International Journal of Pattern Recognition
and Artificial Intelligence
Vol. 22, No. 3 (2008) 445–459


# PERFORMANCE COMPARISONS OF FACIAL EXPRESSION RECOGNITION IN JAFFE DATABASE

FRANK Y. SHIH* and CHAO-FA CHUANG

*Computer Vision Laboratory, College of Computing Sciences*
*New Jersey Institute of Technology*
*Newark, NJ 07102, USA*
**shih@njit.edu*

PATRICK S. P. WANG

*College of Computer and Information Science*
*Northeastern University, Boston, MA 02115, USA*
*pwang@ccs.neu.edu*

Facial expression provides an important behavioral measure for studies of emotion, cognitive processes, and social interaction. Facial expression recognition has recently become a promising research area. Its applications include human-computer interfaces, human emotion analysis, and medical care and cure. In this paper, we investigate various feature representation and expression classification schemes to recognize seven different facial expressions, such as happy, neutral, angry, disgust, sad, fear and surprise, in the JAFFE database. Experimental results show that the method of combining 2D-LDA (Linear Discriminant Analysis) and SVM (Support Vector Machine) outperforms others. The recognition rate of this method is 95.71% by using leave-one-out strategy and 94.13% by using cross-validation strategy. It takes only 0.0357 second to process one image of size 256 × 256.

*Keywords*: Facial expression; feature representation; face recognition; principal component analysis; support vector machine.

## 1. Introduction

Facial expression plays a principal role in human interaction and communication since it contains critical and necessary information regarding emotion. The task of automatically recognizing different facial expressions in human-computer environment is significant and challenging. A variety of systems have been developed to perform facial expression recognition.[2,7,18,19,21–23,27,30,34] These systems possess some common characteristics. First, they classify facial expressions using adult facial expression databases. For instances, the authors in Refs. 4, 10, 17, 24 and 32 used the JAFFE database to recognize seven main facial expressions: happy, neutral, angry, disgust, fear, sad and surprise. Chen and Huang[5] used AR database to classify three facial expressions: neutral, smile and angry. Second, most systems conduct two stages: feature extraction and expression classification. For feature extraction, Gabor filter,[4,17,32] PCA (Principal Component Analysis),[5,10,17] and ICA (Independent Component Analysis)[4] are used. For expression classification, LDA (Linear Discriminant Analysis),[1,10,17] SVM (Support Vector Machine),[4] two-layer perceptron,[32] and HMM (Hidden Markov Model)[35] are used.

Facial expression recognition can also be applied to medical treatment of patients. Dai *et al.*[9] proposed to monitor patients on bed by utilizing the facial expression recognition to detect the status of patients. Gagliardi *et al.*[11] investigated the facial expression ability for individuals with Williams syndrome. Sprengelmeyer *et al.*[26] explored facial expression recognition of emotion for people with medicated and unmedicated Parkinson's disease.

Since JAFFE database is commonly used in measuring the performance of facial expression recognition systems, we concentrate on applying our system on this database and perform comparisons with other systems.[4,10,17,24,32] Lyons *et al.*[17] made use of Gabor filter at different scales and orientations, and applied 34 fiducial points for each convolved image to construct the feature vector for representing each facial image. After that, PCA is applied to reduce the dimensionality of feature vectors, and LDA is used to identify seven different facial expressions. Their recognition rate is 92% for JAFFE database. Zhang *et al.*[32] adopted Gabor wavelet coefficients and geometric positions to construct the feature vector for each image and applied two-layer perceptron to distinguish seven different facial expressions. Their recognition rate is 90.1%.

Buciu *et al.*[4] tested different feature extraction methods such as Gabor filter and ICA combined with SVM using three different kernels: linear, polynomial and radial basis function, to check which combination can produce the best result. From their experiments, the best recognition rate is 90.34% by using Gabor wavelet at high frequencies combined with the polynomial kernel SVM at degree 2. Dubuisson *et al.*[10] combined the sorted PCA as feature extractor with LDA as the classifier to recognize facial expressions. Their recognition rate is 87.6%. Shinohara and Otsu[24] used Higher-order Local Auto-Correlation (HLAC) features and LDA to test the performance. Unfortunately, their system is unreliable since the correct rate is only 69.4%.

In this paper, we present a systematic comparison of feature extraction and classification methods to the problem of fully automatic recognition of facial expressions and to find the optimal solution to it. The rest of this paper is organized as follows. Section 2 describes the JAFFE database and the proposed methods. Section 3 presents the experimental procedure. Section 4 provides the experimental results and performance comparisons. Finally, we draw conclusions in Sec. 5.

