Probabilistic Expression Analysis on Manifolds

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

Abstract

In this paper, we propose a probabilistic video-based facial expression recognition method on manifolds. The concept of the manifold of facial expression is based on the observation that the images of all possible facial deformations of an individual make a smooth manifold embedded in a high dimensional image space. An enhanced Lipschitz embedding is developed to embed the aligned face appearance in a low dimensional space while keeping the main structure of the manifold. In the embedded space, a complete expression sequence becomes a path *on the expression manifold, emanating from a center that corresponds to the neutral expression. Each path consists of several clusters. A probabilistic model of transition between the clusters and paths is learned through training videos in the embedded space. The likelihood of one kind of facial expression is modeled as a mixture density with the clusters as mixture centers. The transition between different expressions is represented as the evolution of the posterior probability of the six basic paths. The experimental results demonstrate that the probabilistic approach can recognize expression transitions effectively. We also synthesize image sequences of changing expressions through the manifold model.*

1. Introduction

Facial expression is one of the most powerful means for people to coordinate conversation and communicate emotions and other mental, social, and physiological cues. While people can recognize facial expressions easily even though the appearance of the expression varies a lot between different individuals, it is a challenging task for a computer to automatically determine 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 non-prototypic expressions such as "raised brows," prototypic expressions such as emotional labels, or facial actions such as the Action Units defined in Facial Action Coding System (FACS) [1]. Psychologists claim that there are six kinds of universally recognized facial expressions: happiness, sadness, fear, anger, disgust, and surprise [2]. Existing expression analyzers [3,28,29] usually classify the examined expression into one of the basic emotion categories. This approach has two major limitations. First, it is not clear whether all facial expressions expressible by the human face can be classified under these six basic categories. For "blended" expressions, it may be more reasonable to classify them 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 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. Nevertheless, there are only five AUs with the option to score intensity on three-level scale (low, medium, and high). It is usually difficult 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 identify a global representation for all possible facial expressions that affords semantic analysis. In this paper, we explore the space of expression images and propose the *manifold of expressions* as a foundation for expression analysis. An image with *N* pixels can be considered a point in an *N*-dimensional image space, and the variability of image classes can be represented as low-dimensional manifolds embedded in the image space. People change facial expressions continuously over time. Thus all images of an individual'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 this 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 expressions, similar expressions are points in the local neighborhood on the manifold. Sequences of basic emotional expressions become

Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04)
1063-6919/04 $20.00 © 2004 **IEEE**

paths on the manifold extended from the reference center, as illustrated in Figure 1. The blends of expressions lie between those paths, so they can be defined analytically by the positions of the basic paths. The analysis of the relationships between different facial expressions is facilitated on the manifold.

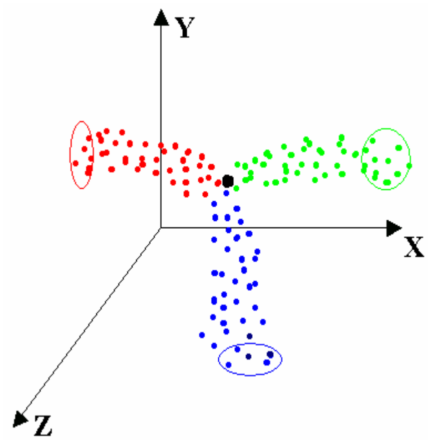

Figure 1: Illustration of a 3D expression manifold. The reference center is defined by the neutral face. Image sequences from three different expressions are shown. The further a point is away from the reference point, the higher is the intensity of that expression.

It is a formidable task to learn the complete structure of the manifold of expressions in a high dimensional image space. To overcome this problem, our core idea is to embed the nonlinear manifold in a low dimensional space and recognize facial expression from video sequences probabilistically. Figure 2 illustrates the overall structure of the system.

