5

Classifying Facial Expressions Based on Topo-Feature Representation

Xiaozhou Wei, Johnny Loi and Lijun Yin
State University of New York at Binghamton
USA

1. Introduction

Facial expression analysis and recognition could help humanize computers and design a new generation of human computer interface. A number of techniques were successfully exploited for facial expression recognition (Chang et al., 2004; Cohen et al., 2004; Cohen et al., 2003; Gu & Ji, 2004; and Littlewort et al, 2004), including feature estimation by optical flow (Mase, 1999; Yacoob & S. Davis, 2006), dynamic model (Essa & Pentland, 1997), eigen-mesh method (Matsuno et all.) and neural networks (Rosenblum et all., 1996). The excellent review of recent advances in this field can be found in (Y. Tian et al., 2001; Pantic & Rothkrantz, 2000; Zhao et al., 2000). The conventional methods on facial expression recognition concern themselves with extracting the expression data to describe the change of facial features, such as Action Units (AUs) defined in Facial Action Coding System (FACS) (Donato et al., 1999). Although the FACS is the most successful and commonly used technique for facial expression representation and recognition, the difficulty and complexity of the AUs extraction limit its application. As quoted by most previous works (Essa & Pentland, 1997; Yacoob & Davis, 1996), capturing the subtle change of facial skin movements is a dif cult task due to the difficulty to implement such an implicit representation. Currently, feature-based approaches (Reinders et al., 1995; Terzopoulos & Waters, 1993) and spatio-temporal based approaches (Essa & Pentland) are commonly used. Yacoob & Davis, 1996 integrated spatial and temporal information and studied the temporal model of each expression for the recognition purpose, a high recognition rate was achieved. Colmenarez et al. used a probabilistic framework based on the facial feature position and appearances to recognize the facial expressions, the recognition performance was improved, but only the feature regions other than the surface information were explored. Recently, Tian, Kanade and Cohn (Tian et al., 2001) noticed the importance of the transient features (i.e., furrow information) besides the permanent features (i.e, eyes, lips and brows) in facial expression recognition. They explored the furrow information for improving the accuracy of the AU parameters, an impressive result was achieved in recognizing a large variety of subtle expressions. To our knowledge, little investigation has been conducted on combining texture analysis and surface structure analysis for modeling and recognizing facial expressions. A detailed higher level face representation and tracking system is in high demand. In this paper, we explore the active texture information and facial surface features to meet the challenge – modeling the facial expression with sufficient accuracy.

Facial texture appearance plays an important role in representing a variety of facial expressions. It is possible to classify a facial expression by noting the change of facial texture. On the other hand, the facial expression change is reflected by the variation of the facial topographic deformation. It is thus of interest for us to investigate the relationship between these features and the corresponding expressions in order to model facial expressions in an intuitive way.

Facial texture consists of four major active regions: eyebrow, eye, nose and mouth. A set of active textures of a person is a class of textures with the same statistical property, which in general represents a statistical resemblance determined by the action of different facial expressions. In this paper, we propose a facial expression analysis system based on the integration of the topographic feature and the active texture. The system is composed of three major components: texture enhancement by increasing the image resolution; face surface region representation using topographic analysis; and the similarity measurement by a topographical masking method. Figure 1 outlines the system composition.

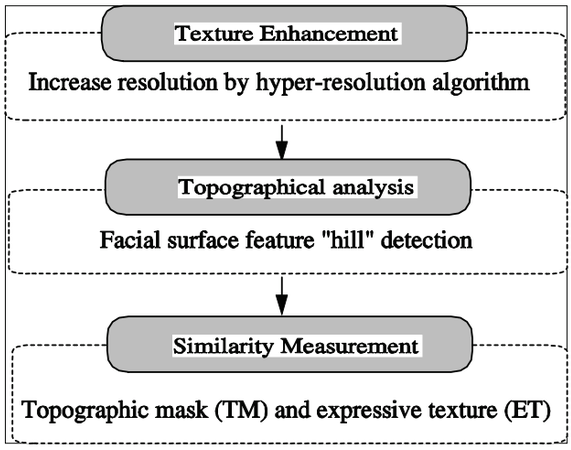

Fig. 1: Facial expression analysis system

Analysis of the facial surface features relies on the texture details. In order to detect the surface feature in a detailed level, in the first stage, we enhance the image resolution by a so called hyper-resolution algorithm. The topographical labeling process can then be carried out in the second stage. In the third stage, the texture similarity and topographic disparity between the test image and the reference image are measured for classifying different facial expressions. The method outputs a score to signify the level of similarity. In the end, a simple classification rule is defined for distinguishing six universal expressions.

In Section 2, the algorithm for texture detail recovery is described. Section 3 describes the method for facial surface representation using topographic analysis. The algorithm for similarity measurement and expression classification will be described in Section 4, followed by the experiment result presented in Section 5. Finally, the concluding remark will be given in Section 6.