## 2. The JAFFE Database and the Proposed Methods

The image database we use in our experiments is the JAFFE (Japanese Female Facial Expression) database.[17] This dataset is used as the benchmark database for

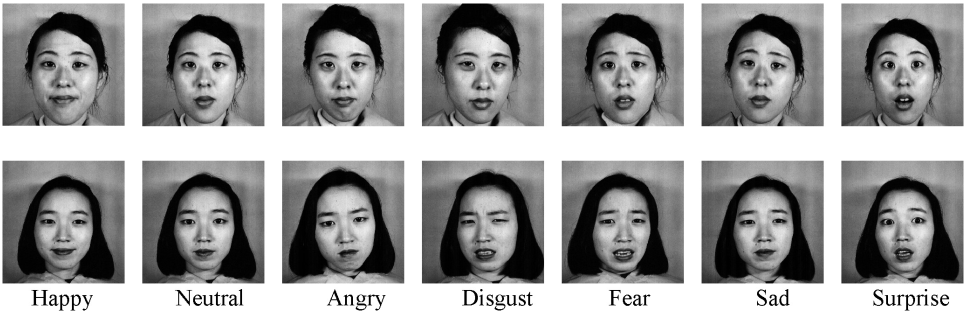



Fig. 1. Samples of two expressors containing seven different facial expressions.

researchers in Refs. 4, 10, 17, 24 and 32. The database contains ten Japanese females. There are seven different facial expressions, such as neutral, happy, angry, disgust, fear, sad and surprise. Each female has two to four examples for each expression. Totally, there are 213 grayscale facial expression images in this database. Each image is of size $256 \times 256$. Figure 1 shows two expressors comprising seven different facial expressions from the JAFFE database.

In the following, we briefly describe the methods used in facial expression recognition.

### 2.1. *Discrete Wavelet Transform (DWT)*

*Discrete Wavelet Transform* (DWT)[6,20,33] is a suitable tool for extracting image features because it allows the analysis of images on various levels of resolution. Typically, low-pass and high-pass filters are used for decomposing the original image. The low-pass filter results in an approximation image and the high-pass filter generates a detail image. The approximation image can be further split into a deeper level of approximation and detail according to different applications.

Suppose that the size of an input image is $N \times M$. At the first filtering in the horizontal direction of down-sampling, the size of images will be reduced to $N \times (M/2)$. After that, further filtering and down-sampling in the vertical direction, four subimages are obtained, each being of size $(N/2) \times (M/2)$. Figure 2 shows the sub-band decomposition of an $N \times M$ image, where H and L respectively denote high-pass and low-pass filters, and $\downarrow 2$ denotes down-sampling by a factor of 2.

The outputs of these filters are given by Eqs. (1) and (2).

$$a_{j+1}[p] = \sum_{n=-\infty}^{+\infty} l[n-2p]a_j[n] \tag{1}$$

$$d_{j+1}[p] = \sum_{n=-\infty}^{+\infty} h[n-2p]a_j[n] \tag{2}$$

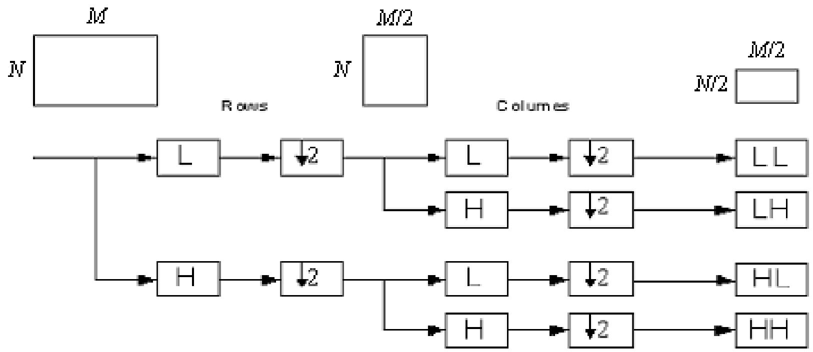



Fig. 2. Sub-band decomposition of an $N \times M$ image.

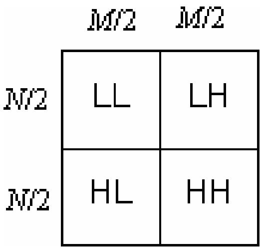


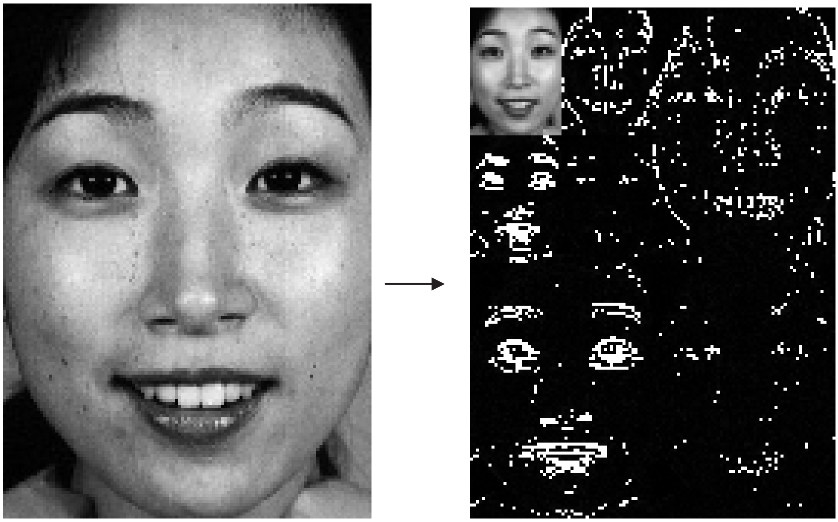

Fig. 3. First-level decomposition.

