

# A Comprehensive Empirical Study on Linear Subspace Methods for Facial Expression Analysis

Caifeng Shan, Shaogang Gong, and Peter W. McOwan  
Department of Computer Science, Queen Mary, University of London  
Mile End Road, London E1 4NS  
{cfshan, sgg, pmco}@dcs.qmul.ac.uk

## Abstract

*Automatic facial expression analysis is a vital component of intelligent Human-Computer Interaction (HCI). In this paper, we present a extensive empirical study on linear subspace methods for facial expression analysis. Locality Preserving Projections (LPP) and Orthogonal Neighborhood Preserving Projections (ONPP) are first time applied to facial expression analysis. We systematically examine a number of linear subspace methods, and show that, in our comparative studies, the Supervised LPP (SLPP) is superior in supervised methods, while ONPP performs best in unsupervised learning.*

## 1. Introduction

Automatic facial expression analysis is a vital component of intelligent Human-Computer Interaction (HCI), and has attracted much attention in recent years (see, for instance, the surveys [14, 6]).

Many techniques have been proposed to analyze facial expressions, such as Neural Networks [19], Bayesian Networks [4], and Support Vector Machines [1], to name just a few. One of the successful classes of methods is appearance-based statistical methods. Linear subspace methods including Principal Component Analysis (PCA) [20], Linear Discriminant Analysis (LDA) [2], and Independent Component Analysis (ICA) [5] have been introduced for facial expression analysis. In Lyons et al's work [12], PCA + LDA was adopted to classify facial images based on the Gabor wavelet features. Donato et al [5] explored PCA, LDA, ICA, and other techniques for facial action classification. Recently some new linear subspace methods [7, 11] have been developed, and have been applied for face recognition [8, 10]. However, it is still unknown whether they are powerful for facial expression analysis. Moreover, until now, there is no comprehensive comparison of these linear subspace methods for facial expression analysis, although it

is very necessary and important for further study.

In this paper, we survey and compare a number of linear subspace methods for facial expression analysis, which include PCA [20], LDA [2], Locality Preserving Projections (LPP) [7] (unsupervised and supervised manners), and Orthogonal Neighborhood Preserving Projections (ONPP) [11] (unsupervised and supervised manners). These methods are investigated using different appearance features on two public databases.

## 2. Linear Subspace Methods

The generic problem of linear dimensionality reduction is the following. Given a set  $x_1, x_2, \dots, x_m$  in  $R^n$ , find a transformation matrix  $W$  that maps these  $m$  points to  $y_1, y_2, \dots, y_m$  in  $R^l$  ( $l \ll n$ ), where  $y_i = W^T x_i$ .

### 2.1. Principle Component Analysis

PCA [20] aims to extract a subspace in which the variance is maximized. The objective function is as follows:

$$\max_{\mathbf{w}} \sum_i (y_i - \bar{y})^2 \quad \text{where} \quad \bar{y} = \frac{1}{m} \sum_i y_i \quad (1)$$

The solution can be obtained by taking the eigenvector of the sample covariance matrix associated with leading eigenvalues.

### 2.2. Linear Discriminant Analysis

LDA [2] seeks directions that are efficient for discrimination. Suppose the data samples belong to  $l$  classes, The objective function is as follows:

$$\max_{\mathbf{w}} \frac{\mathbf{w}^T S_B \mathbf{w}}{\mathbf{w}^T S_W \mathbf{w}} \quad (2)$$

$$S_B = \sum_{i=1}^l n_i (\mathbf{m}^{(i)} - \mathbf{m})(\mathbf{m}^{(i)} - \mathbf{m})^T \quad (3)$$

$$S_W = \sum_{i=1}^l (\sum_{j=1}^{n_i} (x_j^{(i)} - \mathbf{m}^{(i)})(x_j^{(i)} - \mathbf{m}^{(i)})^T) \quad (4)$$

where  $\mathbf{m}$  is the mean of all the samples,  $n_i$  is the number of samples in the  $i$ th class,  $\mathbf{m}^{(i)}$  is the average vector of the  $i$ th class, and  $x_j^{(i)}$  is the  $j$ th sample in the  $i$ th class.

