

# Robust Symbolic Dual-View Facial Expression Recognition With Skin Wrinkles: Local Versus Global Approach

Yizhen Huang, Ying Li, *Senior Member, IEEE*, and Na Fan

**Abstract**—Simple cartoon facial expressions can be represented by emoticons, that is, a special sequence of symbols. This inspires us that a sketch of facial feature contour may be adequate to recognize expressions. Metrics of such sketches are easier to be calibrated under varying illumination and head pose. While skin wrinkles such as nasolabial folds, eye pouches, dimples, forehead, and chin furrows are not salient facial features, they may convey crucial subtle signals about an individual's emotion. Our experiments have shown that the side-view profile plus skin wrinkles can correctly differentiate nearly 70% expressions, and it contributes to the increase of overall recognition rate. Finally, we compare the accuracy and robustness of various local and global processing schemes, especially under the condition of partial occlusion.

**Index Terms**—Expression intensities, facial expression recognition, side-view profile, skin wrinkles.

## I. INTRODUCTION

OVER the last two decades, face recognition has been a heavily studied topic in computer vision (refer to [1] for a recent survey). Paralleled to this, facial expression recognition has also become an active research area [2]–[5] with potential applications in human-computer interaction and cognition of human emotion. Despite the fact that the recognition of face and facial expression are two independent fields, they facilitate the development of each other as they share many of their methodologies and even theories. For instance, facial feature extraction is commonly applied in both areas. Also, as there is a facial expression space corresponding to each face space, their classifiers are likely to be comparable. Finally, it is also a research direction to track [6] and recognize [7], [8] faces in presence of expression.

Multi-view face recognition methods have been proposed and developed for over a decade [9]–[11]. Many of them use different variations of linear discriminant analysis (LDA), independent component analysis (ICA), or principal component

analysis (PCA) to project faces into subspaces, then categorize faces with support vector machines (SVMs) or neural networks (NNs). In most cases, these methods treat information from multiple views as a challenge as they have to handle faces with a single non-frontal view.

On the other hand, research on facial expression recognition using multiple views is still a virgin territory. This inspires us to use an additional side view to improve recognition accuracy where recognition of one particular expression would need a pair of simultaneously captured photos of the face. The reason for such research proposal is twofold. First, it is difficult to extract facial expression cues from non-frontal views, which explains the lack of prior work on multi-view expression recognition; second, expression recognition is not as mature as face recognition; thus, recognition rate is still the first priority instead of the applicability.

Scientific study of facial expressions can date back to the seminal work done by Ekman *et al.* [12], [13], who analyzed six fundamental facial expressions and encoded them into the so-called facial action coding system (FACS). FACS enumerates action units (AUs) of a face that cause facial movements. FACS and AUs are still frequently adopted in recent literature [24]–[26]; nevertheless, FACS has different versions [14] and lacks the expressive power to describe different variations of possible facial expressions [15].

Padgett *et al.* [16] exploited neural models to capture and encode nonlinear mappings among different expressions. The original intention for people who invented NN is to mimic the rationale of biological neurons. It has been proved to be useful in many applications of the cognitive science [17]. Neural networks have been reported to have biological explanation corresponding to the human vision system due to their intrinsic flexibility and have been successfully applied for intelligent recognition system [18]–[20].

In the last decade, the field of expression analysis has grown immensely (see [5] for a most recent survey). Pantic *et al.* [24], [25] used computer vision techniques such as edge detection, active contours, and NNs, combined with rule-based forward reasoning, to recognize AUs from static frontal- and/or profile-view face images. Its fiducial points were extracted from sampled contours that may be sensitive to the change of illumination, head pose, or background. Tian *et al.* [26] regarded facial wrinkles as transient facial features and approximated them with one or two line segments.

In this work, we propose to use a more robust active wavelet network (AWN) [27] and the deformable template model

Manuscript received December 12, 2009; revised March 25, 2010; accepted May 14, 2010. Date of current version September 15, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Nadia Magnenat-Thalmann.

Y. Huang is with the Computer Sciences Department, University of Wisconsin-Madison, Madison, WI 53706 USA (e-mail: huang.yizhen@gmail.com).

Y. Li is with the IBM T. J. Watson Research Center, Hawthorne, NY 10532 USA (e-mail: yingli@us.ibm.com).

N. Fan is with the Department of Electronic Engineering, East China Normal University, Shanghai 200241, China (e-mail: fanna.cn@gmail.com).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMM.2010.2052792

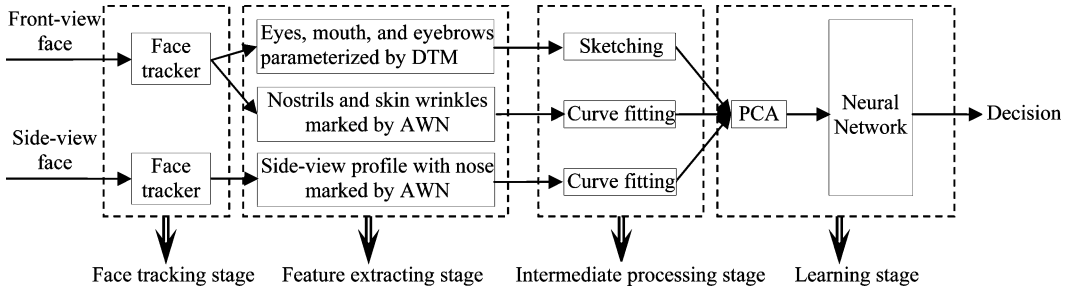


Fig. 1. Overall architecture using global approach.

