In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June, 1998

A Multi-Method Approach for Discriminating Between Similar Facial Expressions, Including Expression Intensity Estimation

James Jenn-Jier Lien
Robotics Institute
Carnegie Mellon University

Takeo Kanade
Departments of Computer Science
and Electrical Engineering,
Robotics Institute
Carnegie Mellon University

Jeffrey F. Cohn
Department of Psychology
University of Pittsburgh

C. C. Li
Department of Electrical Engineering
University of Pittsburgh

Abstract

Facial expression provides sensitive cues about emotion and plays a major role in interpersonal and human-computer interaction. Most facial expression recognition systems have focused on only six basic emotions and their concomitant prototypic expressions posed by a small set of subjects. In reality, humans are capable of producing thousands of expressions that vary in complexity, intensity, and meaning. To represent the full range of facial expression, we developed a computer vision system that automatically recognizes individual action units (AUs) or AU combinations using Hidden Markov Models and estimates expression intensity. Three modules are used to extract facial expression information: (1) facial feature point tracking, (2) dense flow tracking with principal component analysis (PCA), and (3) high gradient component detection (i.e. furrow detection). The average recognition rate of upper and lower face expressions is 85% and 88%, respectively, using feature point tracking, 93% (upper face) using dense flow tracking with PCA, and 85% and 81%, upper and lower face respectively, using high gradient component detection.

1. Introduction

The face is a rich source of information about human behavior. Facial expression displays emotion [7], regulates social behavior [5], signals communicative intent [9], is computationally related to speech production [16], and reveals brain function and pathology [18]. To make use of the information afforded by facial expression, automated reliable and valid measurement is critical.

Most facial expression recognition systems use either complicated three-dimensional (3-D) wireframe face models to recognize and reproduce facial expressions [8,21] or analyze averaged optical flow within local regions (*e.g.*, forehead, brows, eyes, nose, mouth, cheek, and chin). A limitation of wireframe face models is that the initial adjustment between the 3-D wireframe and the 2-D surface images is manual, which affects the accuracy of the recognition results. Additionally, it is impractical and difficult to use 3-D wireframe models when working with high-resolution images, large databases (*i.e.*, number of subjects or image sequences), or faces with complex geometric motion properties.

In contrast to the complex 3-D geometric models, optical flow-based approaches treat the facial expression recognition problem as 2-D. These approaches have been shown to track motion and classify prototypic emotion expressions [3,4,15,19,24]. A problem, however, is that the flow direction of each individual local face region is changed to conform to the flow plurality of the region [3, 19, 24] or averaged over an entire region [14, 15]. Parameter thresholds are often used [3,4]. These systems are relatively insensitive to subtle motion because information about small deviations is lost when the flow pattern is removed or thresholds are imposed. The recognition ability and accuracy of these systems may be reduced further when presented with less stylized expressions.

Most research in facial expression recognition is limited to six basic emotions (*i.e.*, joy, fear, anger, disgust, sadness, and surprise) posed by a small set of subjects. These stylized expressions are classified into emotion categories rather than facial action [3,4,19,24]. In everyday life, however, prototypic expressions occur relatively infrequently. Emotion or intention more often is communicated by small changes in the face. For example, disagreement or anger may be communicated to an interactant by furrowed eyebrows (AU 4). The degree of anger experienced may be communicated by the intensity of the brow motion. Our goal is to develop a system that robustly extracts and recognizes subtle facial feature motion, discriminates between similar facial expressions and quantifies intensity variation [11].

