Manifold Based Analysis of Facial Expression

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

Abstract

We propose a novel approach for modeling, tracking and recognizing facial expressions. Our method works on a low dimensional expression manifold, which is obtained by Isomap embedding. In this space, facial contour features are first clustered, using a mixture model. Then, expression dynamics are learned for tracking and classification. We use ICondensation to track facial features in the embedded space, while recognizing facial expressions in a cooperative manner, within a common probabilistic framework. The image observation likelihood is derived from a variation of the Active Shape Model (ASM) algorithm. For each cluster in the low-dimensional space, a specific ASM model is learned, thus avoiding incorrect matching due to non-linear image variations. Preliminary experimental results show that our probabilistic facial expression model on manifold significantly improves facial deformation tracking and expression recognition.

1. Introduction

Computational facial expression analysis is an active and challenging research topic in computer vision, impacting important applications such as in human-computer interaction and data-driven animation. Approaches for automatic modeling and recognition of facial expressions are generally classified as static (processing still images) and dynamic (tracking and analyzing facial deformations in video sequences).

In the past decade, many techniques have been proposed to automatically classify expressions in still images, using methods based on Neural Networks [1,16], Gabor wavelets [2] and rule-based methods [3], to mention just a few. However, in recent years, more attention has been given to modeling facial deformation in dynamic scenarios [4,5,18], which allows the integration of information temporally across the video sequence, potentially increasing recognition rates over single-image approaches. For such dynamic scenarios, current methods work in two separate stages: tracking and recognition. The tracking module extracts features over time, while the recognition module processes this information for expression classification.

Many systems obtain facial motion information by computing dense flow between successive image frames. But flow estimates are easily disturbed by the variation of lighting and non-rigid motion, and they are also sensitive to the inaccuracy of image registration and motion discontinuities [5].

Model-based approaches, such as Active Shape Models (ASM) [6] and Active Appearance Models (AAM) [7] have been successfully used for tracking facial deformation. The ASM method detects facial landmarks through a local-based search constrained by a global shape model, statistically learned from training data. The AAM algorithm elegantly combines shape and texture models, assuming a linear relationship between appearance and parameter variation. Both methods, however, tend to fail in the presence of non-linear image variations such as those caused by large facial expression changes.

Nonlinear embedding methods such as ISOMAP [8], local linear embedding (LLE) [20], charting a manifold [21], and global coordinate of local linear models [22] are promising in handling high dimensional nonlinear data. Recently, researchers have applied manifold methods to face recognition [10,19,23] and facial expression representation [15, 24].

This paper presents a novel representation for dynamic facial expression analysis, as well as a probabilistic framework for tracking and recognizing facial deformation in a cooperative manner. Our assumption is that video sequences of a person undergoing different facial expressions define a smooth and relatively low dimensional manifold in a feature space described by a set of facial landmarks.

Initially, we use Isomap embedding [8] to project our training video data into the low dimensional expression manifold. Then, a Gaussian mixture model is applied to cluster data in the low dimensional expression space. For each cluster, a specific ASM model is learned, since tracking by online probabilistic model is more robust to non-linear image variations. In addition, we learn the dynamics for each cluster in the manifold to improve tracking and recognition.

Based on this representation, a particle filter tracker is used to track facial deformation in the embedded space, while recognizing facial expressions. Differing from traditional methods that consider expression tracking and recognition in separate stages, we address these tasks in a common probabilistic framework, which enables them to be solved in a cooperative manner.

The remainder of this paper is organized as follows: in Section 2 we discuss related work. Section 3 covers the learning of our proposed representation, while Section 4 describes the framework to track and recognize facial expressions. Section 5 reports our experimental results and Section 6 presents conclusions and future work.

2. Related Work

Recently, Wang et al. [9] demonstrated the importance of applying non-linear dimensionality reduction in the field of non-rigid object tracking. In fact, representing the object state as a globally coordinated low dimensional vector improves tracking efficiency and reduces local minimum problems in optimization. They learn the object's intrinsic structure in a low dimension manifold with density modeled by a mixture of factor analyzers. Our work also models the intrinsic structure of facial expressions for tracking, but extends it to include recognition in a unified probabilistic framework.