2. Texture enhancement

In order to analyze and detect the face surface features from low resolution images, facial details recovery by enhancing the image resolution is necessary due to the various imaging

environment where only few effective pixels could be captured in the face region. Another reason for increasing resolution is that our subsequent analysis using topographical feature relies on a surface fitting procedure, which requires accurate pixel representation within the face region. Here, we propose a super-resolution method to recover details in the facial region. Traditional methods for scaling a low resolution image to a high resolution version depend on interpolation algorithms, which suffer problems of blurred images lacking distinct edges and fine textures. Recently, *training based* super-resolution algorithms have been developed by (Baker et al., 2002; Freeman. et al., 2002; Sun et al., 2003), where an image database containing the high resolution detail of many different subjects is used. The visually plausible high resolution details for one low resolution *target image* are reconstructed based on the pattern recognition and substitution of "similar" details in a potentially large database of high resolution model images (also called *source images*.) Most existing algorithms work on either the pixel or pixel patch level. However, both humans and computers judge image sharpness and detail by the quality of an image's strong edges. This observation leads us to a novel idea: instead of matching patches in the spatial domain, we can first transform each image into a new parametric vector space structured by the image's edges. Our "patches" are then composed of only the texture details (in the form of high-pass filtered image pixels) sampled on and around image edges, with coordinates relative to these edges. Thus, we simply need to match and replace low resolution data around each target edge with the appropriate high resolution detail from a database of source edges. Central to this new method is the novel transform of image content from the orthogonal pixel space to a parametric space structured around edges.

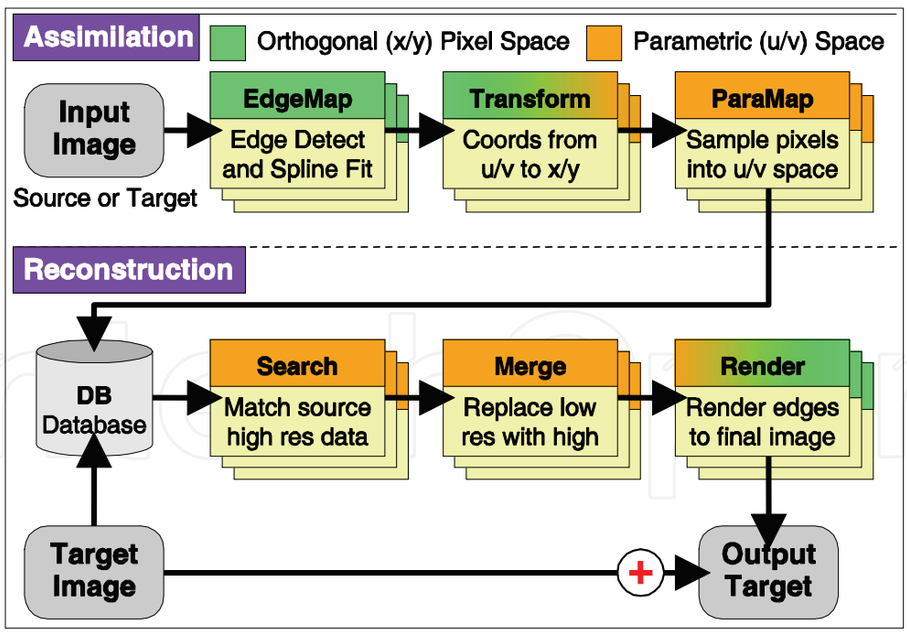

Fig. 2: HyperRes Pipeline Overview.

This so-called *hyper-resolution (HyperRes)* algorithm is illustrated in Figure 2, where the pipeline is logically split into two parts: the *Assimilation* phase, through which both source and target images are transformed and added to a database, and the *Reconstruction* phase, in

which a single target image is reconstructed at a higher resolution than its original pixels provided.

2.1 Assimilation

Prior to starting the true HyperRes algorithm, the target image must be interpolated up to the desired output size (e.g., 200% or more) using the simple bi-cubic interpolation. To construct an image around its edges, we first detect those edges by using Canny edge detector, then apply the cubic splines to t the edges. This parametric curve representation allows us to establish a bidirectional mapping between two coordinate spaces: the *orthogonal space* in which the image pixels reside along *x*, *y* coordinates, and the *parametric space,* a nonlinear transformed space relative to the curvature of each edge. Each edge has its own local parametric space in the form of a warped rectangle, with the *u* coordinate running parallel to the edge and the *v* coordinate running out along the edge's normal as evaluated at *u*. The coordinate of a point **P**(u) at a parametric coordinate u in the span between any two known edge points $\mathbf{E}_n$ and $\mathbf{E}_{n+1}$ is defined as

$$P(u) = \frac{1}{2}\begin{bmatrix} 1 \\ (\frac{u-U_n}{S_n}) \\ (\frac{u-U_n}{S_n})^2 \\ (\frac{u-U_n}{S_n})^3 \end{bmatrix} \cdot M \cdot \begin{bmatrix} E_{n-1} \\ E_n \\ E_{n+1} \\ E_{n+2} \end{bmatrix} \tag{1}$$