### 2.3. Locality Preserving Projections

LPP [7] optimally preserves local neighborhood information by building a neighborhood graph. It employs the same objective function with Laplacian Eigenmaps [3], a nonlinear technique for dimensionality reduction:

$$\min_{\mathbf{w}} \sum_{i,j} (\mathbf{w}^T x_i - \mathbf{w}^T x_j)^2 S_{ij} \quad (5)$$

where  $S$  is the local similarity matrix. The objective function can be reduced to:

$$\begin{aligned} & \frac{1}{2} \sum_{i,j} (\mathbf{w}^T x_i - \mathbf{w}^T x_j)^2 S_{ij} \quad (6) \\ &= \sum_i \mathbf{w}^T x_i D_{ii} x_i^T \mathbf{w} - \sum_{i,j} \mathbf{w}^T x_i S_{ij} x_j^T \mathbf{w} \\ &= \mathbf{w}^T X(D - S)X^T \mathbf{w} = \mathbf{w}^T X L X^T \mathbf{w} \quad (7) \end{aligned}$$

where  $X = [x_1, x_2, \dots, x_m]$ ,  $L = D - S$  is a Laplacian matrix, and  $D_{ii} = \sum_j S_{ji}$ . Matrix  $D$  provides a natural measure on the data points. A constraint is imposed as follows:

$$\mathbf{y}^T D \mathbf{y} = 1 \Rightarrow \mathbf{w}^T X D X^T \mathbf{w} = 1 \quad (8)$$

The solution is given by the minimum eigenvalue solution to the generalized eigenvalue problem:

$$X L X^T \mathbf{w} = \lambda X D X^T \mathbf{w} \quad (9)$$

When class information is available, LPP can be performed in a supervised manner. In our previous work [16], we proposed a Supervised LPP (SLPP) to encode more discriminative power than the original LPP for improved classification capability. SLPP preserves class information when constructing the neighborhood graph such that the local neighborhood of a sample is composed of samples from the same class only.

### 2.4. Orthogonal Neighborhood Preserving Projections

OONP [11] aims to preserve the intrinsic geometry of the local neighborhoods, as does Locally Linear Embedding (LLE) [15], a nonlinear dimensionality reduction technique. In the first step, ONPP constructs a weighted  $k$ -nearest neighbor graph which models explicitly the data topology, and computes the optimal weights in each neighborhood. Each data sample  $x_i$  is reconstructed by a linear combination of its  $k$  nearest neighbors. The reconstruction errors are measured by minimizing the objective function

$$\epsilon(V) = \sum_i |x_i - \sum_j V_{ij} x_j|^2 \quad (10)$$

In the second step, ONPP computes an explicit linear mapping from the input space to the reduced space. ONPP imposes the constraint that each data sample in the reduced space is reconstructed from its neighbors by the same weights used in the input space. This leads to the solution of the following optimization problem, where  $Y = [y_1, y_2, \dots, y_m]$  and  $M = (I - V^T)(I - V)$

$$\min_Y \sum_i |y_i - \sum_j V_{ij} y_j|^2 \quad (11)$$

$$= \min_W |W^T X(I - V^T)|^2 \quad (12)$$

$$= \min_W \text{tr}(W^T X M X^T W) \quad (13)$$

By imposing the additional constraint that the columns of  $W$  are orthonormal, the solution to the above optimization problem is the basis of the eigenvectors associated with the  $d$  smallest eigenvalues of  $X M X^T$ .

ONPP can be performed in either an unsupervised or a supervised setting [11]. In the supervised setting, the adjacent data samples in the nearest neighbor graph belong to the same class. So there is no need to set parameter  $k$ , and the supervised ONPP (SONPP) becomes fully automatic.

