Analysis and Synthesis of Facial Expressions Using Decomposable Nonlinear Generative Models

Chan-Su Lee
LED-IT Fusion Research Center and
Department of Electronic Engineering
Yeungnam University, Korea
chansu@ynu.ac.kr

Dimitris Samaras
Computer Science Department
Stony Brook University
Stony Brook, NY 11794, USA
samaras@cs.sunysb.edu

Abstract— **This paper presents a new framework that models facial expressions in multiple people with different expressions and synthesize new *stylized subtle facial expressions* using the generative models. As facial expressions pass through nonlinear shape deformations during facial expressions, we model the facial expression in nonlinear mapping space based on low dimensional embedding and kernel mapping. Characteristics of different type of expressions and variances in different people are decomposed by analyzing the nonlinear mapping between the Euclidean space of facial motion and a low dimensional embedding of these expressions. Using high resolution tracking of densely sampled 3D data, the generative model can control subtle facial expression characteristics of different person in different expression by low dimensional person dependent *style factor* and expression type dependent *expression factor*. The temporal characteristics of the motion can also be controlled by the trajectory sampling on the low dimensional embedding manifold which is independent of person style and expression type. Our experimental results are shown for subtle differences in different smile expressions in different people from dense 3D tracking.**

I. INTRODUCTION

In this paper we present a new framework for analysis and synthesis of facial expressions that exhibit visually subtle but semantically important differences. Facial expression is one of the main inherent non-verbal communication channels for humans. The human face expresses not only basic emotions such as anger, surprise, or happiness, but also subtler thoughts or emotions. If we take the variations of a smile as an example, they communicate not only happiness in smile but many other cognitive/emotional states: a smile of enjoyment, a scornful smile, a pleased smile, a flirtatious smile, a heart-warming smile, a smirk and so on. There are morphological differences in these different forms of smiles [1]. There are also differences in the temporal and geometric characteristics of the facial motion of each individual, even though sometimes the semantic content of a smile is clear to the observer.

In this paper we focus on *expression style*, individual characteristics of expressions, and *expression type*, subtle variations in expressions, and try to learn them from actual 3D data. A person's facial expression style is formed by the differences in facial expression that an individual exhibits, compared to other individuals performing the same expression. The differences are reflected not only in the muscle movement of the face but also in the trajectory and speed of the motion. On the other hand, the type of an expression is the group of expressions that convey the same semantic meaning across different persons. The expression type can be a basic expression category like "smile" but it also can be defined as a subcategory of the same basic expression. In our data-set, we captured smile expressions with subtle differences such as "soft affection", "coy flirtation", and "devious smirk". There are three problems that need to be solved in order to be able to produce animations with such subtle facial expressions.

- Tracking subtle expression details with accurate correspondences.
- Generating new expression sequences using the tracking data.
- Controlling/editing new expressions.

Our goal of analyzing the subtle differences of expression both across individuals and across a very small range of distinct expression such as the types of smile mentioned above, place onerous requirements on the accuracy of the tracking method, which needs to acquire very dense and very accurate intraframe correspondences. Marker based motion capture cannot offer the required resolution. Recent technological advances in digital projection display and computer vision are allowing 3D shape acquisition in real time [26], [16]. However, it is still hard to achieve correspondences between data of the same face in different frames, as well as between different faces. In this paper we rely on a fully automatic method for tracking subtle details of facial expression using harmonic maps presented in [24].

Facial expression undergoes complicated global and local nonlinear deformation between frames. A facial expression sequence lies on a low-dimensional manifold, although it is represented by a high dimensional vector e.g. a collection of 3D nodal points [17]. Recently nonlinear dimensionality reduction algorithms like local linear embedding (LLE) [15] and Isometric feature mapping (Isomap) [20] were applied to find manifold from face datasets [25]. However, these data-driven approaches fail to find manifold representations when there are large variations in the expression data by different type of expression in addition to different people style. In this paper we present a new method for the manifold embedding in different people facial expressions using conceptual manifold embedding.

