Evaluation of Face Resolution for Expression Analysis

Ying-li Tian

IBM T. J. Watson Research Center,
PO Box 704, Yorktown Heights, NY 10598
Email: yltian@us.ibm.com

Abstract

Most automatic facial expression analysis (AFEA) systems attempt to recognize facial expressions from data collected in a highly controlled environment with very high resolution frontal faces (face regions greater than 200 x 200 pixels). However, in real environments, the face image is often in lower resolution and with head motion. It is unclear that the performance of AFEA systems for low resolution face images. The general approach to AFEA consists of 3 steps: face acquisition, facial feature extraction, and facial expression recognition. This paper explores the effects of different image resolutions for each step of facial expression analysis. The different approaches are compared for face detection, face data extraction and expression recognition. A total of five different resolutions of the head region are studied (288x384, 144x192, 72x96, 36x48, and 18Xx24) based on a widely used public database [16]. The lower resolution images are down-sampled from the originals.

1. Introduction

Facial expression analysis includes both measurement of facial motion and recognition of expression. The general approach to Automatic Facial Expression Analysis (AFEA) systems consists of 3 steps (see Figure 1): face acquisition, facial feature extraction and representation, and facial expression recognition.

Face acquisition is a processing stage to automatically find the face region for the input images or sequences. It can be a face detector to detect a face in each frame or just detect face in the first frame and then track the face in the remainder of the video sequence. In order to handle large head motion, head finding, head tracking and pose estimation can be applied to a facial expression analysis system.

After the face is located, the next step is to extract and represent the facial changes caused by facial expressions. In facial feature extraction for expression analysis, there are mainly two types of approaches: geometric feature-based methods and appearance-based methods. The geometric facial features present the shape and locations of facial components (including mouth, eyes, brows, nose etc.). The facial components or facial feature points are extracted to form a feature vector that represents the face geometry. In appearance-based methods, image filters, such as Gabor wavelets, are applied to either the whole face or specific regions in a face image to extract a feature vector. Depending on the different facial feature extraction methods, the effects of in-plane head rotation and different scales of the faces can be eliminated, either by face normalization before the feature extraction or by feature representation before the step of expression recognition.

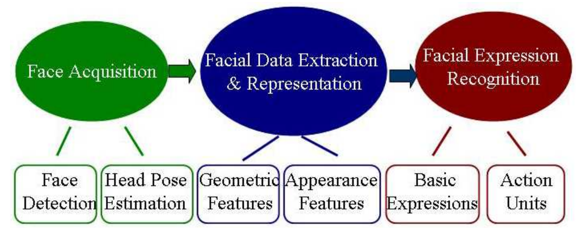


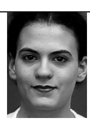
Figure 1: Basic Structure of Facial Expression Analysis Systems.

Facial expression recognition is the last stage of AFEA systems to identify facial changes as facial action coding system (FACS) action units (AUs) [9] or prototypic emotional expressions.

Many recent advances and successes in AFEA have been achieved [2, 3, 5, 8, 10, 12, 19, 21, 27, 30, 32]. With a few exceptions [7, 8, 11, 18, 27], most systems only recognized the basic expressions. We [27, 28] reported some of the most extensive experimental results on AU recognition. Their system can cope with a large change of appearance and limited out-of-plane head motion. To increase the robustness and accuracy of the feature extraction, multi-state face-component models were devised. The system recognized 16 of the 30 AUs whether they occured alone or in combinations. Recently, Bartlett *et al.* [2] and Cohn *et al.* [6, 20] attempted a comparative study for FACS AU recognition (in the upper face) in spontaneously occurring behavior by using same database [13]. In that database, subjects were ethnically diverse, AUs occurred during speech, and out-of-plane motion and occlusion from head motion and glasses were common.

While many recent advances and successes in automatic facial expression analysis have been achieved as described above, many questions remain open. For example, how do we recognize facial expressions in real life? Real-life facial expression analysis is much more difficult than the posed actions studied predominantly to date. Head motion, low resolution input images, absence of a neutral face for comparison, and low intensity expressions are among the factors that complicate facial expression analysis. This paper focuses on evaluation of AFEA system for low resolution input images. Most (AFEA) systems attempt to recognize facial expressions from data collected in a highly controlled laboratory situation with very high resolution faces (face regions are greater than 200 x 200 pixels). However, in real applications, the face region is often in lower resolutions. We [26] were the first attempt to recognize facial expressions in compressed images with lower resolution (the face region is around 50x70 to 75x100 pixels). To handle the full range of head motion, we detected the head instead of the face. Then the head pose was estimated based on the detected head. For frontal and near frontal views of the face, the location and shape features were computed for expression recognition. Our system ran in real-time and successfully dealt with complex real world interactions.