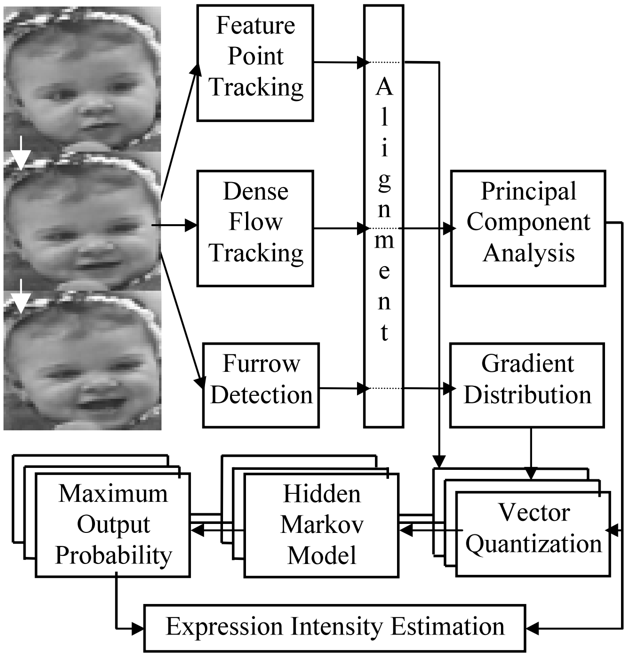
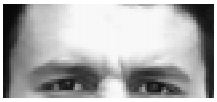
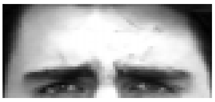
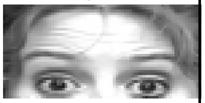
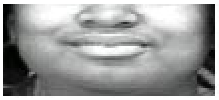
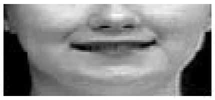
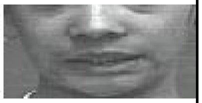
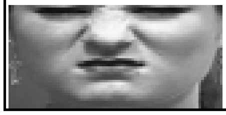
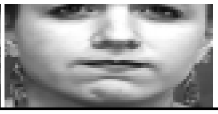
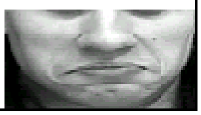


Figure 1. The flow of the facial analysis algorithm.

Table 1. Facial Action Coding System Action Units [6].

Upper Face Expressions		
AU4	AU1+4	AU1+2
		
Lower Face Expressions		
AU12	AU12+25	AU20+25
		
AU9+17	AU17+23+24	AU15+17
		

2. Extraction and Recognition System

Three convergent modules are used to extract expression information (Figure 1). Feature-point and dense flow tracking are used to track facial motion since our goal is to recognize expressions varying in intensity in the spatio-temporal domain. The use of optical flow to track motion is optimized in the face because facial skin and features naturally have a great deal of texture. Facial feature point tracking is especially sensitive to subtle feature motion. Dense flow tracking with principal component analysis (PCA) includes motion information from the entire face. Low-dimensional weighted vectors represent the high-dimensional pixel-wise optical flows of each frame. These weighted vectors are used to estimate expression intensity.

High gradient component detection is used to recognize expressions by the presence of furrows. Facial motion produces transient wrinkles and furrows perpendicular to the motion direction of the activated muscle. The facial motion associated with a furrow produces gray-value change in the face image, which can be extracted by use of high gradient component detectors.

Because analysis of dynamic images produces more accurate and robust recognition than that of a single static image [2], expressions are recognized in the context of entire image sequences of arbitrary length. Hidden Markov Models (HMMs) [20] are used for facial expression recognition because they perform well in the spatio-temporal domain, robustly deal with the time warping problem (compared with [14]), and are analogous to human performance (*e.g.*, for facial expression [11], gesture [25] and speech recognition [20]).

2.1 Facial Action Coding System (FACS)

Our approach to facial expression analysis is based on the Facial Action Coding System (FACS) [6], which is an anatomically based coding system that enables discrimination between closely related expressions. FACS divides the face into upper and lower regions and subdivides motion into action units (AUs). AUs are the smallest visibly discriminable muscle actions that combine to perform expressions. In the present study, three sets of similar facial expressions are recognized and their intensities are estimated (Table 1).

2.2 Normalization

Though all subjects are viewed frontally in our current research, some out-of-plane head motion occurs with face motion (facial expressions). Additionally, face size varies among individuals. To eliminate rigid head motion from nonrigid facial expression, an affine transformation, which includes translation, scaling and rotation factors, is adequate to normalize the face position and maintain face magnification invariance. In an initial processing step, the images are automatically normalized to ensure that flows of each frame have close geometric correspondence. Face position and size are kept constant across subjects so that these variables do not interfere with expression recognition.

The positions of all tracking points or image pixels in each frame are automatically normalized by warping them to a standard 2-D face model based on three facial feature points: the medial canthus of both eyes and the uppermost point on the philtrum (Figure 2).