The method in [25] could factorize expression style or expression type but not both, which lead us to another contribution of this paper, the use of multilinear analysis in the mapping space to extract multiple factors and the decomposition of the mapping space into conceptually orthogonal sub spaces: *expression style space* and *expression type space*. Thus we can factorize both expression style and expression type, especially subtlety in addition to the contents or common temporal characteristics. The proposed generative model which is compact in representation using a low dimensional manifold and can be decomposed conceptually into orthogonal subspaces, gives flexibility in control in addition to capturing the nonlinearity in the mapping space. Analysis based on the proposed model provides us with the ability to distinguish visually subtle but semantically different sub-categories of the same expression, such as a flirtatious from a fake smile. Our system can control of facial expressions with conceptually decomposed facial expression type and personal style by weighting of the type of expressions and personal styles.

A. Related Work

Facial animation has long been an active research area in computer graphics [12]. The two main approaches are either physics/anatomy-based or data-driven. Physics-based models are used to simulate the surface deformations caused by muscle forces [10]. Past efforts to simulate facial muscle actions [8] did not always produce convincing facial expressions. It is hard to model subtlety and personal characteristics in facial expressions in this manner.

Data-driven approaches have attracted increased interest recently because of advances in capture systems. Recently, both static 3-D scans of expressions [3], [2] and time-sequences of 3-D motion [6], [7] have been used to collect 3-D facial expressions. Expression cloning [11] can produce facial animations by reusing existing motion data. Most current methods for capturing 3-D motion data either require the use of 100-200 markers (e.g. [6]) which then need to be manually removed from the images, or model fitting to multiple photographs. Using such methods the recovered geometry of the expressions is rather crude. Template fitting in space-time stereo [26], and hierarchical registration of multiresolution generic model [25] capture dense spatial and temporal facial expression. However, the captured model is limited by the templates or the generic model in capturing subtlety in the facial expression.

Synthesis of facial expressions is preformed by modeling the facial expressions and controlling parameters. Researchers have used linear models (PCA [3]) and variations such as bilinear models [19] and multilinear tensor models for facial expression analysis and synthesis [5], [4], [23], [27]. However, a major limitation of such models is modeling the facial expression in linear subspace. In addition, most of them did not model temporal characteristics of facial expressions. In this paper, we analyze expression in the nonlinear mapping space with preserving temporal characteristics.

In Section II we propose a low dimensional embedding of facial expression and decomposition of facial expression mapping space to learn a generative model for analysis of facial expression. Section III explains how to synthesize new facial expressions and the expression synthesis results are presented in Section IV and conclusions in Section V.

II. Nonlinear Generative Model: Multilinear Analysis of Expression Mapping Space

Facial motion can be described by the collection of the displacements of 3D facial nodal points attached on the skin of the tracked face. As a result of high resolution tracking with one-to-one intra-frame correspondence, we can represent facial expression by the motion vectors for the vertices of the generic face mesh. Let $v_t \in R^{3N\times 1}$ be locations of 3D points at time instance t representing N facial nodal points in a 3-dimensional space, where N is the number of nodal points in a dense generic facial model. The trajectory of the 3D nodal points is the combination of rigid head motion and facial motion, which can be described as

$$v_t = T_{\alpha_t} y_t = T_{\alpha_t}(y_0 + m_t) = T_{\alpha_t}(g + m_t)\,, \qquad (1)$$

where $T\alpha_t$ is the head motion at time t, y_t is the $3\times N$ face nodal point location at time t in face coordinates and $y_0 = g$ is the facial geometry at the initial frames. We assume that the captured facial expression starts from neutral face. The global rigid transformation parameter α comes from the tracking results.

Tracking data are collected from multiple people with different expressions type for learning stylized facial expression for multiple types of subtle facial expression. The displacement in local coordinate of every facial nodal point from the tracking data of expression style s with expression type e can be described as

$$v_t^{se} = T_{\alpha^{se}}(g^s + m_t^{se})\,, \qquad (2)$$

where α^{se} is the global transformation by head motion which depends on expression style and type, g^s is the facial geometry of each person, and m_t is the facial expression motion.

