

# 3D face modeling from single image based on discrete shape space

Dan Zhang<sup>1</sup> | Chenlei Lv<sup>1</sup> | Na Liu<sup>1</sup> | Zhongke Wu<sup>1</sup> | Xingce Wang<sup>1</sup>

School of Artificial Intelligence, Beijing Normal University, Beijing, China

## Correspondence

Xingce Wang, School of Artificial Intelligence, Beijing Normal University, Beijing China.  
 Email: wangxingce@bnu.edu.cn

## Funding information

Beijing Natural Science Foundation of China, Grant/Award Number: No.4172033; National Key R&D Program of China, Grant/Award Number: No. 2017YFB1002604; the National Key Cooperation between the BRICS of China, Grant/Award Number: No.2017YFE0100500

## Abstract

In this article, we propose a novel 3D face modeling method which constructs a new 3D face model from a low-dimensional feature space consisted of a large set of blend shapes based on the discrete shape space theory. The details of original face features are completely retained during the modeling process and a large number of new natural faces are constructed by several face samples. The optimization process of our method is independently decoupled for different facial attributes (identity, expression, and head pose), which improves the application flexibility and reduces the probability of it falling into a local optimal situation. The new facial data with new attributes are constructed based on the geodesic path search in discrete shape space with sufficient freedom and accuracy. In experiments and applications based on public databases (Helen, LFW, and CUFS), the modeling results show our method can provide high-quality 3D face model, with enough freedom for face expression editing and natural facial expression animation from a small facial sample set.

## KEY WORDS

3D face modeling, discrete shape space, facial landmarks, geodesic path

## 1 | INTRODUCTION

3D face modeling is an important issue for many applications such as game production, movie making, and cartoon animation creation. The face data can be edited using the reconstruction face model, which is a powerful tool for imaginative creative activities and business applications (DAZ3D, Maya, and Meitu). It is generally difficult for building a high-quality 3D face model from a single image, because the faces in images have different head poses, face scales and the geometric features are incomplete, which is termed “ill-posed.” To remove the influence of ill-posed features, the prior knowledge of the 3D face data should be used to guide the modeling process. Simultaneously, to synthesize the 3D face object from an image, the modeling method is required to match the geometric features (facial landmarks or contours) between the 3D face data and 2D facial image. In general, a standard 3D face object (from Facewarehouse or BosphorusDB) has more than 10,000 vertices, it cause that the high computational complexity of directly 3D face modeling in original 3D face objects, which it is unattainable. Therefore, we propose a 3D face modeling method which represents 3D face by facial landmarks can quickly and correctly construct 3D face model based on the discrete space in this article.

Chenlei Lv and Dan Zhang are co-first authors.

[Correction added on 22 September 2020, after first online publication: Chenlei Lv author name has been corrected.]

## 1.1 | Related works

3D face modeling has been investigated for many years. According to the form of input data and feature analysis algorithms, 3D face modeling methods can be divided into several categories: 3D geometric modeling, 3D static modeling, 3D dynamic modeling, and local deformation modeling.

*3D geometric modeling* methods extracted geometric features from multifacial images or 3D face objects to construct a facial model. Decarlo et al.<sup>1</sup> used anthropometric facial features to build a facial model. Lee et al.<sup>2</sup> constructed a facial model using multiviews of facial data. Ansari and Abdel-Mottaleb<sup>3</sup> proposed an automatic facial modeling method using two orthogonal views of facial data. Kurtek and Drira<sup>4</sup> used the elastic measure in shape space to analyze different 3D faces. Alashkar et al.<sup>5</sup> mapped the 3D facial data flow into a Grassmann manifold for facial recognition. For 3D face reconstruction from a single image, these methods cannot extract complete geometric information and difficult to achieve facial representation with accurate facial attributes.