where $l[n]$ and $h[n]$ are coefficients of low-pass and high-pass filters, respectively. Accordingly, we can obtain four images denoted as LL, HL, LH and HH. The LL image is generated by two continuous low-pass filters; HL is filtered by a high-pass filter first and a low-pass filter later; LH is created using a low-pass filter followed by a high-pass filter; HH is generated by two successive high-pass filters. Figure 3 illustrates the first-level decomposition.

Every subimage can be decomposed further into smaller images by repeating the above procedure. The main feature of DWT is the multiscale representation of a function. By using the wavelets, a given image can be analyzed at various levels of resolution. Since the LL part contains most important information and discards the effect of noises and irrelevant parts, we adopt the LL part for further analysis in this paper. We extract features from the LL part of the second-level decomposition. The reasons are the LL part keeps the necessary information and the dimensionality of the image is reduced sufficiently for computation at the next stage. Figure 4 shows the two levels of DWT for a happy image from the JAFFE database.

### 2.2. *PCA and 2D-LDA*

*Principal Component Analysis* (PCA) is widely used for feature representation. In this paper, we use PCA as a comparison with the *2D Linear Discriminant Analysis*



(a) (b)

Fig. 4. Two-level decomposition of a happy image. (a) The original image and (b) the decomposed subimages.

(2D-LDA).[16] The central idea behind PCA is to find an orthonormal set of axes pointing in the direction of maximum covariance in the data.[14,28] In terms of facial images, the idea is to find the orthonormal basis vectors or the eigenvectors of the covariance matrix of a set of images, with each image being treated as a single point in a high-dimensional space.[15]

Since each image contributes to each of the eigenvectors which resemble ghost-like faces when displayed, it is referred to as *holon*[8] or *eigenface*,[28] and the new coordinates system is referred to as the *face space*. Individual images can be projected onto the face space and represented exactly as weighted combinations of the eigenface components. The resulting vector of weights that describes each face can be used in data compression and face classification. Data compression relies on the fact that the eigenfaces are ordered, with each one accounting for a different amount of variation among the faces. Compression is achieved by reconstructing images using only those few eigenfaces that account for the most variability.[25] It results in dramatic reduction of dimensionality. Classification is performed by projecting a new image onto the face space and comparing the resulting weight vector with the weight vectors of a given class.

Li and Yuan[16] introduced a new approach for image feature extraction and representation called 2D-LDA. The difference between LDA and 2D-LDA is that in LDA, we use the image vector to compute the between-class and within-class scatter matrices, but in 2D-LDA, we use the original image matrix to compute the two matrices. They claimed that the 2D-LDA can achieve better results than other feature extraction methods, such as PCA, LDA and 2D-PCA.[28] The idea of 2D-LDA is described below.

Suppose that we have $M$ training samples belonging to $L$ classes ($L_1$, $L_2, \ldots, L_L$). The training samples in each class are denoted as $N_i$ ($i = 1, 2, \ldots, L$). The size of each training image $\mathbf{A}_j$ ($j = 1, 2, \ldots, M$) is $m \times n$. Our purpose is to find a good projection vector, $\mathbf{x}$, such that when $\mathbf{A}_j$ is projected onto $\mathbf{x}$, we can obtain the projected feature vector, $\mathbf{y}_j$, of the image $\mathbf{A}_j$.

$$\mathbf{y}_j = \mathbf{A}_j\mathbf{x} \quad j = 1, 2, \ldots, M. \tag{3}$$

Similar to LDA, we can find the between-class scatter matrix, $\mathbf{TS}_{\mathrm{B}}$, and the within-class scatter matrix, $\mathbf{TS}_{\mathrm{W}}$, of the projected feature vectors by using training images. The criterion is to project the images onto a subspace that maximizes the between-class scatter and minimizes the within-class scatter of the projected data. Since the total scatter of the projected samples can be represented by the trace of the covariance matrix of the projected feature vectors, the Fisher linear projection criterion can be described as

$$\mathbf{J}(\mathbf{x}) = \frac{\mathrm{tr}(\mathbf{TS_B})}{\mathrm{tr}(\mathbf{TS_W})} = \frac{\mathbf{x}^T\mathbf{S_B}\mathbf{x}}{\mathbf{x}^T\mathbf{S_W}\mathbf{x}} \tag{4}$$

where $\mathbf{S}_B = \sum_{i=1}^{L} N_i(\bar{\mathbf{A}}_i - \bar{\mathbf{A}})^T(\bar{\mathbf{A}}_i - \bar{\mathbf{A}})$ and $\mathbf{S}_W = \sum_{i=1}^{L}\sum_{\mathbf{A}_k\in L_i}(\mathbf{A}_k - \bar{\mathbf{A}}_i)^T \times (\mathbf{A}_k - \bar{\mathbf{A}}_i)$.