Where

$$M = \begin{bmatrix} \frac{1}{6} & \frac{2}{3} & \frac{1}{6} & 0 \\ \frac{-1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & -1 & \frac{1}{2} & 0 \\ \frac{-1}{6} & \frac{1}{2} & \frac{-1}{2} & \frac{1}{6} \end{bmatrix} \tag{2}$$

$$S_n = \mid E_{n+1} - E_n \mid \tag{3}$$

To completely define the parametric space, we must also find the *v* coordinate, which runs parallel to the curve's normal at a given value of *u*. The complete transform from a parametric coordinate *u, v* to an orthogonal coordinate $Q(u,v) \rightarrow x, y$ is thus:

$$Q(u,v) = P(u) + v \cdot \frac{\frac{\partial P(u)}{\partial u}}{\frac{\partial P(u)}{\partial u}} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \tag{4}$$

where *Q(u, v)* is the final coordinate in the orthogonal space. $-v_\sigma \leq v \leq v_\sigma$ (v_σ is set as 8 pixels). After fitting each edge to a parametric curve representation, we warp the curve area around each edge to a standard rectangle (u, v) area, thus forming a "parametric map" (called *ParaMap*). Figure 3 shows an example of the *uv-xy* transformation and ParaMap formation.

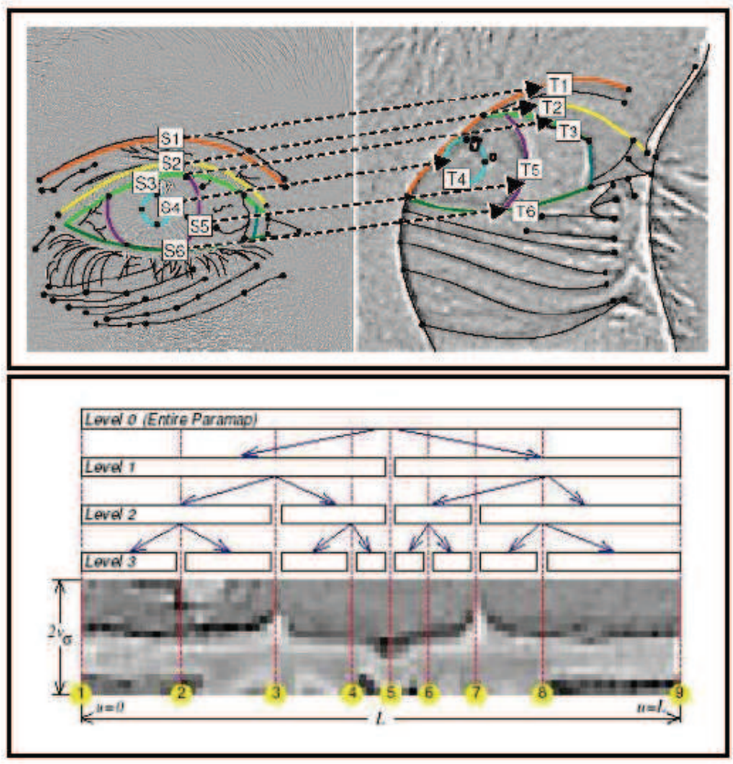

Fig. 3: Top: source-target edges match using an eye image as an example (S: source; T: target). Bottom: Paramap formation (example of S1)

2.2 Reconstruction

To reconstruct the high resolution version of a given target edge, we search for similar edge textures in the database by matching the contents of paramaps between the source and target images. To facilitate efficient operation, each edge is recursively split into two segments, such that we can always attempt to match the longest possible edge first, then subdivide as needed to optimize for accuracy.

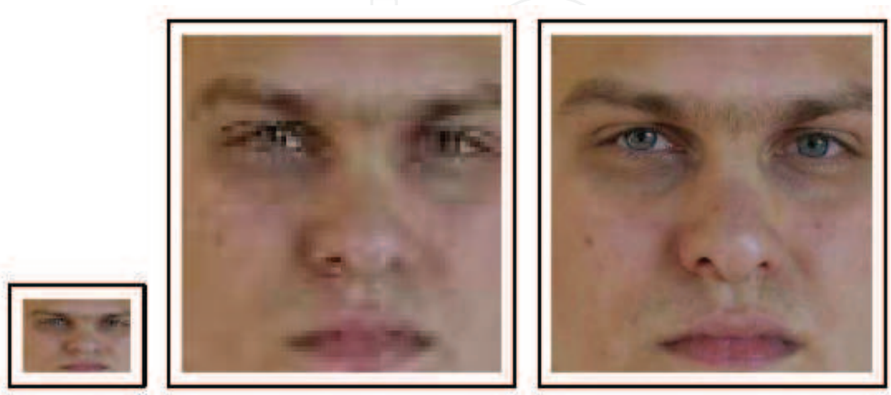

Fig. 4: left: original face region 64*64 pixels; middle: enlarged by cubic interpolation (256*256); right: enlarged by HyperRes (256*256).

The results of the Search stage are in the form of key pairs: a target key and its matching source key. In the Merge stage, we map the source key back to the corresponding subsection of the paramap owning it, and *merge* that parametric pixel data back into the correct place in the target edge's paramap. The reconstructed paramap is further rendered onto the curvature of the appropriate target edge, which is implemented by projecting the u, v coordinates back into orthogonal x, y space. As a result, the detail-reconstructed high-frequency image is created. Finally, we add this high-frequency image back into the original low resolution image to generate the new resolution-increased image. Figure 4 illustrates one example of the HyperRes results.