We were also inspired by the work of Lee et al. [10], who present a method for modeling and recognizing human faces in video sequences. They use an appearance model composed of pose manifolds and a matrix of transition probabilities to connect them. In our work, we consider transition probabilities among clusters in the embedded space, effectively capturing the dynamics of expression changes and exploiting the temporal information for recognition.

Zhou, Krueger and Chellapa [11] proposed a generic framework to track and recognize human faces simultaneously by adding an identity variable to the state vector in the sequential importance sampling method. The posterior probability of the identity variable is then estimated by marginalization. Their work, however, does not consider tracking and recognition of facial deformation, the main focus of this paper.

Cootes et al. [6] proposed the Active Shape Model algorithm, which detects facial landmarks through a local-based search constrained by a global shape model, statistically learned from training data. This method was extensively used for facial deformation tracking, but may fail under large expression transitions. In our approach, we use specific ASM models for each cluster in the embedded space. On-line model selection is done probabilistically in a cooperative manner with expression classification, thus improving tracking reliability.

3. Learning Dynamic Facial Deformation

Non-linear dimensionality reduction has attracted attention for a long time in computer vision and visualization research. Images lie in a very high dimensional space, but a class of images generated by latent variables lies on a manifold in this space. For human face images, the latent variables may be the illumination, identity, pose and facial deformations. In this paper, we are interested in embedding the facial deformations of a person in a very low dimensional space, which reflects the intrinsic structure of facial expressions. From training video sequences of different people undergoing different expressions, a low dimensional manifold is learned, with a subsequent probabilistic modeling used for tracking and recognition.

3.1. The Training Database

For preliminary testing, we collected a database of two subjects who were asked to perform six basic facial expressions multiple times. To reduce the influence of illumination variation, we preprocessed the training data video sequence by detecting a set of 2D facial landmarks in each image, which defines the shape of a face in each particular frame. We use the Active Shape Model algorithm to accomplish this task. With a good manual initialization and separate training models prepared specifically for each expression image set, we can extract the face shape precisely. Figure 1 shows the facial points in our shape model.

The whole training dataset, comprising different video sequences of different people undergoing different facial expressions, is then specified by a set $X = \{x_1, ..., x_n\}$, where $x_i \in R^{2D}$ notes a set of D facial points in a particular frame, and n denotes the total number of images in the training data. Unlike traditional manifold embedding papers, where data can be in any order, our training images are ordered according to the video sequences, thus allowing the learning of dynamics on the manifold, as we will show later.

3.2. Isomap Embedding

To embed the high dimension data set $X = \{x_1, ..., x_n\}$ with $x_i \in R^{2D}$ to a space with low dimension d < 2D, we use the Isomap embedding algorithm [8]. Our goal is to find the latent variable

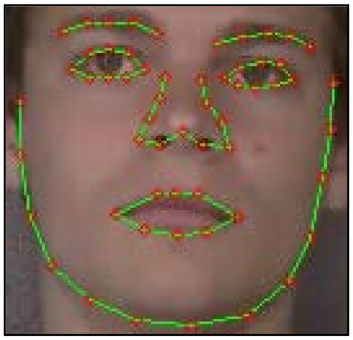

Figure 1. The shape model, defined by a set of facial landmarks.

$Y = \{y_1, y_2, \ldots y_n\}$, where $y_i \in R^d$. This latent variable encodes the knowledge of the data set and controls the data variations.

In the Isomap algorithm, the local geometry of the high dimensional manifold is initially measured through the distances between neighboring data points. For each pair of non-neighboring data points, Isomap finds the shortest path through the data set connecting them, subject to the constraint that the path must hop from neighbor to neighbor. The length of this path is an approximation to the distance between its end points, as measured within the underlying manifold. Finally, the classical method of multidimensional scaling [14] is used to find a set of low dimensional points with similar pairwise distances.

Figure 2 shows the result of projecting our training data (set of facial shapes) in a three dimensional space using Isomap embedding.

3.3. Mixture Model on the Embedded Space

In the lower dimensional embedded space, we describe the distribution of the data using a Gaussian Mixture Model (GMM). The Expectation-Maximization (EM) algorithm is used to estimate the distribution. The following equation describes the density model, where $p(y)$ is the probability that a point in the low dimensional space is generated by the model, k is the number of Gaussians, $p(\omega = i)$ constitutes the mixture coefficients and $N(\mu_i, C_i)$ describes each Gaussian distribution with mean μ_i and covariance matrix C_i:

$$p(y) = \sum_{i=1}^{K} p(\omega = i) N(\mu_i, C_i) \tag{3.1}$$

Figure 3 is an example to illustrate this GMM in the embedded space. Ellipsoids centers and sizes show the mixture centers and the covariance respectively.

