Appearance Manifold of Facial Expression

Caifeng Shan, Shaogang Gong and Peter W. McOwan

Department of Computer Science
Queen Mary, University of London, London E1 4NS, UK
{cfshan, sgg, pmco}@dcs.qmul.ac.uk

Abstract. This paper investigates the appearance manifold of facial expression: embedding image sequences of facial expression from the high dimensional appearance feature space to a low dimensional manifold. We explore Locality Preserving Projections (LPP) to learn expression manifolds from two kinds of feature space: raw image data and Local Binary Patterns (LBP). For manifolds of different subjects, we propose a novel alignment algorithm to define a global coordinate space, and align them on one generalized manifold. Extensive experiments on 96 subjects from the Cohn-Kanade database illustrate the effectiveness of the alignment algorithm. The proposed generalized appearance manifold provides a unified framework for automatic facial expression analysis.

1 Introduction

The ability to recognize affective states of a person is indispensable and very important for successful interpersonal social interaction. Human-Computer Interaction (HCI) designs need to include the ability of affective computing, in order to become more human-like, more effective, and more efficient [13]. Affective arousal modulates all nonverbal communication cues such as facial expressions, body postures and movements. Facial expression is perhaps the most natural and efficient means for humans to communicate their emotions and intentions, as communication is primarily carried out face to face. Therefore, automatic facial expression analysis has attracted much attention [12, 5, 20] in recent years. Though much progress has been made [19, 4, 3], recognizing facial expression with a high accuracy remains to be difficult due to the complexity and variety of facial expressions.

A face image with N pixels can be considered as a point in the N-dimensional image space, and the variations of face images can be represented as low dimensional manifolds embedded in the high dimensional image space [17, 14, 7, 6, 8]. It would be desired to analyze facial expressions in the low dimensional subspace rather than the ambient space. However, research on the manifold of facial expression has been very limited as far as it goes. Chang et al. [2] made first attempt to apply two types of embedding, Locally Linear Embedding (LLE) [14] and Lipschitz embedding, to learn the structure of the expression manifold. In [3], they further proposed a probabilistic video-based facial expression recognition method on manifolds. A complete expression sequence becomes a path on

the expression manifold, and the transition between basic expressions is represented as the evolution of the posterior probability of the six basic paths. Based on a expression manifold obtained by Isomap embedding [17], they also proposed an approach for facial expression tracking and recognition [9]. However, the existing research learned the expression manifold in the feature space described by a set of facial landmarks such as 58 facial points [2, 3]; the appearance manifold of facial expression is still unknown. Moreover, the existing research was conducted on data sets containing only several subjects [2, 3]; there is no published work on the expression manifold carried out on a large number of subjects.

A number of nonlinear techniques have been proposed to learn the structure of a manifold, e.g., Isomap [17], LLE [14], and Laplacian Eigenmap (LE) [1]. However, these techniques yield maps that are defined only on the training data, and it is unclear how to evaluate the maps for new test data, although some mapping methods were discussed in [14]. Therefore, they may not be suitable for expression recognition tasks. Recently He and Niyogi [7] proposed a general manifold learning method called Locality Preserving Projections (LPP) (Section 2), which are obtained by finding the optimal linear approximations to the eigenfunctions of the Laplace Beltrami operator on the manifold. Different from PCA, which implicitly assumes that the data space is Euclidean, LPP assumes that the data space is a linear or nonlinear manifold. LPP shares some similar properties with LLE and LE, such as locality preserving. More crucially, LPP is defined everywhere in the ambient space rather than just on the training data, and so it has significant advantage over LLE and LE in locating and explaining new test data in the reduced subspace. LPP was shown to have superior discriminating power than PCA and LDA in face recognition [8].