We first apply Active Wavelets Networks [5] on image sequences to normalize the variation due to scaling and face pose. An enhanced Lipschitz embedding [6,7] is developed to embed the aligned face appearance in the high dimensional space to a low dimensional space while keeping the main structure of the manifold. In Lipschitz embedding, a coordinate space is defined such that each axis corresponds to a reference set ***R***, drawn from the input data set. Lipschitz embedding leads to good preservation of clusters in practical cases [8,9]. After Lipschitz embedding, the expression sequences in the gallery become paths emanating from the center, which is defined by the neutral expression. We learn the probabilistic model of transition between those paths from the gallery videos. The probe set includes videos of random expression changes, which may not begin or end with neutral expression. The duration and the intensity of the expression are varied. The transition between different expressions is represented as the evolution of the posterior probability of the basic paths. Our empirical study demonstrates that the probabilistic approach can recognize expression transitions effectively. We also synthesize image sequences of changing expressions through the manifold model.

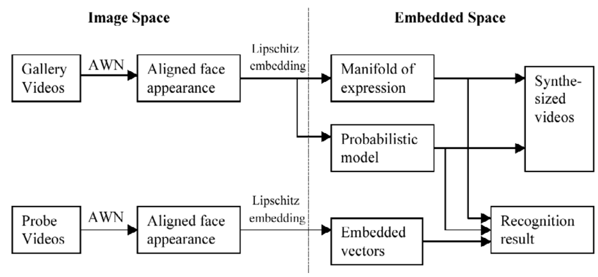

Figure 2: System diagram

The remainder of this paper is organized as follows. We present the related work in Section 2. We then discuss the properties of Lipschitz embedding in Section 3. The probabilistic model built on expression manifolds is described in Section 4. Section 5 presents the experiments we conducted on the manifold of facial expression. Section 6 concludes the paper with discussion.

2. Related Literature

Many researchers have explored the nature of the space of facial expressions. Shalif [10] examined the principal emotional variables in daily life by letting 202 subjects judge the extent of emotional state expressed in photographs. He found that happiness and sadness are opposite directions of a single dimension which is the most prominent dimension in expression space, followed by fear, anger, etc. Schmidt and Cohn [11] measured 195 spontaneous smiles from 95 individuals through facial electromyographic (EMG) data and found consistency in zygomaticus major muscle activity over time. 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. [12] showed that the space of facial expression can be modeled with a bilinear model. Two formulations of bilinear models, asymmetric and symmetric, were fit to facial expression data.

More recently, Seung [25] suggested representing the variability of images as low-dimensional manifolds embedded in image space. Roweis [26] showed that Locally Linear Embedding is able to learn the global structure of nonlinear manifolds, such as those generated by images of faces with only pose and illumination change. Tenenbaum et al. [18] introduced

Isomap to find meaningful low-dimensional structures hidden in the high-dimensional data that is guaranteed to converge asymptotically to the true structure. Kimmel et al. [32] used the invariant signature of manifolds for object recognition.

Facial expression analysis can be performed from static images [3,30] or video sequences [28,29]. Bassili [15] suggested that motion in the image of a face would allow expression to be identified with minimal information about the spatial arrangement of features. Cohen et al. [16] proposed a new architecture of HMMs to segment and recognize facial expression from video sequence automatically. Probabilistic video analysis has gained significant attention since the seminal work of Isard and Blake [13]. They introduced a time series state space model parameterized by a tracking motion vector. Zhou and Chellappa [14] proposed a generic framework to track and recognize human face simultaneously by adding an identity variable to the state vector in the sequential importance sampling method. Lee et al. [4] proposed a video-based face recognition method using probabilistic appearance manifold. The nonlinear appearance is approximated by piecewise linear subspace and the connectivity between the subsets encodes the transition probability between images in them.

In this paper, we analyze the space of expression through the manifold of expression. The manifold is learned from video sequences of basic facial expressions. We embed the manifold of the aligned face appearance from the high dimensional space to a low dimensional space through an enhanced Lipschitz embedding. The probabilistic model of the manifold is learned from gallery videos in the embedded space. The expression transition is represented as the evolution of the posterior probability of the basic paths on the manifold.

