Manifold of Facial Expression

Ya Chang, Changbo Hu, Matthew Turk
Computer Science Department, University of California, Santa Barbara, CA 93106
{yachang, cbhu, mturk}@cs.ucsb.edu

Abstract

In this paper, we propose the concept of Manifold of Facial Expression based on the observation that images of a subject's facial expressions define a smooth manifold in the high dimensional image space. Such a manifold representation can provide a unified framework for facial expression analysis. We first apply Active Wavelet Networks (AWN) on the image sequences for facial feature localization. To learn the structure of the manifold in the feature space derived by AWN, we investigated two types of embeddings from a high dimensional space to a low dimensional space: locally linear embedding (LLE) and Lipschitz embedding. Our experiments show that LLE is suitable for visualizing expression manifolds. After applying Lipschitz embedding, the expression manifold can be approximately considered as a super-spherical surface in the embedding space. For manifolds derived from different subjects, we propose a nonlinear alignment algorithm that keeps the semantic similarity of facial expression from different subjects on one generalized manifold. We also show that nonlinear alignment outperforms linear alignment in expression classification.

1. Introduction

Facial expressions help coordinate conversation and communicate emotions and other meaningful mental, social, and physiological cues. People can recognize facial expression easily even though the appearance of the expression varies a lot between different individuals. However, it is a challenging task to automatically determine the expression due to the variation of facial expression across the human population and to the context-dependent variation even for the same individual.

Facial expressions can be classified in various ways – in terms of some non-prototypic expressions such as "raised brows," in terms of some prototypic expressions such as emotional expressions, or in terms of facial actions that cause an expression such as the Action Units defined in Facial Action Coding System (FACS) [13]. Psychologists claim that there are six kinds of universally recognized facial expressions: happiness, sadness, fear, anger, disgust and surprise [1]. Existing expression analyzers [3,17] generally classify the examined expression into one of the basic emotion categories. This approach has two main limitations. First, it is not clear that all facial expressions expressible by the human face can be classified under the six basic emotion categories. For "blended" expression, it is more reasonable to classify facial expression quantitatively into multiple emotion categories. Second, this approach does not consider the intensity scale of the different facial expressions. In addition, each person has his/her own maximal intensity of displaying a particular facial action. Many behavioral scientists perform facial expression classification through FACS [13] encoding. FACS provides a linguistic description of all visually detectable facial changes in terms of 44 Action Units (AU). Using these rules, an expression can be decomposed into the specific AUs. But there are only five AUs with the option to score intensity on three-level scale (low, medium, and high). Also it is very hard to connect the combinations of AUs with the emotional expression in an analytical way due to the discrete nature of AUs. A key challenge in automatic facial expression analysis is to find some global, analytical, and semantic-related representation for all possible facial expressions.

An image with N pixels can be considered as a point in an N-dimensional image space, and the variability of image classes can be represented as low-dimensional manifolds embedded in image space. Since people change facial expression continuously over time, it is a reasonable assumption that all images of someone's facial expressions make a smooth manifold in the N-dimensional image space with the "neutral" face as the central reference point. The intrinsic dimension of the manifold is much lower than N. If we were to allow other factors of image variation, such as face pose and illumination, the intrinsic dimensionality of the manifold of expression would increase accordingly.

On the manifold of expression, similar expressions are points in the local neighborhood on the manifold. The basic emotional expressions with increasing intensity become curves on the manifold extended from the center. The blends of expressions will lie between those curves, so they can be defined analytically by the positions of the main curves. The analysis of the relationships between different facial expressions will be facilitated on the manifold. More generally, the manifold of facial expression shows promise as a unified framework for facial expression analysis.

It is a formidable task to learn the structure of the manifold of expression in a high dimensional image space. Therefore, we first apply Active Wavelets Networks [12] on the image sequences to reduce the variation due to scaling, illumination condition, and face pose. We seek to embed the manifold from the high dimensional feature space of AWN to a low dimensional space while keeping the main structure of the manifold. In this paper, we investigate two types of embeddings to perform this task. The first is locally linear embedding (LLE) [4]. LLE is an unsupervised learning algorithm that computes low-dimensional, neighborhood-preserving embedding of high-dimensional inputs. LLE maps its inputs into a single global coordinate system of lower dimensionality, and its optimizations do not involve local minima. The second is Lipschitz embedding [5,6]. In Lipschitz embedding, a coordinate space is defined such that each axis corresponds to a reference set ***R***, drawn from the input data set. With a suitable definition of ***R***, we can establish the bounds on the distortion for all pairs of data points in the embedding space. Lipschitz embedding leads to good preservation of clusters in some practical cases [7,8].