The optimal projection, $\mathbf{x}_{\mathrm{opt}}$, can be decided when the criterion is maximized. That is, $\mathbf{x}_{\mathrm{opt}} = \arg\max\limits_{x} \mathbf{J}(\mathbf{x})$. We can find out the solution by solving the generalized eigenvalue problem. Therefore, if we choose $d$ optimal projection axes $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_d$ corresponding to the first $d$ largest eigenvalues, then we can extract features for a given image $\mathbf{I}$. The following equation can be used for feature extraction.

$$\mathbf{y} = \mathbf{I}[\mathbf{x}_1, \mathbf{x}_2, \ldots,\ \mathbf{x}_d]. \tag{5}$$

Since $\mathbf{I}$ is an $m \times n$ matrix and $[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_d]$ is an $n \times d$ matrix, we can form an $m \times d$ matrix, $\mathbf{y}$, to represent the original image, $\mathbf{I}$. We use the 2D-LDA representation of the original image as the input feeding into the classifier at the next stage.

### 2.3. *RBF and SVM*

*Radial Basis Function Network* (RBFN) is a two-layer, hybrid feed forward learning network. It is a fully connected network and generally used as a classification tool. It was first used for discrimination by Broomhead and Lowe[3] in 1988. In a RBF model, the first layer from input nodes to hidden neurons is an unsupervised layer and the second layer from hidden neurons to output nodes is the supervised layer. This means that the first layer connecting input nodes and hidden neurons has a unit weight and does not need training; however, the weights between hidden neurons and output nodes must be trained. The biggest difference between RBFN and traditional neural networks is that RBFN only computes its optimal weights one time but traditional neural networks have to adjust their weights for a couple of times by using back propagation algorithm. In RBFN, each hidden neuron has a symmetric radial basis function as an activation function. Typically, Gaussian-like and direct or inverse multiquadric-like radial basis functions are mostly used as activation functions. The purpose of the hidden neurons is to cluster the input data and reduce dimensionality. The objective of the RBF network is to train input data in order to minimize the sum of square errors and find the optimal weights between hidden neurons and output nodes. Those optimal weights can classify effectively the test data into correct classes. Figure 5 shows the architecture of a traditional radial basis function network.

*Support Vector Machine* (SVM)[29] is a learning system that separates a set of pattern vectors into two classes with an optimal separating hyperplane. The set of vectors is said to be optimally separated by the hyperplane if it is separated without an error and the distance between the closest vector to the hyperplane is maximal. SVM produces the pattern classifier by applying a variety of kernel functions (e.g. linear, polynomial and radial basis function) as possible sets of approximation functions by optimizing the dual quadratic programming problem, and by using structural risk minimization as the inductive principle, as opposed to classical statistical algorithms that maximize the absolute value of an error or a squared error.

The SVM is designed to handle dichotomic classes of input vectors. Recently, researchers have expanded from two-class classification to multiclass classification.[12,13] Different types of SVM are used depending upon the distribution of input patterns. A linear maximal margin classifier is used for linearly separable data, a linear soft margin classifier is used for overlapping classes, and a nonlinear classifier is used for classes that are overlapped as well as separated by nonlinear hyperplanes. Since we intend to recognize seven different facial expressions, we use tree-based one-against-one SVMs to perform multiclasses classification.

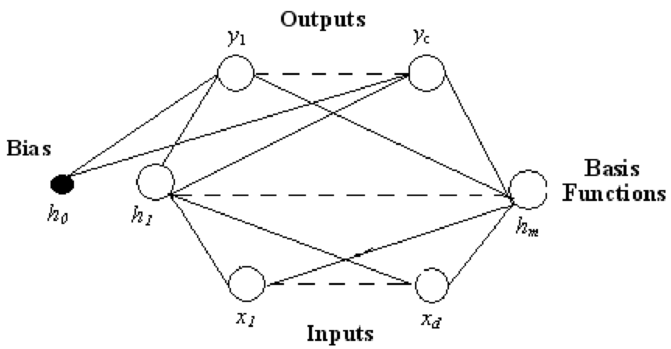



Fig. 5. The architecture of a radial basis function neural network.

## 3. Experimental Procedure

Our experimental procedure is categorized into three stages: preprocessing, feature extraction and expression classification, as illustrated in Fig. 6.

### 3.1. *Preprocessing*

For the purpose of comparisons with the methods in Refs. 4, 10, 17, 24 and 32, we crop the original image of size $256 \times 256$ into $168 \times 120$ by removing the background influences. Since the illumination condition is varied to the images in JAFFE database, we apply histogram equalization to eliminate lighting effects.

### 3.2. *Feature extraction*

We apply DWT to the cropped images two times and extract the LL area. Then 2D-LDA is used to extract important features from each image. We also test other feature representations such as PCA, LDA, ICA and 2D-PCA for comparisons.