The main problem in facial expression animation is how to model and control the facial expression motion, m_t^{se}. The facial expression motion under goes nonlinear deformations and it is of high dimension and it depends both of on a person style and on the expression type. We derive a low dimensional representation for facial motion using a conceptual embedding of facial expression in Sec. II-A, nonlinear mappings between manifold representation and facial motion and their decomposition are learned in Sec. II-B. and decomposition of conceptually orthogonal space are modeled in Sec. II-C Decomposition of other factor like geometry and head motion are explained in Sec. II-D.

A. Facial expression manifolds and conceptual embedding

A good embedding space for facial motion synthesis should achieve two properties. One is to preserve distance in the embedding space and the original facial motion space.

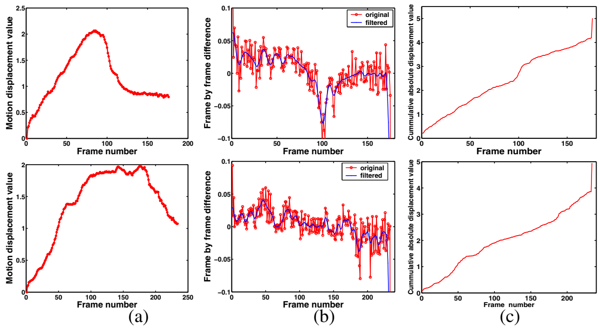

Fig. 1. Temporal characteristics in facial expressions: Col.(a): Facial displacement characteristics from initial face. Col.(b): Frame by frame facial displacement. Col.(c): Cumulative absolute displacement

The distance of two embedding points has to be proportional to the dissimilarity of facial motion in two corresponding frames. This property helps synthesis or generation of new facial expressions using an interpolation function and it helps control of new motion intuitively in the embedding space where distance in this space is similarity measurement in the original space. The other is to provide animator with parameters or an interface to easily specify facial motion properties. It is good to have low dimensional representation as it provides compact parameterization of the motion. In addition, the displacement characteristics or velocity in the motion are important in facial motion synthesis with characteristics of individual facial expressions. Fig. 1 shows facial expression displacement characteristics in two different subjects with the same type of facial expressions. The displacement trajectories also contain individual characteristics of facial motion. We pursue low dimensional manifold representation for the facial motion.

We investigate low dimensional manifolds to find good representation for facial expression synthesis using dimensionality reduction techniques. Nonlinear dimensionality reduction algorithm, LLE [14] and Isomap [18], are applied in two kinds of facial expression data sequences; one is 3D tracking data from range data, the other is 2D texture image sequence. We extract only normalized facial motion, from which global transformation has been removed and normalized based on the mean of all the facial points in each frames. Fig. 2 shows sample example data of 3D tracking and 2D image sequences after alignment. Fig. 3 shows the manifold found by LLE and Isomap from these data sets. They are elliptical one dimensional manifolds in 3-dimensional space. The manifolds are elliptical curves with distortions according to the face geometry and expression types. Sometimes the manifold does not show smooth curves due to noise in the tracking data and images. In the embedding space, distances in local area are related to dissimilarity in actual data. However, it is hard to get proper temporal sampling for synthesis. In addition, the manifold are far from each other for different persons and it is hard to find a uniform representation when we analyze sequences of facial expression data from different persons.

We propose conceptual manifold embedding for facial expressions as a uniform representation of facial expression

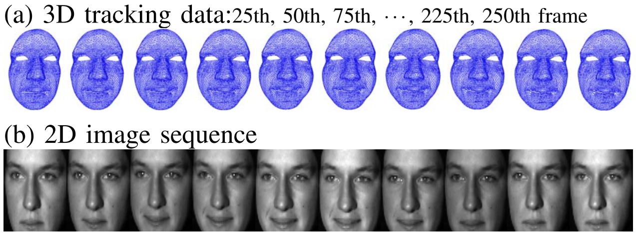

Fig. 2. Examples of 3D tracking data and 2D image sequences

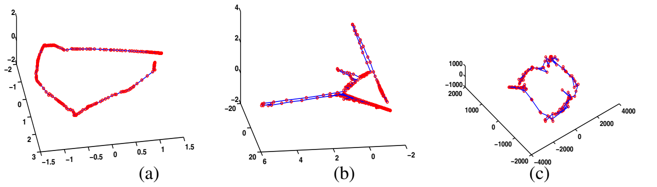

Fig. 3. Facial expression manifolds: (a): LLE. (b):Isomap for 3D tracking data from range data. (c):Isomap for 2D image sequence.