In this paper, we investigate the appearance manifold of facial expression, which provides a unified framework for automatic facial expression analysis. We explore Locality Preserving Projections to learn the structure of the expression manifold from two kinds of feature space: raw image data and Local Binary Patterns (LBP) [16]. For manifolds of different subjects, we propose a novel alignment method to keep the semantic similarity of facial expression from different subjects on one generalized manifold (Section 3). We show in Section 4 the experimental results on the Cohn-Kanade Database [10]. Expression manifolds of 96 subjects are successfully aligned on the generalized manifold. Expression recognition performed on the generalized manifolds further demonstrate the effectiveness of the alignment method. Conclusions are drawn in Section 5.

2   Locality Preserving Projections (LPP)

The generic problem of linear dimensionality reduction is the following. Given a set $x_1, x_2, \ldots, x_m$ in R^n, find a transformation matrix W that maps these m points to $y_1, y_2, \ldots, y_m$ in $R^l (l \ll n)$, such that y_i represent x_i, where $y_i = W^T x_i$. Let $\mathbf{w}$ denote the transformation vector, the optimal projections preserving lo-

cality can be obtained by solving the following minimization problem [7]:

$$\min_{\mathbf{w}} \sum_{i,j} (\mathbf{w}^T x_i - \mathbf{w}^T x_j)^2 S_{ij} \qquad (1)$$

where S_{ij} evaluate the local structure of data space. It can be defined as follows:

$$S_{ij} = \begin{cases} e^{-\frac{\|x_i - x_j\|^2}{t}} & \text{if } x_i \text{ and } x_i \text{ are "close"} \\ 0 & \text{otherwise} \end{cases} \qquad (2)$$

or in a simpler form as

$$S_{ij} = \begin{cases} 1 & \text{if } x_i \text{ and } x_i \text{ are "close"} \\ 0 & \text{otherwise} \end{cases} \qquad (3)$$

where "close" can be defined by $\|x_i - x_j\|^2 < \epsilon$, or x_i is among k nearest neighbors of x_j or x_j is among k nearest neighbors of x_i. The objective function with the choice of symmetric weights $S_{ij}(S_{ij} = S_{ji})$ incurs a heavy penalty if neighboring points x_i and x_j are mapped far apart. Therefore, minimizing it is an attempt to ensure that if x_i and x_j are "close" then $y_i(= \mathbf{w}^T x_i)$ and $y_j(= \mathbf{w}^T x_j)$ are close as well. S_{ij} can be seen as a similarity measure between objects. The objective function can be reduced to:

$$\frac{1}{2} \sum_{ij} (\mathbf{w}^T x_i - \mathbf{w}^T x_j)^2 S_{ij} = \sum_i \mathbf{w}^T x_i D_{ii} x_i^T \mathbf{w} - \sum_{ij} \mathbf{w}^T x_i S_{ij} x_j^T \mathbf{w}$$
$$= \mathbf{w}^T X (D - S) X^T \mathbf{w} = \mathbf{w}^T X L X^T \mathbf{w} \qquad (4)$$

where $X = [x_1, x_2, \ldots, x_m]$ and D is a diagonal matrix whose entries are column (or row, since S is symmetric) sums of S, $D_{ii} = \sum_j S_{ji}$. $L = D - S$ is the Laplacian matrix. The bigger the value D_{ii} (corresponding to y_i) is, the more important is y_i. Therefore, a constraint is imposed as follows:

$$\mathbf{y}^T D \mathbf{y} = 1 \Rightarrow \mathbf{w}^T X D X^T \mathbf{w} = 1 \qquad (5)$$

The transformation vector $\mathbf{w}$ that minimizes the objective function is given by the minimum eigenvalue solution to the generalized eigenvalue problem:

$$X L X^T \mathbf{w} = \lambda X D X^T \mathbf{w} \qquad (6)$$

Note that the two matrices $X L X^T$ and $X D X^T$ are both symmetric and positive semi-definite. The obtained projections are actually the optimal linear approximation to the eigenfunctions of the Laplace Beltrami operator on the manifold [7]. Therefore, though it is still a linear technique, LPP recovers important aspects of the intrinsic nonlinear manifold structure by preserving local structure. A more detailed derivation and justifications of LPP can be found in [7].