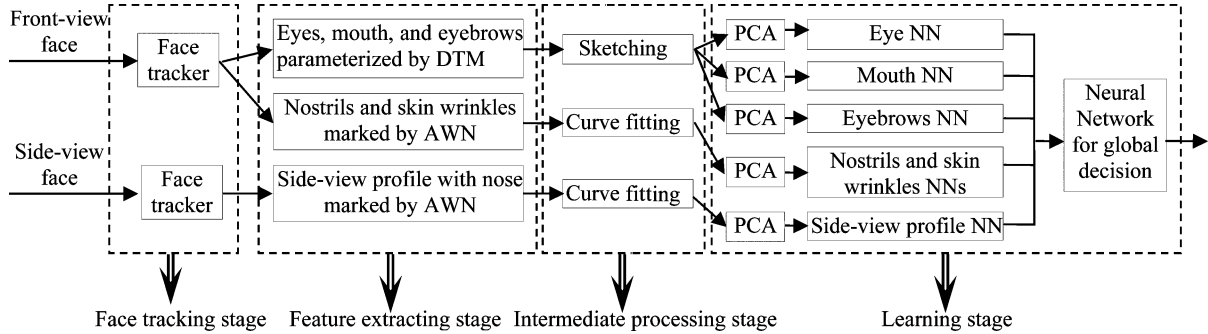


Fig. 2. Overall architecture using local approach.

(DTM) [28] to model facial features, from which some stable measurements such as fitted polynomial curves, area, and aspect ratio are taken. Figs. 1 and 2 illustrate the overall architecture of our system using global and local approaches, respectively. Specifically, the first three stages of both approaches are the same; yet in the last stage, the local approach sends data to different NNs with respect to each facial feature, and the output of these NNs are fed to another NN for a global decision.

The rest of this paper is arranged as follows: Section II describes the process of face tracking; Section III details how to extract facial features using DTM and AWN; Section IV discusses the metrics to quantify facial feature sketches; and Section V presents the underlying principle of the proposed global and local approaches. We report experimental results in Section VI and draw conclusion and point out future direction in Section VII.

## II. FACE TRACKING

The face tracking component of our system is based on a modified version of the ratio template algorithm (RTA) [29]. RTA, originally described by Sinha [30], operates by matching ratios of averaged luminance using a spatial face model. It is able to detect faces under a range of lighting conditions, although it is ineffective when the subject is illuminated from beneath. Moreover, it is able to handle limited changes in scale, yaw, pitch, and tilt of the head. The strengths of this approach lie in its speed, its ease to implement, and its tolerance to varying light conditions in an unstructured indoor environment. A detailed description of the ratio template approach can be found in [29].

The version of RTA described in [29] is modified in our system by including biological proportions, i.e., the golden ratio, into the face model, and by examining higher order relationships in addition to the original ratio measures.

The information provided by the modified RTA is then combined with simple morphological eye/mouth detection, image motion, and density matching to allocate a single face probability to each location in the scene. As shown in Fig. 4, the morphology-based edge image [31] is robust under various illumination conditions.

Our revision to the ratio template is inspired by the naturally occurring golden ratio. The use of eye/mouth identification is supported by studies on human face detection strategies [32], [35]. According to [32], face detecting cells respond to the relative positions of features within a face and any change to this arrangement of facial features reduces their response. There have also been suggestions that humans possess specialized detectors for eyes [35]. The morphological eye/mouth detection mimics these effects by checking the presence of eyes and mouth in the correct relative positions.

## III. FACIAL FEATURE EXTRACTION

Once a face has been located in the scene by the face tracker module, facial features are extracted for further analysis as follows. First, as eyes, mouth, and eyebrows typically have regular geometric shapes where prior knowledge about faces can be applied, they are parameterized with the DTM [28]. Second, as skin wrinkles, nose, and side-view profile are hard to be modeled geometrically [28], [33], we use AWN [27] to extract landmark points on their boundaries.

The AWN is a new face alignment method, which is robust to illumination variation and partial occlusion with strong generalization ability. The key idea of the AWN is to replace the PCA-based texture model in the ASM [21] with a wavelet network representation. The constituents of the wavelet network are single wavelets and their associated coefficients. Particularly, the odd-Gabor function is adopted as the mother wavelet,

which provides the best trade-off between spatial and frequency resolution [22]. Because of the spatially localized property of wavelets, the AWN outperforms the ASM when there are illumination changes or partial occlusions. To achieve real-time process, landmark points of feature boundaries in the AWN are only updated by searching around the edges that have the first and second strongest strength. This eliminates fairly complex routines in deriving point distribution models (PDMs) and greatly speeds up the updating process of landmarks without inducing error points.

Below we elaborate on the process of extracting various facial features.