3. Lipschitz Embedding

Lipschitz embedding [6,7] is a powerful embedding method used widely in image clustering and image search. For a finite set of input data S, Lipschitz embedding is defined in terms of a set R of subsets of S, $R = \{A_1, A_2, ..., A_k\}$. The subsets A_i are termed the reference sets of the embedding. Let $d(o; A)$ be an extension of the distance function d to a subset $A \subset S$, such that $d(o, A) = \min_{x \in A}\{d(o, x)\}$. An embedding with respect to R is defined as a mapping F such that $F(o) = (d(o; A_1); d(o; A_2); ..., d(o; A_k))$. In other words, Lipschitz embedding defines a coordinate space where each axis corresponds to a subset $A_i \subset S$ of the objects, and the coordinate values of object o are the distances from o to the closest element in each of A_i.

With a suitable definition of the reference set R, the distance of all pairs of data points in the embedding space is bounded [17]. So Lipschitz embedding works well when there are multiple clusters in the input data set [8,9]. In our algorithm, we preserve the intrinsic structure of the expression manifold by combining Lipschitz embedding and the main feature of Isomap [18]. Given a video gallery covering six basic facial expressions, there are six "paths" from the neutral image to the six sets of images with the basic expressions at apex on the manifold. In Figure 1, the apex sets in 3D space are illustrated as the points within the circles. Each path is composed of many small steps (the difference between consecutive frames). Different "paths" contain information on how the expressions evolve. The comparative positions between those paths correspond to the relationship between different expressions.

The distance function in Lipschitz embedding reflects the distance between points on the manifold. The crucial property that we aim to retain is which points are close to each other and which are far from each other. Due to the essential nonlinear structure of the expression manifold, the classical approaches of multidimensional scaling (MDS) and PCA fail to detect the true degrees of freedom of the face data set. Tenenbaum et al. [18] seek to preserve the intrinsic geometry of the data by capturing the geodesic manifold distance between all pairs of data points. For neighboring points, input-space distance provides a good approximation to geodesic distance. For faraway points, geodesic distance can be approximated by adding up a sequence of "short hops" between neighboring points. This shortest path can be computed efficiently by the Dijkstra Algorithm [19]. The details of geodesic distance computation can be found in [18].

For our experiments, we used six reference sets, each of which contains only the images of one kind of basic facial expression at its apex. The embedded space is six dimensional. The distance function is the geodesic manifold distance. After we apply the enhanced Lipschitz embedding to the gallery set, there are six basic paths in the embedded space, emanating from the center that corresponds to the neutral image. The images with blended expression lie between the basic paths. In the embedded space, expressions can be recognized by using the probabilistic model described in the next section.

4. Probabilistic model on the expression manifold

In this section, we present the details of the probabilistic model on the manifold of expression. The goal of the probabilistic model is to exploit the temporal information in video sequences. Expression recognition is performed on the manifold constructed for each individual.

4.1. Learn the priors from the gallery

We apply an enhanced Lipschitz embedding to the gallery. The gallery contains videos with only one kind of expression each. Assume there are K image sequences for each kind of basic expression $S, S=\{1,\ldots,6\}$. The embedded vector for the *i*th image in the *j*th video sequence for expression S is $I_{s,j,i} \in R^6$, $j=\{1,\ldots,K\}$. By K-means clustering technique, all points are grouped into clusters $c^n, n=1,\ldots,r$. We compute a cluster frequency measure

$T_{n1,n2} = \#(I_{s,j,i} \in c^{n1} \,\&\, I_{s,j,i+1} \in c^{n2}, j=1,\ldots,K, S=1,..6)$

The prior $p(c^{n2} \mid c^{n1})$ is learned as

$$p(c^{n2} \mid c^{n1}) = \begin{cases} \delta, T_{n1,n2} = 0 \\ T_{n1,n2} * scale \;, otherwise \end{cases}$$

where δ is a small empirical number. Scale and δ are selected such that $\sum_{n2} p(c^{n2} \mid c^{n1}) = 1$.