3.4. Cluster-Based Active Shape Models

If we were to train an Active Shape Model from all the images in a data set together, the significant variation in the data set would not be modeled well and the tracking performance would be poor. Instead, we train a set of ASM models for each image cluster; that is, for each set of images corresponding to a mixture center (with a defined covariance) of the GMM in the embedded space.

We also propose a method to select and probabilistically integrate the ASM models in ICondensation framework. We will show in Section 4 that online model selection allows tracking to be robust under large expression variations.

In ASM, a shape vector S is represented in the space spanned by a set of eigenvectors learned from the training data. As a result, S may be expressed as:

$$S = \overline{S} + Us \tag{3.2}$$

where $\overline{S}$ is the mean shape, U is the matrix consisting of eigenvectors and s constitutes the shape

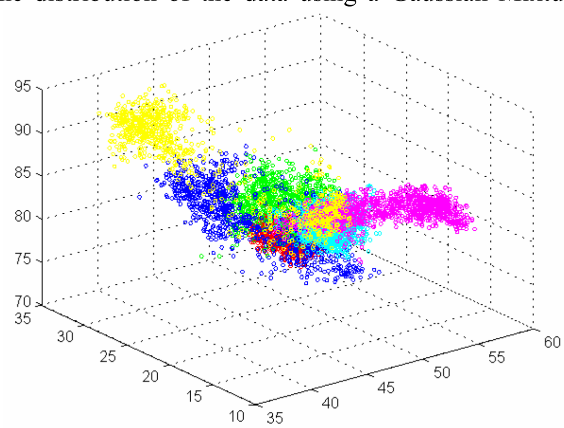

Figure 2: Training data in the embedded space. Different colors correspond to different expressions.

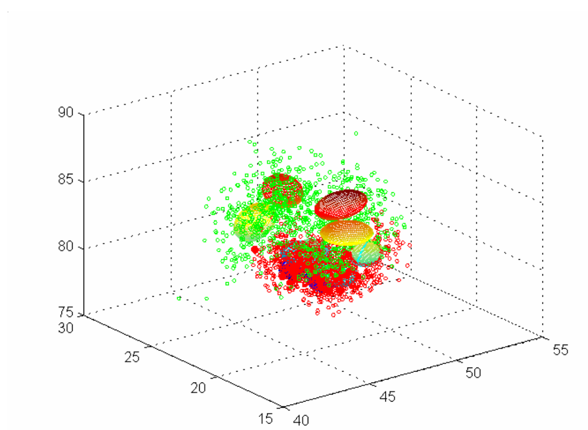

Figure 3: GMM in the embedded space. Ellipsoids centers show the mixture centers; sizes show the covariance.

parameters, which are estimated during ASM search. In Section 4, we will describe how tracking is achieved using the learned ASM models.

3.5. Learning Dynamics on the Manifold

Based on the manifold representation, we can learn a dynamic model, defined as the transition probability $p(y_t \mid y_{t-1})$. Let $\omega \in \{1,..., k\}$ be a discrete random variable denoting the cluster center and let $r \in \{1,..., n_r\}$ be a discrete random variable denoting the expression class. For this work, n_r = 6, meaning that r can assume six basic expressions. We have been using the prototypical universal expressions of fear, disgust, happiness, sadness, anger and surprise, though the method does not depend on this particular grouping.

The dynamic model can be factorized in the following way:

$$p(y_t \mid y_{t-1}) = \sum_{w_t} p(y_t \mid y_{t-1}, \omega_t)\, p(\omega_t \mid y_{t-1})$$

$$= \sum_{\omega_t, \omega_{t-1}} p(y_t \mid y_{t-1}, \omega_t)\, p(\omega_t \mid \omega_{t-1})\, p(\omega_{t-1} \mid y_{t-1}) \qquad (3.3)$$

where $p(\omega_t \mid \omega_{t-1}) = \sum_{r_{t-1}} p(\omega_t \mid \omega_{t-1}, r_{t-1})\, p(r_{t-1})$

This assumes that ω_t and y_{t-1} are conditionally independent given ω_{t-1}.