- **Eyes, mouth, and eyebrows:** All integrals to calculate potential energy functions in the DTM are done in a discrete way by a loop which enumerates all pixels within the range. The template matching process is to iteratively minimize an energy function with parameter adjustment at each step, which not only ensures a convergence by acting as Lyapunov function but also measures the fitness of the template. The minimization is achieved by steepest descent of the energy function in its parameter space. Prior knowledge of different template locations is incorporated as preprocess for faster convergence. Fig. 3(a) exhibits an example of matched eyes, mouth, and eyebrows.
- **Nostrils:** Because the upper part of the front-view nose usually presents low intensity in the edge image, and rarely change in different expressions, we propose to only mark its lower part near nostrils with 14 points using the AWN. Several AWN-labeled examples of nostrils are shown in Fig. 4(a)–(c).
- **Nasolabial folds:** These are the deep folds which run from the side of the nose to the corner of the mouth. It is one of the first signs of aging, yet younger people with fatty cheek pads may also have prominent nasolabial folds. The implicit emotional messages delivered by these folds are vital clues to recognize facial expressions. In a neutral expression, nasolabial folds are inconspicuous, while with different emotions, their intensities in the edge image could vary a lot. We take this as an indicator and extract five landmark points from each nasolabial fold. An AWN-labeled example is delineated in Fig. 5(a).
- **Eye pouches:** The intensity of an eye pouch in the edge image may be strong in exaggerated expressions while weak at other time. We take this as a cue and extract four landmark points from each eye pouch. Fig. 5(b) is an AWN-labeled example of eye pouches.
- **Dimples, forehead, and chin furrows:** The shape and related PDMs of dimples, forehead, and chin furrows vary a lot and are quite complicated. In the AWN, we assume them to appear as a small circle, with three and two line segments, respectively. After the process of the AWN, all positional information of landmarks is discarded. Only their intensity sum is kept for further analysis.
- **Side-view profile with nose:** As accurately extracting facial features from side view is a tough task, we allocate six and nine landmark points to represent the profile of the side-view nose and mouth with the jaw, respectively. Three additional points lying on the border between the nose

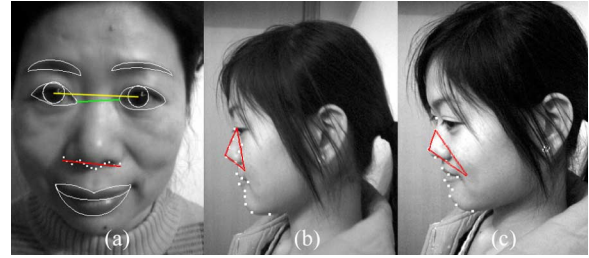


Fig. 3. (a) Example of eyes, mouth, and eyebrows matched by the DTM. (b) and (c) Left-right direction rotation of the head.

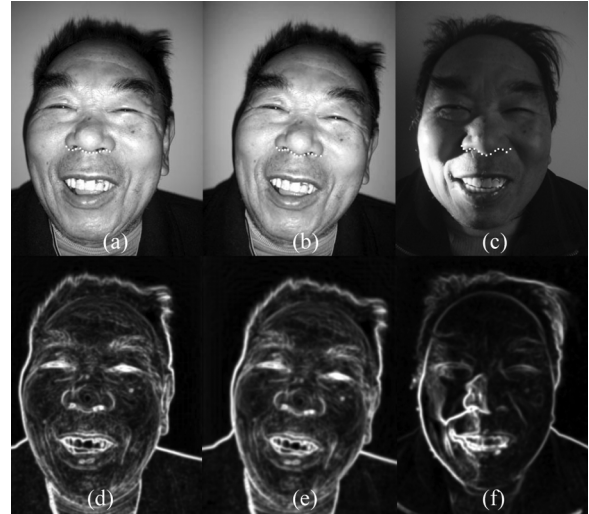


Fig. 4. (a), (b), and (c) Landmark points on nostrils with different head poses under ordinary uniform illumination, stronger uniform illumination, and directional spotlight, respectively. (d), (e), and (f) Morphology-based edge image [31] of (a), (b), and (c), respectively.



Fig. 5. Landmark points and fitted curves of (a) nasolabial folds, (b) eye pouches, and (c) side-view profile with nose.

and the face skin are only collected for feature correction. Three AWN-labeled examples are shown in Figs. 3(b), 3(c), and 5(c).

#### IV. METRICS OF FEATURE SKETCHES

Due to various positions, faces captured at different scenarios are subjected to scaling and rotation that should be corrected. For simplicity, an origin  $(X_c, Y_c)$ , a horizontal scaling ratio  $s_h$ , a vertical scaling ratio  $s_v$ , and a 2-D rotation angle  $\Theta$  are taken into account. Once their values are determined, a simple 2-D linear transformation is carried out for the front and side view separately.

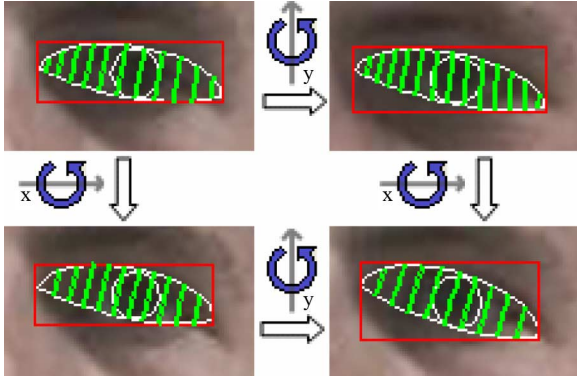


Fig. 6. Matched eye templates rotate  $15^\circ$  along  $x$  and  $y$  axis.

Below we discuss how to calculate these parameters in both front view and side view scenarios.