The prior $p(c \mid S)$ is assigned according to the expression intensity of the cluster center, varying from 0 to 1. For example, the index of the "anger" expression is 1. When the image of a cluster center shows anger intensely, there is less ambiguity for the frames in the cluster to be classified as anger. Therefore the corresponding cluster has a higher $p(c \mid S=1)$.

By Bayes' rule,

$$p(S \mid c) = \frac{p(c \mid S)p(S)}{\sum_S p(c \mid S)p(S)}.$$

For time series $t=0,1,\ldots$, the transition between different expressions can be computed as the transition between the clusters:

$$p(S_t \mid S_{t-1}) = \sum_{n1,n2} p(S_t \mid c_t = c^{n2}) p(c_t = n2 \mid c_{t-1} = c^{n1}) p(c_{t-1} = c^{n1} \mid S_{t-1})$$

Due to the small variation within a cluster, S_{t-1} and S_t are conditionally independent given c_{t-1}.

4.2. Expression Recognition

Given a probe video sequence in the embedded space $I_t, t=0,1,\ldots$, the expression recognition can be represented as the evolution of the posterior probability $p(S_{0:t} \mid I_{0:t})$.

We assume statistical independence between prior knowledge on the distributions $p(c_0 \mid I_0)$ and $p(S_0 \mid I_0)$. Using the overall state vector $x_t = (S_t, c_t)$, the transition probability can be computed as:

$$p(x_t \mid x_{t-1}) = p(S_t \mid S_{t-1})\, p(c_t \mid c_{t-1}) \qquad (1)$$

We define the likelihood computation as follows

$$\begin{aligned} & p(I \mid c, S) \\ &= p(I \mid c)\, p(c \mid S) \\ &\propto \exp[-\frac{1}{2\sigma_c^2} d(I, u_c)]\, p(c \mid S) \end{aligned}$$

where u_c is the center of cluster c, σ_c is the variation of cluster c.

Given this model, our goal is to compute the posterior $p(S_t \mid I_{0:t})$. It is in fact a probability mass function (PMF) since S_t only takes values from 1 to 6. The marginal probability $p(S_t, c_t \mid I_{0:t})$ is also a PMF for the same reason.

Using Equation (1), the Markov property, statistical independence, and time recursion in the model, we can derive:

$$\begin{aligned} & p(S_{0:t}, c_{0:t} \mid I_{0:t}) = p(x_{0:t} \mid I_{0:t}) \\ &= p(x_{0:t-1} \mid I_{0:t-1}) \frac{p(I_t \mid x_t) p(x_t \mid x_{t-1})}{p(I_t \mid I_{0:t-1})} \\ &= p(S_{0:t-1}, c_{0:t-1} \mid I_{0:t-1}) \frac{p(I_t \mid c_t, S_t) p(S_t \mid S_{t-1}) p(c_t \mid c_{t-1})}{p(I_t \mid I_{0:t-1})} \\ &= p(S_0, c_0 \mid I_0) \prod_{i=1}^{t} \frac{p(I_i \mid S_i, c_i) p(S_i \mid S_{i-1}) p(c_i \mid c_{i-1})}{p(I_i \mid I_{0:i-1})} \end{aligned}$$

By marginalizing over $c_{0:t}$ and $S_{0:t-1}$, we obtain Equation (2):

$$p(S_t \mid I_{0:t}) = \int_{c_0} \int_{S_0} \cdots \int_{c_{t-1}} \int_{S_{t-1}} \int_{c_t} p(c_0 \mid I_0) p(S_0 \mid I_0) * \prod_{i=1}^{t} \frac{p(I_i \mid S_i, c_i) p(S_i \mid S_{i-1}) p(c_i \mid c_{i-1})}{p(I_i \mid I_{0:i-1})} dc_t dS_{t-1} dc_{t-1} \ldots dS_0 dc_0 \qquad (2)$$

which can be computed by the priors and the likelihood $p(I_i \mid S_i, c_i), i=1,\ldots,t$. This gives us the probability distribution of the expression categories, given the image sequence.