manifolds. Conceptually, each expression sequence forms an one-dimensional cyclic embedding as the expression starts from a neutral face and comes back to the neutral face. Low dimensional manifolds driven by captured data vary in different people and in different expression types. Sometimes the embedding manifolds are very different and it is hard to find unified representations for multiple expression styles and expression types. Conceptually, however, all data driven manifolds are equal. They are all topologically equivalent, i.e., homeomorphic to each other, and to a circular curve. We use the unit circle as a conceptual embedding space for facial expression.

We embed each captured facial motion frame on a unit circle proportional to the displacement in each adjacent frame. Thus preserving distances in adjacent frames. As we embed frames in the unit circle by pair-wise displacement, the cumulative absolute displacement values are proportional to the coordinate value in the manifold parameterization. Distance in our embedding space is close to the dissimilarity of two motion frames. Fig. 1 (b) shows frame by frame variation of displacement. Considering constant capturing time, displacement is proportional to velocity and frame by frame variation to acceleration. Fig. 1 (c) shows accumulation of absolute displacement. When we describe the embedding coordinate by $u_t \in [0,1]$, we find embedding coordinate in unit circle as $x_t = (\cos(2\pi u_t), \sin(2\pi u_t))$. Let be $c_{t_i} = |m_{t_i} - m_{t_{i-1}}|$ be the adjacent frame displacement and q_t be a cumulative displacement function, then

$$u_{t_i} = \frac{\sum_{j=1}^{t_i} c_{t_i}}{\sum_{k=1}^{k=N} c_{t_k}} = \frac{q_{t_i}}{Q}, \tag{3}$$

where Q is the total displacement value in the sequence. We learn the interpolation function $q(t)$ which passes from q_{t_i} at normalized time instance $t_i = \frac{i \times T}{M \times T} = \frac{i}{M}$, where T is the cycle period and M is total number of sample at equal intervals sequence. Now we can find manifold embedding parameters u_t for any instance of time $t \in [0,1]$ by $u_t = q(t)$. The animator can select equal interval points $[0, 1/M, 2/M, \cdots, (M-1)/M, 1]$ and generate a sequence with original displacement characteristics using the embedding parameter $[0, q(1/M), q(2/M), \cdots, q((M-1)/M), 1]$. In addition, by modifying function $q(t)$, the animator can generate different facial expressions with different displacement characteristics. Our embedding manifold not only keeps pairwise distance in adjacent frame but also gives a mechanism to control facial expression displacement characteristics easily. In addition, our embedding space is very smooth as all the manifold points are on the unit circle.

B. Decomposable Nonlinear Generative Model

In this section, we show how to learn nonlinear mapping between embedding space and original facial motion. Given a set of distinctive facial motion sequence $M^{se} = [m_1^{se} m_2^{se} \cdots m_{N_{se}}^{se}]^T$ and its embedding $X^{se} = [x_1^{se} x_2^{se} \cdots x_{N_se}^{se}]^T$, we can learn nonlinear mapping function $f^{se}(x)$ that satisfies $f^{se}(x_i) = m_i^{se}$, $i = 1 \cdots N_{se}$, where N_{se} is the number of captured motion frame for style s and expression e. In particular we consider nonlinear mapping functions $f(x)$ of the form

$$f^{se}(x) = B^{se} \cdot \psi(x) \tag{4}$$

where B is a $d \times N$ coefficient matrix and $\psi(\cdot) : R^2 \to R^N$ is a nonlinear mapping where N kernel functions can be used to model the manifold in the embedding space, i.e.