3. Topographic analysis

To find an explicit representation for the fundamental structure of facial surface details, the topographic primal sketch theorem is exploited (Haralick et al., 1983), where the grey level image is treated as a topographic terrain surface. Each pixel is assigned one of the topographic label peak, ridge, saddle, hill, at, ravine, or pit, as shown in Figure 5. Hill-labeled pixels can be further specified as one of the labels convex hill, concave hill, saddle hill or slope hill (Trier et al., 1997). We furthermore distinguish saddle hills as *concave saddle hill, convex saddle hill*, distinguish saddle as *ridge saddle* and *ravine saddle*.

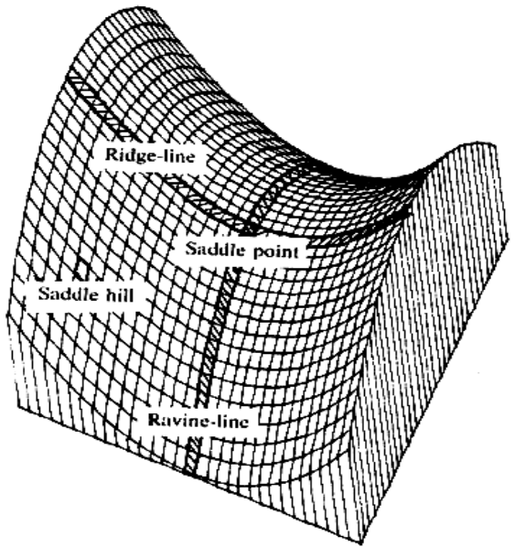

Fig. 5: Topographic labels (Haralick et al., 1983) used for facial skin details representation. The 3D topographical structure can be classified by a number of features, such as peak, pit, ridge, ravine, ridge saddle, ravine saddle, convex hill, concave hill, convex saddle hill, concave saddle hill, slope hill and at (Trier et al., 1997).

Light intensity variations on an image are caused by an object's surface orientation and its reflectance. In visual perception, exactly the same visual interpretation and understanding of a pictured scene occurs no matter how the imaging condition is. This fact suggests that topographic features can be expected to have the robustness associated with human visual perception because they are inherently capable of invariance under monotonic transformations. The topographic categories peak, pit, ridge, valley, saddle, at, and hill can

reveal the three dimensional intrinsic surface of the object, and thus be possible for us to extract the facial surface features. With surface feature classification, the facial image can be segmented into a number of feature areas. The different composition of these basic primitives will give a fundamental representation of different skin surface details. The primitive label classification approach is determined by the estimation of the first-order and second-order directional derivatives. The gradient vector is

$$\nabla f = \left(\frac{\partial f}{\partial x} \quad \frac{\partial f}{\partial y} \right) \tag{5}$$

The second directional derivatives can be calculated by forming the *Hessian matrix* [17].

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} \tag{6}$$

The eigenvalues (λ_1 and λ_2) of the Hessian are the values of the extrema of the second directional derivative, and their associated eigenvectors (ω_1 and ω_2) are the directions in which the second directional derivatives have greatest magnitude. The feature labeling is based on the values of λ_1, λ_2, ω_1, ω_2, and ∇f. For example, a pixel is labeled as a convex hill if the values in this pixel satisfied the following condition:

$$\lambda_1 < 0; \lambda_2 > 0; |\nabla f| > T_G \tag{7}$$

where T_G is a predefined threshold.

Since a temporal skin "wave" is associated with the movement of the expression, the skin surface with a certain expression at a different time will have different shape, resulting in the different label changes. The skin detail shape follows the expression's change while the expressions occur in three distinct phases: *application (from the beginning), release (from the apex)* and *relaxation (to the end)*. For example, from the concave hill to convex hill, from the ridge saddle to ravine saddle. This is known as dynamic labeling along with shape change. We can imagine that the skin surface is represented as a topographic label "map", this "map" is changed along with the skin movement. Figure 6 shows one example of facial surface features labeled by topographic analysis.

4. Facial expression representation and classification

4.1 Representation

The topographic analysis of facial surface outputs a group of topographic labels. Currently, two types of labels, convex hill and convex saddle hill, are used for feature detection. These two types of labels form a topographic mask (TM) of a specific expression, as shown in Figure 6a and Figure 6b (red and pink labels). As we can see, the areas (eye, eyebrow, nose, mouth, nasolabial, etc.) encompassed by the topographic mask are the significant regions for representing an expression. We call these textures as expressive textures (ET) or active textures (AT). Apparently TM and ET are changing along with the expression change. In terms of a specific subject, TM represents the facial expression pattern (i.e., six universal expressions have six types of TMs.) Therefore, we choose these topographic masks and

expressive textures to classify different expressions. Given a specific subject A, by comparing the set of (TM, ET) of A's neutral face with the set of (TM, ET) of A's expressive face, the expression could be identified.

Fig. 6a: Example of topographic facial analysis on sample images of Cohn-Kanade facial expression database (Kanade et al., 2000). (top row: original faces; bottom row: labeled faces with convex hill (in red) and convex saddle hill (in pink).