For each state of r_{t-1} (i.e., each expression class), the cluster transition dynamics $P(\omega_t \mid \omega_{t-1}, r_{t-1})$ can be learned from the training data. $P(y_t \mid y_{t-1}, \omega_t)$ is the dynamic model for a known cluster center. For simplification, we assume the dynamics in a fixed cluster is the same for each expression.

Similar to Wang, et al. [9], we also model the within cluster transition as a first order Gaussian Auto-Regressive Process (ARP) by:

$$p(y_t \mid y_{t-1}, \omega_t) = N(A_{\omega_t} y_{t-1} + D_{\omega_t}, BB^T) \qquad (3.4)$$

which can be represented in generative form as

$$y_t = A_{\omega_t} y_{t-1} + D_{\omega_t} + Bw_k \qquad (3.5)$$

where A_{ω_t} and D_{ω_t} are the deterministic parameters of the process, BB^T is the covariance matrix, and w_k is independent random white noise.

For AR parameter learning, we use the same method as Blake and Isard [12]. Combining equations (3.3), (3.4) and (3.5), we get:

$$p(y_t \mid y_{t-1}) = \sum_{\omega_t \omega_{t-1} r_{t-1}} p(y_t \mid y_{t-1}, \omega_t) p(\omega_t \mid \omega_{t-1}, r_{t-1}) p(r_{t-1}) p(\omega_{t-1} \mid y_{t-1}) = \sum_{\omega_t} N(A_{\omega_t} y_{t-1} + D_{\omega_t}, BB^T) \alpha(\omega_t; y_{t-1}) \qquad (3.6)$$

where

$$\alpha(\omega_t; y_{t-1}) = \sum_{\omega_{t-1}, r_{t-1}} P(\omega_t \mid \omega_{t-1}, r_{t-1}) P(r_{t-1}) P(\omega_{t-1} \mid y_{t-1}) \qquad (3.7)$$

Wang et al. [9] pointed out that the equations above model a Mixture of Gaussian Diffusion (MGD), whose mixture term is controlled by the random variable ω_t. In our work, the mixture term is also controlled by the expression recognition random variable.

4. Probabilistic Tracking and Recognition

In the previous section, we showed how to learn a facial expression model on the manifold as well as its associated dynamics. Now, we show how to use this representation to achieve robust online facial deformation tracking and recognition. Our probabilistic tracking is based on the ICondensation algorithm [13], which is described next, followed by expression classification. Both tracking and recognition are described in the same probabilistic framework, which enables them to be carried out in a cooperative manner.

4.1. ICondensation Tracking

Our object state is composed of rigid and non-rigid parts, defined by $s = (x, y, \theta, sc\,; y_1 ... y_d)$. The rigid part (x, y, θ, sc) represents the rigid face motion (position, orientation and scale), while the non-rigid part ($y_1 ... y_d$) is the low dimensional representation of facial deformation obtained by Isomap embedding, as described in Section 3.

At time *t*, the conditional object state density is represented as a weighted set of samples $\{(s_t^{(n)}, \pi_t^{(n)}), n = 1,... N\}$, where $s_t^{(n)}$ is a discrete sample with associated weight $\pi_t^{(n)}$, where $\sum_n \pi_t^{(n)} = 1$. Below we illustrate one step of a sample's evolution.

After this step, the state with largest weight describes the tracking output in each frame, consisting of face pose (x, y, θ, sc) and deformation, which is obtained by projecting ($y_1 ... y_d$) back to the original shape space, through a nearest-neighbor scheme.

Sequential Importance Sampling Iteration:

Main Objective: Generate sample set $S_t = \{(s_t^{(n)}, \pi_t^{(n)}), n = 1,... N\}$ at time t from sample set $S_{t-1} = \{(s_{t-1}^{(n)}, \pi_{t-1}^{(n)}), n = 1,... N\}$ at time $t-1$.

Algorithm:

For each sample, n = 1 to N:

1) Create samples $\tilde{s}_t^{\,n}$

Choose one of the following sampling methods with a fixed probability:

(1) Generate sample from initialization prior.