- 1) In front view, the line connecting the rightmost point of left eye and the leftmost point of right eye is denoted as the eye baseline; the line connecting the pupil centers in each eye is denoted as the pupil baseline. See the green and yellow lines in Fig. 3(a) as an example. Moreover, we fit the landmark points along the side of the nostrils with a straight line using the least square error method, and denote it as the nose baseline. One such example (see the red line) is shown in Fig. 3(a). Our system uses the arithmetic average of the rotation angles of the above three baselines as  $\Theta$ . The average lengths of the eye baseline and the pupil baseline are used to compute the horizontal scaling  $s_h$ . The nose baseline, however, cannot be used to calculate  $s_h$  as it could be elongated during some facial expressions such as laughing. Finally, the  $y$ -offset between the lowest point of the jaw and the nostrils is used to calculate the vertical scaling  $s_v$ .
- 2) In side view, the eye baseline can hardly be seen. Consequently, we interpolate the three edges of the nose triangle, e.g., the red triangle in Fig. 3(b) and (c), as the three baselines, and calculate  $s_h$ ,  $s_v$ , and  $\Theta$  as their average length and average angle with the horizontal axis. In fact, there exist at least two orientations of the head, a left-right direction [see the slight differences among Figs. 3(b), 3(c), and Fig. 5(c)] and an up-down direction. But the cognitive information available today is too few to correct the former orientation.

As head rotates, the parameterized shapes of all facial features from the DTM will be distorted (see Fig. 6). It would be difficult to correct them without accurate head pose estimation and 3-D face reconstruction. However, if the rotation angle is within certain limit, scaling via evident baselines can effectively offset the changes to the area, the aspect ratio, and the bounding box of facial features, which are used as our metrics for facial feature sketches.

Landmark points acquired from the AWN tend to cluster around image patches with large gradient, and can be easily affected by external factors such as camera errors and lighting conditions. Inspired by the usage of polynomial [33] and Bezier [34] curve fitting in face recognition, we fit different groups of landmark points with different types of curves (as tabulated in

TABLE I  
POLYNOMIAL CURVES TO FIT LANDMARK POINTS

Name of facial feature	Polynomial curve
Each nasolabial fold	A quintic curve
Each eye pouch	A quartic curve
Side-view profile of nose	A quintic curve
Side-view profile of mouth with jaw	A sextic curve
Nostrils	A straight line

Table I), so as to capture the essential shapes of facial features and enhance the robustness. Note that, to ensure numerical instability, we have chosen to use piecewise low order curves for approximating side-view profile. Figs. 3(a) and 5 show some illustrative examples.

## V. LOCAL VERSUS GLOBAL APPROACH

We now elaborate our two approaches, namely, local and global, for facial expression recognition. These two approaches mainly differ in the learning stage. We start with the global approach in Sections V-A and V-B, and move on to the local approach in Section V-C.

### A. Principal Components Analysis

During the learning, all parameters that we have extracted from above are aligned to form an  $N - D$  vector. Denote it as  $\beta$ . As  $\beta$  tends to contain data redundancy, we propose to apply PCA as the preprocess. Specifically, assume there are  $L$  samples to form a training set  $\{\beta_1, \beta_2, \dots, \beta_L\}$  with average vector  $\bar{\beta}$ . We further denote the difference vectors as  $\beta_i^* = \beta_i - \bar{\beta}$  ( $1 \leq i \leq L$ ), and the covariance matrix as  $C = [\beta_1^*, \beta_2^*, \dots, \beta_L^*][\beta_1^*, \beta_2^*, \dots, \beta_L^*]^T / (L - 1)$ . We then compute the eigenvalues and eigenvectors of  $C$ , and denote them as  $\{\lambda_1, \lambda_2, \dots, \lambda_N\}$  and  $\{\Xi_1, \Xi_2, \dots, \Xi_N\}$  ( $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$ ), respectively. Next, we choose  $M$  eigenvectors corresponding to the  $M$  largest eigenvalues to form feature vector  $V = [\Xi_1, \Xi_2, \dots, \Xi_M]^T$ . Experiments have shown that  $M = N/2$  would be enough for our application. Finally, we multiply the feature vector  $V$  with  $\beta_i^*$  ( $i = 1, 2, \dots, L$ ), and obtain  $\Gamma_i = V\beta_i^*$ .  $\Gamma_i$  thus has reduced dimensionality, with redundancy removed. This could significantly save computing cost for NNs as detailed in the next section.

### B. Neural-Network Design

Basically we use three-layer feed-forward back propagation (BP) NNs, where the input and hidden layers have  $M$  and  $1.5M$  neurons, respectively. The output layer has 1 neuron representing the degree of difference between training vectors and current input vector. The Sigmoid transfer functions for the hidden and output layers are  $\text{tansig}(x) = 2/(1 + \exp(-2x)) - 1$ , and  $\text{logsig}(x) = 1/(1 + \exp(-x))$ , respectively. We use scaled conjugate gradient method [23] as the training algorithm. The NN takes  $\{\Gamma_1, \Gamma_2, \dots, \Gamma_L\}$  as the training set, and its target value is 0. The value of  $L$  is determined empirically. Experiments show that  $L = 50$  gives a good tradeoff between training complexity and performance.



Fig. 7. Typical samples of happy, sad, scared, angry, surprised, disgusted, arrogant, and bored expressions of three subjects from our database. Note that each expression has two corresponding images with one frontal view and the other side view.