Fig. 6b: Facial expression representation by the topographic Mask from the JEFFE database (Lyons, 2005). (a) original image (KL); (b) Labeling after resolution increased from 256*256 to 512*512; (c) Background label removal; (d) Topographic Mask of "disgust" expression.

4.2 Classification

For an individual person, there is a unique topographic mask for his/her individual expression. TM is represented in a binary format, in which '1' denotes the labeled region (red and pink region) and '0' for the non-labeled region. The disparity of two masks between the neutral face and the expressive face is formulated as:

$$D_{TM} = \frac{1}{N} \sum TM_n \; xor \quad TM_e \tag{8}$$

where N is the total number of pixels within the TM of neutral expression. TM_n is the neutral mask; TM_e is the expressive mask for one of the six universal expressions. The similarity between the neutral texture and the current expressive texture is measured by the correlation C_{ET} (Yin et al., 2003):

$$C_{ET} = \frac{E(t_n t_e) - m_{t_n} m_{t_e}}{\sigma(t_n)\sigma(t_e)} \tag{9}$$

where t_n and t_e are the neutral texture and the expressive texture, respectively. E() is a mean operation, m_{tn}, $\sigma(t_n)$ and m_{te}, $\sigma(t_e)$ are the means and variances of the two textures to be compared. In general, the similarity of two expressions is characterized by the similarity score S_{exp}, which is defined as:

$$S_{\exp} = (1 - D_{TM}) + C_{ET} \tag{10}$$

In order to differentiate the different expressions, we take a statistic method to calculate the similarity scores through a training set which contains 20 video sequences performed by 20 different subjects, each subject performed a neutral expression and six universal expressions. The average similarity score for each typical expression is obtained in Figure 7. It shows that the similarity score is within the range of [0,2] in a decreasing order from neutral to sad, fear, angry, disgust, surprise, and to happy. By the property of monotonicity of similarity curve shown in Figure 7, we classify the expressions into seven categories by the similarity scores range: [T1, 2] for neutral expression; [T2, T1) for sad; [T3, T2) for angry; [T4, T3) for fear; [T5, T4) for disgust; [T6, T5) for surprise and [0, T6) for happy. \ The thresholds T1 - T6 are obtained by the average values of two "neighbor" expressions, as shown in Figure 7 and Table 1.

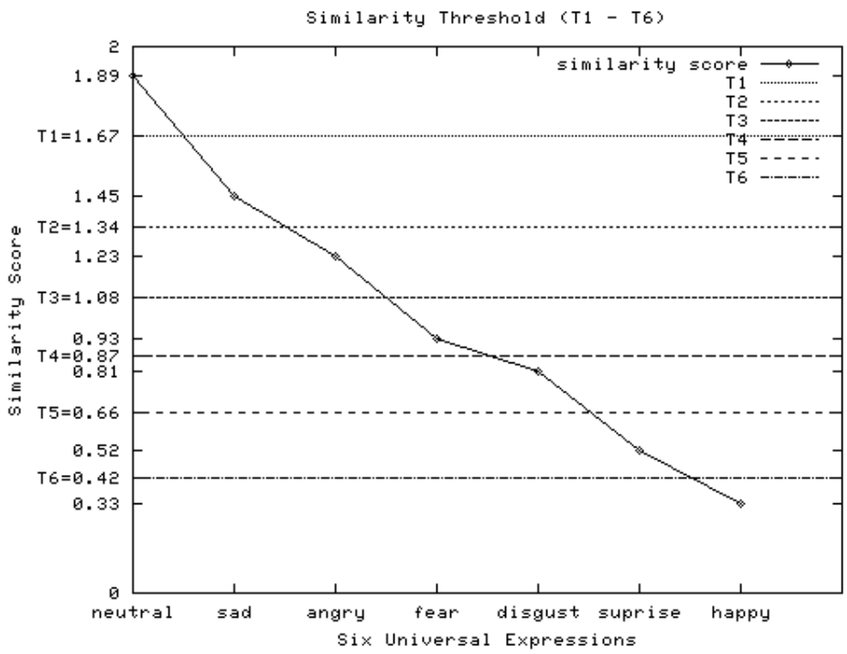

Fig. 7: Similarity scores and threshold values

T1	T2	T3	T4	T5	T6
1.67	1.34	1.08	0.87	0.66	0.42

Table 1: Thresholds for distinguishing six expressions plus a neutral expression. As shown in Figure 7, T1 is the threshold between neutral and sad; T2 between sad and angry; T3: angry and fear; T4: fear and disgust; T5: disgust and surprise; T6: surprise and happy.

5. Experiments

5.1 The JAFFE facial expression database (Lyons, 2005) is used in our experiment. The database contains 212 frontal facial images including ten female subjects performing seven types of expressions (six universal expressions plus one neutral expression). After increasing the resolution to the double size, the face images are labeled by convex hill and convex saddle hill features. Figure 8 shows two sets of topographic masks obtained from two female subjects (KL and TM). The results show that topographic labeling method can accurately capture the facial surface regions.