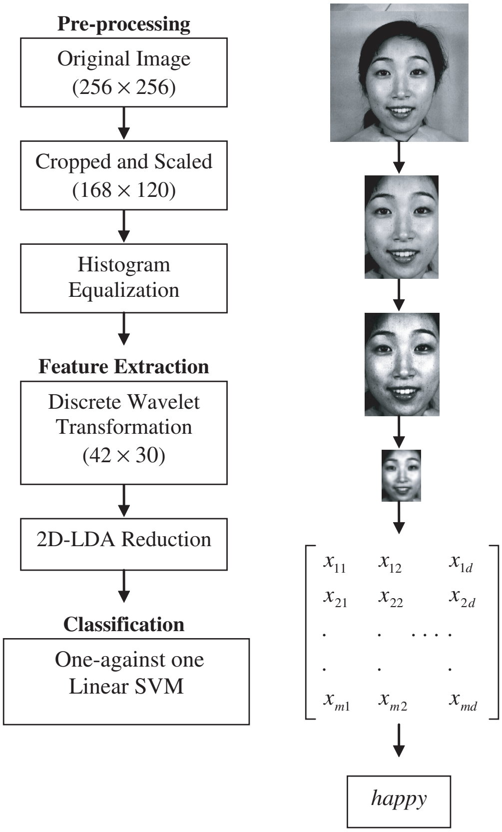



Fig. 6. The experimental procedure.

### 3.3. *Expression classification*

We adopt the linear SVM to identify seven facial expressions in the JAFFE database. To handle the multiclass problem, we construct the tree-based one-against-one SVMs as in Ref. 13. Different kernel types of SVM such as linear, polynomial and radar basis function are also tested at this stage to compare the performance with the linear SVM.

## 4. Experimental Results and Performance Comparisons

We apply cross-validation strategy as in Refs. 17, 24 and 32 and leave-one-out strategy as in Ref. 4 to perform comparisons with other existing systems. For the cross-validation strategy, we randomly divide the database into ten segments in terms of different facial expressions. Each time, we train nine out of the ten segments and test the remaining segment. We perform the same procedure of training and testing repeatedly for 30 times. At last, we average all the 30 recognition rates to obtain the final performance of the proposed system. For the leave-one-out strategy, each time we only test one image in each class, and the remaining images are used for training.

The experimental results show that our system can successfully meet the criteria of accuracy and efficiency for identifying different facial expressions. For accuracy, our proposed method can outperform other existing systems based on the same database. The recognition rate of the proposed system is 95.71% by using leave-one-out strategy and 94.13% by using cross-validation strategy. For efficiency, it takes only 0.0357 second to process one input image of size $256 \times 256$.

Table 1 shows the performance comparisons among our proposed system and the existing systems using the same JAFFE database. From Table 1, we observe that no matter which strategy is used, the proposed system outperforms the others. We also test the effects of different kernels of SVM, such as polynomial and radial basis functions. In our experiments, linear SVM is most suitable for the JAFFE database. It is because the feature vectors extracted by 2D-LDA are clustered in each class and can be separated by the linear SVM. The results are shown in Table 2. From our experimental results, we observe that 2D-LDA is superior to PCA, LDA and 2D-PCA as the feature extraction method. We also compare the performance among different feature extractions with different classifiers.

Table 1. The performance comparisons in JAFFE database.

| The Existing Systems | Strategy | Generalization Rate |
|---|---|---|
| Lyons *et al.*[17] | Cross-validation | 92.00% |
| Zhang *et al.*[32] | Cross-validation | 90.10% |
| Buciu *et al.*[4] | Leave-one-out | 90.34% |
| Dubuisson *et al.*[10] | | 87.60% |
| Shinohara and Otsu[24] | Cross-validation | 69.40% |
| Our proposed system | Cross-validation | 95.71% |
| | Leave-one-out | 94.13% |

Table 2. The performance comparisons of using different kernels of SVM.

| Kernel Functions | Recognition Rates | |
|---|---|---|
| | Cross Validation | Leave-One-Out |
| Linear | 94.13% | 95.71% |
| Polynomial with degree 2 | 91.43% | 92.38% |
| Polynomial with degree 3 | 92.86% | 94.29% |
| Polynomial with degree 4 | 92.22% | 93.33% |
| Radial basis function | 85.71% | 87.14% |

Table 3. The comparisons of PCA, LDA, 2D-PCA, ICA and 2D-LDA.

| Feature Extraction Methods | Recognition Rates | | Testing Speed per Image (second) |
|---|---|---|---|
| | Cross Validation | Leave-One-Out | |
| LDA + SVM | 91.27% | 91.90% | 0.0367 |
| 2D-PCA + SVM | 92.06% | 93.33% | 0.0357 |
| ICA + SVM | 93.35% | 93.81% | 0.0359 |
| PCA + SVM | 93.43% | 94.84% | 0.0353 |
| 2D-LDA + SVM | 94.13% | 95.71% | 0.0357 |