Table 1: An example in five different resolutions from Cohn-Kanade database. The head region is cropped for display purpose only. The lower resolution images are down-sampled from the originals.

				
288x384 (Original)	144x192	72x96	36x48	18x24

In this paper, we investigate the effects of different image resolutions for each step of facial expression analysis. As shown in Table 1, total five different resolutions of the head region are studied (288x384, 144x192, 72x96, 36x48, and 18x24) based on Cohn-Kanade AU-Coded Face Expression Image Database [16]. The lower resolution images are down-sampled from the originals. As shown in Figure 1, two face detectors [23, 29] and a head pose estimator [26] are studied for the step of face acquisition. The effects of resolutions on the extraction of both geometric features and appearance features are investigated in stage 2. For the step of expression recognition, both emotional-specific expressions and selected action units recognition are investigated for different face resolutions.

In Section 2 we describe the implementation details for the methods used in our evaluation in each step of facial expression analysis. Section 3 gives the experimental results. In the last section, we conclude and discuss the results.

2. Experimental Setup

2.1. Face Acquisition

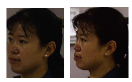
Most research of AFEA attempts to recognize facial expressions only from frontal-view or near frontal-view faces. Since the frontal-view face is not always available in real environments, the face acquisition methods should detect both frontal and non-frontal view faces in an arbitrary scene. To handle out-of-plane head motion, the face can be obtained by face detection, 2D or 3D face tracking, or head pose detection. Non-frontal view faces can be warped or normalized to frontal view for expression analysis.

Face Detection: Many face detection methods have been developed to detect faces in an arbitrary scene [15, 22, 23, 17, 25, 29]. In this paper, we evaluate two face detectors for different image resolutions[23, 29]. One is a neural network based face detector which was developed by Rowley *et al.* [23]. Another was developed by Viola and Jones [29] based on a set of rectangle features. Both of them can only detect frontal and near-frontal views of faces. Details of the face detectors can be found in papers [23, 29].

Head Pose Estimation: In order to handle the full range of head motion for expression analysis in real environments, we also evaluate head pose detection for different image resolutions. We [26] detect the head instead of the face. Head detection uses the smoothed silhouette of the foreground object as segmented using background subtraction, and computes the *negative curvature minima* (NCM) points of the silhouette. Other head detection techniques that use silhouettes can be found in papers [14].

After the head is located, the head image is converted to gray scale, histogram equalized and resized to 32x32. Then a three-layer neural network (NN) is employed to estimate the head pose. The inputs to the network are the processed head image. The outputs are the 3 head poses: 1) frontal or near frontal view, 2) side view or profile, 3) others such as back of the head or occluded face (see Table 2). In the frontal or near frontal view, both eyes and lip corners are visible. In side view or profile, at least one eye or one corner of the mouth becomes self-occluded because of the head. The expression analysis process is applied only to the frontal and near frontal view faces. More details about head pose estimation can be found in our paper [4].

Table 2: The definitions and examples of the 3 head pose classes: 1) frontal or near frontal view, 2) side view or profile, 3) others such as back of the head or occluded faces. The expression analysis process is applied to only the frontal and near frontal view faces.

Poses	Frontal or near frontal	Side view or profile	Others
Definitions	Both eyes and lip corners are visible	One eye or one lip corner is occluded	Not enough facial features
Examples			

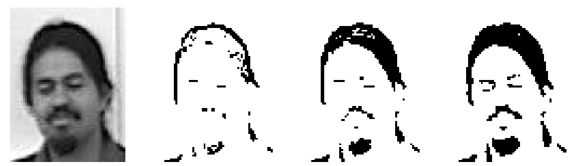
2.2. Facial Feature Extraction and Representation

In our study, we extract two types of features: geometric features and appearance features. Geometric features present the shape and locations of facial components (including mouth, eyes, brows, nose etc.). The facial components or facial feature points are extracted to form a feature vector that represents the face geometry. The appearance features present the appearance (skin texture) changes of the face such as wrinkles and furrows. The appearance features can be extracted on either the whole face or specific regions in a face image.