$$\psi(\cdot) = [\psi_1(\cdot), \cdots, \psi_N(\cdot)]^T \tag{5}$$

Generalized radial basis function (GRBF) interpolation [13] is used in the mapping where each row in the matrix B represents the interpolation coefficients for corresponding element in the input. i.e., we have d simultaneous interpolation functions each from $2D$ to $1D$. The mapping coefficients can be obtained by solving the linear system

$$[m_1^{se} \cdots m_{N_{se}}^{se}] = B^{se}[\psi(x_1^{se}) \cdots \psi(x_{N_{se}}^{se})] \tag{6}$$

Given the facial motion sequence with N_s expression styles and N_e expression types, we obtain $N_s \times N_e$ mapping coefficients. The mapping function contains all the information to generate new interpolated motion for given x_t. For a given kernel $\psi(x)$, the matrix B^{se} captures the facial motion characteristics for expression style s and expression type e. The vectorized mapping coefficients b^{se} are arranged into high order tensor.

C. Multiple factor decomposition in the mapping space

Multilinear tensor analysis make it possible to decompose the facial motion mapping space into orthogonal factors. Multilinear tensor analysis is an extension of principal component analysis (PCA) (one orthogonal factor), and bilinear models (two orthogonal factors) [21]. Tensor decomposition can be achieved by higher-order singular value decomposition (HOSVD), which is a generalization of SVD [9][22]. All the coefficient vectors can be arranged in an order-three facial motion coefficient tensor $\mathscr{B}$ with dimensionality $N_s \times N_e \times N_c$. The coefficient tensor can be decomposed as

$$\mathscr{B} = \mathscr{A} \times_1 S \times_2 E \times_3 F \tag{7}$$

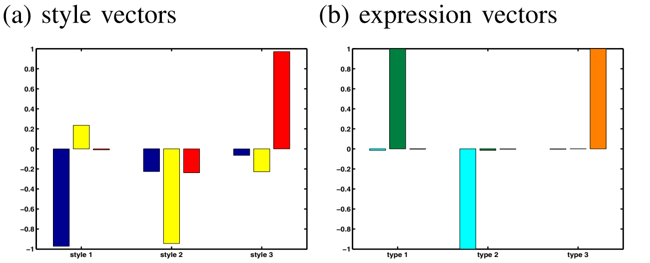

Fig. 4. Examples of expression style vectors and expression type vectors

where S is the collection of the orthogonal basis for expression style subspace. E represents the orthogonal basis of the expression type space and F represents the basis of the mapping coefficient space. $\mathscr{A}$ is a core tensor which governs interactions among different mode basis. For a given b^{ij},the mapping coefficient for expression style i and expression type j, we compute an expression style vector s^i and expression type vector e^j. For any given manifold point x_t we can reconstruct a facial motion as

$$b^{ij} = \mathscr{A} \times_1 s^i \times_2 e^j \times_3 F \tag{8}$$

$$m_t^{ij} = B^{ij} \psi(x_t) \, , \tag{9}$$

where B^{ij} is the matrix representation of the vectorized b^{ij}. Therefore for any given a new style vector s^p, and new expression type vector e^q, we can generate new facial motion for any x_t on the manifold. Fig. 4 shows example representation of expression style vectors and expression type vectors when we analyzed three smiles ($N_e = 3$), “soft affection”, “coy flirtation”, and “devious smirk” for three people ($N_s = 3$).

D. Decomposition of geometry

Head motion in animation increase naturalness of the expression. We can catch the meaning of the expression without head motion. However, the animation without head motion looks nervous or fake expressions. We acquire head motion trajectory T_α as a result of high resolution tracking of facial expression. Different person has different characteristics in head motion. In addition, in the same person, different expression has variance in head motion trajectory.

We decompose head motion parameters by expression style and expression type. We convert rotation matrix into quaternion representation for better numerical interpolation. Therefore the transformation described by

$$\alpha^{se} = [l_x^{se} \, l_y^{se} \, l_z^{se} \, \omega_1^{se} \, \omega_2^{se} \, \omega_3^{se} \, \omega_4^{se}]^T \, , \tag{10}$$

where l_x, l_y and l_z is global translation. Based on conceptual embedding and its mapping function, we decompose head motion into 'style', 'expression', and a trajectory basis and generate new trajectory similar to facial motion.

