

Understanding Effects of Image Resolution for Facial Expression Analysis

YingLi Tian, *Senior Member, IEEE* and Shizhi Chen, *Student Member, IEEE*

Abstract—Recognizing spontaneous expressions in real environments often works on face images captured in lower resolutions or with larger variation of head poses. In this paper, we analyze the effects of face image resolution for performance of automatic facial expression analysis (AFEA) systems. Different approaches are compared for face acquisition, facial feature extraction and representation, and facial expression recognition. Our evaluations are conducted at five different resolutions of the head region on two databases: 1) Cohn-Kanade expression database; and 2) FABO database. There is no obvious difference in the performance of expression analysis when the head region resolution is higher than 72x96 pixels. The performance of AFEA is acceptable for head regions with a resolution of 36x48. For head regions with lower resolutions than 36x48, the performance decreases quickly and AFEA systems become impractical.

Index Terms— Facial expression analysis, face acquisition, feature extraction, low resolution, action units, six basic expressions.

I. INTRODUCTION

Recent advances in algorithms, sensors, and embedded computing hold the promise to enable computer vision technology of facial expression recognition that can address real-life applications such as smile detections in cameras or phones [31] and intelligent tutoring systems [19]. In these real applications, either the face images are often captured in a lower resolution with a larger variation of head poses, or the face images must be down-sampled to a lower resolution due to the limited processing power. However, most existing facial expression analysis (AFEA) systems attempt to recognize facial expressions from data collected in a highly controlled environment with high resolution frontal faces (face regions greater than 200 x 200 pixels) [6,7, 10, 12, 13, 15, 20, 27].

While many recent advances and successes in automatic facial expression analysis have been achieved, 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. Recently, some researchers attempt to recognize spontaneous expressions in real-life environments [1, 22, 26, 31]. Surveys of facial expression analysis can be found in papers [14, 23, 28, 34]. While papers [14], [23], and [28] summarize the AFEA methods before year 2004, paper [34] describes the most recent AFEA systems (before 2008) on multi-modal and spontaneous expressions. In this paper, we focus on understanding the effects of different levels of image resolution at each step of general AFEA systems.

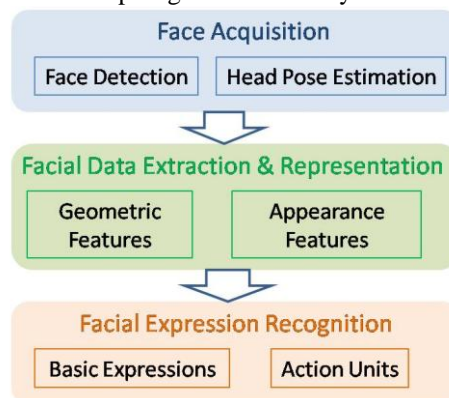


Figure 1: Three basic steps for general automatic facial expression analysis systems.

Facial expression analysis includes both measurement of facial motions and recognition of expressions. The general approach to Automatic Facial Expression Analysis (AFEA) systems consists of 3 steps as shown in 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 video sequences. It can be a face detector that detects a face in each frame or just to detect a face in the first frame and then track the face in the remainder of the video sequence [15, 21]. In addition to face detection, head finding, head tracking and pose estimation can be applied to handle large head motions for facial expression analysis systems [1, 26, 32].

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

Manuscript received January 5, 2012. This work was supported in part by NSF grant IIS-0957016, EFRI-1137172, NOAA CREST Grant NA11SEC4810004, and DHS Summer Research Team Program for Minority Serving Institutions Follow-on Award.

YingLi Tian is with the City College, City University of New York, New York, NY 10031 USA (phone: 212-650-7046; fax: 212-650-8249; e-mail: ytian@ccny.cuny.edu). Prior to joining the City College in September 2008, she was with IBM T.J. Watson Research Center, Yorktown Heights, NY 10598 USA.








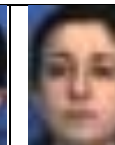
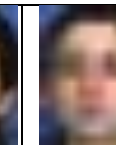
Shizhi Chen is with the City College, City University of New York, New York, NY 10031 USA (e-mail: schen21@ccny.cuny.edu).

methods and appearance feature-based methods. The geometric facial features represent 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 facial features. 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.