4.3. Synthesis of dynamic expressions

The manifold model can also be used to synthesize an image sequence with changing expressions. Given expression S, we keep the indexing $l_1,..., l_r$, $l = 1,...,6$, and r is the number of the clusters, such that:

$$p(c^{l_1} \mid S = l) < p(c^{l_2} \mid S = l)... < p(c^{l_r} \mid S = l)$$

For expression l, there are k gallery videos that begin from the neutral expression, pass the apex, and end with the neutral expression. We set the first video sequence as a template. Then we apply dynamic time warping [22] to the following $k-1$ image sequences. Thus we have a standard time index for all k videos. For every cluster along the path l, we can measure the duration of the cluster by the range of time index of the images within the cluster. Note we compute the time range for increasing and decreasing expression separately since a cluster may cover both types of images at the same time. The time range for each cluster is $w_i, i = 1,..., r$. The average time range of all clusters is $\overline{w}$.

The algorithm for synthesizing an image sequence from expression A to expression B is listed in Figure 3. The critical part is to find a trajectory that maximizes the probability of the transitions between the clusters A_r and B_r. The optimal trajectory is computed by dynamic programming [31]. The correlations between consecutive frames are maximized locally at the same time. To eliminate the jitter and redundancy in the image sequence, we keep a cache for recently appeared frames. If the same frame from the gallery appearances more than twice in the passed n frames, it should be removed from the final video sequence. n is an empirical window width.

Input:

The beginning expression category: $A \in \{1,...,6\}$

The ending expression category: $B \in \{1,..., 6\}$

The length of synthesized video sequence: $fnum$

The embedded vectors of r cluster centers:

$d_i \in R^6, i = 1,..., r$

Output:

Image sequence P

Function:

$floor\ (x)$: return the maximum integer no more than x .

$findnear\ (d, z)$: return the nearest z points to the embedded vector d on the learned manifold.

$GetRaw\ (d)$: return the corresponding face image to the embedded vector d .

correlation(x, y): return the correlation between two images.

$T = floor\ (fnum\ / \overline{w})$;

$n_1 = A_r$; {the cluster with strongest expression A}

$n_T = B_r$; {the cluster with strongest expression B}

$[n_2,..., n_{T-1}] = \arg\max(P(c^{n_T} \mid c^{n_{T-1}})...P(c^{n_2} \mid c^{n_1}))$;

$count = 0$;

for i = 1 **to** $T-1$

$betweenc = (w_{n_i} + w_{n_{i+1}}) / 2$;

$fbegin = findnear(d_i, 1)$;

$fend = findnear(d_{i+1}, 1)$;

$count = count + 1$;

$P_{count} = GetRaw(fbegin)$;

for j **=** *1* **to**

$dist = (fend - fbegin)/(betweenc - j)$;

$candi = findnear(fbegin + dist, 5)$;

$comp = GetRaw(fbegin)$;

for k **=** 1 **to** 5

$candi\_im(k) = GetRaw(candi(k))$;

corr(k)=correlation(comp, candi_im(k));

if $(k == 1)$

then $se = 1$;

$\max = corr\ (k)$;

else if $(corr\ (k) > \max)$

$se = k$;

$\max = corr\ (k)$;

end

end

end

$count = count + 1$;

$P_{count} = candi\_im(se)$;

$fbegin = candi\ (se)$;

end

end

return $P_i, i = 1,..., count$

Figure 3: Expression Synthesis algorithm

5. Experimental results

In this section, we present our empirical studies on five subjects. In our experiments, subjects were instructed to perform a series of facial expressions representing happiness, sadness, anger, surprise, fear, and disgust. The subjects repeated the series seven times for each gallery set. The probe set includes video sequences in which the subjects can change their expression randomly. It was not necessary to begin or end with a neutral expression. The duration and the intensity of the expression are varied. We maintained constant illumination during the experiments and the subjects' faces are all near frontal view.