Geometric Feature Extraction

(1) Feature Tracking: For feature tracking, we use the same method as that was developed in paper [27]. To detect and track changes of facial components in near frontal face images, multi-state models are developed to extract the geometric facial features. A three-state lip model describes the lip state: open, closed, and tightly closed. A two-state model (open or closed) is used for each of the eyes. Each brow and cheek has a one-state model. Some appearance features, such as *nasolabial furrows* and *crows-feet wrinkles*, are represented explicitly by using two states: present and absent. Given an image sequence, the region of the face and approximate location of individual face features are detected automatically in the initial frame. The contours of the face features and components then are adjusted manually in the initial frame. Then, all face feature changes are automatically tracked in the image sequence. Details of geometric feature extraction and representation can be found in paper [27].

(2) Feature Detection: In order to deal with low resolution face images, we also evaluate a simple geometric feature detection method [26]. In our method, six location features are extracted for expression analysis. They are eye centers (2), eyebrow inner endpoints (2), and corners of the mouth (2). The feature extraction approach presented here is similar to that taken by Yang *et al.* [33] and is an attempt to make the extraction of the eye centers more accurate and robust.



(a) (b) (c) (d)

Figure 2: Iterative thresholding of the face to find eyes and brows. (a) grey-scale face image, (b) threshold = 30, (c) threshold = 42, (d) threshold = 54.

To find the eye centers and eyebrow inner endpoints inside the detected frontal or near frontal face, we have developed an algorithm that searches for two pairs of dark regions which correspond to the eyes and the brows by using certain geometric constraints such as position inside the face, size and symmetry to the facial symmetry axis. Similar to paper [33], the algorithm employs an iterative thresholding method to find these dark regions under different or changing lighting conditions.

Figure 2 shows the iterative thresholding method to find eyes and brows. Generally, after *five* iterations, all the eyes and brows are found. If satisfactory results are not found after 20 iterations, we think the eyes or the brows are occluded or the face is not in a near frontal view. Unlike the work of Yang *et al.* to find one pair of dark regions for the eyes only, we find two pairs of parallel dark regions for both the eyes and eyebrows. By doing this, not only are more features obtained, but also the accuracy of the extracted features is improved. As shown in Figure 2(b), the right brow and the left eye is wrongly extracted as the two eyes in Yang's approach. Figure 2(d) shows that the correct positions are extracted for all the eyes and eyebrows by our method. Then the eye centers and eyebrow inner endpoints can be easily determined. If the face image is continually in the frontal or near frontal view in an image sequence, the eyes and brows can be tracked by simply searching for the dark pixels around their positions in the last frame.

After finding the positions of the eyes, the location of the mouth is first predicted. Then, the vertical position of the line between the lips is found, using an integral projection of the mouth region proposed by Yang *et al.* [33]. Finally, the horizontal borders of the line between the lips are found, using an integral projection over an edge-image of the mouth. After Yang *et al.*, the following steps are used to track the corners of the mouth: 1) Find two points on the line between the lips near the previous positions of the corners in the image. 2) Search along the darkest path to the left and right, until the corners are found. Finding the points on the line between the lips can be done by searching for the darkest pixels in search windows near the previous mouth corner positions. Because there is a strong change from dark to bright at the location of the corners, the corners can be found by looking for the maximum contrast along the search path. The details of the tracking method of the mouth corners can be found in the original paper [33].

(3) Geometric Feature Representation: After extracting the location features, the geometric facial features can be represented by a set of parameters for expression recognition based on the line connecting the two eyes (*eye-line*) [27].

For the geometric features estimated by feature tracking, a total of 24 parameters were grouped for the whole face which describe shape, motion, and the state of face components and furrows. To remove the effects of variation in planar head motion and scale between image sequences in face size, all parameters are computed as ratios of their current values to that in the reference frame (neutral frame).

For the geometric features estimated by feature detection, we represent the face location features by 5 parameters. These parameters are the distances between the *eye-line* and the corners of the mouth, the distances between the *eye-line* and the inner eyebrows, and the width of the mouth (the distance between two corners of the mouth). Again, all the parameters are computed as ratios of their current values to that in the reference frame.