Facial expression recognition is the third step of AFEA systems to identify subtle changes in one or a few discrete facial features as facial action coding system (FACS) action units (AUs) [11] or a small set of prototypic emotional expressions (i.e., disgust, fear, joy, surprise, sadness, anger).

We made the first attempt to recognize facial expressions in compressed images with lower resolution (the face region is around 50x70 to 75x100 pixels) [26]. 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 I: EXAMPLES OF FACE IMAGES IN FIVE DIFFERENT RESOLUTIONS FROM THE COHN-KANADE DATABASE [18] AND THE FABO DATABASE [16]. THE HEAD REGION IS CROPPED FOR DISPLAY PURPOSE ONLY. THE LOWER RESOLUTION IMAGES ARE DOWN-SAMPLED FROM THE ORIGINALS.

Example images from Cohn-Kanade Database				
				
288x384	144x192	72x96	36x48	18x24
Example images from FABO Database				
				
205x260	103x130	52x65	26x33	13x16

In this paper, we investigate the effects of different image resolutions for each step of facial expression analysis. As shown in Table I, a total of five different resolutions of the head region are studied based on the Cohn-Kanade AU-Coded Face Expression Image Database [18] and the FABO database [16]. The lower resolution images are down-sampled from the originals. For the face acquisition step, we evaluate face detection [30] and head detection and pose estimation [26] at different resolution levels (Section II). The effects of resolutions on the extraction of both geometric features and appearance features are investigated in the feature extraction

and representation step (Section III). For the step of expression recognition, both emotional-specific expressions and selected action units recognition are investigated for different face resolutions (Section IV).

Section V presents the experiment setup and evaluation results on two public available databases. Section VI presents conclusions and discussion.

II. 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. In this paper, we evaluate methods of face detection and head localization and head pose estimation.

A. Face Detection

Many face detection methods have been developed to detect faces in an arbitrary scene. In this paper, we employ the widely used face detector of Viola and Jones based on a set of Harr-like features with integral image approach to detect frontal and near-frontal views of faces for different image resolutions [30]. Details of the face detector can be found in papers [30].

B. Head Localization and Head Pose Estimation

In order to handle the full range of head motion for expression analysis in real environments, we evaluate head detection and head pose estimation for different image resolutions. For spontaneous expression analysis in real world environments, it is sufficient to estimate the coarse head pose to obtain frontal or near-frontal view faces. No expression analysis is needed for the other head poses. In most real world applications, only the coarse head pose can be estimated because of the limitations of the quality and resolution of the incoming data. A survey for head pose estimation in computer vision can be found in paper [21]. In this paper, we evaluate our method of head pose estimation [26] for facial expression analysis.

For videos with large head motions, face detection becomes unreliable. Instead of detecting faces, we develop a silhouette based method to detect the head [26]. The head detection uses the smoothed silhouette of the foreground object as segmented using background subtraction. Based on human intuition about the parts of an object, a segmentation into parts generally occurs at the *negative curvature minima* (NCM) points of the silhouette [17] as shown with small circles in Figure 2. The boundaries between parts are called cuts (shown as the line L in Figure 2). Singh *et al.* [25] noted that human vision prefers the partitioning scheme which uses the shortest cuts. They

proposed a shortcut rule which requires a cut: 1) be a straight line, 2) cross an axis of local symmetry, 3) join two points on the outline of a silhouette and at least one of the two points is NCM, 4) be the shortest one if there are several possible competing cuts. To obtain the correct head region, we first calculate the head width W , then the head height H is enlarged to $\alpha \times W$ from the top of the head. In our system, $\alpha = 1.4$.

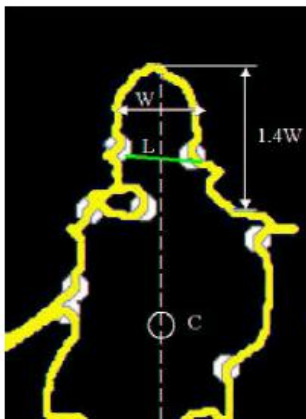


Figure 2. Head localization by calculating the cut of the head part.