By applying LPP to LBP appearance feature space, image sequences of facial expressions of an individual are mapped into the embedded space as shown in Fig 1. The embedded manifolds of another three subjects are shown in Fig 2. It is observed that image sequences representing basic expressions with increasing intensity become curves on the manifold extended from the center (neutral faces) to the typical expressions.
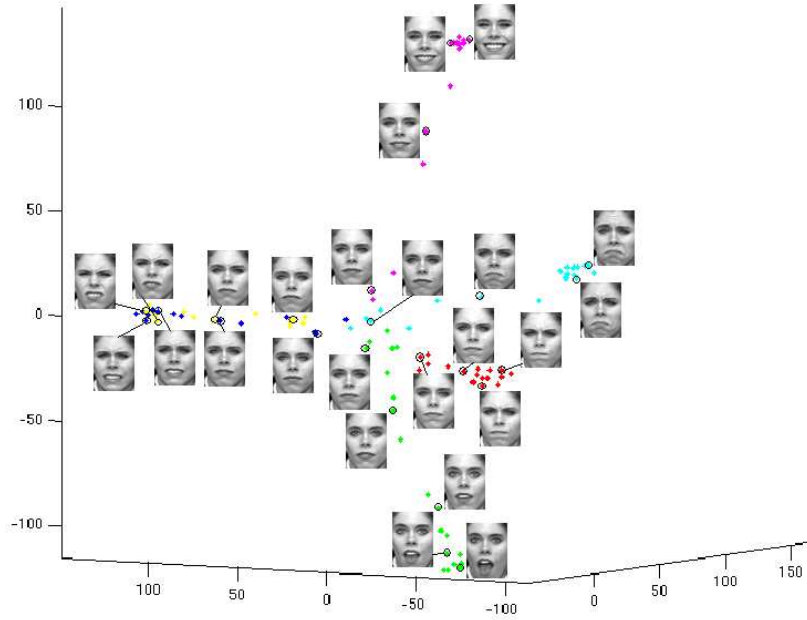
Fig. 1. Six image sequences of basic expressions of an individual mapped into the embedding space described by the first 3 coordinates of LPP. Different sequences are represented by different colors: red: Anger; yellow: Disgust; blue: Fear; magenta: Joy; cyan: Sadness; green: Surprise. (Note: the meaning of colors keeps same in all figures.)

3   Alignment of Manifolds of Different Subjects

Image sequences of facial expressions of an individual makes a continuous manifold in the embedding space; however, due to significant appearance variation across different subjects, the manifolds of different subjects vary much in the covered regions and the stretching directions. Fig 3 shows the embedded manifold of image sequences from six subjects, which clearly shows that different subjects correspond to different clusters. Manifolds of different subjects should be aligned in a way that the images from different subjects with semantic similarity can be mapped to the near region. Chang et al [2] proposed a nonlinear method to align the manifolds of different subjects in the space of Lipschitz embedding. Their alignment method was evaluated on image sequences from two subjects. Here we propose a novel algorithm to align manifolds of different subjects in a global space, and verify its effectiveness on $O(10^2)$ subjects.

As shown in Fig 1, an image sequence representing facial expression with increasing intensity is embedded as a curve on the manifold, from the neutral face to the typical expression. If we define a global coordinate space, in which different typical expressions (including neutral faces and six basic expressions) from multiple subjects are well clustered and separated, the image sequences from

different subjects with the same expression will be embedded as curves between the same two clusters: neutral faces and the typical expression. In this way, the manifolds of different subjects will be aligned on one generalized manifold.