3. Three Extraction Approaches

In our system, three approaches are used to extract the expression information: (1) facial feature point tracking, (2) dense flow tracking with PCA, and (3) high gradient component detection.

3.1 Facial Feature Point Tracking

Because facial features have high texture and represent underlying muscle activation, optical flow may be used to track movement of feature points. Facial expressions are recognized based on the motion of these points. Feature points located around the contours of the brows, eyes, nose, mouth, and below the lower eyelids are manually marked in the first frame of each image sequence using a computer mouse (Figure 3). Each feature point is the center of a 13 x 13 flow window (image size: 490 x 640; row x column pixels but cropped to 417 x 385 pixels) that includes the horizontal and vertical flows.

The movement of facial feature points is automatically tracked across an image sequence using Lucas-Kanade's optical flow algorithm, which has high tracking accuracy [13] (Figure 3). The pyramidal (5 level) optical flow method is used for tracking because it robustly manages large facial feature motion displacement, such as mouth opening or brows raised suddenly. This method deals well with large feature point movement (100 pixel displacement between two frames) while maintaining its sensitivity to subtle (sub-pixel) facial motion.

In this study, upper face expressions are recognized based on the displacements of 6 feature points at the upper boundaries of both brows, and lower face expressions are recognized based on the displacements of 10 feature points around the mouth. The displacement of each feature point is calculated by subtracting its normalized position in the first frame from its current normalized position. The 6- and 10-dimensional horizontal displacement vectors and 6- and 10-dimensional vertical displacement vectors are concatenated to form 12- and 20-dimensional displacement vectors for the upper and lower facial expressions, respectively. The 12- and 20-dimensional displacement vectors of the upper and lower face represent the facial motion of each frame.

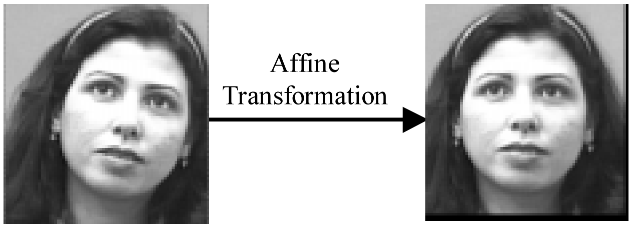


Figure 2. Normalization.

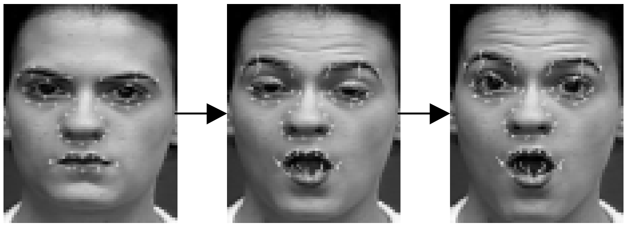


Figure 3. Facial feature point tracking.

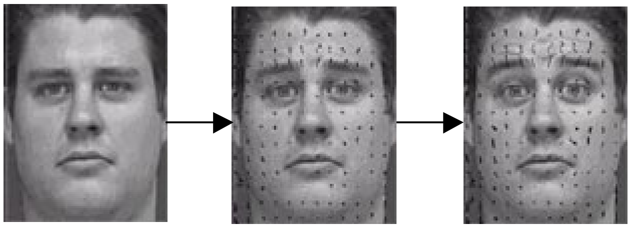


Figure 4. Dense flow tracking. (Compared with Figure 3: same upper face expression but different lower face expression.)

3.2 Dense Flow Tracking with Principal Component Analysis

Though feature-point tracking is sensitive to subtle feature motion and tracks large displacement well, the forehead, cheek and chin regions also have important expression information. To include more detailed motion information, each pixel of the entire face image is tracked using dense flow [23] (Figure 4).

Because we have a large image database in which the motion of consecutive frames in a sequence is strongly correlated, the high-dimensional pixel-wise flows of each frame need to be compressed to their low-dimensional representations without losing significant characteristics and inter-frame correlation. PCA has excellent properties for our purposes, including image data compression and maintenance of a strong correlation between two consecutive motion frames. Since our goal is to recognize expression rather than identify individuals or objects [10, 17, 22], facial motion is analyzed using dense flow- not gray value- to ignore differences across individual subjects (compared with [1]). To ensure that the pixel-wise flows of each frame have relative geometric correspondence, an affine transformation is used to automatically warp the pixel-wise flows of each frame to the 2-D face model.