We found that LLE is suitable for the purpose of visualizing manifolds. After applying Lipschitz embedding, the manifold of expression can be approximately considered as a super-spherical surface in the embedding space. Therefore, expression classification can be performed effectively on the manifold. In the space of the Lipschitz embedding, we propose a nonlinear method to align the manifolds of different subjects while keeping the images of different subjects with similar expressions in the near region. The experiments show that the nonlinearly aligned manifold outperforms the linearly aligned one in many ways.

The remainder of this paper is organized as follows. We describe the related work in next section. We then discuss the properties of LLE in Section 3. The Lipschitz embedding and the nonlinear alignment of manifolds will be presented in Section 4. Section 5 presents the experiments we conducted on the manifold of facial expression. Finally, we present conclusions and future research direction in Section 6.

2. Related Work

Many scientists have explored the nature of the space of facial expressions. Shalif [2] examined the principal emotional variables in daily life by letting 202 subjects judge the extent of emotional state expressed in photographs. He found that the happiness and sadness dimensions are bipolar and are the most prominent dimension in expression space, followed by fear, anger dimension, etc. Schmidt and Cohn [15] measured 195 spontaneous smiles from 95 individuals through facial electromyographic (EMG) data and found consistency in zygomaticus major muscle activity over time. Black and Yacoob [17] classified the facial expression through the movements of prominent facial features. By setting a threshold for the mid-level representation, they can detect the beginning, apex, and ending of four kinds of expressions. Zhang et al. [3] used a two-layer perceptron to classify facial expressions. They found that five to seven hidden perceptrons are probably enough to represent the space of feature expressions. Chuang et al. [9] showed that the space of facial expressions can be modeled with a bilinear model. Two formulations of bilinear models, asymmetric and symmetric, were fit to facial expression data.

In this paper, we explore the space of expression through the manifold of expression. The manifold is learned from image sequences of basic facial expressions. To reduce the variation due to scaling, illumination condition, and face pose, we first apply Active Wavelets Networks [12] on the image sequences. Then we apply LLE and Lipschitz embedding to learn the distribution of different expressions on the manifold of facial expression.

3. Locally Linear Embedding

Locally Linear Embedding (LLE) [4] is a method that maps the high dimensional input to low dimensional, neighbor-preserving embedding. Previous approaches to this problem, based on multidimensional scaling (MDS) [18], have computed embeddings that attempt to preserve pairwise distances between data points. Some methods measure these distances in *p* metric space, while other methods measure these distances along the shortest paths confined to the manifold of observed inputs, such as Isomap [19]. However, LLE recovers global nonlinear structure from locally linear fits. It is based on the following geometric intuitions. If the data points are drawn from some underlying manifold uniformly, each data point and its neighbors are expected to lie on or close to a locally linear patch of the manifold. The overlapping local neighborhoods, collectively analyzed, can provide information about global geometry.

The algorithm of LLE is presented in Fig. 1. LLE aims to minimize the locally linear reconstruction error for every data point in the embedding space. The local geometry of the patches around every data point is characterized as the weight matrix $W_{N,N}$. It can be considered as a least squares problem to compute W subject to the constraint $\sum_{j=1}^{N} w_{i,j} = 1$. This least squares problem can be solved by the Lagrange multiplier algorithm [21]. The low dimensional vectors Y_i represent the global internal coordinates on the manifold. Y_i are chosen to minimize the embedding cost function

$\Phi(Y) = \sum_i \left| \vec{Y}_i - \sum_j w_{i,j} \vec{Y}_j \right|^2$ with fixed weight *W*. By the Rayleigh-Ritz theorem [22], Y_i are the smallest *d* eigenvectors of matrix $M = (I - W)^T (I - W)$ after discarding the bottom eigenvector [4].

LLE is able to learn the global structure of nonlinear manifolds. For data visualization in two and three dimensions, it works well when the data set has only one cluster. This assumption holds when the image sequences of facial expressions are from the same subject. When the data come from multiple subjects, there are many manifolds with different centers (neutral face) and stretching directions. We need to build "global coordinates" for the mixture of locally linear projection from samples to coordinate space [11], or decompose the sample data into some patches, then merge them into one global coordinates in an optimal way [10].

Input:

$\vec{X}_i, i = 1,..., N : \vec{X}_i \in R^D$

k: the number of nearest neighbors for every data point

Output:

$\vec{Y}_i, i = 1,..., N : \vec{Y}_i \in R^d$

for i=*1* **to** *N* **do**

$[\vec{X}_{t_1},..., \vec{X}_{t_k}] = \vec{X}_i$'s *k* nearest neighbors .