We propose to define the global coordinate space based on images of typical expressions. For the data set containing images of typical expressions from different subjects, as appearance varies a lot cross different subjects, there is significant overlapping among different expression classes. Therefore, the original LPP, which performs in an unsupervised manner, fails to embed the data set in low dimensional space in which different expression classes are well clustered. Here we proposed a Supervised Locality Preserving Projections (SLPP) algorithm to solve the problem, which not only preserves local structure, but also encodes class information in the embedding. SLPP preserves class information when constructing the neighborhood graph. The local neighborhood of a sample x_i from class c should be composed of samples belonging to class c only. This can be achieved by increasing the distances between samples belonging to different classes, but leaving them unchanged if they are from the same class. Let $Dis(i,j)$ denote the distance between x_i and x_j, the distance after incorporating class information is defined as

$$SupDis(i,j) = Dis(i,j) + \alpha M \delta(i,j) \qquad \alpha \in [0,1] \qquad (7)$$

where $M = \max_{i,j} Dis(i,j)$, and $\delta(i,j) = 1$ if x_i and x_j belong to the same class, and 0 otherwise. SLPP introduces an additional parameter α to quantify the degree of supervised learning. When $\alpha = 0$, one obtains the unsupervised LPP; when $\alpha = 1$, the result is fully supervised LPP. For fully supervised LPP, distances between samples in different classes will be larger than the maximum distance in the entire data set; this means neighbors of a sample will always be picked from that class it belongs to. Varying α between 0 and 1 gives a partially supervised LPP, where a embedding is found by introducing some separation between classes. SLPP ($\alpha = 1$) is used in this paper. By preserving local structure of data belonging to the same class, SLPP obtains a subspace in which different image classes can be well separated.

By applying SLPP to the data set of images of typical expressions, a subspace is derived, in which different expression classes are well clustered and separated (as shown in Fig 5). The subspace provides global coordinates for the manifolds of different subjects, which are aligned on one generalized manifold. Image sequences representing facial expressions from beginning to apex are mapped on the generalized manifold as the curves from the neutral faces to the cluster of the typical expressions. For comparison, Fig 3 and Fig 6 show the unaligned manifolds and the aligned manifolds of six subjects. The generalized manifold map the images with sematic similarity but from different subjects in the near region; so it provides a unified framework for automatic facial expression analysis.

4 Experiments

The optimal data set for expression manifold learning should contain $O(10^2)$ subjects, and each subject has $O(10^3)$ images that cover basic expressions. How-

ever, until now, there is no such database that can meet this requirement. Chang et al [2, 3] conducted experiments on a small data set builded themselves, e.g., only two subjects (one male and one female) were used in [2]. Here we conduct experiments on the Cohn-Kanade database [10] which consists of 100 university students in age from 18 to 30 years, though each subject only has several tens frames of basic expressions. Image sequences from neutral to target expression were captured, and the duration of the expression varied. In our experiments, 316 image sequences (5,876 images in total) of basic expressions were selected from the database, which come from 96 subjects, with 1 to 6 emotions per subject.

Following Tian [18], we normalized the faces to a fixed distance between the centers of the two eyes. Facial images of 110×150 pixels were cropped from original frames based on the two eyes location. No further alignment of facial features such as alignment of mouth, or remove of illumination changes [18] were performed in our experiments. So variations due to illumination, and pose exist in our data. In [2], Active Wavelets Networks were applied on the image sequences to reduce these variations.

Two kinds of appearance features were used: raw image data (IMG) and Local Binary Patterns (LBP). LBP was proposed originally for texture analysis [11]. Face images can be seen as a composition of micro-patterns which can be effectively described by the LBP features. In our previous research [16, 15], LBP features were shown to be effective and efficient for facial expression analysis. Each face image was represented by a LBP histogram with length of $2, 478$ (see [16] for details). When considering IMG features, for computational efficiency, we down-sampled face images to 55×75 pixels, and represented each image with a $4, 125$-dimensional vector.