After the head is located, the head image is converted to gray scale, histogram equalized and resized to 32×32 . Then the processed head images are the input for a head pose estimation classifier with the outputs as the head poses. We classify head poses to 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 face. In the frontal or near frontal view, both eyes and lip corners are visible. In the 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 of head pose detection can be found in paper [26].

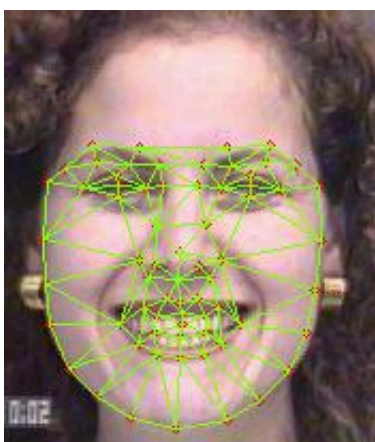


Figure 3. Facial feature tracking by active appearance models (AAMs).

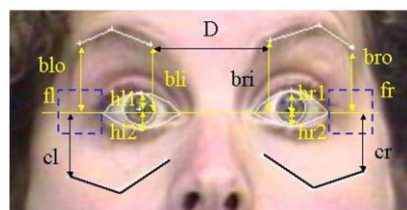
III. FACIAL FEATURE EXTRACTION AND REPRESENTATION

In this study, we evaluate two types of features: geometric features and appearance features. Geometric features present the shape and locations of facial components (including

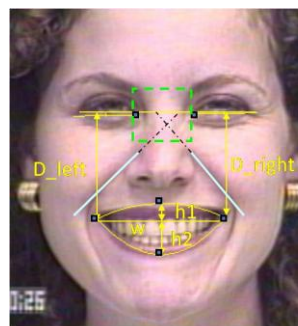
mouth, eyes, brows, nose etc.). The facial components or facial feature points are extracted to form a feature vector that represents the face geometry. The appearance features present the appearance (skin texture) changes of the face such as wrinkles and furrows. The appearance features can be extracted on either the whole face or specific regions in a face image. For appearance feature extraction and representation, we study two different methods: 1) Gabor wavelets based method; and 2) Motion History Image HOG (MHI-HOG) based method. Both methods can be applied to images at different resolution levels.

A. Tracking based Geometric Feature Extraction and Representation (G1)

Active appearance models (AAMs) [8] are generative parametric models commonly used to track facial features in video sequences. In our system, the region of the face and approximate location of individual face components are automatically tracked in the image sequence by performing AAMs to extract 68 facial points as shown in Figure 3.



(a) Upper Face Feature Representation



(b) Lower Face Feature Representation

Figure 4. Tracking based geometric facial feature representation for upper face (a) and lower face (b) respectively (G1).

Based on the locations of these facial points, the extracted features are transformed into a set of parameters based on the inner corners of the eyes which are most reliably detected and their relative position is unaffected by muscle contraction [27]. As shown in Figure 4(a), we represent the upper face features by 15 parameters. Of these, 12 parameters describe the motion and shape of the eyes, brows, and cheeks, 2 parameters describe the state of the crows-feet wrinkles, and 1 parameter describes the distance between the brows. Similarly, we use 9 parameters to represent the lower face features as shown in Figure 4(b). Of these, 6 parameters describe lip shape, state

and motion, and 3 describe the furrows in the nasolabial and nasal root regions. 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 initial frame.

B. Frame based Geometric Feature Extraction and Representation (G2) for Images with Low Resolutions

The above method of extraction of detailed geometric facial features will fail when the resolution of face images decrease. 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).

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. 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 by using an integral projection of the mouth region. Finally, the horizontal borders of the line between the lips are found, using an integral projection over an edge-image of the mouth. 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.

For the geometric features estimated by feature detection, we represent the face location features by 5 parameters: 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 with neutral expression.

C. Gabor Wavelets based Appearance Feature Extraction and Representation (AP1)

We use Gabor wavelets to extract the facial appearance changes as a set of multi-scale and multi-orientation coefficients. 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.