## 3. Experiments

### 3.1. Data Set

We conducted experiments on two public databases: (1) The **Cohn-Kanade Database** [9] consists of 100 university students in age from 18 to 30 years, of which 65% were female, 15% were African-American, and 3% were Asian or Latino. Subjects were instructed to perform a series of 23 facial displays, six of which were prototypic emotions. Image sequences from neutral face to target display were digitized into  $640 \times 490$  pixel arrays. (2) The **JAFFE Database** [12] consists of 213 images of Japanese female facial expression. Ten expressers posed 3 or 4 examples for each of the seven basic expressions (six emotional expressions plus neutral face). The image size is  $256 \times 256$  pixels.

Three data sets were constructed: (1) **S1**: 320 image sequences were selected from the Cohn-Kanade Database. The only selection criterion is that a sequence can be labeled as one of the six basic emotions. The sequences come from 96 subjects, with 1 to 6 emotions per subject. The three peak frames of each sequence were used. (2) **S2**: for the selected 320 image sequences, the neutral face and three peak frames of each sequence were used. (3) **S3**: all 213 images of the JAFFE database were used. The three data sets are summarized in Table 1. As **S3** has much fewer images than **S1** and **S2**, we will investigate the effect of the training set size in our experiment.

Following the methodology of Tian [18], we normalized the faces to a fixed distance between the two eyes. Facial images of  $110 \times 150$  pixels were cropped from original

Data Set	Images	Subjects	Expressions
<b>S1</b>	960	96	six
<b>S2</b>	1280	96	seven
<b>S3</b>	213	10	seven

Table 1. Three data sets for facial expression analysis.

frames based on the two eyes location. No further alignment of facial features such as alignment of mouth, or removal of illumination changes [18] was performed in our experiments.

### 3.2. Appearance Features

To perform facial expression analysis, it is necessary to derive an effective facial representation from original face images. Gabor-wavelet representations have been widely adopted to describe appearance changes of faces [12, 18, 1]. However, the computation is both time and memory intensive. We previously introduced boosted Local Binary Patterns (LBP) features as low-cost appearance features for facial expression analysis [17]. The most important properties of the LBP operator [13] are its tolerance against illumination changes and its computational simplicity.

In our experiments, three kinds of appearance features were compared: raw image data (IMG), LBP features extracted from equally divided sub-regions (LBP) [17], and Boosted LBP features (BoostLBP) [17]. On IMG features, for computational efficiency, we down-sampled face images to  $55 \times 75$  pixels, and represented each image with a  $4,125(55 \times 75)$ -dimensional vector. For LBP features, we divided facial images into 42 sub-regions in total. The 59-bin  $LBP_{8,2}^{u_2}$  operator [17] was applied to each sub-region; so each image was represented by a LBP histogram with length of 2,478 ( $59 \times 42$ ). With regard to BoostLBP features, by shifting and scaling a sub-window, 16,640 LBP features in total were extracted from each face image; then AdaBoost was adopted to select the effective LBP features. AdaBoost training continued until the classifier output distribution for the positive and negative samples were completely separated (See Fig. 1 for the selected LBP features in S1).

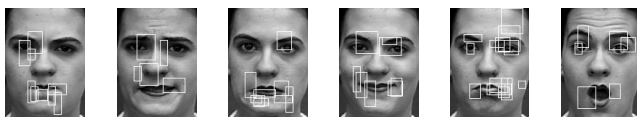


Figure 1. The LBP features selected by AdaBoost for each basic emotion: Anger, Disgust, Fear, Joy, Sadness, and Surprise.