TABLE II  
NNS CONFIGURATION FOR DIFFERENT FACIAL FEATURES

Name	Num. input neurons	Num. hidden layer neurons
NN for 2 eyes	6	15
NN for mouth	4	11
NN for 2 eyebrows	6	13
NN for 2 nasolabial folds	6	13
NN for 2 eye pouches	4	11
NN for all other skin wrinkles	5	9
NN for side-view profile	9	17

TABLE III  
COMPARISON OF RECOGNITION RATE

Method	Recognition rate
Local approach	87.1%
Local approach without sketching & curve fitting	84.9%
Global approach	85.7%
Global approach without sketching & curve fitting	83.5%
Pantic's system [25]	86.3%

Compared to the approach that only a single network is trained with all classes of facial expressions, the multi-network approach used here can achieve better recognition rate. This is because each network is specialized to learn a single class of facial expressions.

Note that we have also tried SVM as the classifier. Nevertheless, its performance is way below than that of NNs as also reported by [24]–[26]. This is because that the highly nonlinear nature of the extracted parameters requires SVM to function at very high dimension in order to capture such correlation.

### C. Local Approach

As most expressions present on face locally, information from a single facial feature is not sufficient to attribute the face to a specific expression. Nevertheless, it can be used to exclude the face from being assigned to some expressions. Based on this observation, principal components of data corresponding to each facial feature from the intermediate processing stage, as described in Section V-C, are sent to separate NN (see Fig. 2). Table II shows the specific setting for each NN. Note that we have tried other NN topologies with similar size, yet only a slight performance difference is observed. Moreover, investigation of the optimal topology is a pure machine learning problem, and is out of the scope of this paper. Throughout this work, we consider the recognition rate as the first priority, and the NN topologies are designed based on the principle that a stable and satisfactory recognition rate is achieved, and further enlargement of the network cannot help to significantly improve the recognition rate.

The NN for global decision (noted as  $NN_g$ ) is a little different from others: If all of  $NN_g$ 's weight values are restricted to be nonnegative,  $NN_g$  can be considered to be a blackbox acting as a monotone multidimensional function, as long as all sigmoid transfer functions in  $NN_g$  are monotone. The larger the output of either local facial feature NN or  $NN_g$  is, the less likely the given expression belongs to the class of expressions trained for it.  $NN_g$  integrates all local likelihoods to a global one, and is no longer an exact numeric matcher for facial characteristics. Therefore, we opt to normalize the inputs to  $NN_g$  w.r.t. their corresponding facial feature intensities, based on the intuition that the more salient a feature is, the more trustable the information from that feature is.

## VI. EXPERIMENTAL RESULTS

We first test various schemes of our approach with the public MMI database at <http://www.mmifacedb.com> and compare it to the Pantic's system [25] (a detailed state-of-the-art functionality and performance comparison could be found in [3]–[5]) in Table III.

To the best of our knowledge, there is no dual-view facial expression database currently available with variation in illumination or head pose. So we have to build our own database by setting up a two-camera system at 30 fps (frame per second) to collect expressions from eight males and 12 females under 800 Lux uniform illumination intensity. In total, eight different expressions have been captured including the six prototypic expressions defined by Ekman *et al.* [12], [13], and two additional ones (arrogant and bored with chest sucked in). Nevertheless, theoretically, our symbolic approach is capable of distinguishing any kind of expression.

The people involved in the experiments are instructed about how to produce the eight expressions, so our database should also be counted as a database of acted expressions. Each person is asked to sit still with neutral expression for 5 s, then act an expression for 5 s, and again sit still with no expression for 5 s. The ten most important principal components of all extracted facial feature parameters from the sample set of the 300 neutral expression frames constructs the PCA space. We define a rough objective measure to quantify expression intensities, which is the Euclidean distance of each of the 150 frames with expressions to its nearest neighbor (i.e., a neutral expression) in the sample set, measured in the PCA space. The 150 expressive frames are graded into six levels based on their ranks in expression intensities, with the 25 first being level-1, the 26th to 50th being level-2, and so forth. Only level-1 frames are qualified in our database. Such recording and screening process is repeated four times to accumulate 100 photo pairs for each expression, among which 50 are used for NN training and the remaining 50