Appearance Manifold of Facial Expression We selected six subjects from the data set, each of which has six image sequences corresponding to six basic expressions. At first, we applied LPP to image sequences of each subject respectively to learn the expression manifold of each subject. 3-D visualization of the embedded manifold in LBP feature space of one subject is shown in Fig 1. Due to limitation of space, we only show the embedded manifolds of another three subjects in Fig 2. It is observed that images of facial expressions of an individual were embedded as a smooth manifold, and every image sequence is mapped to a curve on the manifold that begins from the neutral face and extends in distinctive direction with varying intensity of expression.

Next we applied LPP to image sequences of all six subjects, and 3-D visualization of the embedded manifold are shown in Fig 3. It is observed that there are six clusters in the embedded space, and image sequences of different subjects are mapped to different regions. As said above, due to the significant appearance variation across subjects, it is very hard for LPP to keep images with similar expression but from different subjects in the near region on the manifold. Fig 4 shows the 3-D embedded manifolds of all image sequences from 96 subjects, which consists of many manifolds with different centers (neutral faces), covering regions and stretching directions.
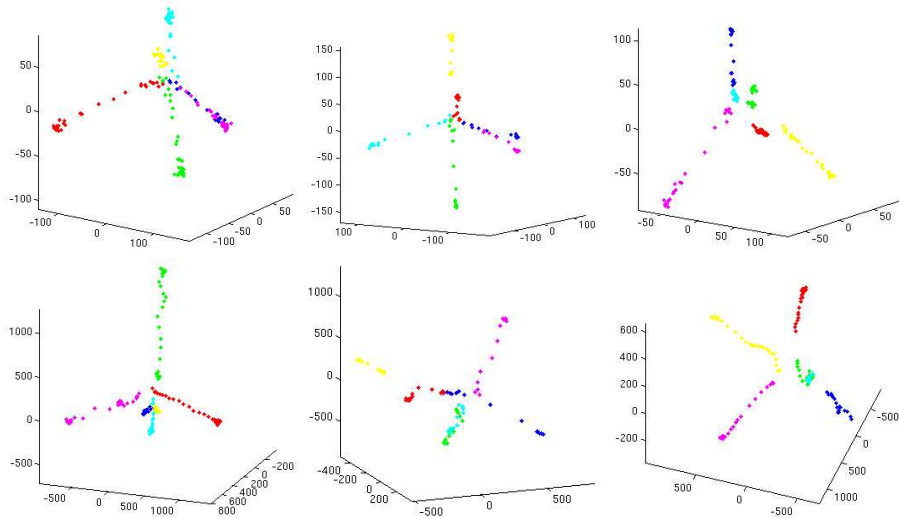
Fig. 2. 3-D visualization of expression manifolds of three subjects (from left to right). The first row: LBP; the second row: IMG.
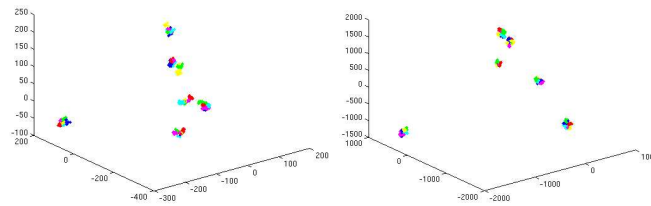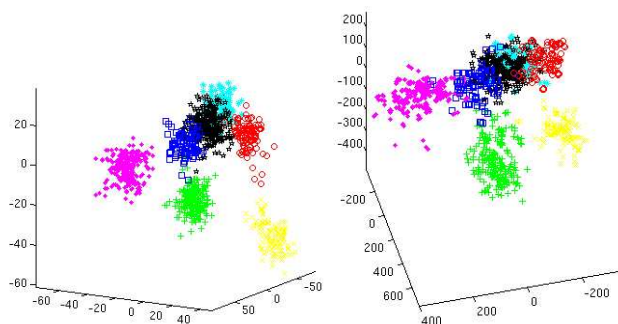

Fig. 3. Image sequences of six subjects mapped into the embedding space described by the first three coordinates of LPP. Left: LBP; Right: IMG