Using PCA and focusing on the (110 x 240 pixels) upper face region, 10 "eigenflows" are created (Figure 5) (10 eigenflows from the horizontal- and 10 eigenflows from the vertical direction flows [11]). These eigenflows are defined as the eigenvectors corresponding to the 10 largest eigenvalues of the 832 x 832-covariance matrix constructed by 832 flow-based training frames from the 44 training image sequences. Compression rate is 83:1.

Each flow-based frame of the expression sequences is projected onto the flow-based eigenspace by taking its inner product with each element of the eigenflow set, producing a 10-dimensional weighted vector (Figure 6). The 10-dimensional horizontal-flow weighted vector and the 10-dimensional vertical-flow weighted vector are concatenated to form a 20-dimensional weighted vector for each flow-based frame.

3.3 High Gradient Component Analysis

Facial motion produces transient wrinkles and furrows perpendicular to the motion direction of the activated muscle.

The facial motion associated with these furrows produces gray-value changes in the face image. High gradient components (*i.e.*, furrows) of the face image are extracted with a variety of line or edge detectors. After normalization of each 417 x 385-pixel image, a 5 x 5 Gaussian filter is used to smooth the image. 3 x 5 (row x column) horizontal line and 5 x 3 vertical line detectors are used to detect horizontal lines (*i.e.*, high gradient components in the vertical directions) and vertical lines in the forehead region, respectively; 5 x 5 diagonal line detectors are used to detect diagonal lines along the nasolabial furrow; and 3 x 3 edge detectors are used to detect high gradient components around the lips and on the chin region.

To verify that the high gradient components are produced by transient skin or feature deformations– and not a permanent characteristic of the individual's face– the gradient intensity of each detected high gradient component in the current frame is compared with corresponding points within a 3 x 3 region of the first frame. If the absolute value of the difference in gradient intensity between these points is higher than the threshold value, it is considered a valid high gradient component produced by facial expression. All other high gradient components are ignored. In the former case, the high gradient component (pixel) is assigned a value of 1. In the latter case, the pixels are assigned a value of 0. An example of the procedure for extracting high gradient components on the forehead region is shown in Figure 7. A gray value of 0 corresponds to black and 255 to white.

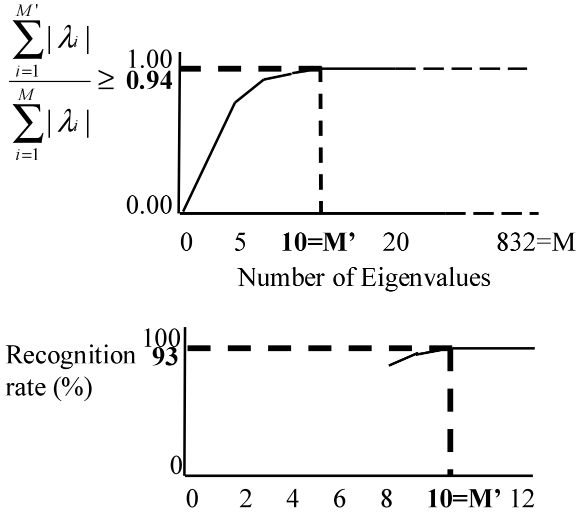


Figure 5. Computation of eigenflow number for vertical direction flows.

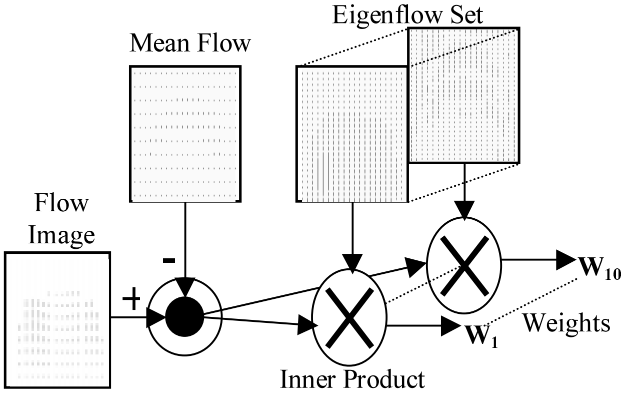


Figure 6. Vertical-flow weighted vector computation for the upper face expressions.