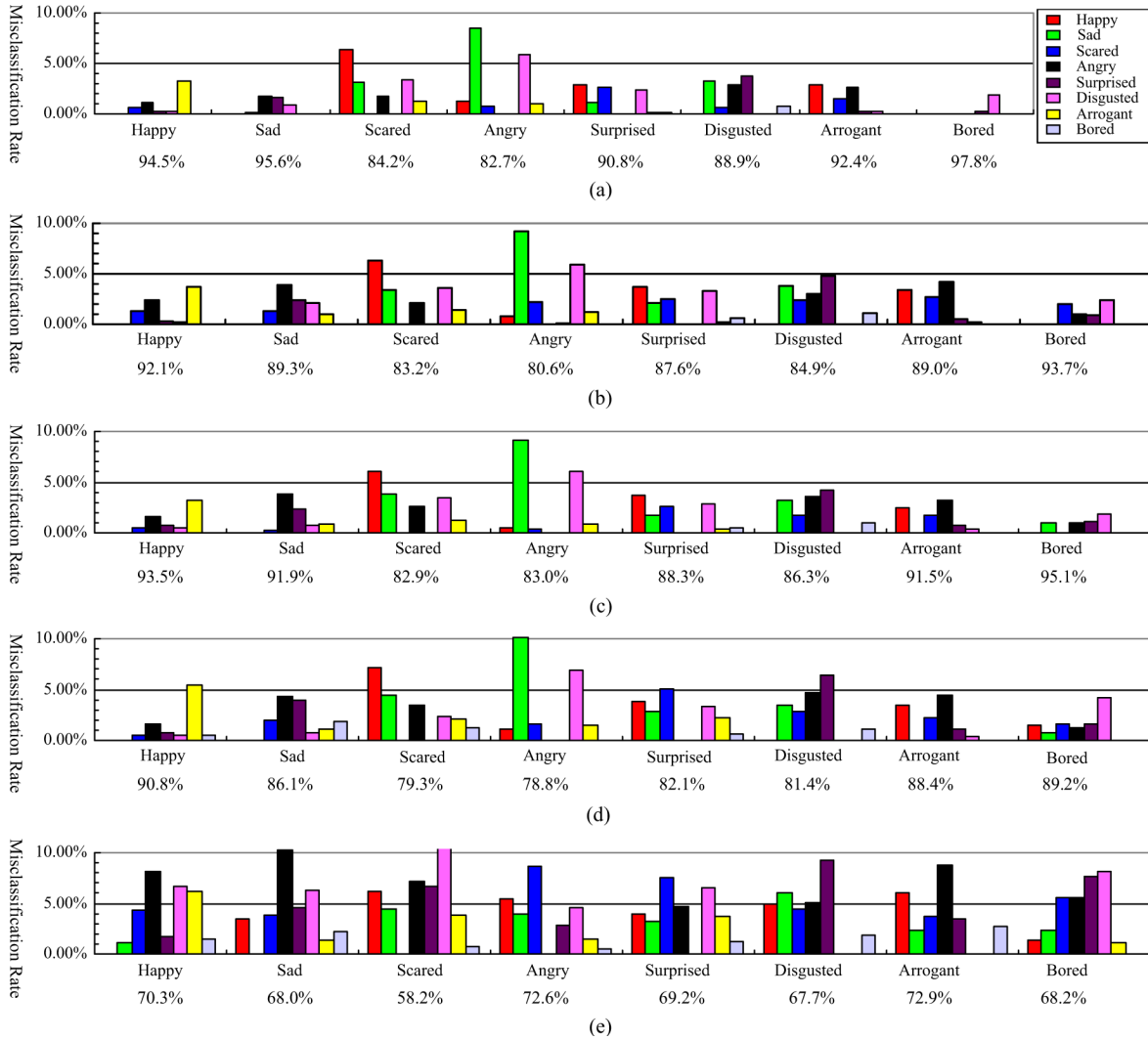


Fig. 8. Distribution of misclassification rate achieved using: (a) local approach, (b) local approach without sketching and curve fitting, (c) global approach, (d) local approach with simply eye, mouth, and eyebrow features, and (e) local approach with only side-view profile and skin wrinkles features. The height of red, green, blue, black, purple, pink, yellow, and gray bars above the name and recognition rate of a given expression indicates the misclassified percentage from this particular expression to happy, sad, scared, angry, surprised, disgusted, arrogant, and bored expressions, respectively.

are used for testing. Fig. 7 displays some sample images taken from three subjects.

Before automatically classifying the collected facial expression images using the proposed system, an interesting experiment would be to evaluate how humans can correctly recognize these posed expressions. The outcome is: ten untrained volunteers have achieved an average recognition rate of 97.3%.

Fig. 8 is a graphical representation of misclassification rates with distinct combinations of aforementioned techniques using our database. It is observed that: 1) the local approach is superior to others in terms of recognition rates, and 2) side-view profile plus skin wrinkles alone can recognize expressions in many cases and are definitely helpful.

Since it is important to quantify the sensitivity and robustness of a method for practical use, the pre-trained networks are simulated by frames with (a) spatially non-uniform additive white Gaussian noise whose strength is inversely proportional to pixel value intensities, (b) under varying illumination or (c) head pose, and (d) all frames of expression intensity level

ranging from level 1 to 6. Fig. 9 plots the results. Interestingly, we find that the local approach is more robust than the global one, and a possible explanation for this is that it breaks unnecessary weight and bias value connections between different facial features so that noises exerted on one feature do not spread to and affect others.

Regarding the speed of the system, the RTA for face tracking is quite fast [29], and the iteration of both AWN and DTM usually converges within 1000 times. On the other hand, while it is very time-consuming to compute eigenvalues and eigenvectors and train the back propagation NN, it is performed offline and is a one-time cost. For each frame, the PCA and NN simulation will need one and two real-number matrix multiplications, respectively. Overall, the system can run at 30 fps on  $352 \times 288$  images on an AMD Athlon 64  $\times$  2 Brisbane 3600+ machine.