$[w_{i,t_1},..., w_{i,t_k}] = \arg\min \sum \left| \vec{X}_i - \sum_{j=1}^{k} w_{i,t_j} \vec{X}_{t_j} \right|^2$

subject to $\sum_{j=1}^{k} w_{i,t_j} = 1$

$w_{i,j} = 0$ when $\vec{X}_j \notin [\vec{X}_{t_1},..., \vec{X}_{t_k}]$

end for

$[\vec{Y}_1,..., \vec{Y}_N] = \arg\min \sum_i \left| \vec{Y}_i - \sum_j w_{i,j} \vec{Y}_j \right|^2$

Return $\vec{Y}_i, i = 1,..., N$

Figure 1: Locally Linear Embedding algorithm

---

4. Lipschitz Embedding

4.1 Algorithm

Lipschitz embedding [5,6] is a powerful embedding method used widely in image clustering and image search. In Lipschitz embedding, a coordinate space is defined such that each axis corresponds to a reference set ***R***, which is drawn from the input data set ***S***. The algorithm of Lipschitz embedding is shown in Fig. 2.

With a suitable definition of the reference set ***R***, the distance of all pairs of data points in the embedding space is bounded [24]. So Lipschitz embedding works well when there are multiple clusters in the input data set [7,8]. Assume that the number of typical facial expressions is *k*, and every reference set contains only the images of one kind of facial expression during apex. If there are only three kinds of facial expressions, the apex of each kind of expression will be mapped on the positive central part of the *x-y*, *x-z*, *y-z* planes, and one kind of expression with different intensity or blended expressions will be mapped on an approximately spherical surface, as illustrated in Fig. 3. Because people can exhibit many more the three facial expressions, the manifold of facial expression will become a *super-spherical* surface in *k*-dimensional space through Lipschitz embedding. The experimental results in Section 5 support this point.

Input:

$S = \{x_1, x_2,..., x_n\}$, $x_i \in R^D$

Output:

$Y = \{y_1, y_2,..., y_n\}$, $y_i \in R^k$

Select *k* subsets $A_1, A_2,..., A_k$ from *S*, $R = \{A_1, A_2,..., A_k\}$.

for *i=1* **to** *n* **do**

for *j=1* **to** *k* **do**

$y_j^i = \min_{x \in A_j} | x_i - x |$

end for

$y_i = (y_1^i, y_2^i,..., y_k^i)$;

end for

Return $Y = \{y_1, y_2,..., y_n\}$

Figure 2: Lipschitz embedding algorithm

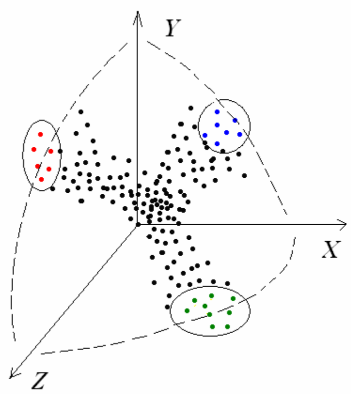


Figure 3: An illustration of Lipschitz embedding for k=3. The points in the circles are in the reference set which represent extreme expressions. The neutral faces are far away from every reference set.

4.2 Alignment of manifolds of different subjects

The data points of a single subject make a continuous manifold in the embedding space, while the neutral face and the intensity of expression that can be displayed by different subjects vary significantly. Correspondingly, the manifolds of different subjects vary a lot in the centers, the stretching directions, and the covered regions. Because of the appearance variation across the subjects, it is very hard to align the manifolds of different subjects in a way that the images from different subjects with semantic similarity can be mapped to the near region. But in the space of Lipschitz embedding, the alignment can be performed in an elegant way.

In Lipschitz embedding, the k reference sets contain typical expressions. The ith coordinate of data points in reference set A_i is zero by definition. So the images in the reference set A_i are mapped to a *compact plane* region R_i. The images that represent a kind of expression from beginning to apex are mapped along the curves from the center (neutral expression) to R_i on the expression manifold. To align the manifolds of different subjects, we need only to align the corresponding region R_i of the different manifolds to the set region one by one. The center of region R_i is $Q_i = (q_1^i, q_2^i, ..., 0_i^i, ..., q_k^i)^T$. The essential idea is to align Q_i to $(1_1, ..., 1_{i-1}, 0_i, 1_{i+1}, ..., 1_k)^T$. The manifold of expression will become a super-spherical surface with approximate radius $\sqrt{k-1}$.