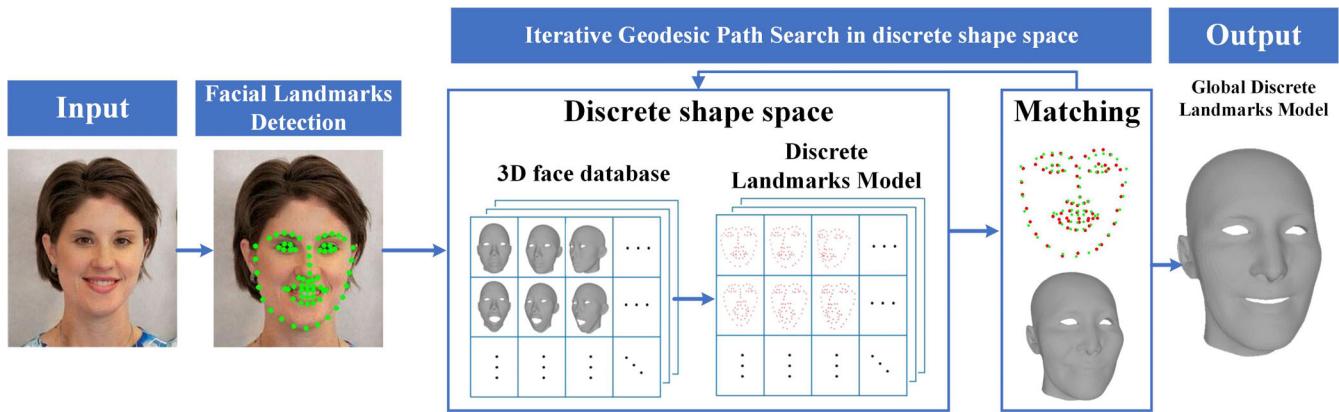
*3D static modeling* methods construct a static face model based on a low-dimensional feature space. The face modeling process is transferred to a linear optimization problem in this space. Blanz<sup>6</sup> proposed the classical facial modeling method 3DMM. 3DMM was used to reconstruct 3D facial data with special characteristics, such as gender and age. Pighin<sup>7</sup> proposed a facial photograph regenerating method by 3D shape morphing. Paysan<sup>8</sup> improved the 3DMM to include pose and illumination invariant face recognitions. Bas<sup>9</sup> proposed a 3DMM fitting method that considered edge information. Luthi et al.<sup>10</sup> proposed Gaussian Process Morphable Models (GPMMs) which represent a class of shapes as a normal distribution of point variables to morph models based on point distribution models (PDMs). Patel and Smith<sup>11</sup> combined the shape analysis in Kendall shape space and 3DMM to construct 3D face model. In feature space, different faces are transferred to a regular representation, which reduces the complexity of the face analysis. However, the construction of a 3D face in a low-dimensional feature space loses some facial feature details, and different face attributes are not decoupled.

*3D dynamic modeling* methods construct a facial model based on various facial samples with different expressions. Lu and Jain<sup>12</sup> proposed a deformation modeling method for 3D face matching. Ichim<sup>13</sup> proposed a 3D facial modeling method in mobile phone. The reconstruction of the 3D facial data included texture and wrinkle. Cao et al.<sup>14</sup> proposed a facial expression database (Facewarehouse) that constructs the multilinear models with facial surface fitting. Jin et al.<sup>15</sup> used the facial frontal and side images to reconstruct the high-fidelity 3D facial model. The new frameworks based on deep learning algorithm were proposed recently. Guo et al.<sup>16</sup> proposed a 3D face modeling method which uses the convolution neural networks to recognize the shape information of facial data. Han et al.<sup>17</sup> used a similar framework in 3D facial sketch modeling with exaggerated expression. For accuracy of 3D face modeling from a single image, such methods have a high probability of becoming trapped in a local optimal. Schönborn et al.<sup>18</sup> proposed a fully probabilistic method to interpret a single-face image with the 3D Morphable Model which is based on Bayesian inference. They inferred the posterior distribution of the model parameters given the target image rather than searching a single optimal solution to avoid falling into local optimum. But the 3D face construction of the methods uses a data-driven Markov chain Monte Carlo (DDMCMC) method and also depends on the blend shapes which are limited by the various training samples provided.

*Local shape modeling* methods have been used in an attempt to improve the accuracy of 3D face modeling by local facial feature matching. Joshi<sup>19</sup> used blend shapes segmented into small regions to model a face in linear space. Tena et al.<sup>20</sup> presented a linear model composed of a collection of PCA submodels and the models were independently trained. Neumann et al.<sup>21</sup> proposed localized deformation models for a face model based on a mesh sequence and sparse matrix decompositions. Bradley and Beeler<sup>22</sup> proposed a new anatomically constrained local deformation model for 3D face modeling from a monocular face capture. The key advantage of local shape modeling methods is the high quality of the 3D face reconstruction which restored the face feature details from the local face regions. However, the local shape models fail at the global characteristic representations of different faces, limiting their performances in related applications such as expression transfer and face animations.