For each subject, we take one of the neutral expressions as a reference expression to which all other expressions of her own can be compared. In order to remove the background hill regions, we detect the connected components of the hill regions, and take the largest region as the face region to form the topographic mask. Based on the obtained TM, we use the similarity score to classify the expressions. Note that before the similarity calculation, the TM and EM are normalized (by the width, height and orientation) to the size of the neutral TM and ET. The similarity score is calculated to classify the input facial expression using thresholds and categories defined in Table 1 and figure 7.

The classification results show that the average correct recognition rate is at 85.8%. Table 2 enumerates the recognition results on seven expressions of JEFFE database.

The experimental result shows that recognition rates are encouraging in view of such a condition that no action units are extracted. The possible extension of this work is to use the similarity score of other topographic labels for classifying richer range of fine expressions.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1)angry	25	1	1	0	1	1	1
(2)disgust	1	26	1	1	0	1	1
(3)fear	1	1	24	0	1	2	2
(4)happy	0	0	1	28	0	0	0
(5)neutral	1	0	1	0	27	1	0
(6)sad	1	2	1	1	1	26	0
(7)suprise	1	0	2	0	0	0	26
RR(%)	83.3	86.7	77.4	93.3	90.0	83.9	86.7

Table 2: Correct expression recognition rate (RR). 212 frontal face images captured from 10 female subjects, including 30 images for each of following expressions: (1) angry, (2) disgust, (4) happy, (5) neutral, (7) surprise; and 31 images for each of following expressions: (2) fear and (6) sad. (Average RR: 85.8%).

5.2 The Cohn-Kanade (CK) facial expression database is used for our second test. We randomly choose 180 sequences performed by 30 subjects, in which each prototypic expression has 30 sequences. Since the first frame of each sequence shows the neutral expression and the last frame shows the peak of a prototypic expression, we select the first frame as a reference frame and the last frame as a probe frame to form an image pair, with a total of 180 pairs in our experiment. In addition, we also generate 30 image pairs from 30 sequences by selecting the first frame and the third frame as the neutral expression pairs. The classification results show that the average correct recognition rate is at 80.9%. Table 3 enumerates the classification results by the confusion matrix. Note that from the experiments, we find that the labeling algorithm works best when the face region has over

256*256 pixels. Although the above two databases contain images with the size larger than 256*256, the valid face region may not have sufficient resolution. Therefore, to keep the operation consistent and efficient, we always increase the image resolution from its original size to the double size using AugRes algorithm before the topographic labeling. The experiment demonstrates the efficiency and the feasibility of this scheme. The possible extension of this work is to use the similarity score of other topographic labels for classifying wide range of fine expressions.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1)angry	23	1	1	1	2	1	2
(2)disgust	1	24	1	1	0	2	0
(3)fear	1	2	22	1	0	2	2
(4)happy	1	1	1	25	0	0	1
(5)neutral	1	0	1	0	27	1	0
(6)sad	2	2	2	0	1	24	0
(7)suprise	1	0	2	2	0	0	25
RR(%)	76.7	80.0	73.3	83.3	90.0	80.0	83.3

Table 3: Confusion matrix and correct expression recognition rate (RR) on CK database. (Average RR: 80.9%).

5.3 Analysis: Identifying facial expression is a challenging problem. However, there is no baseline algorithm or standard database for measuring the merits of new methods. The unavailability of an accredited common database and evaluation methodology make it difficult to compare any new algorithm quantitatively with the existing algorithms. Although the CK database is widely used, some recent work (Cohen et al., 2004; Cohen et al., 2003) utilized a portion of the database because not all the subjects in CK database have six prototypic expressions. This makes the algorithm comparison not feasible without knowing the exact test set in the existing approaches. Here we use the JAFFE database as a common test data to compare the result reported by (Lyons et al., 1999). Lyons et al (1999) developed a elastic graph matching and a linear discriminant analysis approach to classify expressions of JAFFE database. The very impressive results were achieved in classifying facial images in terms of gender, race and expression. In particular, the correct recognition rate for the prototypic expressions is at 92% for a person-dependent test, and at 75% for the person-independent test. Here, we report our average recognition rate for JAFFE database is at 85.8%. Note that this rate is obtained under the person-independent circumstance, which means that the person to be tested has never appeared in the training set. Our system therefore improves the expression classification performance to a certain degree.

Our system has certain pros and cons:

(1) Although our system can work based on the static images, the expression recognition is dependent on the first frame, which means we need a reference face image with neutral expression for each instance. It is frame-dependent. Because the dynamic expression is obtainable by the video capture, it is feasible to select a neutral expression frame in the video.

(2) Our system works on the facial expressions in frontal view. The selection of image pairs with a neutral expression and a test expression is conducted manually. We selected the "peak" (exaggerated) expression as the test data. More spontaneous expressions will be used in the future work.

(3) Our system is person-independent. It appears that topographic structure significantly differs from person to person, in other words, TM is person specific. However, the normalized similarity between a neutral TM and an expressive TM falls in a certain range with respect to the certain expression regardless of subjects to be tested. Therefore, the classification of facial expression is person independent. Moreover, it does not require complete sampling of the expression space for each person. A large training set and a large sampling space for each expression could help improve the accuracy and robustness of the classification.