5.1. Preprocessing

To reduce the variation due to scaling and face poses, we first applied Active Wavelets Networks (AWN) [5] on the image sequence for face registration and facial feature localization. The key idea of AWN is to replace the PCA-based texture model in Active Appearance Models (AAM) [23] with a wavelet network representation. More details about AWN can be found in [5]. Figure 4 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 [24]. From the gallery, about 60 images are selected as the training set, which covers the six kinds of basic expressions. The localizations of all the other images are completely automatic.

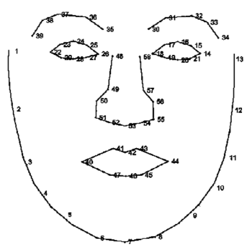

Figure 4: Facial landmarks (58 points)

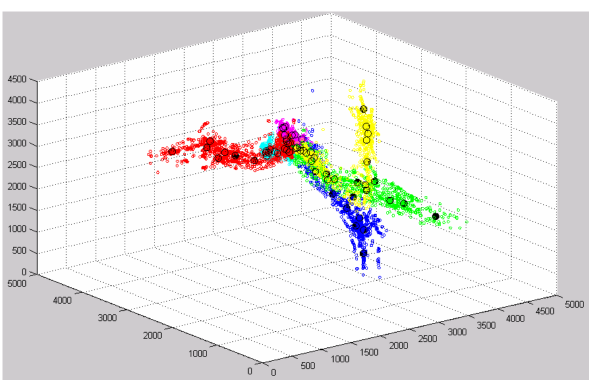

Figure 5: An expression manifold projected on its first three dimensions. Points with different colors represent images with different expression. Anger: red; Disgust: green; Fear: blue; Sad: cyan; Smile: pink; Surprise: yellow. The black points represent 60 cluster centers.

The positions of points on eyebrow and eye contour are used to align the face position. The positions of points on face contour are used to normalize the faces to the same size. The aligned face appearance is used as the input of Lipschitz embedding. Figure 5 is a manifold projected on its first three dimensions for visualization purpose. The number of clusters for K-means algorithm is chosen such that there is no dramatic expression variation within a cluster.

The appearance of different subjects could be aligned through a common 3D face model. Currently, we build a separate manifold for each subject. These manifolds share a similar "skeleton" shape, but vary in reference set positions and path directions. With warped appearance data, the subjects from different subjects can be aligned through linear or nonlinear alignment [27].

5.2. Facial expression recognition

The probe set contains video sequences in which subjects changed their expression randomly. The sequences were recorded at 30 fps and stored at a resolution of 320x240. All results in this paper were obtained on a Xeon 2.8GHz CPU. The complete process, including alignment, embedding, and recognition, runs at 5 fps.

A supplementary video[1] demonstrates the recognition results on the probe set. Overlaid graphical bars indicate the posterior probability of the basic expressions, as shown in Figure 6. The learned manifold is visualized with the embedded vector of the current frame (a black point) at the same time. During the expression transition, the black point "walks" from one expression path to another. The viewpoint of the manifold is changed concurrently for better visualization. Figure 6 shows some sample images from the submitted video. The first image is during a transition from fear to surprise. The second image is during a transition from anger to disgust. The third image and the fourth image are sadness and happiness respectively. The bar figures indicate the expression transition correctly.

5.3. Expression Synthesis

With the manifold model, we synthesize image sequences of aligned face appearance with changing expressions. There are about 6000 images from 42 video sequences (seven for each basic expression) in each gallery set. The lengths of synthesized image sequences are around 200.

Figures 7 and 8 show some selected images (every 20th frame) from the synthesized sequences. The trajectories with the maximum transition probability between clusters reflect the expression change correctly.