Table 4. The comparisons of PCA, LDA, 2D-PCA, ICA and 2D-LDA.

| Feature Extraction Methods | Recognition Rates | | Testing Speed per Image (second) |
|---|---|---|---|
| | Cross Validation | Leave-One-Out | |
| LDA + RBFN | 26.67% | 27.71% | 0.0351 |
| 2D-PCA + RBFN | 25.24% | 26.67% | 0.0349 |
| ICA + RBFN | 27.14% | 27.43% | 0.0347 |
| PCA + RBFN | 36.67% | 37.21% | 0.0346 |
| 2D-LDA + RBFN | 37.14% | 37.71% | 0.0347 |

Table 5. Confusion matrix using cross-validation strategy.

| | Angry | Disgust | Fear | Happy | Neural | Sad | Surprise | Total |
|---|---|---|---|---|---|---|---|---|
| Angry | 30 | | | | | | | 30 |
| Disgust | 1 | 28 | | | | | | 29 |
| Fear | | 2 | 27 | | 1 | | 2 | 32 |
| Happy | | | | 30 | 1 | | | 31 |
| Neural | | | | | 30 | | | 30 |
| Sad | | | 3 | 1 | | 27 | | 31 |
| Surprise | | | | 2 | | | 28 | 30 |
| Total | | | | | | | | 213 |

In Tables 3 and 4, we show performance comparisons of using PCA, LDA, 2D-PCA, ICA and 2D-LDA with SVM and RBFN. We observe that SVM is the best classifier; however, RBFN is unreliable since the recognition rate is unsatisfied. In Tables 5 and 6, we show the confusion matrices of the correct and false numbers under cross-validation and leave-one-out strategies for facial expressions. All the experiments are processed in the Matlab 7 environment under XP and Pentium

Table 6. Confusion matrix using leave-one-out strategy.

| | Angry | Disgust | Fear | Happy | Neural | Sad | Surprise | Total |
|---|---|---|---|---|---|---|---|---|
| Angry | 30 | | | | | | | 30 |
| Disgust | 1 | 28 | | | | | | 29 |
| Fear | | 1 | 29 | | | 1 | 1 | 32 |
| Happy | | | | 29 | | 1 | 1 | 31 |
| Neural | | | | | 30 | | | 30 |
| Sad | | | 2 | 1 | | 28 | | 31 |
| Surprise | | | | | | | 30 | 30 |
| Total | | | | | | | | 213 |

IV with 2.80 GHz. The speed of the proposed method is fast. It only takes about 0.0357 second to process a testing image.

## 5. Conclusions

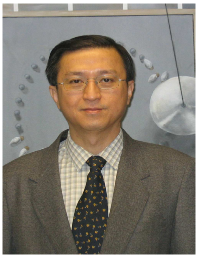
In this paper, we investigate various feature representation and expression classification schemes to recognize seven different facial expressions on the JAFFE database. Experimental results show that the proposed system using DWT, 2D-LDA and linear one-again-one SVMs outperforms others. The recognition rate of the system is 95.71% by using leave-one-out strategy and 94.13% by using cross-validation strategy. It takes only 0.0357 second to process one input image of size $256 \times 256$.

## References

1. B. Abboud, F. Davoine and M. Dang, Statistical modeling for facial expression analysis and synthesis, *Proc. IEEE Int. Conf. Image Processing* (2003), pp. 653–656.
2. M. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel and J. Movellan, Recognizing facial expression: machine learning and application to spontaneous behavior, *Proc. IEEE Conf. Computer Vision and Pattern Recognition* (2005), pp. 568–573.
3. D. S. Broomhead and D. Lowe, Multivariable functional interpolation and adaptive networks, *Compl. Syst.* **2** (1988) 321–355.
4. I. Buciu, C. Kotropoulos and I. Pitas, ICA and Gabor representation for facial expression recognition, *Proc. IEEE Int. Conf. Image Processing* (2003), pp. 855–858.
5. X.-W. Chen and T. Huang, Facial expression recognition: a clustering-based approach, *Patt. Recog. Lett.* **24** (2003) 1295–1302.
6. J.-T. Chien and C.-C. Wu, Discriminant waveletfaces and nearest feature classifiers for face recognition, *IEEE Trans. Patt. Anal. Mach. Intell.* **24** (2002) 1644–1649.
7. C. Chuang and F. Y. Shih, Recognizing facial action units using independent component analysis and support vector machine, *Patt. Recog.* **39** (2006) 1795–1798.
8. G. Cottrell and M. Fleming, Face recognition using unsupervised feature extraction, *Proc. Int. Conf. Neural Networks* (1990), pp. 322–325.
9. Y. Dai, Y. Shibata, T. Ishii, K. Katamachi, K. Noguchi, N. Kakizaki and D. Cai, An associate memory model of facial expression and its application in facial expression recognition of patients on bed, *Proc. IEEE Int. Conf. Multimedia and Expo* (2001), pp. 772–775.