The forehead (upper face) and lower face regions of the normalized face image are divided into 13 and 16 blocks, respectively (Figure 8). The mean value of each block is calculated by dividing the number of pixels having a value of 1 by the total number of pixels in the block. The variance of each block is calculated as well. For upper and lower face expression recognition, mean and variance values are concatenated to form 26- and 32-dimensional mean and variance vectors, respectively, for each frame.

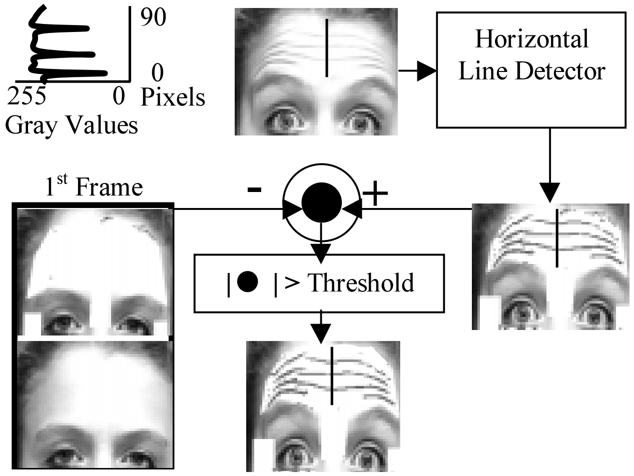


Figure 7. The procedure for horizontal line detection at the forehead (upper face) region.

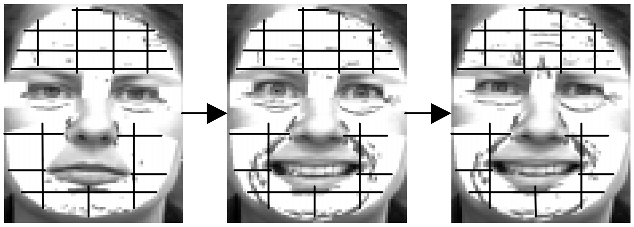


Figure 8. Quantization of the high gradient components.

4. Expression Recognition and Expression Intensity Estimation

The 12- and 20-dimensional training displacement vectors from feature point tracking, the 20-dimensional training weighted vectors from the dense flow tracking with PCA, and the 26- and 32-dimensional training mean and variance vectors from the high gradient component detection are each vector quantized [12]. HMMs are then trained. Because the HMM set represents the most likely individual action unit (AU) or AU combination, it can be employed to evaluate the test-input sequence. The test-input sequence is evaluated by selecting the maximum output probability value from the HMM set.

After recognizing an input facial expression sequence, the expression (motion) intensity of an individual frame in this sequence is estimated using the correlation property of PCA. That is, the minimum distance between two projected points (weighted vectors) in eigenspace has the maximum correlation or motion similarity. The sum-of-squared-difference (SSD) is used to find the frame with the best match in expression (motion) intensity from any training sequence having the same expression as the test frame (Figure 9).

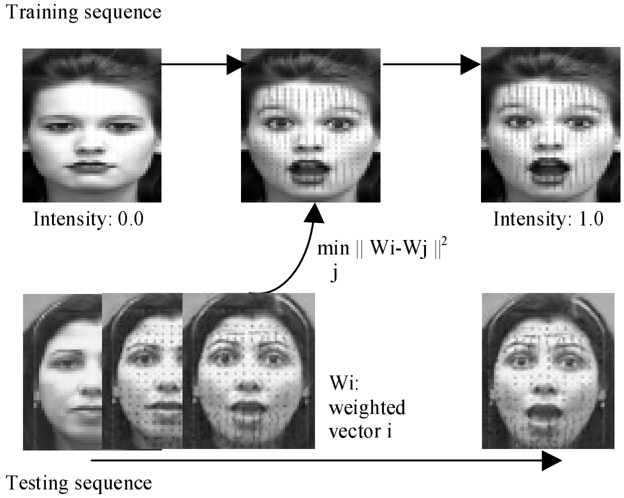


Figure 9. Expression intensity matching.