## VII. CONCLUSION AND FUTURE WORK

This paper presents a robust real-time expression recognition system which exploits the information from the side view of

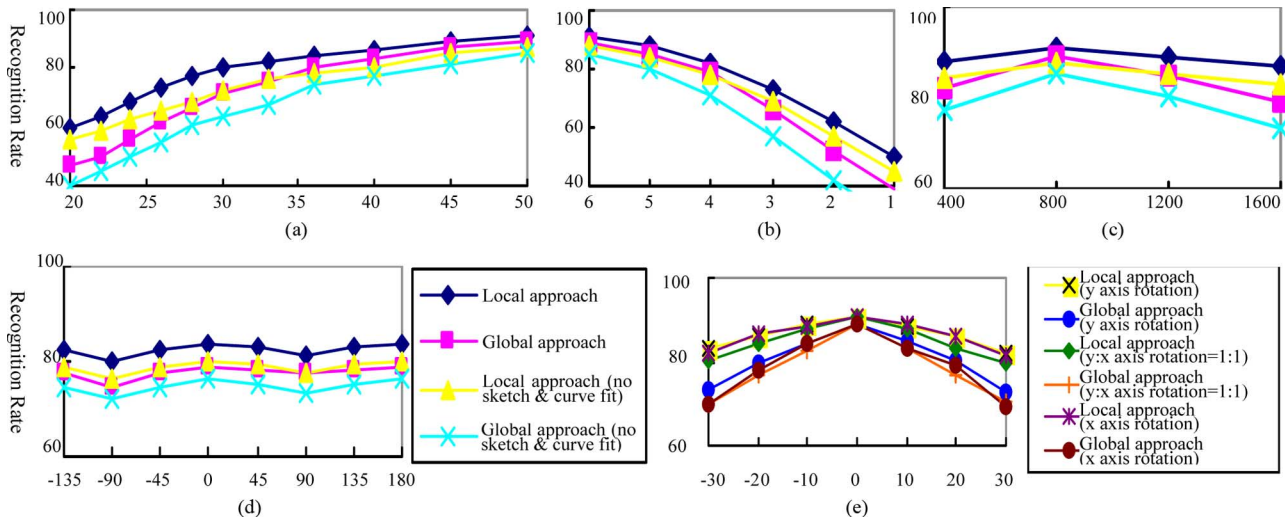


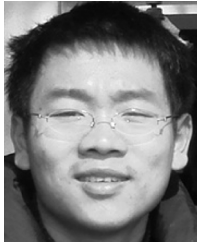
Fig. 9. Recognition rates as a function of: (a) signal-to-noise ratio (SNR) of white Gaussian noise, (b) expression intensity levels, (c) illumination intensity in terms of Lux, (d) directional spotlight incidence angle with horizontal axis, and (e) total head rotation angle.

a face. In recent years, the computer vision community is interested in expression recognition approaches that are robust to illumination, head pose, and partial occlusion. The proposed system is a good engineering solution with balanced robustness of the above three challenges. For the future work, we plan to address the recognition of spontaneous expression under complex illumination condition in dynamic background, instead of dealing with acted facial expressions.

#### REFERENCES

- [1] W. Zhao, R. Chellappa, A. Rosenfeld, and P. J. Phillips, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–458, 2003.
- [2] B. Fasel and J. Luetttin, "Automatic facial expression analysis: A survey," *Pattern Recognit.*, vol. 36, no. 1, pp. 259–275, 2003.
- [3] M. Pantic and L. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1424–1445, Dec. 2000.
- [4] M. Pantic and L. Rothkrantz, "Toward an affect-sensitive multimodal human-computer interaction," *Proc. IEEE*, vol. 91, no. 9, pp. 1370–1390, Sep. 2003.
- [5] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 1, pp. 39–58, Jan. 2009.
- [6] J. M. Buenaposada, E. Munoz, and L. Baumela, "Efficient appearance-based tracking," in *Proc. CVPR 2004 Workshop Articulated and Nonrigid Motion*.
- [7] Y. Wang, G. Pan, and Z. Wu, "3D face recognition in the presence of expression: A guidance-based constraint deformation approach," in *Proc. CVPR 2007*, pp. 1–7.
- [8] A. M. Martinez, "Recognizing imprecisely localized, partially occluded and expression variant faces from a single sample per class," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 6, pp. 748–763, Jun. 2002.
- [9] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," in *Proc. CVPR*, 1994, pp. 84–91.
- [10] S. Z. Li, X. Lu, X. Hou, X. Peng, and Q. Cheng, "Learning multiview face subspaces and facial pose estimation using independent component analysis," *IEEE Trans. Image Process.*, vol. 14, no. 6, pp. 705–712, Jun. 2005.
- [11] Z. Fan and B. Lu, "Fast recognition of multi-view faces with feature selection," in *Proc. ICCV*, 2005, pp. 76–81.
- [12] P. Ekman and W. V. Friesen, *Unmasking the Face*. Englewood Cliffs, NJ: Prentice-Hall, 1975.
- [13] P. Ekman and W. V. Friesen, *The Facial Action Coding System*. San Francisco, CA: Consulting Psychologist, 1978.
- [14] J. C. Hager, P. Ekman, and W. V. Friesen, *FACS Investigator's Guide*. Salt Lake, UT: A Human Face, 2002.
- [15] A. Essa and A. P. Pentland, "Coding, analysis, interpretation, and recognition of facial expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 757–763, Jul. 1997.
- [16] C. Padgett and G. Cottrell, "Representing face images for emotion classification," *Adv. Neural Inf. Process. Syst.*, vol. 9, pp. 894–900, 1997.
- [17] Y. J. Long and Y. Z. Huang, "Image based source camera identification using demosaicking," in *Proc. 8th Int. Workshop Multimedia Signal Processing*, 2006, pp. 419–424.
- [18] Y. Z. Huang and Y. J. Long, "Demosaicking recognition with applications in digital photo authentication based on a quadratic pixel correlation model," *Proc. Computer Vision and Pattern Recognition*, pp. 1–8, 2008.
- [19] N. Fan, C. Jin, and Y. Z. Huang, "A pixel-based digital photo authentication framework via demosaicking inter-pixel correlation," in *Proc. ACM Multimedia and Security Workshop*, 2009, pp. 125–130.
- [20] Y. Z. Huang and N. Fan, "Learning from interpolated images using neural networks for digital forensics," *Proc. Computer Vision and Pattern Recognition*, pp. 1–8, 2010.
- [21] T. F. Cootes, C. J. Taylor, D. Cooper, and J. Graham, "Active shape models—their training and application," *Comput. Vis. Image Understand.*, vol. 61, pp. 38–59, 1995.
- [22] B. S. Manjunath and R. Chellappa, "A unified approach to boundary perception: Edges, textures, and illusory contours," *IEEE Trans. Neural Netw.*, vol. 4, no. 1, pp. 96–107, Jan. 1993.
- [23] A. F. Moller, "A scaled conjugate gradient algorithm for fast supervised learning," *Neural Netw.*, vol. 6, no. 4, pp. 525–533, 1993.
- [24] M. Pantic and L. Rothkrantz, "Expert system for automatic analysis of facial expressions," *Image Vis. Comput.*, vol. 18, no. 11, pp. 881–905, 2000.
- [25] M. Pantic and L. Rothkrantz, "Facial action recognition for facial expression analysis from static face images," *IEEE Trans. Syst., Man, Cybern. B*, vol. 34, no. 3, pp. 1449–1461, Jun. 2004.
- [26] Y. Tian, T. Kanade, and J. Cohn, "Recognizing action units for facial expression analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 97–114, Feb. 2001.
- [27] C. Hu, R. Feris, and M. Turk, "Active wavelet networks for face alignment," in *Proc. British Machine Vision Conf.*, 2003.
- [28] C. L. Huang and C. W. Chen, "Human facial feature extraction for face interpretation and recognition," *Pattern Recognit.*, vol. 25, no. 12, pp. 1435–1444, 1992.
- [29] B. Scassellati, "Eye finding via face detection for a foveated, active vision system," in *Proc. AAAI*, 1998, pp. 969–976.
- [30] P. Sinha, "Perceiving and recognizing three-dimensional forms," Ph.D. dissertation, MIT, Cambridge, MA, 1996.
- [31] P. Maragos, "Tutorial on advances in morphological image processing and analysis," *Opt. Eng.*, vol. 26, no. 7, pp. 623–632, 1987.
- [32] M. J. Tovee, *An Introduction to the Visual System*. West Nyack, NY: Cambridge Univ. Press, 1996.