To remove position noises for appearance feature extraction, the faces are normalized to a fixed distance between the centers 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 lighting changes, the brightness of the face images are linearly rescaled to [0, 255].

D. MHI-HOG and Image-HOG based Motion and Appearance Feature Extraction and Representation (AP2)

To extract appearance features, we propose a new feature descriptor -- MHI-HOG [5, 29], and combining with Image-HOG [9] to capture motion and appearance information of expressions. Here, Image-HOG indicates Histogram of Oriented Gradients (HOG) on the original image (Image-HOG) and MHI-HOG stands for Histogram of Oriented Gradients (HOG) on the Motion History Image (MHI) [3]. MHI-HOG captures motion direction of an interest point as an expression evolves over time. Image-HOG captures the appearance information of the corresponding interesting point. By combining MHI-HOG and Image-HOG, we achieve comparable performance for facial expression recognition with the state of the art on the FABO database [16].

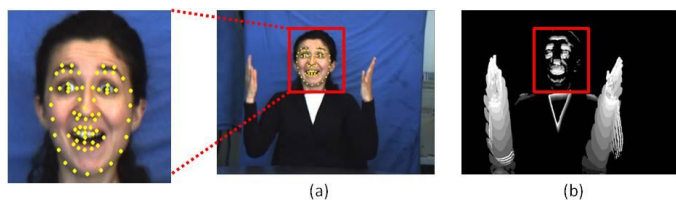


Figure 5. MHI-HOG and Image-HOG based Motion and Appearance Feature Extraction and Representation (AP2). (a) AAM facial landmark points tracking; (b) MHI Image

We first track the facial landmark points using the AAM model [8] as shown in Figure 5(a). The total number of landmark points we use in our system is 53, excluding the face boundary points. Then we extract the Image-HOG and MHI-HOG descriptors of the selected facial landmark points. As shown in Figure 5(b), the MHI image captures motion information of the facial landmark points, while the original image can provide the corresponding appearance information. For each landmark point, the feature dimension of the Image-HOG and MHI-HOG descriptors is 54 and 72 respectively. After concatenating the Image-HOG and MHI-HOG descriptors of all 53 facial landmark points on each frame, the resulted feature vector has 6678 dimensions for each frame. The feature dimension of 6678 is too large for a classifier. Therefore we reduce the feature dimensions down to 80 by applying PCA on both the Image-HOG and the MHI-HOG descriptors respectively. The principal components of the Image-HOG and MHI-HOG are obtained from the training videos. The feature vector with the reduced dimensions is the input for the classifier of the expression recognition.

IV. FACIAL EXPRESSION RECOGNITION

In our study, we investigate expression recognition at different resolution levels: FACS AUs and six basic expressions on the Cohn-Kanade databases, and ten specific expressions including six basic expressions and four more

expressions of “anxiety”, “boredom”, “puzzlement” and “uncertainty” on the FABO database.

A. Expression Recognition on the Cohn-Kanade Database

For evaluations on the Cohn-Kanade database, we compare the recognition accuracy for tracking based detailed geometric features (G1), frame based simple geometric features (G2), Gabor wavelets based appearance features (AP1), and the combinations of geometric and appearance features (G1+AP1 and G2+AP1) by using a neural network-based recognizer to recognize FACS AUs and basic expressions.

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. For AU recognition, we trained one network by using the geometric features alone and one network by using only Gabor wavelets. For using both geometric features and appearance features, these two networks are applied in concert. 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 recognize 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 are happiness, surprise, fear, sadness, disgust, and anger.

B. Expression Recognition on the FABO Database

For evaluations on the FABO database, we employ support vector machines (SVM) with the RBF kernel using one vs. one approach as our multi-class classifier [4]. SVM is to find a set of hyper-planes, which separate each pair of classes of data with maximum margin, then use maximum vote to predict an unknown data’s class. In our experiments, the feature data, i.e. the input features to the SVM, are the feature vector with PCA reduced dimensions on both Image-HOG and MHI-HOG.