(2) Generate sample from importance re-sampling, where the importance function is the posterior from time t-1;

2) Predict s_t^n from $\tilde{s}_t^{\,n}$

a) If $\tilde{s}_t^{\,n}$ was generated from the prior probability, choose s_t^n from $\tilde{s}_t^{\,n}$ adding a fixed Gaussian noise.

b) If $\tilde{s}_t^{\,n}$ was generated from the posterior probability, apply the dynamic model for prediction. For the rigid state part, we use constant prediction, adding a small Gaussian noise. For the non-rigid part, we use the MGD noise model, where the weight of each component is controlled by the cluster center distribution $p(\omega_t)$ and expression classification distribution $p(r_t)$.

3) Update the set of samples. The measurement of the sample s_t^n is $\pi_t^{(n)} = \lambda_t^{(i)} *$ M(s_t^n), where $\lambda_t^{(i)}$ is the importance sampling correction term. M is the sample measurement function, described in the next subsection.

4) After all the samples are generated and measured, normalize $\pi_t^{(n)}$ so that $\sum_n \pi_t^{(n)} = 1$ and store the sample set as $\{(s_t^{(n)}, \pi_t^{(n)}), n = 1,... N\}$

4.1.1. Sample Measurement

In order to measure a sample (function M in the algorithm above), we proceed in the following way. For each mixture center in the embedded space, a specific ASM model is selected to measure image observation. This measure is given by a residual error obtained after applying one step of ASM search (we refer to Cootes et al. [6] for details on the search process). Face pose initialization is given by the sample rigid part (x, y, θ, sc) and shape initialization is computed by projecting the non-rigid part ($y_1 ... y_d$) of the sample back to the original shape space (using a nearest-neighbor scheme).

Once we have a residual error for each one of the mixture centers, the desired sample measurement is obtained by a weighted sum of these residuals, where the weights corresponds to the likelihood of the sample non-rigid part ($y_1 ... y_d$) in each Gaussian model.

This scheme allows tracking to be robust under large facial expression changes, as we will show in Section 5. Next we describe how to update expression classification in each frame, using a common probabilistic framework.

4.2. Expression Recognition Updating

We have already showed that the distribution of the discrete random variable r (the expression recognition variable) directly affects tracking (see sample prediction and dynamic model learning). Now we show how to update the posterior probability $p(r_t \mid y_{0:t})$ in every frame to identify facial deformation.

In the ICondensation tracking, by assuming statistical independence between all noise variables, Markov property and priors of the distributions $p(\omega_0 \mid r_0)$, $p(r_0 \mid y_0)$, $p(y_t \mid y_{t-1})$ on embedded space, our goal is to compute the posterior $p(r_t \mid y_{0:t})$. It is in fact a probability mass function (PMF) as well as a marginal probability of $p(r_t, \omega_t \mid y_{0:t})$. Therefore, the problem is reduced to computing the posterior probability $p(r_{0:t}, \omega_{0:t} \mid y_{0:t})$.

$$
\begin{aligned}
&p(r_{0:t}, \omega_{0:t} \mid y_{0:t}) = \\
&p(r_{0:t-1}, \omega_{0:t-1} \mid y_{0:t-1}) \frac{p(y_t \mid r_{0:t-1}, \omega_{0:t-1}) p(r_t \mid r_{t-1}) p(\omega_t \mid \omega_{t-1})}{p(y_t \mid y_{0:t-1})} \\
&= p(r_0, \omega_0 \mid y_0) \prod_{l=1}^{t} \frac{p(y_l \mid r_l, \omega_l) p(r_l \mid r_{l-1}) p(\omega_l \mid \omega_{l-1})}{p(y_l \mid y_{0:l-1})}
\end{aligned}
$$

By marginalizing over $\omega_{0:t}$ and $r_{0:t-1}$, we obtain:

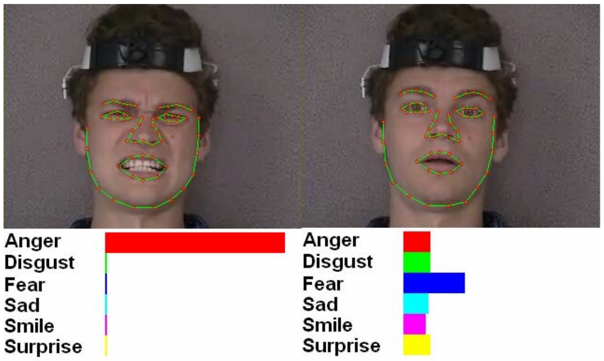

Figure 4: Sample frames of our output tracking and recognition result in a video sequence with more than 10^4 frames.