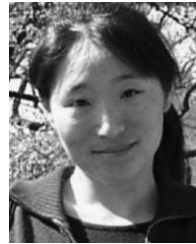
- [33] R. Brunelli and T. Poggio, "Face recognition: Features versus templates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 10, pp. 1042–1052, Oct. 1993.
- [34] P. Sarbajit, P. K. Biswas, and A. Ajith, "Face recognition using interpolated Bezier curve based representation," in *Proc. IEEE Int. Conf. Information Technology*, 2004, vol. 1, pp. 45–49.
- [35] J. K. Hietanen, "Does your gaze direction and head orientation shift my visual attention?," *NeuroReport*, vol. 10, pp. 3443–3447, 1999.



**Yizhen Huang** received the B.S. degree (with highest honors) in computer science from Shanghai Jiaotong University, Shanghai China. He is now pursuing the Ph.D. degree in image processing and computer vision at the Computer Sciences Department of the University of Wisconsin, Madison (UW-Madison).

Before joining UW-Madison, he served in various industrial R&D and consulting tasks for ATI Graphics Division, AMD, Inc., and Intel Asia-Pacific Research & Development Ltd. at the Shanghai ZiZhu site. He published a series of papers on digital forensics, image interpolation, and numerical analysis.

Mr. Huang has served as a reviewer for numerous journals such as the IEEE TRANSACTIONS ON NEURAL NETWORKS, *Journal of Visual Communication and Image Representation*, and *Optics Express*. He was an IEEE ICME 2010 Technical Program Committee member.



**Ying Li** (SM'07) received the B.S. and M.S. degrees in computer science and engineering from Wuhan University, Wuhan, China, and the Ph.D. degree in electrical engineering from the University of Southern California (USC), Los Angeles, in 1993, 1996, and 2003, respectively.

Since March 2003, she has been with IBM T. J. Watson Research Center as a Research Staff Member. Her research interests include digital image processing, content-based image analysis and retrieval, multimodal-based video content analysis,

indexing and representation, pattern recognition, computer vision, multimedia and e-Learning applications. She has authored or co-authored around 35 conference and journal papers, as well as four books and book chapters on various multimedia-related topics. She currently holds 14 U.S. patents.

Dr. Li has served on the Technical Program Committees of tens of IEEE and ACM conferences such as ICME, ICIP, ICASSP, and ACM Multimedia. Besides, she is on the Program Committee of various IEEE conferences and workshops. In 2005, she was invited as the NSF panelist for the Video Analysis Track.



**Na Fan** received the B.E. degree in electronic engineering from East China Normal University, Shanghai, China. She is now pursuing the M.S. degree in image processing, communication, and information theory at East China Normal University, Shanghai.

Her research interests involve computer vision, image processing, and information theory.