[1] Available at http://ilab.cs.ucsb.edu/demos/cvpr04.m2v

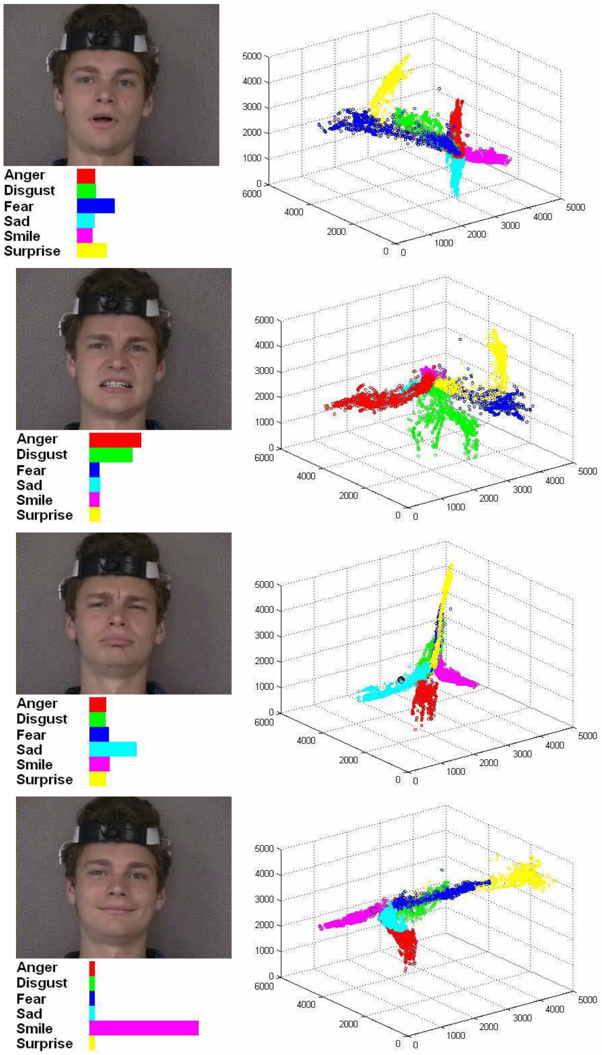

Figure 6: Facial expression recognition result with manifold visualization

6. Conclusions and Discussions

In this paper, we proposed the concept of the manifold of facial expressions. The expression manifold provides a global representation for all possible facial expressions and affords semantic analysis. It shows promise as a unified framework for facial expression analysis. To explore the structure of the expression manifold in high dimensional image space, we apply an enhanced Lipschitz embedding to the aligned face appearance data. In the embedded space, the expression sequences become paths on the expression manifold. Each path consists of several clusters. A probabilistic model of transition between the paths and clusters is learned through training videos. The likelihood of one kind of facial expression

Figure 7: 12 frames selected from a transition from anger to happiness.

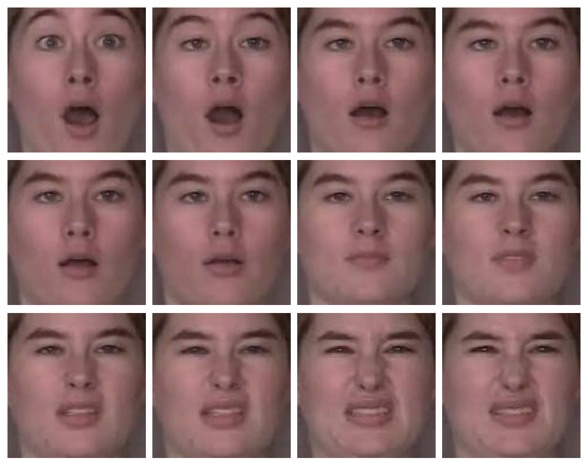

Figure 8: 12 frames selected from a transition from surprise to disgust.

is modeled as a mixture density with the clusters as the mixture center. The transition between different expressions is represented as the evolution of the posterior probability of the six basic paths.