For facial expression recognition, we recognize a total of 10 expressions including both basic and non-basic expressions. Basic expressions are happiness, surprise, fear, sadness, disgust, and anger. Non-basic expressions are anxiety, boredom, puzzlement, and uncertainty.

V. EFFECTS OF IMAGE RESOLUTION FOR FACIAL EXPRESSION ANALYSIS

A. Experimental Results on the Cohn-Kanade Database

The DFAT subset of Cohn-Kanade expression database [18] 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 8-bit gray-scale values. The length of the image sequences are varying from 9 to 47 frames. The size of the head region is about 280x380 pixels. More details about the database can be found at paper [18]. Table II summarizes the effects of different resolutions of face images for each step of expression analysis on Cohn-Kanade expression database.

TABLE II: SUMMARY OF THE EFFECTS OF FACES AT DIFFERENT RESOLUTIONS FOR EXPRESSION ANALYSIS ON COHN-KANADE DATABASE [18]. FOR FACE ACQUISITION, "FD" INDICATES FACE DETECTOR. "HPE" INDICATES HEAD POSE ESTIMATION. FOR FEATURE EXTRACTION, "G1" INDICATES GEOMETRIC FEATURES EXTRACTED BY AAMS BASED FEATURE TRACKING. "G2" INDICATES GEOMETRIC FEATURES EXTRACTED BY FRAME BASED FEATURE DETECTION. "AP1" INDICATES APPEARANCE FEATURES EXTRACTED BY GABOR WAVELETS.

Face Resolution					
Face Acquisition					
FD	100%	100%	100%	100%	No
HPE	98.5%	98%	98.2%	97.8%	98%
Facial Feature Extraction					
G1	Yes	Yes	Yes	No	No
G2	Yes	Yes	Yes	Yes	No
AP1	Yes	Yes	Yes	Yes	Yes
FACS AU Recognition					
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+AP1	92.8%	93%	92.2%	N/A	N/A
G2+AP1	91.2%	90.8%	90%	87.7%	N/A
Basic Expression Recognition					
G1	92.5%	91.8%	91.6%	N/A	N/A
G2	74%	73.8%	72.9%	61.3%	N/A
AP1	91.7%	92.2%	91.6%	77.6%	68.2%
G1+AP1	93.8%	94%	93.5%	N/A	N/A
G2+AP1	93.2%	93%	92.8%	89%	N/A

B. Experimental Results on the FABO Database




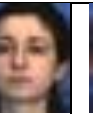

The FABO database was collected by Gunes and Piccardi [16]. The FABO database contains videos of face and body expressions recorded by the face and body cameras, simultaneously. In our experiments, we only use the videos from the body camera, which contains both face and body gesture information. The size of head region is about 200x260 pixels. More details about the database can be found at paper [16].

In our evaluation, we select 284 videos with the same expression labels from both face and body gesture and including both basic and non-basic expressions (i.e. anxiety, boredom, puzzlement, and uncertainty). Each video contains 2 to 4 expression cycles. Videos in each expression category are randomly separated into three subsets. Two of them are

chosen as training data. The remaining subset is used as testing data. No same video appears for both training and testing, but the same subject may appear in both training and testing sets due to the random separation process. Three-fold cross validation is performed over all experiments. The average performances are reported in the paper.

Table III summarizes the effects of different resolutions of face images for each step of expression analysis on the FABO database.

TABLE III: SUMMARY OF THE EFFECTS OF FACES AT DIFFERENT RESOLUTIONS FOR EXPRESSION ANALYSIS ON FABO DATABASE [16]. FOR FACE ACQUISITION, "FD" INDICATES FACE DETECTOR. FOR FEATURE EXTRACTION, "AAM" INDICATES AAMS BASED FEATURE TRACKING. "AP2" INDICATES MHI-HOG AND IMAGE-HOG BASED MOTION AND APPEARANCE FEATURES EXTRACTION. THE RECOGNITION RESULTS WITH "*" INDICATES THAT THE "AP2" FEATURES ARE EXTRACTED ON THE POINTS CORRESPONDING TO THE SAME LOCATIONS OF AAM ON THE ORIGINAL RESOLUTION IMAGES.