$$p(r_t = l \mid y_{0:t}) = \int_{\omega_0}\int_{r_0}\cdots\int_{\omega_{t-1}}\int_{r_{t-1}}\int_{\omega_t} p(r_0, \omega_0 \mid y_0)$$

$$\prod_{l=1}^{t} \frac{p(y_l \mid r_l, \omega_l) p(r_l \mid r_{l-1}) p(\omega_l \mid \omega_{l-1})}{p(y_l \mid y_{0:l-1})} d\omega_t dr_{t-1} d\omega_{t-1} \ldots d\omega_0 dr_0$$

This equation can be computed by prior distributions and the product of the likelihood $\prod_{l=1}^{t} p(y_l \mid r_l, \omega_l)$.

5. Experiments

In this section, we present our preliminary experimental results on facial deformation tracking and recognition.

To learn the structure of the expression manifold, we need $O(10^3)$ images to cover basic expressions for each subject and to enable stable geodesic distance computation. Since there is no database with a sufficiently large amount of subject data available, we built our own small data set for the experiments. In our experiments, subjects were instructed to perform a series of six kinds of prototypical facial expressions, representing happiness, sadness, anger, surprise, fear, and disgust. The subjects repeated the series seven times for the gallery set. The probe set includes a long sequence (more than 10^4 frames) where the subject can change his/her expression randomly. To simplify the problem, we assume constant illumination and near frontal view pose.

To generate the shape sequence from the training data set, we trained ten ASM models for different kinds of deformations. We manually select the model in this offline stage to robustly track facial deformation along the video sequences. The shape space dimension is 90. We used the Isomap algorithm to obtain a space with dimensionality d=3.

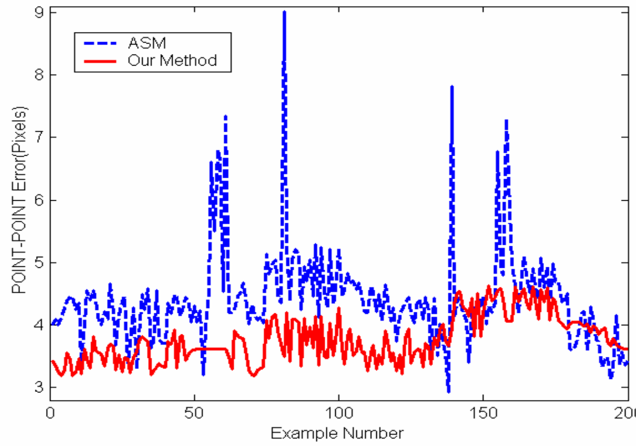

Figure 5: Comparison of tracking precision between an ASM tracker and our method. We have obtained considerably improvement, mainly under the presence of images with large expression changes.

We verified that our probabilistic method is able to track and recognize long sequences of subjects performing subtle and large expression changes. Figure 4 shows two frames from a tracking and recognition test using a video sequence of more than 10^4 frames. The bars for each expression label in the figure indicate their respective recognition probabilities. A complete output video sequence is available at http://ilab.cs.ucsb.edu/demos/.

We also quantitatively analyze the performance of our tracker with a standard ASM tracker. Figure 5 shows a precision comparison, considering as ground truth a manual labeling of eye corners and lip corners. The same images were used to train both trackers. The difference is that our method automatically splits this data to train a set of models, which are probabilistically selected during tracking. This allows more robust performance under large facial expression changes.

6. Conclusions

We proposed a novel framework for dynamic facial expression analysis. We now summarize our main contributions:

(1) A new representation for tracking and recognition of facial expressions, based on manifold embedding and probabilistic modeling in the embedded space.

(2) A robust method for facial deformation tracking based on a set of ASM models, which are probabilistically selected during tracking, improving reliability under large expression changes.

(3) A probabilistic expression classification method, which integrates information temporally across the video sequence. In contrast with traditional

methods that consider expression tracking and recognition in separate stages, we address these tasks in a common probabilistic framework, which enables them to be solved in a cooperative manner.