The experiments results on the probe sets demonstrate that the expression transition can be recognized effectively. We also synthesize image sequences of changing expressions through the manifold model. Our experimental results show that manifold methods provide an analytical way to analyze the relationship between different expressions, and to recognize blended expressions.

We will evaluate and quantify the results more systematically with many more subjects in future work. Another future research direction is to consider variation on face pose and illumination [20, 21], which will add more degrees of freedom to manifold of expression. How these factors affect the intrinsic geometry of expression manifold will be a challenging topic for future study.

Acknowledgements

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

References

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

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

[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", *Third IEEE Intel. Conf. On Automatic Face and Gesture Recognition*, 1998.

[4] K. Lee, J. Ho, M.H. Yang, and D. Kriegman, "Video-based Face Recognition Using Probabilistic Appearance Manifolds", *Conference on Computer Vision and Pattern Recognition*, 2003.

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

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

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

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

[9] 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.

[10] I. Shalif, "The Emotions and the Dimensions of Discrimination Among Them in Daily Life", Ph.D. Thesis, Psychology Dept., Bar-Ilan Univ., Ramat-Gan, Israel, 1991.

[11] K. Schmidt and J. Cohn, "Dynamics of Facial Expression: Normative Characteristics and Individual Difference", *Intl. Conf. on Multimedia and Expo*, 2001.

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

[13] M. Isard, A. Blake, "Contour tracking by stochastic propagation of conditional density", *Proc. Of European Conference on Computer Vision,* 1996.

[14] S. Zhou, V. Krueger, R. Chellappa. "Probabilistic recognition of human faces from video", *Computer Vision and Image Understanding,* 2003.

[15] J.N. Bassili, "Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face", *Journal of Personality and Social Psycholog*y, 37:2049–2059, 1979.

[16] I. Cohen, N. Sebe, A. Garg, L.S. Chen, and T.S. Huang, "Facial Expression Recognition From Video Sequences: Temporal and Static Modeling", *Computer Vision and Image Understanding,* 2003.

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

[18] J.B. Tenenbaum, V. de Silva and J.C. Langford. "A global geometric framework for nonlinear dimensionality reduction", *Science*, vol. 290, pp. 2319--2323, 2000.

[19] E.W. Dijkstra. "A note on two problems in connection with graphs", *Numeriche Math*, 1, pp.269-271, 1959.

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

[21] 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.

[22] C.S. Myers, and L. R. Rabiner, "A comparative study of several dynamic time-warping algorithms for connected word recognition", *The Bell System Technical Journal*, 60(7): 1389-1409, September 1981.

[23] T.F. Cootes, G.J. Edwards and C.J. Taylor. "Active Appearance Models", *IEEE Trans. on* Pattern Analysis and Machine Intelligence, vol.23, No.6, pp.681-685, 2001.

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

[25] H.S. Seung and D.D. Lee, "The Manifold Ways of Perception", *Science*, vol 290, December 2000.

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

[27] Y. Chang, C. Hu, and M. Turk, "Manifold of facial expression," *Proc. IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, Nice, France, Oct. 17, 2003.

[28] 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.

[29] I. Essa and A. Pentland, "Coding, Analysis Interpretation, Recognition of Facial Expressions", *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, pp. 757-763, July 1997.

[30] M. Pantic and L.J.M. Rothkrantz, "Expert System for Automatic Analysis of Facial Expression", *Image and Vision Computing J.*, vol. 18, no. 11, pp. 881-905, 2000.

[31] D. Bertsekas, *Dynamic Programming and Optimal Control*, Athena Scientific, Massachusetts, 2000.

[32] A. Elad, R. Kimmel, "On bending invariant signatures for surfaces", *IEEE Trans. Pattern Analysis and Machine Intelligence,* Vol. 25, Issue: 10, Oct. 2003, pp. 1285 - 1295