Alignment of Appearance Manifolds We selected one neutral face and three peak frames (during the apex of expression) of every sequence to build a data set that consists of images of 7-class basic expressions. The Supervised LPP was explored to embed the data set to a subspace as shown in Fig 5. Different expressions were well clustered and separated in the subspace. It is also observed that different expression classes are better separated with LBP features. The distributions obtained reflect the human observation that Joy and Surprise can be clearly separated, but Anger, Disgust, Fear and Sadness are easily confused. In many existing research such as [18, 4], most confusions also come from Anger, Disgust, Fear and Sadness.

The subspace derived by SLPP provides global coordinates for the manifolds of different subjects. Fig 6 plots appearance manifolds of the six subjects in the global space, which are successfully aligned on one generalized manifold. The
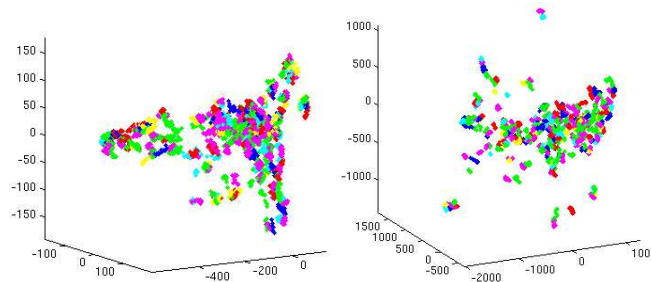
Fig. 4. Image sequences of 96 subjects mapped into the embedding space described by the first three coordinates of LPP. Left: LBP; Right: IMG


Fig. 5. Images of typical facial expressions mapped into the embedding space described by the first three coordinates of SLPP. Left: LBP; Right: IMG.

manifolds of 96 subjects are also aligned on the generalized manifold as shown in Fig 7. We can conclude that the images with semantic similarity but from different subjects are successfully embedded in the near region in the global space. A supplementary video [1] demonstrates image sequences of different subjects are embedded on the generalized manifold.

The global space is learned from images of typical basic expressions. So it is simple and easy to implement. Although only image sequences of basic expressions are discussed until now, the generalized appearance manifold provides a global semantic representation for all possible facial expressions. For example, the blends of expression will lie between the curves of basic expressions, so can be analytically analyzed based on the basic curves. Intensity of expression can also be defined easily on the generalized manifold. Therefore, the analysis of facial expression will be facilitated on the generalized manifold.

---

[1] Available at *http://www.dcs.qmul.ac.uk/~cfshan/demos/manifold_align.avi*
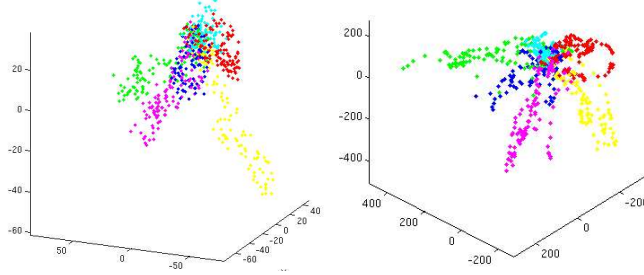
Fig. 6. The aligned manifolds of the six subjects. Left: LBP; Right: IMG.
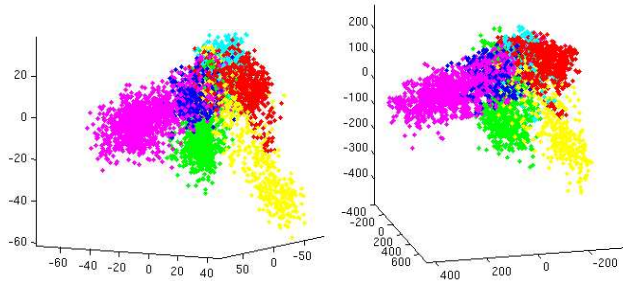
Fig. 7. The aligned manifolds of 96 subjects. Left: LBP; Right: IMG.