## 1.2 | Contributions

Our method belongs to the 3D geometric modeling, which constructs discrete shape space to represent the 3D faces and provides a similarity measure between different faces. The basic idea of our method is similar to Reference 11. The difference is that our face modeling process is computed in the discrete shape space directly rather than in low-dimensional feature space. The advantages of shape space can be used in face modeling from a single image, which includes a rigid transfer robustness and an accurate optimization process. In summary, our main contributions are as follows:



**FIGURE 1** The pipeline of our face modeling method

1. We construct a discrete shape space to represent 3D faces by facial landmarks for face representation, similarity measurement and new face synthesis. The construction of shape space does not require complicated preprocessing, and it does not require reconstruction when the new samples are added in face database.
2. We propose a new optimization process to match a 3D face counterpart and 2D facial image in the discrete shape space. The optimization process is decoupled to different facial attributes and the modeling result is more accurate than methods based on energy gradient descent algorithm.
3. We propose an effective 3D face data synthesis method based on the optimization process which determines the necessary parameters from the geodesic path searching in discrete shape space. The facial details are not lost and a large number of 3D faces with new attributes can be constructed from a small 3D face set.

In Figure 1, we show the pipeline of our method including three steps: (1) the facial landmarks are extracted from the image to construct a 2D discrete landmark model; (2) a discrete shape space is built based on the 3D discrete landmark models of the 3D face database with different identities, expressions, and poses; (3) the 2D discrete landmark model is input as a source into the discrete shape space, and the target 3D discrete landmark model is searched to match. Using the target 3D discrete landmark model, we can recover the target 3D facial object.

The rest of this article is organized as follows. The fundamentals of shape space theory are introduced in Section 2. In Section 3, the details of discrete shape space construction are provided. In Section 4, we describe the face modeling method in discrete shape space. The experiments and applications of our method are shown in Section 5.

## 2 | FUNDAMENTALS

The discrete shape space theory is proposed from Kendall's work<sup>23</sup> that is suitable for measuring the similarity between different curves in a Riemannian framework. A shape can be represented a point in a high-dimensional, nonlinear manifold, called Kendall shape space.

### 2.1 | Riemannian manifold

$M$  is a 2-dimensional symmetric and positive Riemannian manifold equipped with a metric  $d$ . A Riemannian metric determines an inner product on each tangent space  $T_p M$  ( $p \in M$ ).<sup>24</sup>

The geodesic can be defined based on Riemannian manifold. For each  $C^1$  curve  $\gamma : [0, 1] \rightarrow M$ , we define its length:

$$L[\gamma] = \int_0^1 \langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle_{\gamma(t)}^{1/2} d(t) = \int_0^1 \|\dot{\gamma}(t)\| d(t). \quad (1)$$

A geodesic  $\gamma : [0, 1] \rightarrow M$ , is minimal if its length is less than or equal to the length of any other piecewise smooth curve joining its endpoints. In Kendall shape space, the similarity measurement of shapes is achieved using the geodesic.

## 2.2 | Kendall shape space

The shape information are represented by a landmarks' sequence in Kendall shape space which is a Riemannian manifold with a discrete form and rigid transfer removed.<sup>25</sup> To illustrate Kendall shape space, we propose the Riemannian manifold with discrete form first.

$$M = \mathbb{R}^{m \times k} \setminus \{0\}, D \in M, D = (x_{(1)}, \dots, x_{(k)}), \quad (2)$$

where  $k$  is the number of points in a shape and  $m$  is the dimension of one point. The dimension of manifold  $M$  is  $mk$ .  $D$  is the landmarks' sequence. We define two discrete point sets  $D(a)$  and  $D(b)$ . The natural Riemannian metric in  $M$  is proposed in Equation (3).  $V(a), V(b) \in T_D M$  (tangent space of  $M$ ).

$$\eta \langle V(a), V(b) \rangle_D = \text{tr}(V(a)V(b)^t) = \sum_{i=1}^k \langle V(a)_i, V(b)_i \rangle_{\mathbb{R}^m}. \quad (3)$$