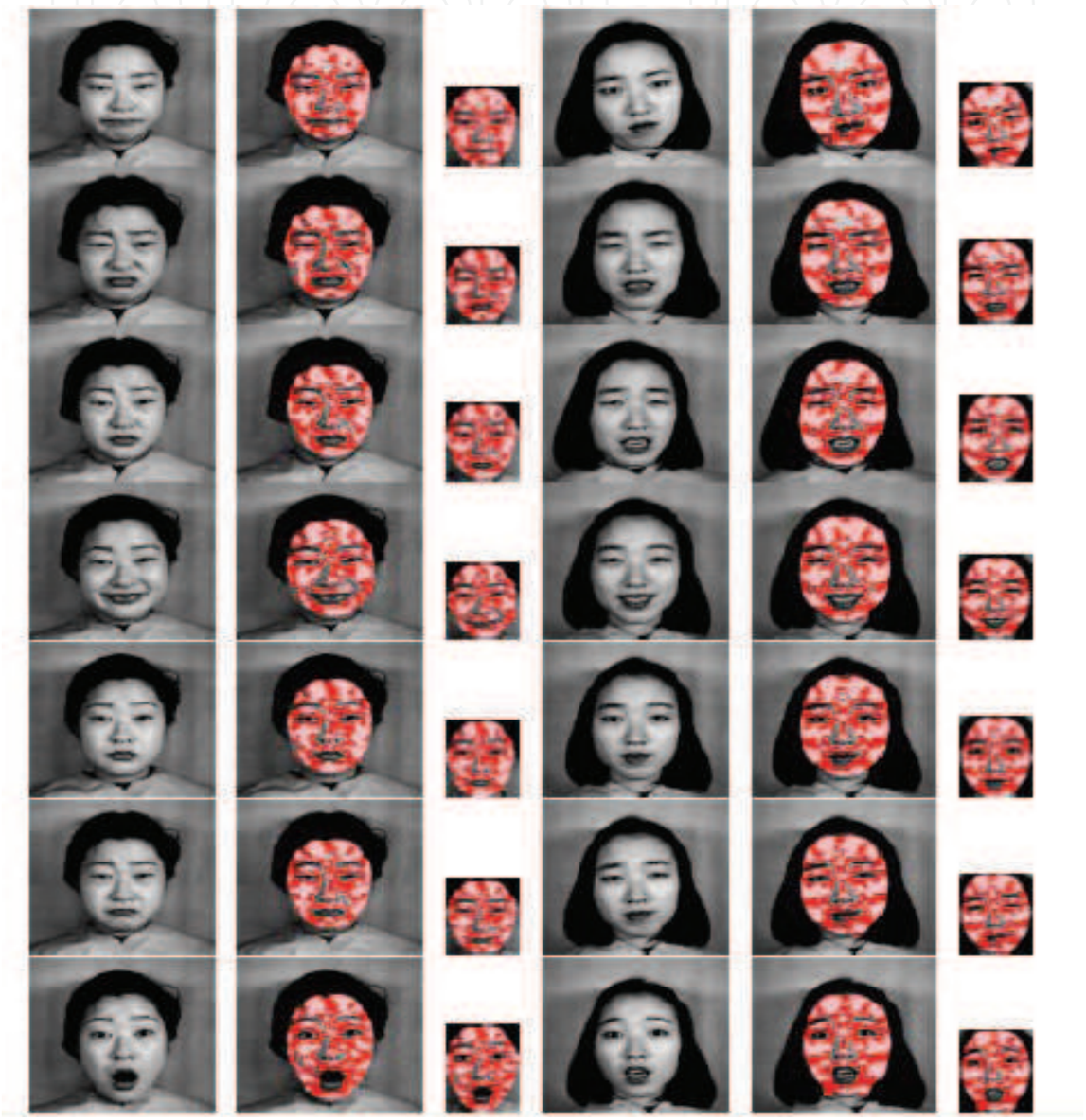

Fig. 8: Example of labeling on JAFFE facial expressions (Lyons et al., 2005) (names: KL for left three columns and YA for right three columns). From top to bottom: angry, disgust, fear, happy, neutral, sad, and surprise. Each row shows the expression, labeled face region, and the Topographic Mask, respectively.

6. Conclusion

We presented a new scheme to model and recognize facial expressions based on the topographic shape structure and the enhanced textures. The encouraging results demonstrate the efficiency and the feasibility of the proposed scheme. This work can be the first step for the coarse classification of the expressions. A fine classification approach will be investigated as the second step by taking the probability of appearance of topographic mask and other labels into account (e.g., ridge and ravine). In the future, the similarity score can be used as an input in order to design the classifier to classify subtle facial expressions recognition, for example, using the Bayesian recognition framework (Colmenarz et al; Gu & Ji, 2004). In addition, the intensive test on a larger amount of facial expression data will be conducted in the future.

7. Acknowledgment

This material is based upon the work supported in part by the National Science Foundation under grant No. IIS-0414029, IIS-0541044, NYSTAR, and AFRL.

8. References

G. Donato, P. Ekman, and et al. Classifying facial actions. *IEEE Trans. PAMI*, 21(10):974-989, 1999.

I. Essa and A. Pentland. Facial expression recognition using a dynamic model and motion energy. In *ICCV95*, pages 360–367.