Appearance Feature Extraction

(1) Gabor Wavelet Representation: We use Gabor wavelets to extract the facial appearance changes as a set of multi-scale and multi-orientation coefficients. Unlike paper [8], which applies Gabor wavelets to the upper face and lower face separately, we apply the Gabor filters to the difference image for the whole face. The difference images are obtained by subtracting a neutral expression frame for each sequence, and were convolved with a bank of 40 Gabor filters with 8 orientations and 5 spatial frequencies.

(2) Face Alignment: To remove position noises for appearance feature extraction, the faces are normalized to a fixed distance between the center of two eyes for each face resolution. For example, the distance between the eyes is 104 pixels for resolution 288x384, 52 pixels for 144x192, 26 pixels for 72x96, 13 pixels for 36x48, and 7 pixels for 18x24 respectively. To remove the light changes, the brightness of the face images are linearly rescaled to [0, 255].

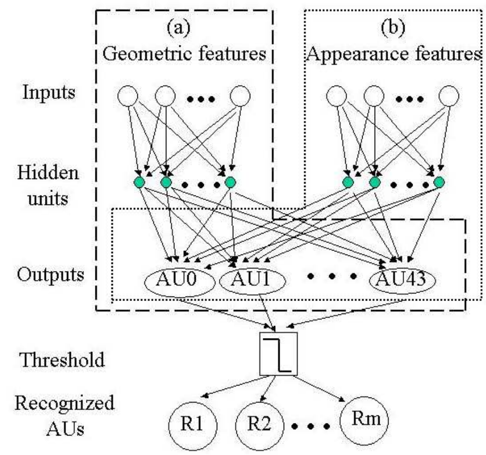


Figure 3: Neural network-based recognizer for expressions.

2.3. Expression Recognition

We investigate two type of expression recognition: FACS AUs and basic expressions. For each type of expression recognition, we compare the recognition accuracy for geometric features, appearance features, and both geometric and appearance features. Figure 3 shows an example of our network-based recognizer to recognize FACS AUs.

We use three-layer neural networks with one hidden layer to recognize expressions by a standard back-propagation method. The inputs can be either the normalized geometric features or the appearance feature or both. The outputs are the recognized action units or six basic expressions. Figure 3(a) is a sun-network for recognizing AUs by using the geometric features alone. The sub-network shown in Figure 3(b) is used for recognizing AUs by using Gabor wavelets. For using both geometric features and regional appearance patterns, these two sub-networks are applied in concert. The outputs either are the recognized AUs or the six basic expressions. If the outputs are AUs, each output unit gives an estimate of the probability of the input image consisting of the associated AUs. The networks are trained to respond to the designated AUs whether they occur singly or in combination. When AUs occur in combination, multiple output nodes are excited. When the outputs are the six basic expressions, only one output node with the highest probability is selected.

For AU recognition, we recognized 14 AUs. There are AU1, AU2, AU4, AU5, AU6, AU9, AU10, AU12, AU15, AU17, AU20, AU23, AU24, and AU25*, where AU25* includes AU25, AU26, and AU27. The six basic expressions which we recognized are happiness, surprise, fear, sadness, disgust, and anger.

3. Effects of Image Resolution for Facial Expression Analysis

3.1. Database

The DFAT subset of Cohn-Kanade expression database [16] is used for our experiments. The database contains 704 image sequences from 97 subjects. Subjects sat directly in front of the camera and performed a series of facial behaviors which were recorded in an observation room. Image sequences with in-plane and limited out-of-plane motion were included. The image sequences began with a neutral face and were digitized into 640x480 pixel arrays with either 8-bit gray-scale values. The length of the image sequences are varying from 9 to 47 frames. The size of head region is about 280$\times$380 pixels. More details about the database can be found at paper [16].

Table 3: Summary of the effects of faces at different resolutions for expression analysis. For face acquisition, ”FD” indicates face detector. ”HPE” indicates head pose estimation. For feature extraction, ”G1” indicates geometric features extracted by feature tracking. ”G2” indicates geometric features extracted by feature detection. ”AP” indicates appearance features extracted by Gabor wavelets.