To construct Kendall shape space, the three rigid influence factors which are defined by group action in  $M$  should be removed. In Equation (4), we provide the group action  $G$  which includes translation  $b$ , scaling  $a$ , rotation  $O$ .  $1_k = (1, 1) \in \mathbb{R}^{1 \times k}$ .

$$G.D \mapsto (b, a, O; D) \mapsto aO.D + b.1_k. \quad (4)$$

In Kendall shape space, the geodesic distance between different shapes is invariant in group action  $G$ . The property is used to shape similarity measure between curves which is represented as  $M/G = \{[p] | p \in M\}$ , and  $D(a), D(b) \in M/G$ . In Riemannian manifold  $M$ ,  $[D]$  is the group achieved from the  $G.D$ . The  $[D]$  in quotient space  $M/G$  share a same reflection. The quotient space can be regarded as a Kendall shape space. The geodesic distance in Kendall shape space can be represented as:

$$d_{Ken}(D(a), D(b)) = \inf_{g \in G} d_M(D(a), g.D(b)). \quad (5)$$

*Removing scaling and translation.* The similarity measure between different shapes in Kendall shape space is invariant to the group action which means the influence of  $G$  should be removed from shape's representation. To remove the scaling and translation of  $D(a)$  and  $D(b)$ , the centroid is aligned and the scaling normalization processed. In Equation (6), we show this process;  $\bar{x}$  is the centroid of  $D(a)$ , and  $s(D(a))$  is the scale of the  $D(a)$ . Using each point in  $D(a)$  minus the centroid and divided by  $s(D(a))$ , the new form  $D_s(a)$  of  $D(a)$  is achieved with the translation and scale factors removed.

$$D_s(a) = (x_{(1)} - \bar{x}, \dots, x_{(k)} - \bar{x}) / s(D(a)),$$

$$\bar{x} = \frac{1}{k} \sum_{j=1}^k x_{(j)}, s(D(a)) = \left( \sum_{j=1}^k \|x_{(j)} - \bar{x}\| \right)^{1/2}. \quad (6)$$

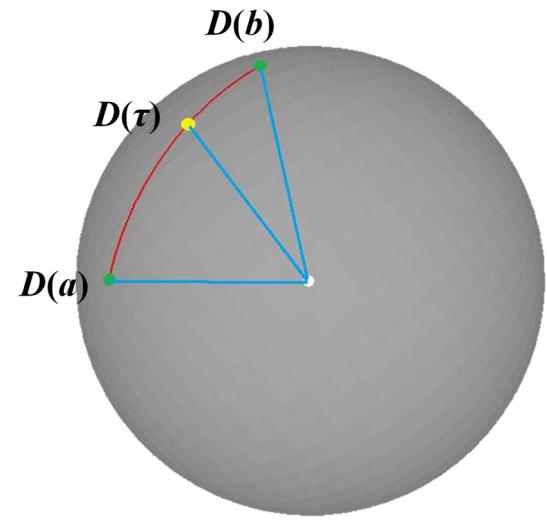
Based on Equation (6), a new shape measurement is proposed in Equation (7) where  $O$  represents the rotation group and  $S_m^k$  represents the preshape space of Kendall shape space with the scaling and translation factors removed.

$$d_{Ken}(D(a), D(b)) = \inf_{O \in G} d_{S_m^k}(D_s(a), O.D_s(b)). \quad (7)$$

*Procrustes analysis.* There are several methods to remove the influence of rotation. In Kendall shape space, the process can be transferred to a singularly valuable decomposition (SVD) problem. The computation can be interpreted as the alignment of eigenvectors in the matrix. Equation (8) provides computation:  $Z$  is the orthogonality of  $D_s$ ;  $U$  and  $V$  are the matrices from the SVD decomposition; and  $R$  is the matrix that represents the rotation action from  $SO(m)$ .

$$d_{S_m^k}(Z(a), O.Z(b)) = \text{arccos}(R\Lambda),$$

$$Z(b)Z(a)^t = U \Lambda V, R \in SO(m). \quad (8)$$

**FIGURE 2** Kendall shape space and geodesic path between shapes

In Kendall theory, the orthogonality is optimally registered for the rotation group  $O$ . Combining the different equations, the final computation for the geodesic distance representation in Equation (9) is achieved, called the procrustes analysis.