5. Experimental Results

For this study, frontal views of all subjects were videotaped under constant illumination using fixed light sources, and none of the subjects wore eyeglasses. Previously untrained subjects were video recorded performing a series of expressions, and the image sequences were coded by certified FACS coders. Facial expressions were analyzed in digitized image sequences of arbitrary length (expression sequences from neutral to peak varied from 9 to 47 frames).

Subjects were 85 males and females (Asian, Euro- and African-American) between the ages of 1 and 35 years. 300 image sequences were analyzed. Recognition accuracy did not vary between males and females, and Euro- and African-Americans.

The average recognition rate of upper face expressions was 85% by feature point tracking, 93% by dense flow tracking with PCA, and 85% by high gradient component detection. These results are based on 60, 44, and 100 training image sequences and 75, 75, and 160 testing image sequences, respectively (Table 2). The average recognition rate of lower face expressions was 88% by feature point tracking and 81% by high gradient component detection (based on 120 and 50 training image sequences, and 150 and 80 testing image sequences, respectively) (Table 2). Results for dense flow tracking with PCA are not yet available for the lower face.

Table 2. Recognition results.

Upper Face Expression Recognition			
Human	**Feature Point Tracking**		
	AU 4	AU 1+4	AU 1+2
AU 4	**22**	3	0
AU 1+4	4	**19**	2
AU 1+2	0	2	**23**
Human	**Dense Flow Tracking with PCA**		
AU 4	**23**	2	0
AU 1+4	3	**22**	0
AU 1+2	0	0	**25**

Human	**High Grade Component Detection**			
	AU 0	AU 4	AU 1+4	AU 1+2
AU 0	**26**	4	0	0
AU 4	5	**43**	2	0
AU 1+4	0	1	**24**	5
AU 1+2	0	0	7	**43**

Lower Face Expression Recognition						
Human	**Feature Point Tracking**					
AUs	12	12+25	20+25	9+17	15+17	17+23+24
12	**25**	0	0	0	0	0
12+25	0	**21**	4	0	0	0
20+25	0	5	**20**	0	0	0
9+17	0	0	0	**22**	2	0
15+17	0	0	0	0	**23**	2
17+23+24	0	0	0	3	1	**25**

Human	**High Grade Component Detection**	
	12+25	9+17
12+25	**42**	8
9+17	7	**23**

6. Conclusion

We have developed a computer vision system, Automated Face Coding, that automatically recognizes facial expressions based on FACS action units. To optimize system performance, three modules extract facial motion: feature point tracking, dense flow tracking with PCA, and high gradient component detection.

The pyramidal optical flow module for feature point tracking is an easy, fast, and accurate way to track facial motion. It tracks large displacement well and is sensitive to subtle feature motion. To track motion across the entire face, dense flow with PCA is used. PCA compresses the high-dimensional pixel-wise flows to low-dimensional weighted vectors. Unlike feature point tracking, dense flow tracking with PCA introduces motion insensitivity and is subject to error due to occlusion (*e.g.*, appearance of tongue or teeth when the mouth opens) or discontinuities between the face contour and background. Additionally, processing time is prolonged in dense flow tracking (98% computing time of this system) because of recursive computation using wavelet approach (multiple basis functions).

High-gradient component detection is sensitive to change in transient facial features (*e.g.*, furrows), but is subject to error from individual differences in subjects. Younger subjects, especially infants, show less furrowing than older ones, which reduces the information value of high gradient component detection.

Though all three modules resulted in some recognition error, the pattern of errors was encouraging. That is, the error results were classified into the expression most similar to the target (*e.g.*, AU 4 is confused with AU 1+4 but not AU 1+2). Because each module has strengths and weaknesses, feature point tracking, dense flow tracking with PCA, and high gradient component detection can be used in combination to produce a more robust and accurate recognition system. A focus of current work is the implementation of a multi-dimensional HMM to integrate these three modules.

In future work, we will recognize more detailed and complex action units, increase the processing speed of dense flow analysis, interpolate expression intensity, and separate rigid and non-rigid motion more robustly. Potential applications include assessment of nonverbal behavior in clinical and research settings, speech recognition in combination with lip-reading, video-conferencing, and human-computer interface/interaction. In addition, automated quantitative assessment of facial expression can inform work in facial animation.