### 3.3. Experimental Results

We applied PCA, LPP, ONPP, LDA, SLPP, and SONPP to learn the subspace of facial expression in the feature space of IMG, LBP and BoostLBP. The 2D visualization of embedded subspaces are shown in Fig. 2 - Fig. 4. It is observed that different expressions are heavily overlapped in

the 2D subspaces generated by unsupervised methods PCA, LPP, and ONPP (with all three facial features). It is easily understood that the projections of PCA are spread out since PCA aims at maximizing the variance. With regard to LPP and ONPP, although they preserve local neighborhood information, as expression images contain complex variations and significant overlapping among different classes, it is difficult for them to yield meaningful projections in the absence of class information. For supervised methods, it is surprising to observe that different expressions are still heavily overlapped in the 2D subspace derived by SONPP. In contrast, LDA and SLPP yields much meaningful projections since images of the same class are mapped close to each other. SLPP seems to provide the best projections since different classes are well separated. This is because SLPP preserves the locality and class information simultaneously in the projections. On the other hand, LDA discovers only the Euclidean structure, and can not see the underlying nonlinear manifold that expression images lie on; so its discriminating power is limited. The results obtained by SLPP also reflect the human observation that Joy and Surprise can be clearly separated, but Anger, Disgust, Fear and Sadness are easily confused, and so reenforce the findings of other published work [18, 4]. On comparing appearance features, BoostLBP seems to provide the best performance, as the clusters appear more cohesive and clear in the SLPP subspace, while IMG's performance is worst.

For facial expression recognition, we adopted the nearest-neighbor classifier for its simplicity. The Euclidean metric was used as our distance measure. The number of nearest neighbors was set to be 7 or 11 according to the size of the training set. To evaluate the algorithms' generalization ability, we adopted a 10-fold cross-validation test scheme. The average recognition results (with the standard deviation) are shown in Table 2 - Table 4. Note that, for LDA, the dimension of the reduced subspace is at most  $c - 1$ , where  $c$  is the number of classes. In general, the performance of the PCA, LPP, ONPP, SLPP and SONPP methods varies with the dimension of subspace. We show the best results obtained by these methods. It is observed that on the three data sets, for unsupervised methods, ONPP performs much better than PCA and LPP, and PCA outperforms LPP, with all three facial features. With regard to supervised methods, on the three data sets with the three facial representations, it is seen that SLPP is the best method, with a clear margin of superiority over LDA (12-38% better) and SONPP (11-64% better), while LDA's performance is 4-19% better than SONPP (an exception is that SONPP performs 2% better than LDA with BoostLBP features on S3). It is observed that SONPP does not always work better than ONPP, e.g., SONPP has inferior performance to ONPP on S1 (with LBP and BoostLBP features) and S2. This reenforces the observation on the embedded subspaces shown in