$$T_{\alpha_t} = \mathscr{H} \times \alpha^s(t) \times \alpha^e(t) \tag{11}$$

We can control facial geometry separately during synthesis since we formulated the whole expressions based on static geometry and dynamic motions by in Eq. 2. The geometry style, which is 3D geometry in neutral face, only depends on individual person. We collected each person style as column g

$$C = [g^1 g^2 \cdots g^{N_g}], \tag{12}$$

where N_g is the number of geometries collected in neutral expression. By applying an asymmetric bilinear model [21], we can represent geometry based on geometry basis and geometry style vector as follows:

$$C^{gs} = D^g s_g^s, \tag{13}$$

where D^g is geometry basis and s_g is geometry style.

The overall generative model supports synthesis of stylized facial expression with characteristic expression type. Now, the generative model Eq. 2 can be expressed as

$$v^{se}(t) = T_{\alpha_t}\left(D^g s_g^s + \mathscr{A} \times s \times e \times F \times \psi(x(t))\right), \tag{14}$$

The generative model captures nonlinearity in the mapping space and provides compact control parameters for style and expression type variations during facial expression synthesis. In addition, it supports generation of stylized geometry and head motion.

III. Facial expression synthesis

There are several steps we need to follow to synthesize a new facial expression using our generative model. First, we select style and expression parameters as in Sec. III-A. Then we generate geometry to add facial motion as in Sec. III-B. We compute the mapping coefficients corresponding to the estimated expression style and expression type and we generate facial motion sequences by tracing manifold points to generate sequential motions. The head motion is superimposed to make realistic expression.

A. style and expression parameterizations

We can generate new style parameters by linearly weighting the existing style parameters that comes from the learning the generative model in Sec. II-B

$$s^{new} = w_{s_1} s_1 + w_{s_2} s_2 + \cdots + w_{s_{N_s}} s_{N_s}, \tag{15}$$

where $\sum_1^{N_s} w_i = 1$. For example, if we want to generate new expression as style with 60% of the first style and 25% of the second style and 15% of the third style, then we can generate new style as $s^{new} = 0.6 s_1 + 0.25 s_2 + 0.15 s_3$. In the same way, we can compute new expression parameters by

$$e^{new} = w_{e_1} e_1 + w_{e_2} e_2 + \cdots + w_{e_{N_e}} e_{N_e} \tag{16}$$

These parameterizations are close to the conceptual categories and can be achieved intuitively by adjusting weighting values.

Another way is to estimate parameters from data. If another person's whole expression tracking data are available, we can compute the mapping coefficient in the same way as for training. Based on the coefficient vector b^{new} we can estimate style and expression parameters iteratively by minimizing the error

$$E(s'_{new}, e'_{new}) = ||b^{new} - \mathscr{A} \times_1 s'_{new} \times_2 e'_{new}||, \tag{17}$$

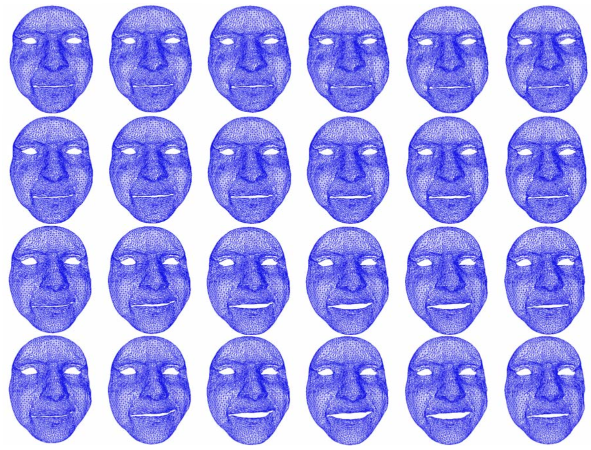

Fig. 5. Synthesis of new facial expression by weighting two different expression type: First row: soft affection (SA), Fourth row: Devious smirk (DS), Second row: 75% SA + 25% DS, Third row: 25% SA + 75% DS. Columns: 20th, 40th, 60th, 100th, 120th frames among 150 synthesized sequences.

where s'_{new}, and e'_{new} are the estimated expression style and expression type, represented by the combination of the weights from the estimated expression style and type basis. By editing the estimated style and expression parameter, we can generate new facial expression based on the new person style and expression type.

