sequences.)

The average recognition rate for three upper facial expressions is 92% based on 60 testing image sequences.









| HMM / Human | AU4 | AU1+4 | AU1+2 | Recognition Rate |
|-------------|-----|-------|-------|------------------|
| AU4 | 21 | 2 | 0 | 93% |
| AU1+4 | 2 | 12 | 1 | 80% |
| AU1+2 | 0 | 0 | 22 | 100% |

The average recognition rate for six lower facial expressions is 92% based on 90 testing image sequences.

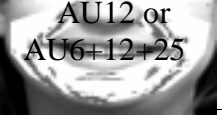

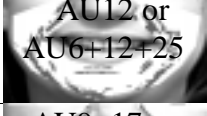
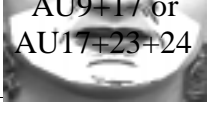
| HMM / Human | AU12 | AU6+12+25 | AU20+25 | AU9+17 | AU17+23+24 | AU15+17 | Recognition Rate |
|-------------|------|-----------|---------|--------|------------|---------|------------------|
| AU12 | 15 | 0 | 0 | 0 | 0 | 0 | 100% |
| AU6+12+25 | 0 | 13 | 2 | 0 | 0 | 0 | 87% |
| AU20+25 | 0 | 2 | 13 | 0 | 0 | 0 | 87% |
| AU9+17 | 0 | 0 | 0 | 15 | 0 | 0 | 100% |
| AU17+23+24 | 0 | 0 | 0 | 0 | 13 | 2 | 87% |
| AU15+17 | 0 | 0 | 0 | 0 | 1 | 14 | 93% |

Table 14 Recognition results of the motion furrow detection method. (The number given in each block is the number of testing image sequences.)

The average recognition rate for three upper facial expressions is 85% based on 160 testing image sequences.

| HMM Human |  |  |  |  | Recognition Rate |
|---|---|---|--|---|-----------------------------------|
|  | 26 | 4 | 0 | 0 | 87% |
|  | 5 | 43 | 2 | 0 | 86% |
|  | 0 | 1 | 24 | 5 | 80% |
|  | 0 | 0 | 7 | 43 | 86% |

The average recognition rate for six lower facial expressions is 81% based on 150 testing image sequences.

| HMM Human |  |  | Recognition Rate |
|---|---|--|-----------------------------------|
|  | 86 | 14 | 86% |
|  | 12 | 38 | 76% |

High gradient component detection is sensitive to changes in transient facial features (*e.g.*, furrows), but is subject to error due to individual differences in subjects. Younger subjects, especially infants, show less furrowing than older ones, which reduces the information value of the high gradient components. Older subjects, in general, have permanent shapes of furrows on their faces. No matter how different their expressions or expression intensities are, the similar shape of furrows still can be seen such as in Figure 63.a (images at the second row). Occasionally, different FACS AUs may have the similar shape of furrows such as between AU12 and AU6+12+25, between AU6+12+25 and AU20+25, or between AU9+17 and AU17+23+24, since there are common facial muscle actions for both facial motions (Figure 68). Furthermore, the crow-feet wrinkles may or may not appear during facial expression AU6+12+25 (Figure 1). With these reasons, explaining the FACS AUs based only on the shape of furrows is not adequate. Furthermore, we used a constant threshold for motion line or edge detection. Since the gray values on each facial image are sensitive to facial motion and lighting, which depend on individual subjects, a dynamic thresholding would be needed.

The optical flow using the pyramid approach is a simple, fast, and accurate method of tracking facial feature points. It tracks large displacement well and is also sensitive to subtle feature motion.

In general, the pattern of errors in all three methods are similar, *i.e.*, errors were resulted from classifying an expression to an expression type which is most similar to the target (*e.g.*, AU4 was confused with AU1+4 but not AU1+2). It appears that the automatic feature point tracking method has given very good performance and its processing was very efficient. Potentially it can be developed into a real time recognition system. Summary of all three different extraction and recognition methods for facial expressions is in Table 15.



Figure 68 Different FACS AUs have the similar shape of furrows such as between AU12 and AU6+12+25, between AU6+12+25 and AU20+25, and between AU9+17 and AU17+23+24, since there are common facial muscle actions for both facial motions.

Table 15 Summary of three different extraction and recognition methods for facial expressions.

| Extraction System | | | |
|---|--|--|---|
| Three Methods | Feature Point Tracking (5-Level Pyramid) | Dense Flow Tracking with PCA (2-Level Pyramid) | High Gradient Component Analysis in the Spatio-Temporal Domain |
| Computing Time | Fast (1%) (70 (13x13-pixel) windows: 20 seconds/frame) (SUN Sparc 5) | Very Slow (98%) (417 x 385 pixels: 20 minutes/frame) (SGI-Irix: 6 times faster than Sparc 5) | Fast (1%) (417 x 385 pixels: 5 second/frame) (SUN Sparc 5) |
| Hair at Forehead | Occlusion | Occlusion | No occlusion |
| Lighting | Sensitive | Sensitive | Sensitive |
| Subtle Motion (< 2 pixel) | Sensitive (Subpixel accuracy) | Insensitive | Sensitive |
| Large Motion (> 15 pixels) | 100 pixels (Subpixel accuracy) | Missed tracking | Sensitive |
| Advantage | Simple and accurate May run in real time | Includes the entire face motion region | May run in real time |
| Disadvantage | Limited to pre-selected features | Time consuming | Detection error from individual differences in subjects (younger > older) |
| Recognition System (Recognition Rate) | | | |
| Different Inputs to HMMs | Displacement Vector Sequence | Weight Vector Sequence | Mean-Variance Vector Sequence |
| Upper Facial Expressions | 85% | 92% | 85% |
| Lower Facial Expressions | 88% | 92% | 81% |