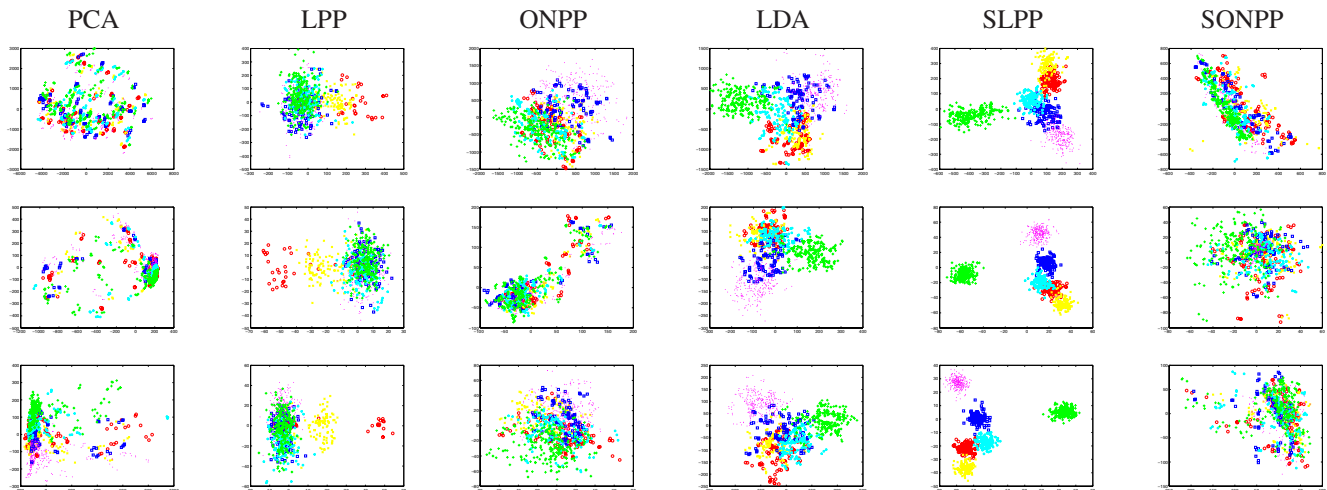


Figure 2. Images of **S1** are mapped into the 2D embedding spaces. (In Fig. 2 - Fig. 4, different expressions are represented by different color: red circle: Anger; yellow x-mark: Disgust; blue square: Fear; magenta point: Joy; cyan star: Sadness; green plus: Surprise; black pentagram: Neutral. Top row: **IMG**; Middle row: **LBP**; Bottom row: **BoostLBP**.)

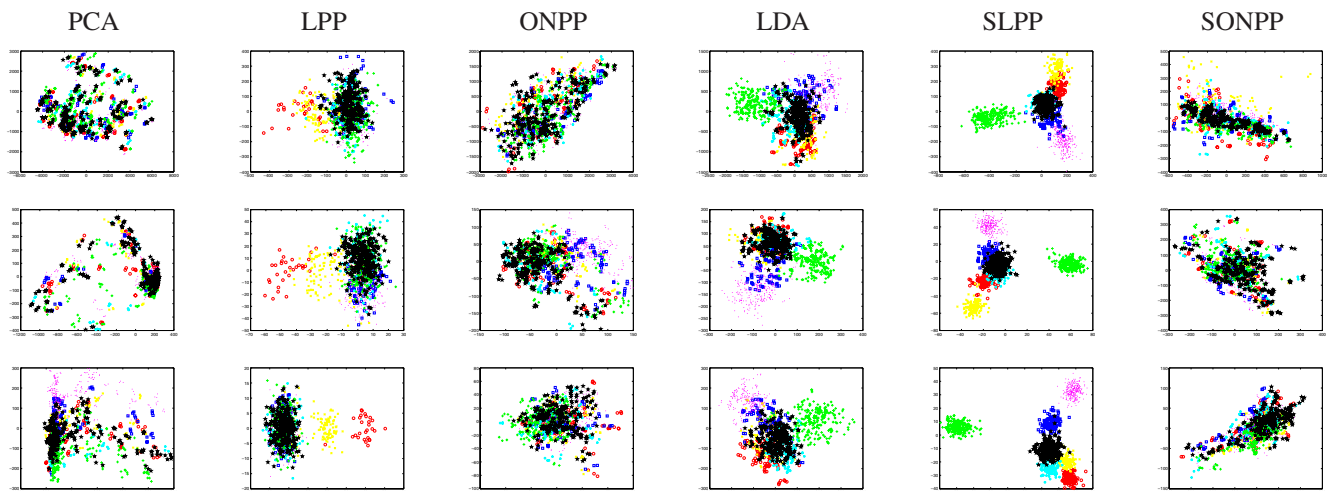


Figure 3. Images of **S2** are mapped into the 2D embedding spaces.