10. S. Dubuisson, F. Davoine and M. Masson, A solution for facial expression representation and recognition, *Sign. Process.: Imag. Commun.* **17** (2002) 657–673.
11. C. Gagliardi, E. Frigerio, D. Buro, I. Cazzaniga, D. Pret and R. Borgatti, Facial expression recognition in Williams syndrome, *Neuropsicologia* **41** (2003) 733–738.
12. G. Guo, S. Li and K. Chan, Support vector machines for face recognition, *Imag. Vis. Comput.* **19** (2001) 631–638.
13. C.-W. Hsu and C.-J. Lin, A comparison of methods for multiclass support vector machines, *IEEE Trans. Neural Networks* **13** (2002) 415–425.
14. M. Kirby and L. Sirovich, Applications of the Karhunen-Loeve procedure for the characterization of human faces, *IEEE Trans. Patt. Anal. Mach. Intell.* **12** (1990) 103–108.
15. M. Kosugi, Human-face search and location in a scene by multi-pyramid architecture for personal identification, *Syst. Comput. Japan* **26** (1995) 27–38.
16. M. Li and B. Yuan, 2D-LDA: a novel statistical linear discriminant analysis for image matrix, *Patt. Recogn. Lett.* **26** (2005) 527–532.
17. M. Lyons, J. Budynek and S. Akamatsu, Automatic classification of single facial images, *IEEE Trans. Patt. Anal. Mach. Intell.* **21** (1999) 1357–1362.
18. M. Monwar, Machine learning techniques in pain recognition, M.S. thesis, Computer Science, The University of Northern British Columbia (December 2006).
19. M. Monwar and S. Rezaei, Pain recognition using artificial neural network, *Proc. IEEE Int. Symp. Signal Processing and Information Technology*, Vancouver, Canada (2006), pp. 28–33.
20. E. Morales and F. Y. Shih, Wavelet coefficients clustering using morphological operations and pruned quadtrees, *Patt. Recogn.* **33** (2000) 1611–1620.
21. F. Y. Shih and C. Chuang, Automatic extraction of head and face boundaries and facial features, *Inform. Sci.* **158** (2004) 117–130.
22. F. Y. Shih, Y. Fu and K. Zhang, Multi-view face identification and pose estimation using B-spline interpolation, *Inform. Sci.* **169** (2005) 189–204.
23. F. Y. Shih, K. Zhang and Y. Fu, A hybrid two-phase algorithm for face recognition, *Patt. Recogn. Artif. Intell.* **18** (2004) 1423–1435.
24. Y. Shinohara and N. Otsu, Facial expression recognition using Fisher weight maps, *Proc. IEEE Int. Conf. Automatic Face and Gesture Recognition* (2004), pp. 499–504.
25. L. Sirovich and M. Kirby, Low-dimensional procedure for the characterization of human faces, *J. Opt. Soc. Amer.* **4** (1987) 519–524.
26. R. Sprengelmeyer, A. Young, K. Mahn, U. Schroeder, D. Woitalla, T. Butter, W. Kuhn and H. Przuntek, Facial expression recognition in people with medicated and unmedicated Parkinson's disease, *Neuropsychologia* **41** (2003) 1047–1057.
27. C. Thomaz, D. Gillies and R. Feitosa, Using mixture covariance matrices to improve face and facial expression recognitions, *Patt. Recogn. Lett.* **24** (2003) 2159–2165.
28. M. Turk and A. Pentland, Eigenfaces for recognition, *J. Cogn. Neurosci.* **3** (1991) 71–86.
29. V. N. Vapnik, *The Nature of Statistical Learning Theory* (Springer-Verlag, New York, 1995).
30. P. S. Wang and S. N. Yanushkevich, Biometric technologies and applications, *Proc. IASTED Int. Conf. Artificial Intelligence and Applications*, Innsbruck, Austria (2007), pp. 226–231.
31. J. Yang, D. Zhang, A. Frangi and J.-Y. Yang, Two-dimensional PCA: a new approach to appearance-based face representation and recognition, *IEEE Trans. Patt. Anal. Mach. Intell.* **26** (2004) 131–137.

32. Z. Zhang, M. Lyons, M. Schuster and S. Akamatsu, Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron, *Proc. IEEE Int. Conf. Automatic Face and Gesture Recognition* (1998), pp. 454–459.
33. B.-L. Zhang, H. Zhang and S. Ge, Face recognition by applying wavelet subband representation and kernel associative memory, *IEEE Trans. Neural Networks* **15** (2004) 166–177.
34. D. Zhang, X. You, Q. Chen, P. S. Wang and Y. Y. Tang, Face representation by using non-tensor product wavelets, *Proc. Int. Conf. Pattern Recognition*, Hong Kong (2006), pp. 503–506.
35. Y. Zhu, L. C. De Silva and C. C. Ko, Using moment invariants and HMM in facial expression recognition, *Patt. Recogn. Lett.* **23** (2002) 83–91.