Face Resolution					
	205x260 (Original)	103x130	52x65	26x33	13x16
Face Acquisition					
FD	100%	100%	100%	100%	No
Facial Feature Extraction					
AAM	Yes	Yes	Yes	No	No
AP2	Yes	Yes	Yes	Yes	Yes
Recognition of Ten Specific Expressions					
AP2	66.53%	66.17%	65.53%	52.83%*	45.6%*

VI. DISCUSSION AND CONCLUSION

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 based on the Cohn-Kanade AU-Coded Face Expression Image Database (288x384, 144x192, 72x96, 36x48, and 18x24) [18] and the Bimodal Face and Body Gesture FABO Database (205x260, 103x130, 52x65, 26x33, and 13x16) [16].

Our empirical studies illustrated following conclusions: (1) Head detection and head pose estimation can detect faces in lower resolutions than face detector. (2) Appearance feature extraction needs face alignment. (3) There is no obvious difference in the performance of expression analysis when the head region resolution is 72x96 or higher. Geometric features and appearance features achieve the same level of recognition rates for both FACS AUs and six basic expressions. (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 objective of this paper is to help us to understand the 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 modeling of spontaneous head motion and action unit recognition in spontaneous facial behavior is exciting developments [1, 24]. 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 [20], but higher parametric generic, or even custom-fitted head models, may prove necessary for more complete action unit recognition.

ACKNOWLEDGMENTS

The author thanks Prof. Jeffray Cohn for providing Cohn-Kanade AU-Coded Face Expression Image Database, and Dr. Hatice Gunes for providing the FABO Database.

REFERENCES

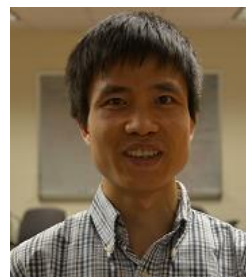
- [1] M. S. Bartlett, G. Littlewort, M. G. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Fully automatic facial action recognition in spontaneous behavior," *Journal of Multimedia*, no. 6, pp. 22–35, 2006.
- [2] S. Basu, I. Essa, and A. Pentland, "Motion Regularization for Model-Based Head Tracking," in *13th Int'l Conf on Pattern Recognition*. Austria, Vienna, August 25-30, 1996.
- [3] A. Bobick and J. Davis, The recognition of human movement using temporal templates. *IEEE Trans. Pattern Anal. Mach. Intell.* 23, 257–267, 2001.
- [4] C. Chang and C. Lin, LIBSVM : a library for support vector machines, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [5] S. Chen, Y. Tian, Q. Liu, and D. Metaxas, Recognizing Expressions from Face and Body *Gesture by Temporal Normalized Motion and Appearance Features*, Fourth IEEE Workshop on Human Communicative Behaviour Analysis, held in conjunction with CVPR, 2011.
- [6] 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.
- [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] T.F. Cootes, G.J. Edwards, and C.J. Taylor. Active appearance models. *PAMI*, 23(6):681–685, June 2001.
- [9] N. Dalal, B. Triggs, "Histogram of Oriented Gradients for Human Detection", *CVPR* 2005.
- [10] 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.
- [11] 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.
- [12] 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.