B. Geometry synthesis

A new geometry style can be parameterized similar to style and expressions as described in Sec. III-A by the conceptual weighting or estimation from data. Our tracking algorithm establishes correspondence between people, we can get proper interpolation between subjects by weighting of the geometry style parameters. In addition, the weighting parameter can be modified as a function of time. For example, we can change geometry from a geometry that has style s_g^1 to geometry that has style s_g^2 during an expression synthesis by linear interpolation

$$s_g^{new}(u_t) = \alpha(u_t) s_g^1 + (1 - \alpha(u_t)) s_g^2, \tag{18}$$

where u_t changes from 0 to 1 during the duration of expression. As a result, the static geometry is modeled in terms of a style basis dynamically based on the function of time-varying weighting parameters.

IV. Results

In order to demonstrate the power of our framework we have focused on the analysis and synthesis of different types of smiles that routinely occur among humans. For improved results we used actors to perform three different types of smile: a) soft affection (SA), b) coy flirtation (CF), and c) devious smirk(DS). Two male actors and two female actor participated in performing the facial expressions that were then captured using our range scanner. In order to achieve relatively consistent facial expressions across actors for each smile type we gave the same scenario to each actor.

We then tracked and captured range data and we analyzed them using our framework. In particular, we learned a nonlinear generative model for three people with three different

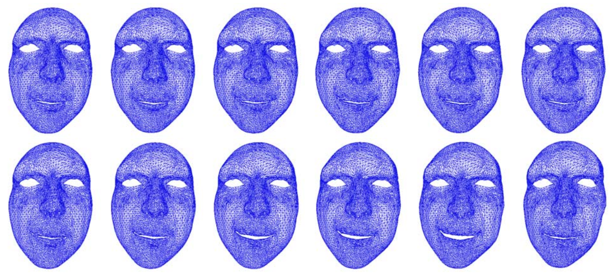

Fig. 6. Subtlety transformations: First row: Transferring from soft affection (SA) to coy flirtation (CA) Second row: Transferring from CA to devious smirk (DS)

types of smile expressions based on our tracking results. Following our approach, the decomposed models for the head motion and geometry are learned separately. Data from one female actor were reserved to evaluate facial expression synthesis performance for new actors.

Fig. 5 shows the generation of two subtle smile expressions SA and DS and the synthesis of a new in-between smile by changing the weight of two different types of smile. Notice that we can generate a convincing combination of two different types of smile without loss of the related subtle expressions with our method. Our generative model can generate in-between expressions as shown in the second and third rows. Thus, we can transfer style and expression details to new subjects.

Fig. 6 shows an example of smile expression transfer. By changing the smile type parameter according to the embedding state we can achieve a smooth transition between subtle smile expressions within a cycle.

Fig. 7 shows the transferring of expressions from one person to another. The generative model generates the necessary in-between smiles styles using linear interpolation of two styles. The smile style variations we have created produce convincing differences in smile styles even when the same facial geometry is used.

V. Conclusion

In this paper we have presented a new framework that can synthesize new subtle facial expressions by capturing and tracking subtle facial expressions from range data. Our approach is based on the use of a decomposable generative model and the learning of a relevant nonlinear mapping space. Using high resolution facial expression tracking, we achieved tracking of subtle facial expressions. Using a nonlinear decomposable generative model we have achieved the synthesis of very subtle facial smile expressions, which is not possible using a linear subspace. Finally, the conceptual decomposition of the mapping space and the low dimensional embedding helps the animator control intuitive parameters for synthesizing new facial expressions.

Fig. 7. Transferring facial expressions with variation of expression style: First column: Subject M soft affection, Fifth column: Subject D soft affection, Second column: 75% M style + 25% D style, Third column: 50% M style + 50% D style, Fourth column: 25% M style + 75% D style soft affection.

References

[1] Millicent H. Abel, editor. *An Empirical Reflection on the Smile*. The Edwin Mellen Press, 2002.