I. Essa and A. Pentland. Coding, analysis, interpretation, and recognition of facial expressions. *IEEE Trans. PAMI*, 19(7), 1997.

A. Colmenarz *et al*. A probalilistic framework for embedded face and facial expression recognition. In *CVPR99*.

S. Baker *et al*. Limits on super-resolution and how to break them. *IEEE PAMI*, 24(9), 2002.

W. Freeman. *et al*. Example-based super-resolution. *IEEE Comp. Graphics and App.*, 22(2):56–65, 2002.

Y. Tian *et al*. Recognizing action units for facial expression analysis. *PAMI*, 23(2), 2001.

R. Haralick and et al. The topographic primal sketch. *The Int. J of Robotics Research*, 2(2):50–72, 1983.

T. Kanade, J.F. Cohn, and Y. L. Tian. Comprehensive database for facial expression analysis. In *IEEE 4th International conference on Automatic Face and Gesture Recognition*, France, 2000. http://vasc.ri.cmu.edu/idb/html/face/facial expression/.

K. Mase. Recognition of facial expression from optical flow. *IEICE Transactions*, E74(10):3474–3483, October 1991.

K. Matsuno, C. Lee, S. Kimura, and S. Tsuji. Automatic recognition of human facial expressions. In *ICCV95*, pages 352–359.

M. Pantic and L. Rothkrantz. Automatic analysis of facial expressions: the state of the art. *IEEE Trans. PAMI*, 22(12), 2000.

M. Reinders, P. Beek, B. Sankur, and J. Lubbe. Facial feature localization and adaptation of a generic face model for model-based coding. *Signal Processing: Image Communication*, 7:57–74, July 1995.

M. Rosenblum, Y. Yacoob, and L. Davis. Human expression recognition from motion using a radial basis function network architecture. *IEEE Trans. on Neural Network*, 7(5):1121–1138, 1996.

J. Sun, N. Zheng, H. Tao, and H. Shum. Image hallucination with primal sketch priors. In *CVPR*, 2003.

D. Terzopoulos and K. Waters. Analysis and synthesis of facial image sequences using physical and anatomical models. *IEEE Trans. PAMI*, 15(6):569, 1993.

O. Trier, T. Taxt, and A.K. Jain. Recognition of digits in hydrographic maps: binary versus topographic analysis. *IEEE Trans. PAMI*, 19(4), April 1997

Y. Yacoob and L. S. Davis. Recognizing human facial expressions from long image sequences using optical flow. *IEEE Trans. PAMI*, 18(6):636–642, June 1996.

L. Yin, S. Royt, et al. Recognizing facial expressions using active textures with wrinkles. In *IEEE Inter. Conf. on Multimedia and Expo 2003*, pages 177–180, Baltimore, MD, July 2003.

W. Zhao, R. Chellappa, A. Rosenfeld, and P. Phillips. Face recognition: A literature survey. *Technique report: CAR-TR-948, CS-TR-4167, University of Maryland, College Park*, October 2000.

Y. Chang, C. Hu, and M. Turk. Probabilistic expression analysis on manifolds. In IEEE Inter. Conf. on CVPR, Washington DC, June 2004.

I. Cohen, F. Cozman, N. Sebe, M. Cirelo, and T. Huang. Semi-supervised learning of classifiers: Theory, algorithms for bayesian network classifiers and application to human-computer interaction. IEEE Trans. PAMI, 26(12), 2004.

I. Cohen, N. Sebe, A. Garg, L. Chen, and T. Huang. Facial expression recognition from video sequences: temporal and static modeling. Computer Vision and Image Understanding, 91(1), 2003.

H. Gu and Q. Ji. Facial event classification with task oriented dynamic bayesian network. In CVPR, 2004.

G. Littlewort, M. Bartlett, I. Fasel, J. Susskind, and J. Movellan. Dynamics of facial expression extracted automatically from video. In CVPR Workshop on FPIV'04, 2004.

M. Lyons. http://www.mis.atr.co.jp/ mlyons/jaffe.html. 2005.

M. Lyons, J. Budynek, and S. Akamatsu. Automatic classi_cation of single facial images. IEEE Trans. PAMI, 21(12):1357{1362, December 1999.

Affective Computing

Edited by Jimmy Or

ISBN 978-3-902613-23-3

Hard cover, 284 pages

Publisher I-Tech Education and Publishing

Published online 01, May, 2008

Published in print edition May, 2008

This book provides an overview of state of the art research in Affective Computing. It presents new ideas, original results and practical experiences in this increasingly important research field. The book consists of 23 chapters categorized into four sections. Since one of the most important means of human communication is facial expression, the first section of this book (Chapters 1 to 7) presents a research on synthesis and recognition of facial expressions. Given that we not only use the face but also body movements to express ourselves, in the second section (Chapters 8 to 11) we present a research on perception and generation of emotional expressions by using full-body motions. The third section of the book (Chapters 12 to 16) presents computational models on emotion, as well as findings from neuroscience research. In the last section of the book (Chapters 17 to 22) we present applications related to affective computing.

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen