

# Personalized 3-D Facial Expression Synthesis based on Landmark Constraint

Haoran Liang\*, Mingli Song<sup>†</sup>, Lei Xie<sup>‡</sup> and Ronghua Liang\*

\*College of Computer Science, Zhejiang University of Technology, Hangzhou, China

E-mail: hrliang1988, ronghualiang2010@gmail.com

<sup>†</sup>College of Computer Science, Zhejiang University, Hangzhou, China

E-mail: brooksong@zju.edu.cn

<sup>‡</sup>School of Computer, Northwestern Polytechnical University, Xi'an, China

E-mail: xielei21st@gmail.com

**Abstract**—With the development of computer technology, 3-D facial expression synthesis has been an important and challenging task in the field of computer animation. Since the faces generated by previous works lack of personalization, we propose a novel approach for 3-D facial expression synthesis based on non-linear learning. Firstly, a pre-process alignment is performed for input 2-D or 3-D faces with landmarks based on cylindrical mapping, and the intrinsic representations of faces are generated using radial basis function network. Secondly, according to the low dimensional representations of input faces, reconstruction operations are carried out to synthesize 3-D face expressions by sharing linear combination coefficients. Finally, the output 3-D face expressions are further optimized by its corresponding landmarks both in 2-D and 3-D spaces using locality-constrained linear coding. The experimental results indicate the robustness and effectiveness of our facial expression synthesis approach.

## I. INTRODUCTION

THREE-DIMENSIONAL (3-D) facial-expression synthesis has been extensively studied over the past two decades for its wide range of applications in movie production, video conference and digital entertainment. Recent researches have shown its potential value in human computer intelligent interactions. Although we can generate 3-D face model with exact expressions from 3-D scanners, the expensive cost is not applicable when facing large numbers of subjects in real time facial expression sequences. We still need efficient methods to transfer captured expressions to models.

The existing approaches on 3-D facial expression synthesis mainly fall into three groups, namely parameter-driven, example-based and learning-based.

As a common technique in computer graphics, parameter-driven synthesis is first used for forming a parametric face in 1970s by Parke [9]. In this approach, different personality traits from individual faces are defined by specific parameters, which contain a set of control parameters for generating facial expression by controlling basic movements of faces. However, it usually uses a low resolution face model with sparse vertex distribution which leads to the failure of mimicking subtle expression details. Afterwards, facial action coding system (FACS) [8] has been raised for generating 3-D facial expressions by using parameters to define generic facial deformations. Although extensions such as facial definition

parameter (FDP) [7] and facial animation parameter (FAP) [6] are developed, they all fail to give expression details.

Example-based synthesis, also known as expression cloning or expression retargeting, is presented to overcome the limitations of parameter-driven methods. Jun-Yong Noh and Ulrich Neumann [10] applied motion vectors on target faces to clone facial expressions. Song et. al. [11] introduced vertex-tent coordinate for modeling local deformations in source faces. In their approach, a consistency constraint is built for transferring local deformations to the target face. However, example-based synthesis is not suitable for real time applications owing to its computational complexity and sensitivity to noise, which may produce flaws and distortions on target face.

Learning-based synthesis is developed by analyzing the information from a set of training 3-D faces to generate the relationships between input data and target faces. Vlasic et. al. [12] presented multi-linear algebra for face modeling. They applied different coefficients obtained by tensor decomposition to synthesize facial expressions. Later, Tao et. al. [19] introduced Bayesian tensor analysis for 3-D facial expression modeling. They constructed the multi-linear model from a probabilistic point of view. Though both of them are with good performances, they cannot handle faces with incomplete data.

In comparison with the complexity of 3-D face model generation, it is quite easier to capture a variety of expressions in 2-D facial images. There already exist a number of efficient methods to reconstruct a 3-D face model from a single 2-D face image. Among those representative algorithms mentioned above, the learning-based methods have drawn much attention in recent researches for 3-D face model reconstruction. These methods recover the 3-D shapes by using the common information shared by the 3-D shape and 2-D image subspace. Thus a coupled training set containing pairs of 2-D faces and corresponding 3-D faces is used. Reiter et. al. [13] presented canonical correlation analysis (CCA) to predict 3-D faces from 2-D faces using a statistical approach. They assumed that both 2-D and 3-D faces are embedded in the corresponding linear subspaces and tried to maximize the correlation. Wang and Yang [14] employed nonlinear manifold embedding and alignment (NMEA) to recover a 3-D shape. They utilized

a nonlinear dimensionality reduction technique to learn the local image models for each patch of facial image and the local surface models for each patch of 3-D shape. Song et. al. [15] introduced a coupled radial basis function network (C-RBF) for 3-D face shape reconstruction. In this approach, the intrinsic representations of 2-D and 3-D faces are generated to build 3-D shapes using linear combination coefficients shared in low dimensional subspace. These methods mainly focus on optimizing global mappings, which leads to the loss of details in results. For facial expression synthesis, it is important to transfer all the appearance details of facial expression from one model to another.

In order to generate personalized 3-D facial expressions with important details, in this paper, we present a novel approach for expression synthesis based on landmark constraint. Song et. al. [15] have proved that the RBF [1] network is effective for reconstructing expressive faces for input single images. Taking account of the fact that RBF cannot handle local details on faces, we present two improvements as our key contributions in this paper:

- 1) We construct the low dimensional representations for 2-D and 3-D faces and share the reconstructional coefficients for both faces and landmarks in order to make local constraints.
- 2) A local constrained deformation is conducted for reconstructed 3-D faces. As a matter of fact, the landmarks form 2-D faces we used for 3-D facial deformation lack z coordinate which is essential in 3-D face modeling. Thus we build a coupled dictionary for coordinate system and generate the z coordinate using locality-constrained linear coding.

The remainder of the paper is organized as follows: Section II introduces the problem statement and an overview of our approach. Section III describes the detail of the approach including RBF network, landmark generation, coordinate optimization, and local deformation. Section IV shows our experimental results and comparison. Finally conclusions are made in Section V.

## II. PROBLEM STATEMENT AND OVERVIEW

A practical approach for 3-D facial expression synthesis should have one key factor: given coefficients, it can accurately generate expressions for an arbitrary neutral face. In our approach, inspired by [15], a C-RBF network is used to reconstruct 3-D faces with expressions. Figure 1 shows the framework of the C-RBF network. The process of landmark generation and coordinate optimization is shown in Figure 2. Our approach is divided into three steps.

- 1) We prepare a set of coupled training data that contains several pairs of 2-D faces and its corresponding 3-D faces with expressions. Let a 2-D face or 3-D face be the input. A pre-process alignment based on cylindrical mapping is performed for the 3-D faces in database. To construct the mapping between input face and 3-D expressional face, we assume that they are

identically distributed in intrinsic representations which can be treated as the common geometric structure [2]. Therefore, we build RBF networks to obtain intrinsic representations of input faces and expressional faces by mapping functions. Then temporary expressional 3-D faces are reconstructed by the shared linear combination coefficients.

- 2) The 3-D expressional faces in training data are combined with their corresponding 3-D landmarks to build a coupled dictionary in order to generate the landmarks of synthesized temporary faces using locality-constrained linear coding. For further optimization, we conduct a same process on 2-D faces and 2-D landmarks as well as the vertices in 3-D faces' coordinate system to build another two coupled dictionaries.
- 3) A local constrained deformation for 3-D expressional faces is performed on the basis of landmarks from 2-D and 3-D spaces. By relocating the locations of vertices in synthesized temporary faces, natural and smooth 3-D expressional faces are finally obtained.

## III. PERSONALIZED 3-D FACIAL EXPRESSION SYNTHESIS BASED ON LANDMARK CONSTRAINT

### A. Radial Basis Function Network

RBF network is effective for handling sparse, high-dimensional, and noisy data. It has been widely used for computer vision applications [3][4]. In general, an RBF network is an artificial neural network with three layers: an input layer, a hidden layer and an output layer. RBFs are used as activation functions. The output of a RBF network can be written as:

$$\varphi(x) = \sum_{i=1}^N a_i \rho(\|x - c_i\|) \quad (1)$$

where  $x$  is the input,  $N$  is the number of hidden units,  $a_i \in [a_1, \dots, a_n]$  encodes the linear combination coefficients,  $c_i$  is the center vector for  $i$ th neuron,  $\rho(\cdot)$  is the RBF. We use Gaussian  $\rho(\|x - c_i\|) = \exp(-\beta \cdot \|x - c_i\|^2)$  in our approach. Here (1) can be normalized in the region of input space:

$$\varphi(x) = \frac{\sum_{i=1}^N a_i \rho(\|x - c_i\|)}{\sum_{i=1}^N \rho(\|x - c_i\|)} \quad (2)$$

As introduced by Song [15], a coupled RBF network is built for the training data set that contains pairs of 2-D faces  $X^{2D} = [x_1^{2D}, x_2^{2D}, \dots, x_n^{2D}]$  and 3-D faces  $X^{3D} = [x_1^{3D}, x_2^{3D}, \dots, x_n^{3D}]$ , where  $x_i^{2D} \in R^L$  and  $x_i^{3D} \in R^H$  are the  $i$ th individual's 2-D face image and the corresponding 3-D face, respectively. C-RBF network outputs the intrinsic representations of input 2-D and 3-D faces. Let  $Y^{2D} = [y_1^{2D}, y_2^{2D}, \dots, y_n^{2D}]$  and  $Y^{3D} = [y_1^{3D}, y_2^{3D}, \dots, y_n^{3D}]$  be the corresponding intrinsic representations of  $X^{2D}$  and  $X^{3D}$ , where  $y_i^{2D} \in R^l$  ( $l \ll L$ ) and  $y_i^{3D} \in R^h$  ( $h \ll H$ ) represent the  $i$ th face's 2-D and 3-D intrinsic representations, respectively.

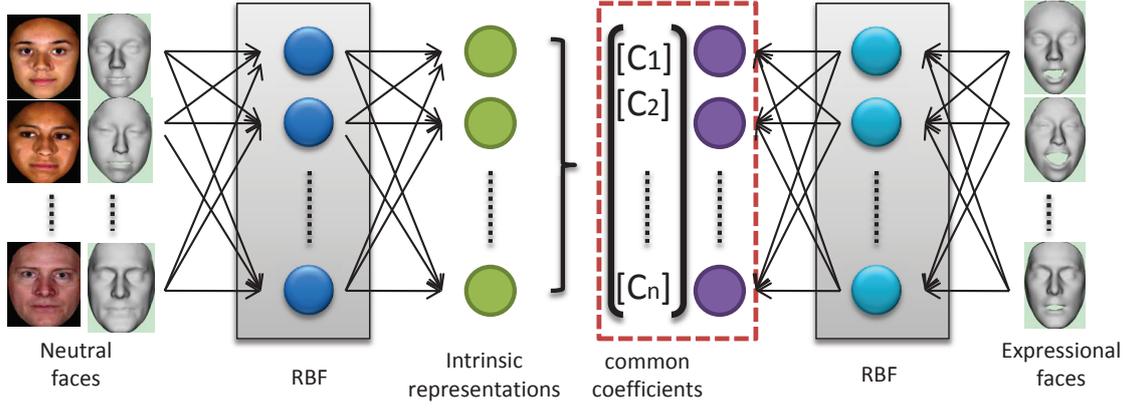


Fig. 1. Framework of the C-RBF network. The network discovers the intrinsic representations of input faces and expressional faces by mapping functions. The coefficients are combined with the intrinsic representations on the right to synthesize temporary expressional faces.

We define the mapping function between a 3-D face and its intrinsic representations as  $F^{3D}$  and  $f^{3D}$ , respectively, as well as  $F^{2D}$  and  $f^{2D}$  for 2-D face. By using  $X^{2D}$ ,  $X^{3D}$ ,  $Y^{2D}$ ,  $Y^{3D}$  as inputs in (2), we can obtain the following normalized RBFs:

$$\begin{aligned}
 F^{2D}(x^{2D}) &= \frac{\sum_i^n b_i^{2D} \rho(x^{2D} - c_{x_i}^{2D})}{\sum_j^n \rho(x^{2D} - c_{x_j}^{2D})} \\
 f^{2D}(y^{2D}) &= \frac{\sum_i^n a_i^{2D} \rho(y^{2D} - c_{y_i}^{2D})}{\sum_j^n \rho(y^{2D} - c_{y_j}^{2D})} \\
 F^{3D}(x^{3D}) &= \frac{\sum_i^n b_i^{3D} \rho(x^{3D} - c_{x_i}^{3D})}{\sum_j^n \rho(x^{3D} - c_{x_j}^{3D})} \\
 f^{3D}(y^{3D}) &= \frac{\sum_i^n b_i^{3D} \rho(y^{3D} - c_{y_i}^{3D})}{\sum_j^n \rho(y^{3D} - c_{y_j}^{3D})}
 \end{aligned} \tag{3}$$

where  $A = [a_1, \dots, a_n]$  and  $B = [b_1, \dots, b_n]$  are the linear combination coefficients. We can obtain synthesized 3-D faces as our temporary faces by seeking solution for (3), the specific method for solving (3) can be found in [15].

### B. Landmark Generation and Coordinate Optimization

Landmark has been widely utilized for computer vision applications for a long time. It can accurately describe the shapes and locations of face organs, which is critical for improving the local details of synthesized 3-D faces. A typical and successful application for landmarks on 2-D face is active shape model (ASM) [16]. However, it's still a challenging task to generate landmarks on 3-D face. We present a novel approach based on locality-constrained linear coding (LLC) [18] to achieve 3-D landmark generation and 3-D vertex coordinate optimization. For learning local nonlinear geometry of data in a semi-supervised way, local coordinate coding (LCC) [17] has shown promising results, and its fast implementation, namely locality-constrained linear coding (LLC), is proposed by Wang [18]. Let  $X$  be a set of D-dimensional input faces, i.e.  $X = [x_1, x_2, \dots, x_n] \in R^{D \times N}$ . Given a dictionary with  $M$  bases,  $B = [b_1, b_2, \dots, b_m] \in R^{M \times N}$ , locality-constrained linear coding finds a best coding  $c \in R^k$  for a sample  $x$  that

minimizes the reconstruction error and the violation of the locality constraint. The process can be written as an objective function as follows:

$$\begin{aligned}
 \min_c \|x - Bc\|^2 + \lambda \sum_{i=1}^M D_i * c_i \\
 s.t. \sum_{i=1}^M c_i = 1
 \end{aligned} \tag{4}$$

where  $D_i = \exp(\frac{\|x - B_i\|^2}{\sigma})$  stands for the locality adaptor that gives different freedom for each basis vector proportional to its similarity to the input data  $X$ . There is an analytical solution in [18] written as follows:

$$c^* = Norm(C_i + \lambda * diag(D_i)) \tag{5}$$

where  $C_i = (B - 1x_i^T)(B - 1x_i^T)^T$  represents the data covariance matrix.

The dictionary  $B$  is assumed to be known in (4). We conduct the LLC coding criteria to train a dictionary that is adapted to the distribution of the samples given a set of training samples. Hence an optimal dictionary can be obtained by minimizing the objective function below:

$$\arg \min_{C, B} \|x_i - Bc_i\|^2 + \lambda \sum_{j=1}^M D_i * c_i^j \tag{6}$$

where  $c_i$  refers to the corresponding coefficient for input data  $x_i$ .

By concatenating the training face and its corresponding landmarks as the input  $x_i$  for (6), we can obtain a coupled dictionary  $B = \{B_1, B_2, \dots, B_n\}$ , where  $B_i = \{B_f^i, B_l^i\}$ . Respectively,  $B_f^i$  stands for the dictionary of  $i$ th face and  $B_l^i$  refers to the dictionary of corresponding landmarks.

For a newly input face, we can approximate it by (7) after obtaining the coupled dictionary:

$$x_f^* = \sum B_f c \tag{7}$$

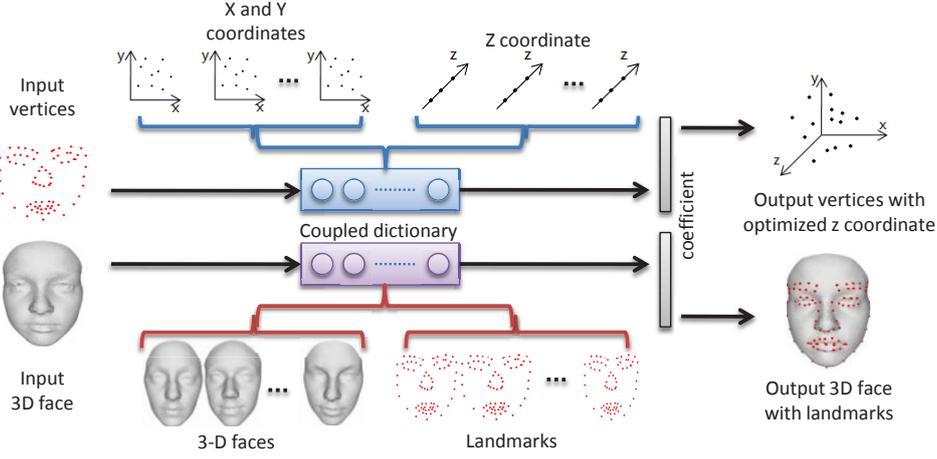


Fig. 2. The learning processes for generating landmarks on the bottom and coordinate optimization on the top. The inputs use the linear combination coefficients discovered from coupled dictionary to get the outputs.

where  $B_f$  denotes the dictionary of face, and  $c$  denotes the coding coefficient for  $x_f$  in  $B_f$ .

Because of the locality property of manifolds,  $c$  is quite sparse with a few non-zero elements. The process above can be seen as selecting the appropriate local bases for input face  $x_f$  and can be written as:

$$\min_c \sum \|x_f - B_f c\|^2. \quad (8)$$

In our approach, we assume that the face models and their corresponding landmarks share the same geometric structure in low dimensional representations. Hence the landmarks corresponding to input face can be approximated by the coding coefficient of  $x_f$ . By replacing the  $B_f$  in (7) with  $B_l$ , we can obtain the landmarks  $x_l$  of input face using the following:

$$x_l^* = \sum B_l c. \quad (9)$$

This approach is effective for both 2-D faces and 3-D faces. And the landmarks from 2-D faces and 3-D faces are combined to make further optimization for the locations of vertices in synthesized 3-D expressional faces. However, the landmarks from 2-D space lack z coordinate, without which we may receive unsmooth results in experiments. Therefore, a coupled dictionary  $B' = \{B_1', B_2', \dots, B_n'\}$  is built for the vertices on 3-D face, where  $B_i' = \{B_{xy}, B_z\}$ .  $B_{xy}$  denotes the dictionary of x and y coordinates of vertices and  $B_z$  refers to the corresponding z coordinate.

Similarly, the x and y coordinates can be represented by

$$x_{xy}^* = \sum B_{xy} c' \quad (10)$$

and z coordinate can be generated by

$$x_z^* = \sum B_z c'. \quad (11)$$

### C. Local Deformation

Since the temporary 3-D expressional faces generated by RBF network lack of personalization, we conduct a local deformation by using landmarks both from 2-D and 3-D faces. Let  $x_l^{2D} = \{v_l^{2D}\}_{1 \leq l \leq L}$  and  $x_l^{3D} = \{v_l^{3D}\}_{1 \leq l \leq L}$  be the landmarks for 2-D faces and 3-D faces, respectively. We define  $S_{li} = v_{li}^{3D}(x, y) - v_{li}^{2D}$ , where  $v_{li}^{2D} \in x_l^{2D}$ ,  $v_{li}^{3D} \in x_l^{3D}$  and  $v_{li}^{3D}(x, y)$  denotes the x/y coordinates of  $v_{li}^{3D}$ . Then for  $x^{3D} = \{v_i^{3D}\}_{1 \leq i \leq M}$ , the optimized locations  $v_i^{3D'}$  of all the vertices on temporary faces can be written as:

$$v_i^{3D'}(x, y) = \frac{\sum_{l=1}^L \mu_l S_{li}}{\sum \mu_l} + v_i^{3D}(x, y) \quad (12)$$

where  $\mu = \exp(-\|v_i^{3D} - v_l^{3D}\|)$  is the weight coefficient which is negatively correlated with the Euclidean distance between two vertices.

Noting that we only obtain the x and y coordinates for vertices, z coordinate can be generated by the aforementioned method using (10)(11). After the local deformation we obtain a local optimized 3-D face model  $x_*^{3D}$  from the temporary 3-D face model  $x^{3D}$ . Considering the global and local optimization, we combine  $x_*^{3D}$  and  $x^{3D}$  for weighted average in order to generate the final result of expression synthesis. The process can be written as follows:

$$x_{output}^{3D} = \eta x^{3D} + (1 - \eta) x_*^{3D} \quad (13)$$

where  $\eta \in (0, 1)$  is the weight parameter to adjust the impact between  $x_*^{3D}$  and  $x^{3D}$ .

## IV. EXPERIMENTAL RESULTS

To validate the effectiveness of the proposed method, we carry out experiments on a 32-bit Windows system with Core2 E6300, 2GB RAM. The 2-D and 3-D faces data are from the BU-3DFE database. We select 100 neutral face models and 400 expressional face models including happy, sad, surprising

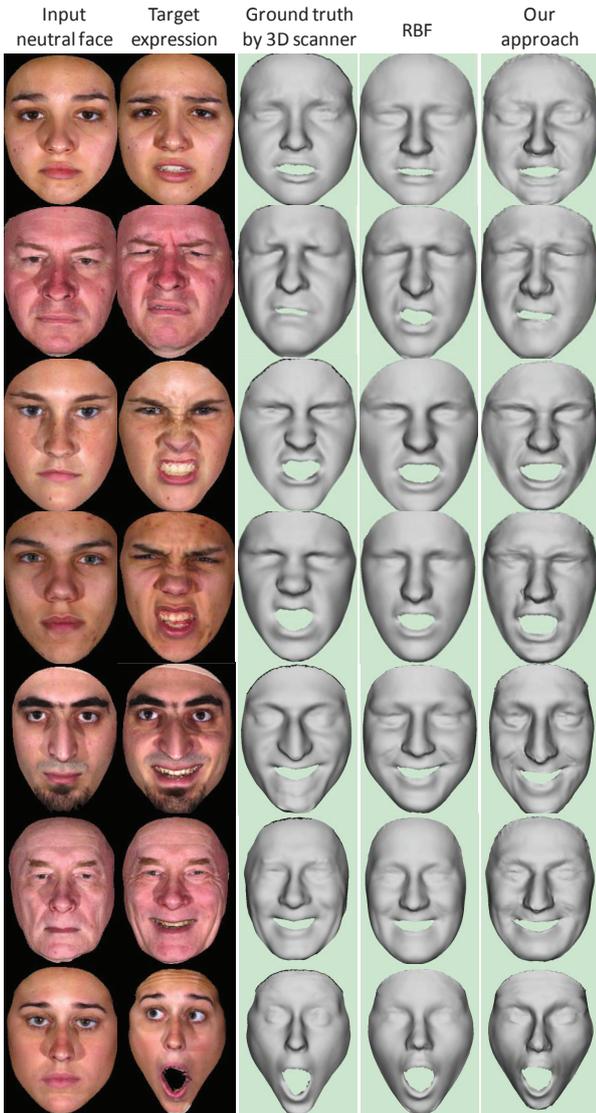


Fig. 3. The synthesis results of seven samples from BU-3DFE database. First column shows the neutral face. The second is the target expression. Third to fifth columns are the shapes from ground truth, RBF, and the proposed approach, respectively.

and disgusting from 100 subjects. These 500 models were obtained by a 3-D scanner and each model has a corresponding 2-D face image. All the faces in this database are with corresponding landmarks generated by our manual calibration. In our experiment, 80% of the models are used for training and the remaining 20% are used for testing.  $\eta$  is set to 0.3.

Considering that the unaligned data may lead to inaccurate results, a preprocess for face alignment is performed for each training 3-D faces. Firstly, feature points are located on face organs of all 3-D faces in the same positions. Then the cylindrical coordinates of 3-D faces are obtained by cylindrical mapping [5]. Finally, all the vertices from 3-D faces are aligned by centric mapping.

In Figure 3, we show the synthesized results of the expressional 3-D faces from the BU-3DFE database and compare

TABLE I  
MEAN SQUARE ERROR OF TWO APPROACHES CORRESPONDING TO THE SUBJECTS IN FIG.3

Subjects	1	2	3	4	5	6	7
RBF	9.80%	6.74%	4.12%	4.81%	5.84%	4.00%	4.97%
Our	8.72%	5.72%	3.14%	4.19%	3.75%	3.23%	4.15%

them with the results generated by RBF. Table I illustrates the estimation results by comparing the mean square errors and the equation can be written as follow:

$$\text{error} = \frac{\sqrt{\frac{|x_{gt} - x_{syn}|^2}{k}}}{\text{Diaglength}(x_{gt})} \quad (14)$$

where  $x_{gt}$  is the 3-D face ground truth,  $x_{syn}$  is the synthesized face,  $\text{Diaglength}(x_{gt})$  represents the diagonal length of ground truth's bounding box and  $k$  denotes the vertex number of 3-D face. From the results, we can find that the faces synthesized by the proposed method are personalized with more details than the faces generated by RBF. This is because the RBF only takes global optimization into account which leads to a smooth result but lacks of personalization.