A straightforward solution is to apply a linear transformation to the embedding space. A linear transformation in the k-dimensional space can be determined by the transformation of k points, i.e. the centers of the k reference regions. Unfortunately, linear transformation does not preserve the semantic similarity of data points well. Our experiments show that images of the different expressions get mixed up after such a linear transformation.

We propose a nonlinear transformation that can align the reference regions while preserving the semantic similarity of data points at the same time. The detailed algorithm is presented in Fig. 4. The goal is to align all nonzero coordinate values of Q_i to 1. To align q_j^i of Q_i to 1, the jth member of each data point is multiplied by a scaling factor. Therefore the critical part is to design a nonlinear continuous function Φ that returns a scaling factor as 1 for the points in all other reference sets except A_i, near $1/q_j^i$ for all points in A_i, and an interpolated number according to their positions for the remaining data points. The nonlinear function Φ is defined in Fig. 5. After every process, one nonzero coordinate value of a region center is normalized to 1 and the structure of the manifold is preserved. Fig. 6 compares the unaligned manifold and the manifold after one step alignment for a simple case $k=3$. The alignment can be achieved by repeating this process $k(k-1)$ times. We will show that the nonlinear alignment also outperforms the linear alignment in expression classification in Section 5.

When the semantic meanings of every reference set are defined in the same way for different subjects, the corresponding reference set will be aligned to the same region by our algorithm. So the aligned manifold will map the images with semantic similarity but from different subjects in the near region. Furthermore, the expression classification can be performed with high accuracy on this generalized manifold.

Input:

$S = \{x_1, x_2, ..., x_n\}$, $x_i \in R^D$

Output:

$Y = \{y_1, y_2, ..., y_n\}$, $y_i \in R^k$

Select k subsets $A_1, A_2, ..., A_k$ from S.

Apply Lipschitz embedding algorithm as Fig. 2. The result is $Y = \{y_1, y_2, ..., y_n\}$, $y_i = (y_1^i, y_2^i, ..., y_k^i)$

for $i=1$ **to** k **do**

$Q_i = mean\ (y_j, if\ \ x_j \in A_i)$

end for

$Q = \{Q_1, Q_2, ..., Q_k\}$, $Q_i = (q_1^i, q_2^i, ..., 0_i^i, ..., q_k^i)$.

for $i=1$ **to** k **do**

for $j=1$ **to** k **do**

if $j \neq i$

for $s=1$ **to** n **do**

$scale = \Phi(y_s, Q_i, j);$

$y_j^s = y_j^s * scale;$

end for

update Q

end if

end for

end for

Return $Y = \{y_1, y_2, ..., y_n\}$

Figure 4: The alignment algorithm

Function $\Phi(y_s, Q_i, j)$

$\{Q_i = (q_1^i, q_2^i, ..., 0_i^i, ..., q_k^i)$, $y_s = (y_1^s, y_2^s, ..., y_k^s)\}$

$b = 1;$

$v = 1;$

for $t=1$ **to** k **do**

if $(t \neq i)\ \&(t \neq j)$

$b = b * q_t^i;$

$v = v * y_t^s;$

end if

end

{if $y_s \in R_p$, $p \neq i$, then $v = 0, scale = 1$;

if $y_s \in R_i$, then $v \cong b, scale \cong 1/q_j^i$}

$scale = (v * (1 - 1/q_j^i)/b + 1) * \exp(y_j^s / c)$;

{ c is an empirical constant}

Return *scale*

Figure 5: Function $\Phi(y_s, Q_i, j)$

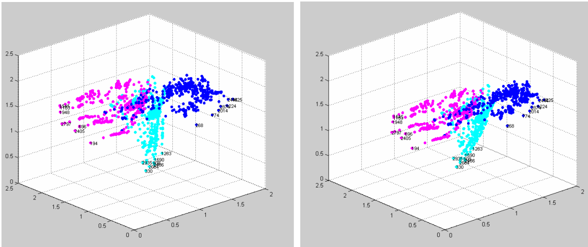


Figure 6: The comparison of the unaligned manifold and the manifold after one step alignment for $k=3$. There are three classes of data points with different colors. The points with labels are in reference sets. After one step alignment, the center of magenta reference region is drawn to right (near 1), while blue and cyan points with labels do not move. The whole structure of the manifold is preserved well.

5. Experimental Results

In this section, we present the experimental results on exploring the manifold of facial expression. To learn the structure of the expression manifold, we need $O(10^3)$ images that cover basic expressions for every subject. There are no standard databases that can meet this requirement, so we build our own data set. Here we present preliminary results on the two subjects (one male, one female).