$$\begin{aligned} d_{Ken}(D(a), D(b)) &= \inf d_{Sk_m}(Z(a), O.Z(b)) \\ &= \text{arccos}(\Lambda), \end{aligned} \quad (9)$$

$$\begin{aligned} D(\tau) &= \frac{1}{\sin(\theta)}(\sin(\theta(1-\tau))D(a) + \sin(\theta\tau)D(b)), \\ \theta &= d(D(a), D(b)), \quad \tau \in [0, 1], \end{aligned} \quad (10)$$

The distance from Equation (9) can be regarded as the geodesic distance in Kendall shape space. Following the geodesic path between two shapes  $D(a)$  and  $D(b)$ , the new shape can be constructed. Equation (10) describes the new shape construction process:  $\theta$  is the intersection angle between two shapes,  $D(a)$  and  $D(b)$ , in the space that is equal to the geodesic distance; and  $\tau$  is a parameter that defines the position in the geodesic path between  $D(a)$  and  $D(b)$ . In Figure 2, we illustrate the geodesic path between  $D(a)$  and  $D(b)$  in Kendall shape space.

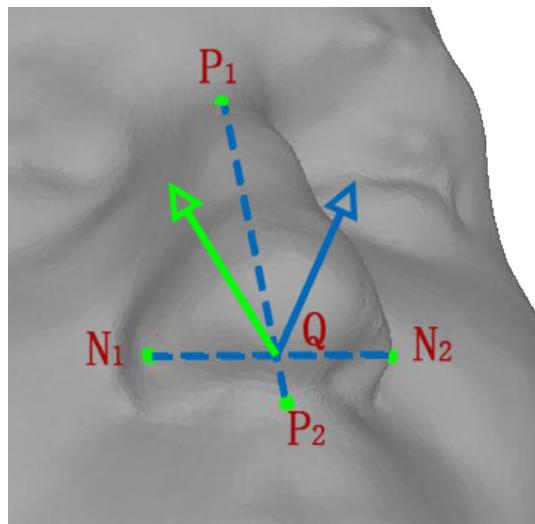
### 3 | 3D FACIAL LANDMARKS BASED DISCRETE SHAPE SPACE

To use the relevant tools of Kendall shape space in 3D face modeling, a facial representation with discrete points that reveals the important facial features is required. We propose a discrete landmarks model (DLM) with a representation based on facial landmarks. The facial landmarks come from different regions, such as the nose, eyes, mouth, and face contour.<sup>26</sup>

The construction process of the DLM includes three parts: (1) translation removal equation; (2) scaling removal equation; (3) rotation alignment equation. Equation (11)–(13) describe construction of DLM. Based on the DLMs, the discrete shape space for 3D face modeling is constructed.

$$\begin{aligned} F &= \{x_1, \dots, x_k\}, \quad B \in F, \\ L &= \{v_1, \dots, v_k\}, \quad v_i = x_i - B. \end{aligned} \quad (11)$$

$F$  is the set of the facial landmarks, and  $B$  is the benchmark of  $F$ . The benchmark is used as a uniform location for the landmarks from different faces. We select the nasal tip landmark as the benchmark (For nasal tip selection in 3D registration face database, we sign the position manually).  $L$  is the set of landmark vectors that are obtained by the landmarks and benchmark. Then we achieve the preliminary landmarks model.



**FIGURE 3** Instances of the local coordinate system using different landmarks. In the 3D scene, the points  $P_1, P_2, N_1$ , and  $N_2$  are landmarks around the nose area (selected manually). The point  $Q$  is the pedal of vector  $P_1P_2$  and vector  $Q$ , which is calculated using by  $P_1P_2 \times N_1N_2$

$$s(L) = \left( \sum_{i=1}^k \|v_i\| \right),$$

$$L_s = v'_1, \dots, v'_k, \quad v'_i = \frac{v_i}{s(L)}, \quad \sum_{i=1}^k \|v'_i\| = 1. \quad (12)$$

In Equation (12),  $s$  is the scale of  $L$ . A different  $L$  scale influences the analysis of facial feature relative positions. The landmark model requires unique scale. To remove the scaling factor, each vector in  $L$  minus centroid and divided by  $s(L)$ . The new landmark model,  $L_s$ , is then constructed. The scale of  $L_s$  is normalized, and the relative positions of the vectors in  $L$  remain. The representation of  $L_s$  follows the discrete shape  $D_s$  in Equation (6). In Kendall theory, the rotation is removed using Equations (8) and (9). In our method, we use a simple method instead of procrustes analysis. The DLM is constructed using facial landmarks that include semantic information. A local coordinate system using certain landmarks to align the facial landmarks can be constructed.