**Frank Y. Shih** received the B.S. degree from National Cheng-Kung University, Taiwan, in 1980, the M.S. degree from the State University of New York at Stony Brook, in 1984, and the Ph.D. degree from Purdue University, West Lafayette, Indiana, in 1987, all in electrical engineering. He is presently a professor jointly appointed in the Department of Computer Science, the Department of Electrical and Computer Engineering, and the Department of Biomedical Engineering at New Jersey Institute of Technology, Newark, NJ. He currently serves as the Director of Computer Vision Laboratory.

Dr. Shih is currently on the Editorial Board of the *Int. J. Pattern Recognition*, the *Int. J. Pattern Recognition Letters*, the *Int. J. Pattern Recognition and Artificial Intelligence*, the *Int. J. Recent Patents on Engineering*, the *Int. J. Recent Patents on Computer Science*, the *Int. J. Internet Protocol Technology*, and the *J. Internet Technology*. Dr. Shih has contributed as a steering member, committee member, and session chair for numerous professional conferences and workshops. He was the recipient of the Research Initiation Award from the National Science Foundation in 1991. He won the Honorable Mention Award from the International Pattern Recognition Society for Outstanding Paper and also won the Best Paper Award in the International Symposium on Multimedia Information Processing. He has received several awards for distinguished research at New Jersey Institute of Technology. He has served several times on the Proposal Review Panel of the National Science Foundation.

Dr. Shih holds the research fellow for the American Biographical Institute and the IEEE senior membership. He has published one book (Digital Watermarking and Steganography, CRC Press, 2007), seven book chapters and over 190 technical papers, including 92 in well-known prestigious journals.

His current research interests include image processing, computer vision, sensor networks, pattern recognition, bioinformatics, information security, robotics, fuzzy logic, and neural networks.

**Chao-Fa Chuang** received MBA degree from National Chung Hsing University, Taiwan, and Ph.D. degree in computer science from New Jersey Institute of Technology in January 2006. He is now working as a system analyst at the enfoTech & Consulting, Inc.

His research interests include image processing, pattern recognition, artificial intelligence and machine learning.

**Patrick S. P. Wang** is tenured full professor of computer science at College of Computer and Information Science of Northeastern University, Boston, MA, USA. He is *Fellow* of IAPR (International Association of Pattern Recognition) and co-founding chief editor of IJPRAI (*International Journal of Pattern Recognition and Artificial Intelligence*) and WSP MPAI Book Series. Dr. Wang has been elected as Otto-Von-Guericke Distinguished Guest Professor at the Graphics and Simulations Lab of the University of Magdeburg (near Berlin) of Deutschland (Germany). Dr. Wang has been visiting at University of Calgary's Biometrics Technology Laboratory, Calgary, Canada as *iCORE* (Informatics Circle of Research Excellence) visiting Professor, where he published his 23rd book in July 2007, entitled: *Image Pattern Recognition — Synthesis and Analysis in Biometrics* by World Scientific Publishing Co. He is now *Zi-Jiang Visiting Chair Professor* at the Department of Computer Science and Technology, East China Normal University, Shanghai, China. Professor Wang has also served as Honorary Advisor Professor of Xiamen University and Sichuan University of China.

Dr. Wang got his B.S. in electronic engineering from National Chiao Tung University (Hsin Chu), M.S. in electrical engineering from National Taiwan University, M.S. in information and computer science from Georgia Institute of Technology, Ph.D. in computer science from Oregon State University. He has been Visiting Scientist at MIT AI Lab, and Research Consultant at MIT Sloan School, as well as adjunct faculty of Computer Science at Harvard University Extension School. For the past decades, he has been invited to give lectures in many countries in America, Asia and Europe. He has served as the Chair of IAPR-SSPR TC-2 (Structural and Syntactic Pattern Recognition) for 4 years (1992–1996). He has also organized many international conferences, including Data Mining (MLDM2003 and 2005) at Leipzig and BT2004, Int. Workshop on Biometric Technology, 2004, Calgary, Canada. Dr. Wang was general conference co-chair of the IAPR's 18th ICPR International Conference on Pattern Recognition, August, 2006 at Hong Kong, China. Prof. Wang serves as Program Vice-Chair and Distinguished International Advisor of the *IEEE-BIBE2007* at Harvard Medical. He gave a keynote plenary lecture on *Biometrics Intelligence Information Systems and Applications* and was honored and received a Plaque of *Outstanding Achievement Award* from Conference General Chair Prof. Jack Yang and IEEE-SMC President Prof. Larry Hall. He also gave a tutorial *on IPR and Biometrics* at the *IEEE-SMC 2007*, Montreal, Canada, October 2007. Dr. Wang was also invited to give a keynote speech and tutorial at IAS 2007, Manchester, UK, and e-Forensics 2008 at Adelaide, Australia.

His other literature works on music and poems have received many awards including the prestigious World Journal Prose Competition Award, and his book *Harvard Meditation Melody* (哈佛冥想曲) has won the prestigious Taiwan Provincial Better Publication Award.