In our experiments, subjects were instructed to perform a series of seven kinds of facial expressions: happiness with closed mouth, happiness with open mouth, sadness, anger, surprise, fear, disgust. (The fact that these are not true emotion-driven expressions is not relevant to the analysis.) The subjects repeated the series seven times. A total of 4851 images were captured (1824 frames for the female subject, 3027 frames for the male subject). To simplify the problem, we only used frontal view.

The experiments consisted of two parts after the preprocessing. In the first part, we applied LLE to visualize the manifold. In the second part, Lipschitz embedding was applied to construct the manifolds for two subjects. We aligned the expression manifolds of the subjects by linear and nonlinear alignment algorithms, then compared their efficiency. Finally, we performed expression classification on the manifold acquired by both LLE and Lipschitz embedding.

5.1. Preprocessing

To reduce the variation due to scaling, illumination condition, and face pose, we applied Active Wavelets Networks (AWN) [12] on the image sequence for face registration and facial feature localization. The inputs for the embedding algorithms are the (x, y) coordinates of 58 localized facial features. This preprocessing increases the generality of the expression manifold and reduces the data dimensionality greatly compared with the raw images.

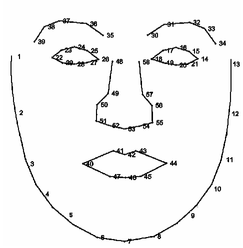


Figure 7: Facial landmarks (58 points)

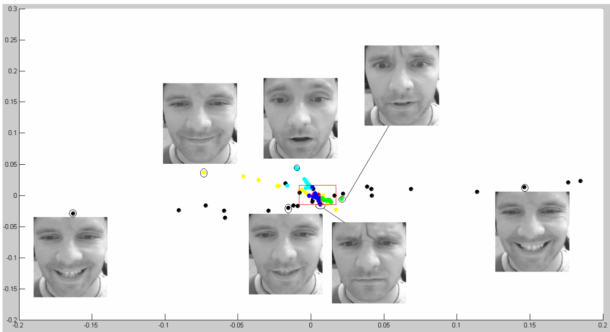
Figure 8: Representative images after applying AWN. From left to right, up to down, the images with expressions: neutral, small smile, big smile, sadness, anger, surprise, fear, and disgust.

AWN is a new face alignment method, which is robust to illumination and partial occlusion and has very strong generalization ability. The key idea of AWN is to replace the PCA-based texture model in Active Appearance Models (AAM) [16] with a wavelet network representation. Because of the spatially localized property of wavelets, AWN has better performance over AAM when there are partial occlusions or illumination changes. More details can be found in [12]. Fig. 7 shows the face model used in this paper. It is slightly modified from the model of AAM-API, a C++ implementation of the Active Appearance Model framework [14]. Among the 4851 face images, 117 images are selected out as the training set, which covers the seven kinds of typical expressions. The localizations of all the test images are completely automatic. The typical facial feature localization results of different expression are shown in Fig. 8. Each raw image is reduced to a 116-dimensional vector (the (x, y)

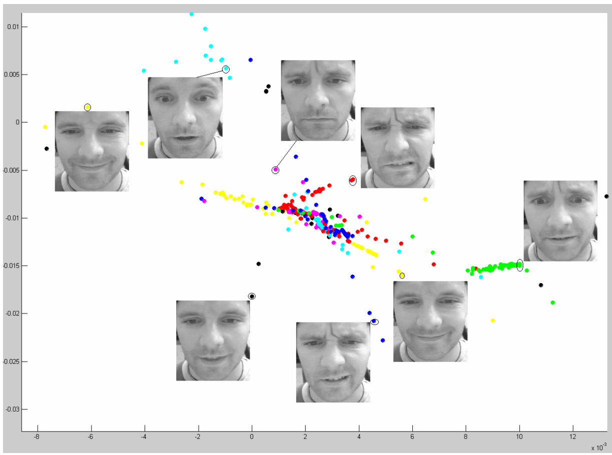
coordinates of 58 feature points after alignment) for further processing.

5.2. Results of LLE

We found that LLE is very sensitive to the selection of the number of nearest neighbors. When the data set contains just one series of seven kinds of facial expressions, every image sequence is mapped to a curve that begins from the center (neutral face) and extends in distinctive direction with varying intensity of expression as in Fig. 9. While there are many series, the sequences with the same kind of facial expression diverge in different directions when the number of nearest neighbors is small. The images of different expressions become mixed up easily when we increase the number of nearest neighbors as shown in Fig. 10.



(a)



(b)

Figure 9: The first 2 coordinates of LLE of 478 images with the number of nearest neighbors k=9. Fig. (b) is the enlarged red rectangular region in Fig. (a). Seven sequences for seven different expressions, which are represented by different colors: small smile: yellow; big smile: black; sadness: magenta; anger: cyan; disgust: red; surprise: green; fear: blue. The representative images are shown next to the circled points.