Our results are preliminary, as our data set is quite small. We plan to perform much more extensive experimentation and provide more substantial quantitative results. For future work we will attempt to extend our framework to include pose and illumination variations.

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

References

[1] C. Padgett and G. Cottrell. "Representing Face Images for Emotional Classification". *Proceedings of Conf. Advances in Neural Information Processing Systems*, pp. 894-900, 1996.

[2] 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 Int. Conf. On Automatic Face and Gesture Recognition*, 1998.

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

[4] I. Cohen, N. Sebe, L. Chen, A. Garg and T. Huang. "Facial Expression Recognition from Video Sequences: Temporal and Static Modeling". *Computer Vision and Image Understanding*, 91(1), pp. 160-187, 2003.

[5] Y. Zhang and Q. Ji. "Facial Expression Understanding in Image Sequences Using Dynamic and Active Visual Information Fusion". *Proceedings of Int. Conf. On Computer Vision*, Nice, France, 2003.

[6] T. Cootes, C. Taylor, D. Cooper and J. Graham. "Active Shape Models – their Training and Applications". *Computer Vision and Image Understanding*, 61(2), 1995.

[7] T. Cootes, G. Edwards and C. Taylor. "Active Appearance Models". *IEEE Transactions on PAMI*, 23(6), pp.681-685, 2001.

[8] J. Tenenbaum, V. de Silva and J. Langford. "A Global Geometric Framework for Nonlinear Dimensionality Reduction". *Science*, vol. 290, pp. 2319-2323, 2000.

[9] Q. Wang, G. Xu and H. Ai. "Learning Object Intrinsic Structure for Robust Visual Tracking". *Proceedings of Int. Conf. on Computer Vision and Pattern Recognition*, June 16-22, 2003.

[10] K. Lee, J. Ho, M. Yang and D. Kriegman. "Video-based Face Recognition Using Probabilistic Appearance Manifolds". *Proceedings of Int. Conf. on Computer Vision and Pattern Recognition*, June 16-22, 2003.

[11] S. Zhou, V. Krueger and R. Chellappa. "Probabilistic Recognition of Human Faces from Video". *Computer Vision and Image Understanding,* 91(1), 2003.

[12] A. Blake and M. Isard. "Active Contours". Springer-Verlag, Cambridge University press, 1998.

[13] M. Isard and A. Blake. "ICondensation: Unifying Low-level and High-level Tracking in a Stochastic Framework". *Proceedings of European Conference on Computer Vision*, 1998.

[14] F. W. Young and R. M. Hamer. *Multidimensional Scaling: History, Theory and Applications*. Erlbaum, New York, 1987.

[15] 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.

[16]Y. Tian, T. Kanade and J. F. Cohn. "Recognizing Action Units for Facial Expression Analysis". *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 23, No. 2, February, 2001.

[17] Y. Zhu, D. Comaniciu, S. Schwartz, V. Ramesh: "Multimodal Data Representations with Parameterized Local Structures". *European Conf. Computer Vision*, Copenhagen, Denmark, Vol. 1, pp. 173-189, 2002

[18] Y. Li, S. Gong and H. Liddell. "Video-based online face recognition using identity surfaces". In *Proc. IEEE International Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-time Systems*, pp. 40-46, Vancouver, Canada, July 2001.

[19] M.-H. Yang. "Face Recognition Using Extended Isomap". *International Conference on Image Processing, 2002.*

[20] S. T. Roweis and L. K. Saul. "Nonlinear Dimensionality Reduction by Locally Linear Embedding". *Science*, vol. 290, pp. 2323--2326, 2000.

[21] M. Brand. "Charting a manifold". *Neural Information Processing Systems 15,* 2002.

[22] S. Roweis, L. Saul and G. Hinton. "Global Coordination of Local Linear Models". *Neural Information Processing Systems 14 (NIPS'2001)* , pp. 889--896, 2002.

[23] Douglas Fidaleo and M. Trivedi, "Manifold analysis of facial gestures for face recognition". *In ACM SIGMM Multimedia Biometrics Methods and Application Workshop*, Nov. 8, 2003.

[24] Y. Chang, C. Hu, M. Turk, "Probabilistic expression analysis on manifolds," International Conference on Computer Vision and Pattern Recognition, Washington DC, June 2004.