References

[1] M.S. Bartlett, *et al.*, "Classifying Facial Action," *Adv. in Neural Info. Proc. Sys. 8*, MIT Press, 1996.

[2] J.N. Bassili, "Emotion Recognition: The Role of Facial Movement and the Relative Importance of Upper and Lower Areas of the Face,"*J. of Personality and Social Psy.*, Vol. 37, pp. 2049-2059, 1979.

[3] M.J. Black and Y. Yacoob, "Recognizing Facial Expressions under Rigid and Non-Rigid Facial Motions," *Intl. Workshop on Automatic Face and Gesture Recognition*, Zurich, pp. 12-17, 1995.

[4] M.J. Black, *et al.*, "Learning Parameterized Models of Image Motion," *CVPR*, 1997.

[5] J.F. Cohn and M. Elmore, "Effect of Contingent Changes in Mothers' Affective Expression on the Organization of Behavior in 3-Month-Old Infants," *Infant Behavior and Development*, Vol. 11, pp. 493-505, 1988.

[6] P. Ekman and W.V. Friesen, "The Facial Action Coding System," *Consulting Psy. Press*, CA, 1978.

[7] P. Ekman, "Facial Expression and Emotion," *American Psychologist*, Vol. 48, pp. 384-392, 1993.

[8] I.A. Essa, "Analysis, Interpretation and Synthesis of Facial Expressions," *Perceptual Computing TR 303*, MIT Media Laboratory, Feb. 1995.

[9] A.J. Fridlund "Human Facial Expression: An Evolutionary View," *Academic Press*, CA, 1994.

[10] M. Kirby and L. Sirovich, "Application of the Karhuneh-Loeve Procedure for the Characterization of Human Faces," *IEEE Trans. on PAMI* 12, No. 1, 1990.

[11] J.J. Lien, A.J. Zlochower, J.F. Cohn, C.C. Li, and T. Kanade, "Automatically Recognizing Facial Expressions in the Spatia-Temporal Domain," *Perceptual User Interface Workshop*, Canada, 1997.

[12] Y. Linde, A. Buzo, and R. Gray, "An Algorithm for Vector Quantizer Design," *IEEE Trans. on Communications*, Vol. COM-28, NO. 1, 1980.

[13] B.D. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," *Proc. of the 7th Intl. Joint Conf. on AI*, 1981.

[14] K. Mase and A. Pentland, "Automatic Lipreading by Optical-Flow Analysis," *Systems and Computers in Japan*, Vol. 22, No. 6, 1991.

[15] K. Mase, "Recognition of Facial Expression from Optical Flow," *IEICE Trans.*, Vol. E74, pp. 3474-3483, 1991.

[16] D. McNeil, "So you think gestures are nonverbal?" *Psychological Review*, *92*, 350-371, 1985.

[17] H. Murase and S.K. Nayar, "Visual Learning and Recognition of 3-D Objects from Appearance," *IJCV*, 14, pp. 5-24, 1995.

[18] W.E. Rinn,. "The neuropsychology of facial expression: A review of the neurological and psychological mechanisms for producing facial expressions." *Psychological Bulletin, 95*, pp. 52-77, 1984.

[19] M. Rosenblum, Y. Yacoob and L.S. Davis, "Human Emotion Recognition from Motion Using a Radial Basis Function Network Architecture," *Proc. of the Workshop on Motion of Non-rigid and Articulated Objects*, Austin, TX, Nov. 1994.

[20] L.R. Rabiner, "An Introduction to Hidden Markov Models," *IEEE ASSP Magazine*, pp. 4-16, Jan. 1986.

[21] D. Terzopoulos and K. Waters, "Analysis of Facial Images Using Physical and Anatomical Models," *ICCV*, pp. 727-732, Dec. 1990.

[22] M. Turk and A. Pentland, "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, pp. 71-86, 1991.

[23] Y.T. Wu, T. Kanade, J. F. Cohn, and C.C. Li, "Optical Flow Estimation Using Wavelet Motion Model," *ICCV*, 1998.

[24] J. Yacoob and L. Davis, "Computing Spatio-Temporal Representations of Human Faces," *CVPR*, pp. 70-75, 1994.

[25] J. Yang, "Hidden Markov Model for Human Performance Modeling," *Ph.D. Dissertation*, University of Akron, August 1994.