The reason is that LLE is an unsupervised learning algorithm. It selects the nearest neighbors to reconstruct the manifold in the low dimensional space. There are two types of variations in the data set: the different kinds of facial expressions and the varying intensity for every kind of facial expression. Generally, LLE can catch the second type of variation – an image sequence is mapped in a "line," and LLE can keep the sequences with different expressions distinctive when there is only one sequence for each expression. When the data set contains many image sequences for the same kind of expression, it is very hard to catch the first kind of variation using a small number of nearest neighbors. But with the increased number of nearest neighbors, the images of different expressions are more prone to be mixed up.

We classify expressions by applying a *k*-Nearest Neighbor method in the embedding space. The result is in Table 1. It can be seen that LLE cannot achieve good expression classification results without building "global coordinates" for the mixture of locally linear projection [11], or performing some decomposing/merging processes [10].

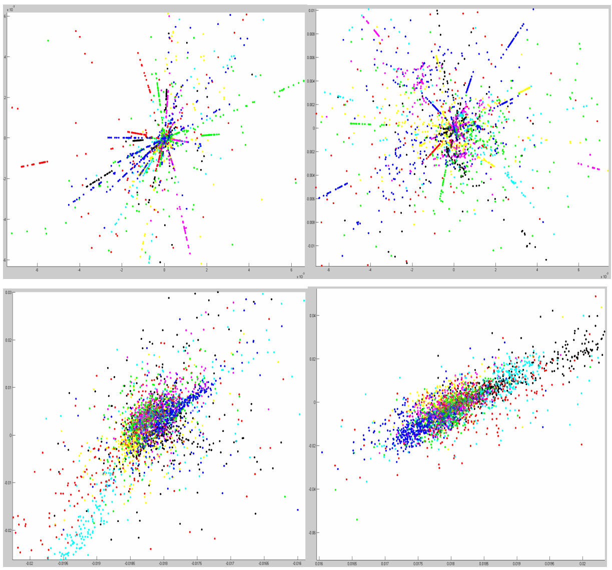


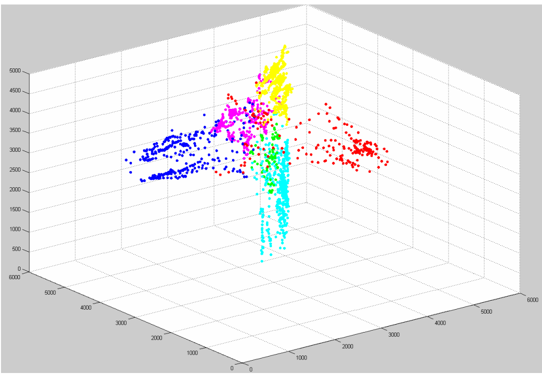
Figure 10: The first 2 coordinates of LLE of 3027 images (the male subject) with the number of nearest neighbors k=10,20,30,50, from left to right, up to down. The meaning of colors is the same as in Fig. 9.

Table 1: Comparison of classification rate on the four manifolds in Fig. 10.

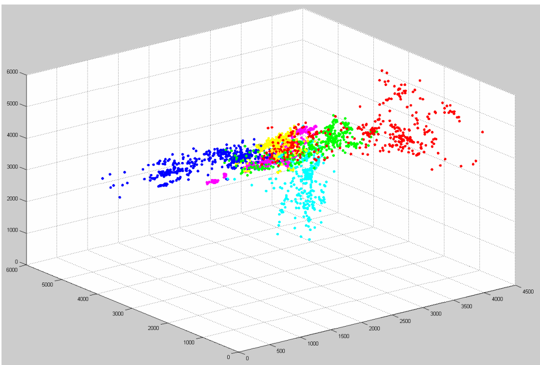
	LLE(k=10)	LLE(k=20)	LLE(k=30)	LLE(k=50)
k_NN(k=9)	15.76%	16.60%	13.53%	16.32%
k_NN(k=13)	15.62%	17.57%	15.90%	16.47%

5.3. Results of Lipschitz embedding

There are six kinds of typical expressions in the data set (we consider the small smile and big smile as the same class). So there are six reference sets. For every sequence, only one image during the apex of expression is selected for the corresponding reference set. Every image is mapped to a six dimensional vector, which represents its distance to each kind of "extreme" expression. For the purpose of visualization, we can project the manifold onto its first three dimensions as shown in Fig. 11. One can see that the expression manifold can be considered approximately as a spherical surface. The alignment of the two manifolds in Fig. 11 by nonlinear alignment is shown in Fig. 12. The alignment by linear alignment is shown in Fig. 13. We can see the clusters of different expressions are preserved well through nonlinear alignment, while images of different expressions become mixed up after linear alignment.