Face Process		 288x384 (Original)	 144x192	 72x96	 36x48	 18x24
Face Acquisition	FD	100%	100%	100%	100%	0%
	HPE	98.5%	98%	98.2%	97.8%	98%
Feature Extraction	G1	Yes	Yes	Yes	No	No
	G2	Yes	Yes	Yes	Yes	No
	AP	Yes	Yes	Yes	Yes	Yes
FACS AUs	G1	90%	90.2%	89.8%	N/A	N/A
	G2	71%	70.8%	72%	54.3%	N/A
	AP	90.7%	90.2%	89.6%	72.6%	58.2%
	G1+AP	92.8%	93%	92.2%	N/A	N/A
	G2+AP	91.2%	90.8%	90%	87.7%	N/A
Basic Expressions	G1	92.5%	91.8%	91.6%	N/A	N/A
	G2	74%	73.8%	72.9%	61.3%	N/A
	AP	91.7%	92.2%	91.6%	77.6%	68.2%
	G1+AP	93.8%	94%	93.5%	N/A	N/A
	G2+AP	93.2%	93%	92.8%	89%	N/A

3.2. Experimental Results

Table 3 summarizes the effects of faces at different resolutions for each step of expression analysis. For face acquisition, ”FD” indicates face detector. ”HPE” indicates head pose estimation. For feature extraction, ”G1” indicates geometric features extracted by feature tracking. ”G2” indicates geometric features extracted by feature detection. ”AP” indicates appearance features extracted by Gabor wavelets. For expression analysis, ”FACS AU” indicates the recognition of FACS AUs. ”Basic Expression” indicates the recognition of six basic expressions.

Effects of Image Resolution for Face Acquisition:

In this investigation, we focus on effects of image resolution for face acquisition. All images in 704 sequences from 97 subjects are used for face detection and head pose estimation. The results show that both face detectors [23, 29] can detect all faces for head region in resolution of 288x384, 144x192, 72x96, and 36x48, but failed for resolution of 18x24. Compared to face detection, head pose estimation can detect 98% faces for all the resolutions. Our previous results of head pose estimation showed that the head pose estimation can achieve 96% accuracy for 5 poses when head region resolution is about 8x8 pixels [4].

Effects of Image Resolution for Feature Extraction

In this experiment, a total of 696 sequences from 97 subjects (the sequences with very low intensity expression changes are removed) are processed. For the feature tracking method on resolution of 288X384, the tracking results are visually observed. We found the main features such as lip corners, eye center, and brow positions can be well tracked for more than 99% of the sequences. Then we use the tracking results for the resolution of 288x384 as ”ground truth”. For other resolutions, we compare the six main features (lip corners, eye center, and brow positions) which obtained by feature tracking or feature detection with the ”ground truth”. We found that the feature tracking method [27] works well for head region in resolution of 72x96 and higher. The feature detection method can extract reasonable features for head region in resolution of 36Xx48 and higher. For the appearance feature extraction, it is difficult to judge the extracted features without going to the step of expression recognition. We process all of the 696 sequences.

Effects of Image Resolution for Expression Recognition

(1) Recognition of FACS AUs: A total of 690 sequences from 97 subjects are used for FACS AU recognition by removing the sequences either with very low intensity expression changes or which we cannot extract geometric features for the highest resolution. For each sequence, the neutral face and three peak frames are selected for recognition. Currently, we do not use sequential information. We randomly separate the sequences into a training set and a test set, based on the total number of each AUs. That means the same subjects may appear in both training and testing. As described above, we recognize 14 AUs. Almost all AUs are in combination with following exceptions: 9 sequences for single AU12, 2 sequences for single AU17, and 63 sequences for single AU25*.

The average accuracy of the AU recognition is shown in Table 3. For resolution of head region in 72x96 or larger, the same level of recognition rates are achieved by feature tracking and appearance features. The feature detection method achieves lower recognition rates since fewer features are used. Also, appearance features work better for lower resolution of face images. But the face images must be well aligned. Smith *et al.* [24] recognized 6 upper AUs by Gabor wavelets and obtained the comparable results .