$$L_{sr} = \{T(v'_1), \dots, T(v'_k)\}, \quad (13)$$

$$d_{Ken}(L_{sr}(a), L_{sr}(b)) = \arccos(L_{sr}(a) \cdot L_{sr}(b)). \quad (14)$$

$L_{sr}$  is the final representation of the DLM. The rotation factors still exist for the vectors in  $L_s$ . This influence comes from the different head poses of the 3D facial data. We introduce a transform function  $T$  which is based on a local coordinate system to remove the rotation from the landmark model. We select the landmarks around the nose to build the coordinate system which are robust to facial expressions. The geodesic distance of different DLMs can be calculated using Equation (14).

In Figure 3, we show the instances of the local coordinate system. Based on an existing 3D face database, we construct the  $L_{sr}$  form for each face, which is used to construct the discrete shape space.

## 4 | 3D FACE MODELING PROCESS

Based on the discrete shape space, we propose a 3D face modeling process from a single image. The modeling process includes two steps: a geodesic path search in discrete shape space (*GPSD*) and 3D face reconstruction. Traditional modeling methods<sup>9</sup> use an energy gradient descent algorithm to reduce the facial landmarks distance between the 3D face and the 2D facial image. In our method, the geodesic path searches for the new DLM construction based on a geodesic distance computation in discrete shape space, and the different facial attributes are independently optimized. The target optimization distance energy,  $E_{synthesis}$ , is given by Equation (15).  $S_{image}$  represents the facial landmarks in a 2D facial image.  $T_r$  represents the facial landmarks from a synthetic 3D face based on a 3D facial database. In the computation process, the coordinates of the Z axis in  $T_r$  are ignored for consistency of the data dimensions.

$$E_{synthesis} = d_{Ken}(L_{sr}(S_{image}), L_{sr}(T_r)). \quad (15)$$

The geodesic path search algorithm achieves a searching path in the discrete shape space. According to the path, the new 3D face can be reconstructed.

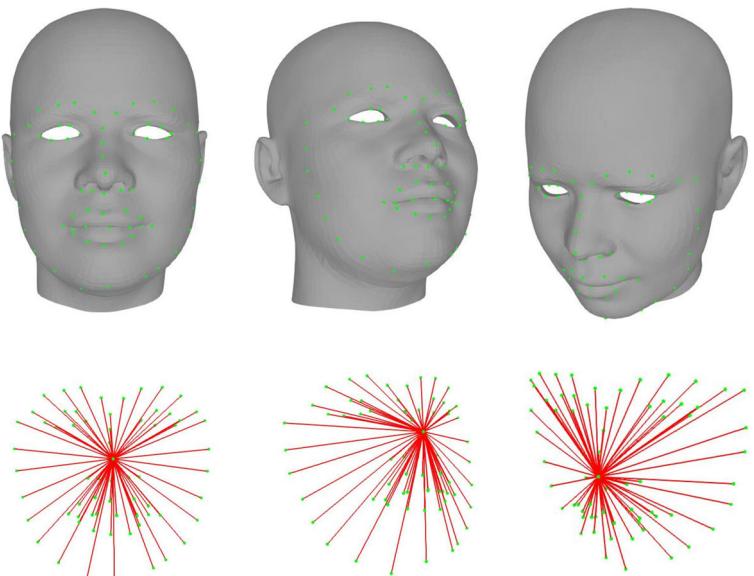
#### 4.1 | Geodesic path search in discrete shape space (GPSD)

The *GPSD* is used to synthesis a new 3D target DLM from the discrete shape space to match the 2D target DLM of facial image. In discrete shape space, the different DLMs from the 3D face database are organized to a tensor that follows the different facial attributes (identity, expression, and head pose). In the 3D face database, the head pose attribute does not exist. To add the head pose into different faces, we rotate the face data to follow different axes and achieve new 3D DLMs (Figure 4). We achieve a set of 3D DLMs that have different poses.