(a)



(b)

Figure 11: The projection on first three dimensions after Lipschitz embedding. (a) the female subject. (b) the male subject. The meaning of colors is the same as in Fig. 9.

Since the points with the same kind of expression can be approximated by a curve on the manifold, we use a least squares line to fit those points. The line of happiness on the manifold of subject 1 should have similar direction with the line of happiness on the manifold of subject 2 after alignment. We compare the angles between the lines on the manifold of both subjects after linear/nonlinear alignment for each kind of expression. The result is shown in Fig. 14. The experimental results show that the nonlinear alignment has better performance in preserving the same kind of expression from different subjects in the near region.

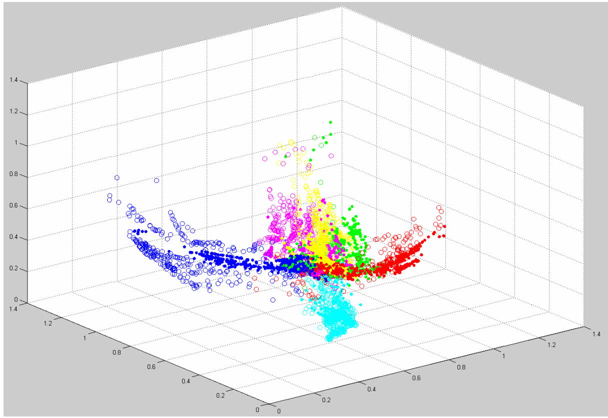


Figure 12: The aligned manifolds after nonlinear alignment. The points from the first manifold are represented as circles. The meaning of colors is the same as in Fig. 9.

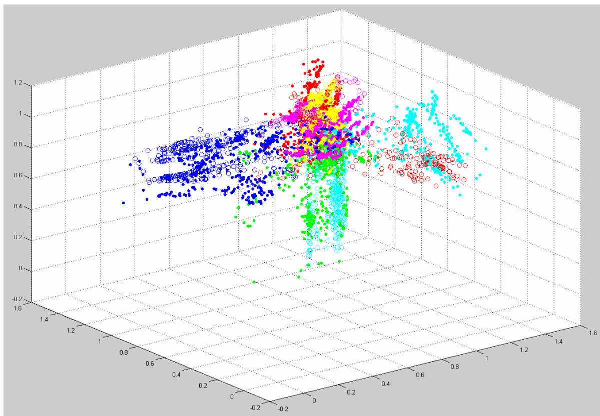


Figure 13: The aligned manifolds after linear alignment. The points from the first manifold are represented as circles. The meaning of colors is the same as in Fig. 9.

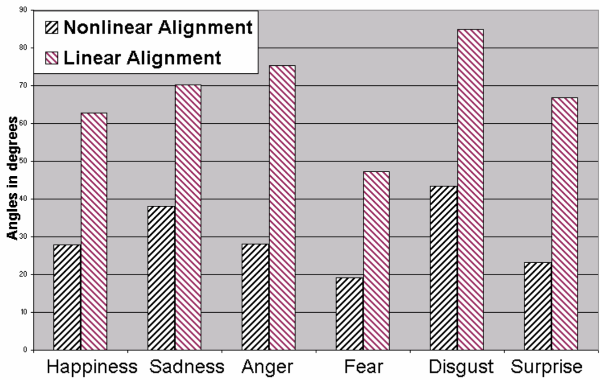


Figure 14: Comparison of angles between expression lines from two subjects after linear/nonlinear alignment.

We apply a *k*-Nearest Neighbor method to classify expressions for all 4851 images on the aligned manifold of facial expression. The comparison of the classification rate between the generalized manifold drawn by nonlinear alignment and linear alignment is presented in Table 2. The experimental results further demonstrate the effectiveness of nonlinear alignment.

Table 2: Comparison of recognition rate on one generalized manifold after linear/nonlinear alignment.

	Manifold by linear alignment	Manifold by nonlinear alignment
k_NN(k=9)	94.42%	96.09%
k_NN(k=13)	94.60%	96.37%

6. Conclusions and Future Work

In this paper, we proposed the concept of the manifold of facial expression. Because the expression manifold can provide a global, analytical representation for all possible facial expressions, it shows promise as a unified framework for facial expression analysis. To explore the structure of the expression manifold in the high dimensional feature space derived by AWN, we investigated two types of embedding methods, LLE and Lipschitz embedding. LLE is suitable for the visualization of manifolds, while Lipschitz embedding is effective for more generalized analysis on the expression manifold. For manifolds of different subjects, we proposed a nonlinear alignment algorithm, which preserves the semantic similarity on one generalized manifold. Nonlinear alignment also outperforms linear alignment in expression classification.