[2] V. Blanz, C. Basso, T. Poggio, and T. Vetter. Reanimating faces in images and video. In *Eurographics'03*, pages 641–650.

[3] V. Blanz and T. Vetter. A morphable model for the synthesis of 3d faces. In *SIGGRAPH'99*, pages 187–194.

[4] E. S. Chunang, H. Deshpande, and C. Bergler. Facial expression space learning. In *Pacific Graphics*, pages 68–76, 2002.

[5] T. Ezzat, G. Geiger, and T. Poggio. Trainable videorealistic speech animation. In *SIGGRAPH'02*, pages 388–398.

[6] B. Guenter, C. Grimm, D. Wood, H. Malvar, and F. Pighin. Making faces. In *SIGGRAPH'98*, pages 55–66, 1998.

[7] G. A. Kalberer and L. Van Gool. Realistic face animation for speech. *Intl. Journal of Visualization and Computer Animation*, 2002.

[8] P. Kalra, A. Mangili, N. Magnenat-Thalmann, and D. Thalmann. Simulation of facial muscle actions based on rational free form deformations. In *Eurographics'92*, pages 59–69.

[9] Lieven De Lathauwer, Bart de Moor, and Joos Vandewalle. A multilinear singular value decomposiiton. *SIAM Journal On Matrix Analysis and Applications*, 21(4):1253–1278, 2000.

[10] Y. Lee, D. Terzopoulos, and K. Walters. Realistic modeling for facial animation. In *SIGGRAPH'95*, pages 55–62.

[11] J.-Y. Noh and U. Neumann. Expression cloning. In *SIGGRAPH'01*, pages 277–288.

[12] F. I. Parke and K. Waters. *Computer facial animation*. A K Peters, 1996.

[13] Tomaso Poggio and Fredrico Girosi. Networks for approximation and learning. *Proceedings of the IEEE*, 78(9):1481–1497, 1990.

[14] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

[15] Sam Roweis and Lawrence Saul. Nonlinear dimensionality reduction by locally linar embedding. *Science*, 290(5500):2323–2326, 2000.

[16] S. Rusinkiewicz, O. Hall-Holt, and M. Levoy. Real-time 3d model acquisition. In *SIGGRAPH*, pages 438 – 446, 2002.

[17] H Sebastian Seung and Daniel D. Lee. The manifold ways of perception. *Science*, 290(5500):2268–2269, 2000.

[18] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

[19] J. B. Tenenbaum and W. T. Freeman. Separating style and content with biliear models. *Neural Computation*, 12:1247–1283, 2000.

[20] J.B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319 – 2323, 2000.

[21] Joshua B. Tenenbaum and William T. Freeman. Separating style and content with biliear models. *Neural Computation*, 12:1247–1283, 2000.

[22] M. Alex O. Vasilescu and Demetri Terzopoulos. Multilinear subspace analysis of image ensembles. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, 2003.

[23] Daniel Vlasic, Matthew Brand, Hanspeter Pfister, and Jovan Popović. Face transfer with multilinear models. *ACM Trans. Graph.*, 24:426–433, 2005.

[24] Yang Wang, Mohit Gupta, Song Zhang, Sen Wang, Xianfeng Gu, Dimitris Samaras, and Peisen Huang. High resolution tracking of non-rigid motion of densely sampled 3d data using harmonic maps.

[25] Yang Wang, Xiaolei Huang, Chan-Su Lee, Song Zhangand Zhiguo Li, Dimitris Samaras, Dimitris Metaxas, Ahmed Elgammal, and Peisen Huang. High resolution acquisition, learning and transfer of dynamic 3d facial expressions. *Computer Graphics Forum*, 23(3):677–686, 2004.

[26] Li Zhang, Noah Snavely, Brian Curless, and Steven M. Seitz. Spacetime faces: high resolution capture for modeling and animation. *ACM Trans. Graph.*, 23(3):548–558, 2004.

[27] Qingshan Zhang, Zicheng Liu, Baining Guo, Demetri Terzopoulos, and Heung-Yeung Shum. Geometry-driven photorealistic facial expression synthesis. *IEEE Transactions on Visualization and Computer Graphics*, 12:48–60, 2006.