- [13] B. Fasel and J. Luttin. Recognition of asymmetric facial action unit activities and intensities. In *Proceedings of International Conference of Pattern Recognition*, 2000.
- [14] B. Fasel and J. Luetin, "Automatic Facial Expression Analysis: A Survey," *Pattern Recognition*, Vol. 36, pp. 259-275, 2003.
- [15] H. Gunes and M. Piccardi, Automatic Temporal Segment Detection and Affect Recognition from Face and Body Display, *IEEE Transactions on Systems, Man, and Cybernetics-Part B*, Special Issue on Human Computing, Vol. 39, No. 1, pp. 64-84, Feb. 2009.
- [16] H. Gunes and M. Piccardi, "A bimodal face and body gesture database for automatic analysis of human nonverbal affective behavior", International Conference Pattern Recognition, 2006.
- [17] D.D. Hoffman and W.A. Richards. Parts of recognition. *Cognition*, 18:65-96, 1984
- [18] 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.
- [19] S. D'Mello, R. W. Picard, and A. Graesser, "Toward an affect-sensitive autotutor," *IEEE Intelligent Systems*, vol. 22, pp. 53-61, 2007.
- [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] E. Murphy-Chutorian and M. M. Trivedi, Head Pose Estimation in Computer Vision: A Survey, *IEEE Trans. On Pattern Analysis and Machine Intelligence*, Vol. 31 no. 4, 2009. pp. 607-626
- [22] M. A. Nicolaou, H. Gunes and M. Pantic, Audio-visual Classification and Fusion of Spontaneous Affective Data in Likelihood Space, Proc. of ICPR, the 20th IAPR Int. Conf. on Pattern Recognition, 23-26 Aug. 2010, Istanbul, Turkey, pp. 3695-3699.
- [23] M. Pantic and L. J. M. Rothkrantz, "Automatic Analysis of Facial Expressions: The State of the Art," *IEEE Trans. On Pattern Analysis and Machine Intelligence*, Vol. 22, No. 12, pp. 1424-1445, 2000.
- [24] T. Simon, M. H. Nguyen, F. De la Torre and J. F. Cohn. Action Unit Detection with Segment-based SVMs. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.
- [25] M. Singh, G. D. Seyranian, and D.D Hoffman. Parsing silhouettes: the short-cut rule. *Perception and Psychophysics*, 61(4):636-660, May 1999.
- [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. F. Cohn, "Facial Expression Analysis" In book *Handbook of Face Recognition*, Edited by Stan Li and A.K. Jain, Springer-Verlag January, 2004.
- [29] Y. Tian, L. Cao, Z. Liu, and Z. Zhang, "Hierarchical Filtered Motion for Action Recognition in Crowded Videos", *IEEE Transactions on Systems, Man, and Cybernetics--Part C: Applications and Reviews*, 2011.
- [30] 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.
- [31] J. Whitehill, G. Littlewort, I. Fasel, M. Bartlett, and J. Movellan, "Toward Practical Smile Detection", *IEEE Trans. on Pattern Analysis and Machine Intell.*, VOL. 31, NO. 11, Nov. 2009.
- [32] J. Xiao, S. Baker, I. Matthews, and T. Kanade, Real-Time Combined 2D+3D Active Appearance Models, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June, 2004, pp. 535 - 542.
- [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.
- [34] Zeng, Z., Pantic, M., Roisman, G.I., and Huang, T.S., A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 31, No. 1, 2009.



YingLi Tian (M'99-SM'01) received her BS and MS from TianJin University, China in 1987 and 1990 and her PhD from the Chinese University of Hong Kong, Hong Kong, in 1996. After holding a faculty position at National Laboratory of Pattern Recognition, Chinese Academy of Sciences, Beijing, she joined Carnegie Mellon University in 1998, where she was a postdoctoral fellow at the Robotics Institute. Then she worked as a research staff member in IBM T. J. Watson Research Center from 2001 to 2008. She is one of the inventors of the IBM Smart Surveillance Solutions.

She is currently an associate professor in Department of Electrical Engineering at the City College of New York. Her current research focuses on a wide range of computer vision problems from motion detection and analysis, to human identification, facial expression analysis, and video surveillance. She is a senior member of IEEE.



Shizhi Chen (S'11) is a Phd student in the Department of Electrical Engineering at the City College of New York. His research interests include facial expression recognition, scene understanding, machine learning and related applications. He received the BS degree of Electrical Engineering from SUNY Binghamton, New York in 2004, and the MS degree of Electrical Engineering and Computer Science from UC Berkeley, California in 2006. From 2006 to 2009, he worked as an engineer in several companies including Altera, Supertex Inc., and US Patent and Trademark Office. He is a member of Eta Kappa Nu (electrical engineering honor society), and a member of Tau Beta Pi (engineering honor society). He also received numerous scholarships and fellowships, including Beat the Odds scholarship, Achievement Rewards for College Scientists (ARCS) Fellowship, and NOAA CREST Fellowship.