(2) Recognition of Basic Expressions: Since the database is coded as FACS AUs, 375 sequences are selected for six basic expression recognition, based on the guide line from FACS to basic expressions, by removing those sequences do not clearly meet the guide line. For example, AU6+12 or AU12 alone with intensity C/D interpret happiness and AU1+2+5B interprets a surprise reaction. The average accuracy of the basic expression recognition is shown in Table 3. Same as AU recognition, for resolution of head region in 72x96 or higher, the same level of recognition rates are achieved by feature tracking and appearance feature. The feature detection method achieves lower recognition rates since fewer features are used. Also as reported by Cohen *et al.*[5], we observe that good recognition results are achieved for happiness and surprise, but most confusions comes from anger, disgust, fear, and sadness.

4. Conclusion and Discussion

In this paper we have presented an experimental evaluation of different face resolutions for each step of facial expression analysis: face acquisition, facial feature extraction, and facial expression recognition. A total of five different resolutions of the head region were studied (288x384, 144x192, 72x96, 36x48, and 18x24) based on Cohn-Kanade AU-Coded Face Expression Image Database [16].

Our empirical studies illustrated following conclusions: (1) Head detection and head pose estimation can detect faces in lower resolution than face detectors. (2) Appearance feature extraction needs face alignment. (3) There is no difference in the recognition of expression analysis when the head region resolution is 72x96 or higher. Geometric features and appearance features achieve same level of recognition rates for both FACS AU recognition and six basic expression recognition. (4) When the resolution of the head region is about 36x48 or lower, appearance features achieve better recognition results than geometric features, but the faces must be well aligned. (5) When the resolution of the head region is lower than 36x48, more reliable results can be obtained for recognizing emotional-specific expressions than for recognizing finer levels of expressions (e.g. FACS AUs.)

The goal was to help understand question: how do we recognize facial expressions in real life?

Real-life facial expression analysis is much more difficult than the posed actions studied predominantly to date. Head motion, low resolution input images, absence of a neutral face for comparison, and low intensity expressions are among the factors that complicate facial expression analysis. Recent work in 3-D modeling of spontaneous head motion and action unit recognition in spontaneous facial behavior are exciting developments. How elaborate a head model is required in such work remains a research question. A cylindrical model is relatively robust and has proven effective as part of a blink detection system [31], but higher parametric generic, or even custom-fitted head models, may prove necessary for more complete action unit recognition.

Most work to date has used a single, passive camera. While there are clear advantages to approaches that require only a single passive camera or video source, multiple cameras are feasible in a number of settings and can be expected to provide improved accuracy. Active cameras can be used to acquire high resolution face images. Also, the techniques of super-resolution can be used to get higher resolution images from multiple low resolution images [1]. At present, it is an open question how to recognize expressions in situations in which a neutral face is unavailable, expressions are of low intensity, or other facial or nonverbal behaviors, such as occlusion by the hands, are present.

Acknowledgements

We would like to thank T. Sim, S. Baker, and M. Bsat at CMU for the creation and public access of the CMU Pose, Illumination, and Expression (PIE) Database of Human Faces which we used for head pose detection. We would also like to thank J. Cohn at University of Pittsburgh to provide Cohn-Kanade AU-Coded Face Expression Image Database.

References

[1] S. Baker and T. Kanade. Limits on super-resolution and how to break them. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(9):1167 – 1183, September 2002.

[2] M. S. Bartlett, B. Braathen, G. Littlewort-Ford, J. Hershey, I. Fasel, T. Marks, E. Smith, T. Sejnowski, and J. R. Movellan. Automatic analysis of spontaneous facial behavior: A final project report. Technical Report INC-MPLab-TR-2001.08, Machine Perception Lab, Institute for Neural Computation, University of California, San Diego, San Diego, CA, October 2001.

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

[4] L. Brown and Y. li Tian. Comparative study of coarse head pose estimation. In *IEEE Workshop on Motion and Video Computing, Orlando FL, Dec. 5-6*, 2002.

[5] I. Cohen, N. Sebe, F. G. Cozman, M. C. Cirelo, and T. S. Huang. Coding, analysis, interpretation, and recognition of facial expressions. *Journal of Computer Vision and Image Understanding Special Issue on Face recognition*, 2003.

[6] J. Cohn, T. Kanade, T. Moriyama, Z. Ambadar, J. Xiao, J. Gao, and H. Imamura. A comparative study of alternative facs coding algorithms. Technical Report CMU-RI-TR-02-06, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, November 2001.