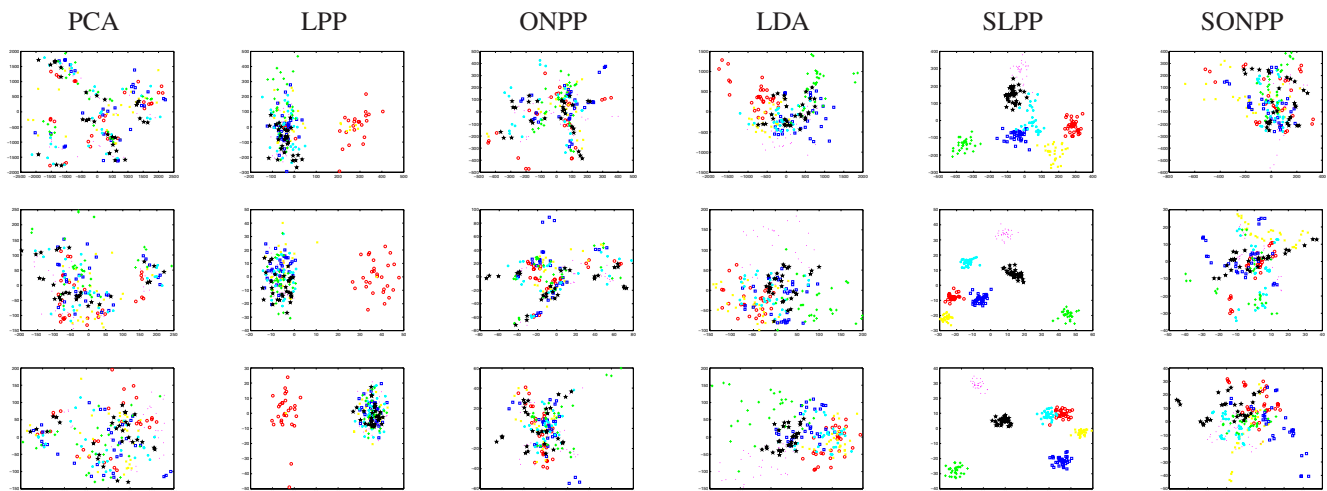


Figure 4. Images of **S3** are mapped into the 2D embedding spaces.



Fig 2 - Fig. 4. We can draw the conclusion that the label information used in SONPP does not provide it with more discriminative power than ONPP for facial expression analysis. In contrast, LDA and SLPP performs robustly better than the unsupervised methods PCA, LPP, and ONPP (LDA and ONPP perform comparably on S3).

The best result of 94.7% in 6-class expression recognition, achieved by BoostLBP based SLPP, is the best performance reported so far on the database. Previously Tian achieved 94% performance using a three-layer neural networks when combining geometric features and Gabor wavelet features [18]. With regard to 7-class expression, BoostLBP based SLPP achieves the best performance of 92.0%, which is also very encouraging given that previously published recognition rates on this database were 81-83% [4]. The confusion matrix of 7-class facial expression on S2 is shown in Table 5, which shows that most confusion occurs between Anger, Fear, Sadness, and Neutral.

	Anger	Disgust	Fear	Joy	Sadness	Surprise	Neutral
Anger	84.3%	2.8%	0	0	5.5%	0	7.4%
Disgust	0	97.5%	2.5%	0	0	0	0
Fear	0	0	76.8%	13.1%	1.0%	2.0%	7.1%
Joy	0	0	0	96.5%	0	0	3.5%
Sadness	2.4%	0	0.8%	0	83.3%	0	13.5%
Surprise	0	0	1.3%	0	0.4%	95.6%	2.7%
Neutral	0.8%	0	0	2.0%	2.4%	0	94.8%

Table 5. Confusion matrix of 7-class expression recognition on S2.

Recognition performance on S3 is much poorer than that on S1 and S2, as there are fewer images in the data set resulting in a poor sampling of the underlying latent space. The effect of the training set size is also reflected on the standard deviation of 10-fold cross-validation. The standard deviations of S3 are much larger than those of S1 and S2. So the recognition performance of linear subspace methods on the small training sets is not robust and reliable. On comparing the standard deviation of 10-fold cross validation on S1 and S2, SLPP produces the smallest deviation (an exception is LDA has the smallest deviation with IMG features on S2). This demonstrates that SLPP is much robust than other methods on the relatively large data set.

On comparing facial features, as shown in Fig. 5, BoostLBP features performs consistently and robustly better than LBP and IMG features. LBP features outperforms IMG feature; however, an exception is that, in LPP subspaces, IMG features have slightly better performance than LBP features.

The plots in Fig. 6 show the averaged recognition rates versus dimensionality reduction of different subspace methods with BoostLBP features. The performance difference between SLPP and LDA are conspicuous when the dimension of subspace is small. But when the dimension continues increasing, their performance become similar. For PCA, ONPP, and SONPP, when the dimension continues increasing, their performance become stable, and gradually converge to nearly the same level. LPP's performance de-

grades when dimension increases, and is the worst overall. The plots of S3 has greater variations compared to those of S1 and S2, and this may be due to the small size of the training set.