## V. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed a novel approach to synthesize personalized 3-D expressional face based on nonlinear learning. By using C-RBF network, the intrinsic representations of 2-D and 3-D faces are discovered to generate temporary 3-D expressional faces by sharing linear combination coefficients. We utilized LLC for locating the landmarks on temporary faces and optimized the z coordinate for vertices in contrast with the landmarks form 2-D space. Finally, the faces and landmarks are combined to conduct a local deformation to obtain synthesized 3-D expressional faces.

In comparison with existing algorithms, the experimental results have validated the effectiveness of the proposed method. However, it is noticeable that some of our synthesized results are not smooth enough in details, such as the eyes. Hence, more efforts will be paid on synthesizing smooth 3-D expressional faces in the future work.

## REFERENCES

- [1] A. G. Bors and I. Pitas, "Object classification in 3-D images using alpha-trimmed mean radial basis function network," *IEEE Transactions on Image Processing*, vol. 8, no. 12, pp. 1744-1756, 1999.
- [2] S. T. Roweis and L.K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323-2326, 2000.
- [3] J. Park and W. I. Sandberg, "Universal approximation using radial-basis-function networks," *Neural computation*, vol. 3, no. 2, pp. 246-257, 1991.
- [4] M. J. Er, S. Wu, J. Lu and H. L. Toh. "Face recognition with radial basis function (RBF) neural networks," *IEEE Transactions on Neural Networks*, vol. 13, no. 30, pp. 697-710, 2002.
- [5] Challis and H. John, "A procedure for determining rigid body transformation parameters," *Journal of Biomechanics*, vol. 28, no. 6, pp. 733-737, 1995.
- [6] P. S. Aleksic and A. K. Katsaggelos, "Automatic facial expression recognition using facial animation parameters and multistream HMMs," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 1, pp. 3-11, 2006

- [7] M. Escher, I. Pandzic and N. Thalmann, "Facial deformations for MPEG-4," *Computer Animation 98*, 1998, pp. 56-62.
- [8] P. Ekman and W. V. Friesen, "Facial action coding system," *Consulting Psychologists Press*, Palo Alto: Stanford University, 1977.
- [9] Parke and I. Frederick, "A parametric model for human faces," *DTIC Document*, Salt Lake City, Utah: Dept of Computer Science, 1974.
- [10] J. Noh and U. Neumann, "Expression cloning," *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, 2001, pp. 277-288.
- [11] M. Song, Z. Dong, C. Theobalt, H. Wangm, Z. Liu and H. Seidel "A generic framework for efficient 2-D and 3-D facial expression analogy," *IEEE Transactions on Multimedia*, vol. 9, no. 7, pp. 1384-1395, 2007.
- [12] D. Vlastic, M. Brand, H. Pfister and J. Popvoi "Face transfer with multilinear models," *ACM Transactions on Graphics (TOG)*, vol. 23, no. 3, pp. 426-433, 2005.
- [13] M. Reiter, R. Dormer, G. Langs and H. Bischof, "3d and infrared face reconstruction from rgb data using canonical correlation analysis," *18th International Conference on Pattern Recognition*, vol. 1, pp. 425-428, 2006
- [14] X. Wang and R. Yang, "Learning 3-D shape from a single facial image via non-linear manifold embedding and alignment," *IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 414-421.
- [15] M. Song, D. Tao, X. Huang, C. Chen and J. Bu "Three-Dimensional Face Reconstruction From a Single Image by a Coupled RBF Network," *IEEE Transactions on Image Processing*, vol. 21, no. 5, pp. 2887-2897, 2012.
- [16] A. Blake, M. Isard, "Active shape models," *Springer*, 1998.
- [17] K. Yu, T. Zhang and T. Gong, "Nonlinear learning using local coordinate coding," *Advances in Neural Information Processing Systems*, vol. 22, pp. 2223-2231, 2009.
- [18] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang and Y. Gong "Locality-constrained linear coding for image classification," *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 3360-3367.
- [19] D. Tao, M. Song, X. Li, J. Shen, X. Wu, C. Faloutsos and S. J. Maybank "Bayesian tensor approach for 3-D face modeling," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 10, pp. 1397-1410, 2008.