Facial Expression Recognition Following Chang et al [2], we applied a k-Nearest Neighbor method to classify expressions on the aligned expression manifold. Since there is no clear boundary between neutral face and the expression of a sequence, we manually labelled neutral faces, which introduced noise in our recognition. The recognition results are presented in Table 1. The experimental results further demonstrate the effectiveness of our alignment method.

	k-NN $(k = 9)$	k-NN $(k = 11)$	k-NN $(k = 13)$
IMG	92.04%	91.27%	89.98%
LBP	90.71%	90.79%	90.67%

Table 1. Expression recognition results on the generalized appearance manifold of facial expression.

5 Conclusions

This paper investigates the appearance manifold of facial expression, which provide a general framework for automatic facial expression analysis. Locality Preserving Projections (LPP) is explored to learn expression manifolds from two kinds of feature space: raw image data and Local Binary Patterns (LBP). For

manifolds of different subjects, we propose a novel alignment algorithm by learning a global space from images of typical expressions. The semantic similarity of facial expression from different subject is well kept on the generalized manifold. Extensive experiments on 96 subjects from the Cohn-Kanade database illustrate the effectiveness of the alignment algorithm.

References

1. M. Belkin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *NIPS*, 2001.
2. Y. Chang, C. Hu, and M. Turk. Mainfold of facial expression. In *IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, 2003.
3. Y. Chang, C. Hu, and M. Turk. Probabilistic expression analysis on manifolds. In *CVPR*, 2004.
4. I. Cohen, N. Sebe, A. Garg, L. Chen, and T. S. Huang. Facial expression recognition from video sequences: Temporal and static modeling. *CVIU*, 2003.
5. B. Fasel and J. Luettin. Automatic facial expression analysis: a survey. *Pattern Recognition*, 36:259–275, 2003.
6. D. Fidaleo and M. Trivedi. Manifold analysis of facial gestures for face recognition. In *ACM SIGMM Multimedia Biometrics Methods and Application Workshop*, 2003.
7. X. He and P. Niyogi. Locality preserving projections. In *NIPS*, 2003.
8. X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang. Face recognition using laplacianfaces. *IEEE PAMI*, 27(3):328–340, Mar 2005.
9. C. Hu, Y. Chang, R. Feris, and M. Turk. Manifold based analysis of facial expression. In *CVPR Workshop on Face Processing in Video*, 2004.
10. T. Kanade, J.F. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. In *IEEE FG*, 2000.
11. T. Ojala, M. Pietikinen, and T. Menp. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE PAMI*, 2002.
12. M. Pantic and L. Rothkrantz. Automatic analysis of facial expressions: the state of art. *IEEE PAMI*, 22(12):1424–1445, 2000.
13. M. Pantic and L. Rothkrantz. Toward an affect-sensitive multimodal human-computer interaction. In *Proceeding of the IEEE*, 2003.
14. L. K. Saul and S. T. Roweis. Think globally, fit locally: Unsupervised learning of low dimensional manifolds. *Journal of Machine Learning Research*, 2003.
15. C. Shan, S. Gong, and P. W. McOwan. Conditional Mutual Information Based Boosting for Facial Expression Recognition. In *BMVC*, 2005.
16. C. Shan, S. Gong, and P. W. McOwan. Robust facial expression recgnition using local binary patterns. In *IEEE ICIP*, 2005.
17. J. B. Tenenbaum, V. Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290, Dec 2000.
18. Y. Tian. Evaluation of face resolution for expression analysis. In *CVPR Workshop on Face Processing in Video*, 2004.
19. Y. Tian, T. Kanade, and J. Cohn. Recognizing action units for facial expression analysis. *IEEE PAMI*, 23(2), 2001.
20. Y. Tian, T. Kanade, and J. Cohn. Facial Expression Analysis, *Handbook of Face Recognition*. Springer, 2003.