## 4. Conclusions

This paper presented a comprehensive study on a number of linear subspace methods for facial expression analysis. We first time applied LPP, SLPP, ONPP, and SONPP to facial expression analysis. In our comparative studies, SLPP was shown to perform best in supervised methods for facial expression analysis, while in unsupervised methods, ONPP produces the best results.

**Acknowledgement** — The authors would like to thank E. Kokiopoulou for providing the source code of ONPP.

## References

- [1] M. S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan. Recognizing facial expression: Machine learning and application to spontaneous behavior. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005. 1, 3
- [2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, July 1997. 1
- [3] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003. 2
- [4] I. Cohen, N. Sebe, F. Cozman, M. Cirelo, and T. Huang. Learning bayesian network classifiers for facial expression recognition using both labeled and unlabeled data. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2003. 1, 3, 5
- [5] G. Donato, M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski. Classifying facial actions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1999. 1
- [6] B. Fasel and J. Luetttin. Automatic facial expression analysis: a survey. *Pattern Recognition*, 36:259–275, 2003. 1
- [7] X. He and P. Niyogi. Locality preserving projections. In *International Conference on Advances in Neural Information Processing Systems (NIPS)*, 2003. 1, 2
- [8] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang. Face recognition using laplacianfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3):328–340, Mar 2005. 1
- [9] T. Kanade, J. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. In *IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, 2000. 2
- [10] E. Kokiopoulou and Y. Saad. Face recognition using oprafaces. In *IEEE International Conference on Machine Learning and Applications*, 2005. 1

	PCA (%)	LPP (%)	ONPP (%)	LDA (%)	SLPP (%)	SONPP (%)
IMG	49.3 (7.1)	48.9 (6.2)	61.1 (5.9)	67.5 (7.6)	<b>86.4 (5.0)</b>	61.9 (6.8)
LBP	49.6 (8.9)	42.6 (5.5)	65.7 (8.8)	73.4 (7.1)	<b>90.4 (2.9)</b>	63.1 (9.4)
BoostLBP	74.9 (9.1)	48.9 (9.0)	79.2 (6.0)	84.2 (6.7)	<b>94.7 (3.5)</b>	75.9 (6.9)

Table 2. The averaged recognition results (with the standard deviation) of 6-class expression recognition on **S1**.

	PCA (%)	LPP (%)	ONPP (%)	LDA (%)	SLPP (%)	SONPP (%)
IMG	41.4 (7.0)	37.9 (8.4)	54.1 (5.9)	59.5 (5.4)	<b>82.2 (6.2)</b>	50.0 (6.8)
LBP	50.6 (6.4)	36.6 (5.0)	61.5 (6.1)	67.9 (5.5)	<b>87.1 (3.0)</b>	57.8 (6.4)
BoostLBP	65.9 (8.8)	44.0 (7.3)	71.4 (6.6)	75.9 (5.8)	<b>92.0 (3.9)</b>	68.3 (6.4)

Table 3. The averaged recognition results (with the standard deviation) of 7-class expression recognition on **S2**.

	PCA (%)	LPP (%)	ONPP (%)	LDA (%)	SLPP (%)	SONPP (%)
IMG	39.5 (6.3)	31.6 (14.5)	52.3 (19.0)	51.2 (11.0)	<b>69.2 (15.7)</b>	48.5 (12.2)
LBP	51.8 (13.2)	31.0 (10.5)	52.3 (16.2)	56.4 (16.1)	<b>72.4 (18.1)</b>	54.2 (16.0)
BoostLBP	62.6 (17.5)	32.4 (9.8)	65.4 (14.1)	65.4 (15.5)	<b>74.2 (16.1)</b>	66.8 (13.7)

Table 4. The averaged recognition results (with the standard deviation) of 7-class expression recognition on **S3**.