To verify the generality of expression manifolds, images from more subjects are needed for further testing. Only images with six basic emotional expressions were used to learn the structure of the manifold. We will learn more properties of expression manifolds with the images of blended expressions. It has been shown that the change of face pose will add additional dimensionality to the manifold [20,23]. How the pose variations affect the manifold of expression will also be a topic of future research.

Acknowledgements
This work has been supported by NSF ITR grant # 0205740.

References

[1] P. Ekman, *Emotion in the Human Face*, Cambridge University Press, New York, 1982.

[2] I. Shalif, "The Emotions and the Dimensions of Discrimination Among Them in Daily Life", Ph.D. Thesis, 1991.

[3] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu, "Comparison Between Geometry-based and Gabor-wavelets-based Facial Expression Recognition Using Multi-layer Perceptron", *Proceedings of Third IEEE Intel. Conf. On Automatic Face and Gesture Recognition*, 1998.

[4] S. Roweis and L. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding", *Science*, 290; 2323-2326, December 2000.

[5] J. Bourgain, "On Lipschitz Embedding of Finite Metric Spaces in Hilbert Space", *Israel J. Math.*, vol. 52,nos. 1-2, pp. 46-52, 1985.

[6] W. Johnson and J. Lindenstrauss, "Extension of Lipschitz Mapping into a Hilbert Space", *Contemporary Math.*, vol. 26, pp.189-206, 1984.

[7] G. Hristescu and M. Farach-Colton, "Cluster-Preserving Embedding of Proteins", technical report, Rutgers Univ., Piscataway, New Jersey, 1999.

[8] M. Linial, N. Linial, N. Tishby, and G. Yona, "Global Self Organization of All Known Protein Sequences Reveals Inherent Biological Signatures", *J. Molecular Biology*, vol. 268, no.2, pp. 539-556, May 1997.

[9] E. Chuang, H. Deshpande, and C. Bregler, "Facial Expression Space Learning", *Pacific Graphics*, 2002.

[10] M. Brand, "Charting a Manifold", *Proceedings of Neural Information Processing System*, 2002.

[11] S. Roweis, L. K. Saul, and G. E. Hinton, "Global Coordination of Local Linear Models", *Proceedings of Neural Information Processing System*, 2002.

[12] C. Hu, R. Feris, and M. Turk, "Active Wavelet Networks for Face Alignment", *British Machine Vision Conference*, 2003.

[13] P. Ekman and W.V. Friesen, *Facial Action Coding System: Manual*, Palo Alto: Consulting Psychologist Press, 1978.

[14] http://www.imm.dtu.dk/~aam/

[15] K. Schmidt and J. Cohn, "Dynamics of Facial Expression: Normative Characteristics and Individual Difference" *Proceedings of Intel. Conf. On Multimedia and Expo*, 2001.

[16] T.F. Cootes, G.J. Edwards and C.J. Taylor. "Active Appearance Models", *IEEE Transactions on PAMI*, vol.23, No.6, pp.681-685, 2001.

[17] M. Black and Y. Yacoob, "Recognizing Facial Expressions in Image Sequences Using Local Parameterized Models of Image Motion", *Intel. J. of Computer Vision*, 25(1), pp. 23-48, 1997.

[18] T. Cox and M. Cox, *Multidimensional Scaling*, Chapman & Hall, London, 1994.

[19] J. Tenebaum, V. Silva, and J. Langford, "A Global Geometric Framework for Nonlinear Dimensionality Reduction", *Science*, 290; 2319-2323, December, 2000.

[20] Y. Li, S. Gong, and H. Liddell. "Recognizing Trajectories of Facial Identities Using Kernel Discriminate Analysis", *Proceedings of British Machine Vision Conference*, 2001.

[21] D. Bertsekas, *Nonlinear Programming*, Athena Scientific, Massachusetts, 1999.

[22] R.A. Horn, C.R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1990.

[23] S.Z. Li, R. Xiao, Z. Li, and H. Zhang, "Nonlinear Mapping from Multi-View Face Patterns to a Gaussian Distribution in a Low Dimensional Space", *IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems (RATFG-RTS)*, 2001.

[24] N. Linial, E. London, and Y. Rabinovich, "The Geometry of Graphs and Some of Its Algorithmic Applications", *Combinatorica*, vol. 15, pp. 215-245, 1995.