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

[8] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski. Classifying facial actions. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 21(10):974–989, Oct. 1999.

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

[10] I. Essa and A. Pentland. Coding, analysis, interpretation, and recognition of facial expressions. *IEEE Trans. on Pattern Analysis and Machine Intell.*, 19(7):757–763, July 1997.

[11] B. Fasel and J. Luttin. Recognition of asymmetric facial action unit activities and intensities. In *Proceedings of International Conference of Pattern Recognition*, 2000.

[12] G. L. Ford. Fully automatic coding of basic expressions from video. Technical Report INC-MPLab-TR-2002.03, Machine Perception Lab, Institute for Neural Computation, University of California, San Diego, San Diego, CA, September 2002.

[13] M. Frank and P. Ekman. The ability to detect deceit generalizes across different types of high-stake lies. *Personality and Social Psychology*, 72:1429–1439, 1997.

[14] H. Fujiyoshi and A. Lipton. Real-time human motion analysis by image skeletonization. In *proc. of the Workshop on Application of Computer Vision*, 1998.

[15] B. Heisele, T. Serre, M. Pontil, and T. Poggio. Component-based face detection. In *Proc. IEEE Conf. on Computer Vision and Pattern Recogn. (CVPR)*, 2001.

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

[17] S. Z. Li and L. Gu. Real-time multi-view face detection, tracking, pose estimation, alignment, and recognition. In *IEEE Conf. on Computer Visioin and Pattern Recognition Demo Summary*, 2001.

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

[19] M. Lyons, S. Akamasku, M. Kamachi, and J. Gyoba. Coding facial expressions with gabor wavelets. In *Proceedings of International Conference on Face and Gesture Recognition*, 1998.

[20] T. Moriyama, T. Kanade, J. Cohn, J. Xiao, Z. Ambadar, J. Gao, and M. Imanura. Automatic recognition of eye blinking in spontaneously occurring behavior. In *Proceedings of the 16th International Conference on Pattern Recognition (ICPR '2002)*, volume 4, pages 78 – 81, August 2002.

[21] M. Pantic and L. J. M. Rothkrantz. Expert system for automatic analysis of facial expression. *Image and Vision Computing*, 18(11):881–905, 2000.

[22] A. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 84–91, 1994.

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

[24] E. Smith, M. S. Bartlett, and J. Movellan. Computer recognition of facial action: A study of co-articulation effects. In *Proceedings of 8th Joint Symposium on Neural Computation*, May, 2001.

[25] K. K. Sung and T. Poggio. Example-based learning for view-based human face detection. *IEEE Transactions On Pattern Analysis and Machine intelligence*, 20(1):39–51, January 1998.

[26] Y. Tian, L. Brown, A. Hampapur, S. Pankanti, A. W. Senior, and R. M. Bolle. Real world real-time automatic recognition of facial expressions. In *Proceedings of IEEE workshop on performance evaluation of tracking and surveillance, Graz, Austria*, March, 2003.

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

[28] Y. Tian, T. Kanade, and J. Cohn. Evaluation of gabor-wavelet-based facial action unit recognition in image sequences of increasing complexity. In *Proceedings of the 5th IEEE International Conference on Automatic Face and Gesture Recognition (FG'02), DC*, May, 2002.

[29] P. Viola and M. Jones. Robust real-time object detection. In *International Workshop on Statistical and Computational Theories of Vision - Modeling, Learning, Computing, and Sampling*, 2001.

[30] Z. Wen and T. S. Huang. Capturing subtle facial motions in 3d face tracking. In *Proc. Of Int. Conf. On Computer Vision*, 2003.

[31] J. Xiao, T. Kanade, and J. F. Cohn. Robust full motion recovery of head by dynamic templates and re-registration techniques. In *Proceedings of International Conference on Automatic Face and Gesture Recognition*, pages 163–169, 2002.

[32] Y. Yacoob and L. Davis. Recognizing human facial expression from long image sequences using optical flow. *IEEE Trans. on Pattern Analysis and Machine Intell.*, 18(6):636–642, June 1996.

[33] J. Yang, R. Stiefelhagen, U. Meier, and A. Waibel. Real-time face and facial feature tracking and applications. In *Proceedings of Auditory-Visual Speech Processing (AVSP 98), New South Wales, Australia*, 1998.