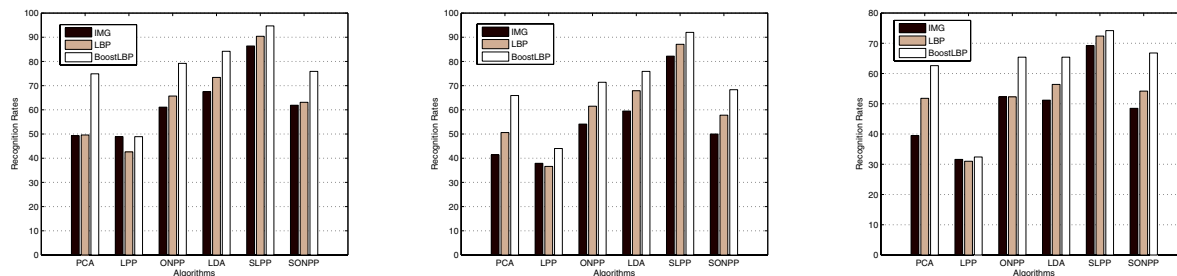


Figure 5. Comparison of different subspace methods with different features. (from left to right: **S1**, **S2**, and **S3**.)

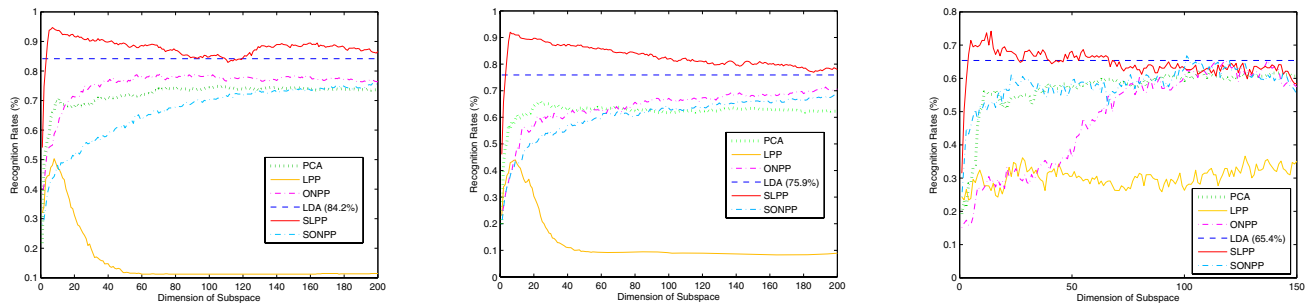


Figure 6. Averaged recognition accuracy versus dimensionality reduction. (from left to right: **S1**, **S2**, and **S3**.)

- [11] E. Kokiopoulou and Y. Saad. Orthogonal neighborhood preserving projections. In *IEEE International Conference on Data Mining (ICDM)*, 2005. 1, 2
- [12] M. J. Lyons, J. Budynek, and S. Akamatsu. Automatic classification of single facial images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(12):1357–1362, December 1999. 1, 2, 3
- [13] T. Ojala, M. Pietikinen, and T. Menp. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002. 3
- [14] M. Pantic and L. Rothkrantz. Automatic analysis of facial expressions: the state of art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1424–1445, 2000. 1
- [15] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290, Dec 2000. 2
- [16] C. Shan, S. Gong, and P. W. McOwan. Appearance manifold of facial expression. In *IEEE ICCV workshop on Human-Computer Interaction (HCI)*, 2005. 2
- [17] C. Shan, S. Gong, and P. W. McOwan. Conditional mutual information based boosting for facial expression recognition. In *British Machine Vision Conference (BMVC)*, 2005. 3
- [18] Y. Tian. Evaluation of face resolution for expression analysis. In *CVPR Workshop on Face Processing in Video*, 2004. 2, 3, 5
- [19] Y. Tian, T. Kanade, and J. Cohn. Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2), 2001. 1
- [20] M. Turk and A. P. Pentland. Face recognition using eigenfaces. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1991. 1