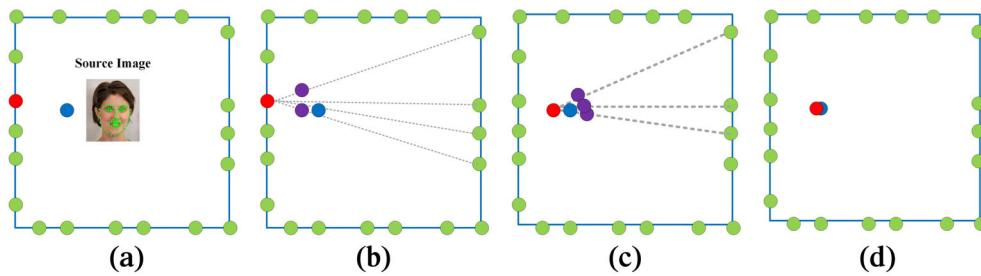
To explain the implementation details, we show an instance of *GPSD* for a certain facial attribute (expression) in Figure 5. The initialization of the *GPSD* is searching the start DLM with minimum distance energy  $E_{synthesis}$  from the discrete shape space to the 2D target DLM in the facial image. In Figure 5a, the red point represents the start DLM with three attribute parameters (identity, expression, and head pose) in the space. The start DLM in this place is the initial DLM in global optimization, so it is also called the global start DLM. Next, we compute a set of new DLMs using geodesic path searching. The *GPSD* selects DLMs with the same identity and head pose as the global start DLM, then these become the direction DLMs. Using the start DLM and direction DLMs, a new set of DLMs are obtained, which are called middle DLMs and calculated using Equation (10) (The start DLM represents  $D(a)$ , and the direction DLMs represent the  $D(b)$ ). The parameter  $\tau$  is fixed at a small value, 0.1.). Each middle DLM has an expression transfer from the start DLM to a direction DLM. Then we update the new start DLM from the middle DLMs with a minimum distance energy, to the target DLM. In Figure 5b, the purple points represent the middle DLMs. In Figure 5c, the red point represents the new start point that is selected from the middle DLMs. The *GPSD* iteratively updates the start DLM. The iteration of *GPSD* is stopped when the  $E_{synthesis}$  between the start DLM and target DLM is lower than a certain value(0.05) or the  $E_{synthesis}$  cannot be reduced by the start DLM update. The start DLM after the final iteration (red point in Figure 5d) is the new constructed DLM  $L_{sr}(T_{expression})$ , with a new synthesized expression.

$$List_{expression} = \{e_1, \dots, e_\phi\}. \quad (16)$$

In Equation (16), the process of *GPSD* is recorded as a list of expression indexes (the expression index  $e$  represents the exist expression data in the original 3D face database) of direction DLMs' selections in each iteration, and  $List_{expression}$  is the expression indexes sequence.



**FIGURE 4** New DLMs with different rotation. The first and second lines show are the original face data and DLMs, respectively



**FIGURE 5** Instance of expression optimization by *GPSD*. The red point is the start DLM; green points are the direction DLMs; blue point is the 2D target DLM from the facial image; purple points are the middle DLMs

The update process is represented in Equation (17).

$$L_{sr}(T_{expression}) = GPSD(List_{expression}, L_{sr}(S_{image})). \quad (17)$$

Using the list of expression indexes from the discrete shape space, the new DLM can be obtained with a new expression for each person. The process of identity optimization is similar to that of the expression optimization. First, we compute the new set of direction DLMs which are achieved by a new DLM with new expression following  $List_{expression}$  of each person. Next, we select the start DLM from the direction DLMs; using the same method to update the start DLM and achieve the new construction DLM with the new identity attribute. Then the new list is obtained for identity index  $i$ , which is represented by  $List_{identity}$  in Equation (18). The update process is described by Equation (19).

$$List_{identity} = \{i_1, \dots, i_\varphi\}, \quad (18)$$

$$L_{sr}(T_{identity}) = GPSD(List_{identity}, L_{sr}(S_{image})). \quad (19)$$

Based on  $List_{expression}$  and  $List_{identity}$ , the new DLM with new expression and identity attributes is reconstructed from the discrete shape space. For now, the new DLM has the same head pose as that of the global start DLM. After the *GPSD* process, the head pose should be updated for accurate matching. We rotate the new DLM and achieve new attributes for rotation. In Equation (20), the rotation parameters (the rotation angle  $r$  includes  $x_r, y_r$ , and  $z_r$  around  $X, Y$ , and  $Z$ , respectively) are recorded in  $List_{rotation}$ .

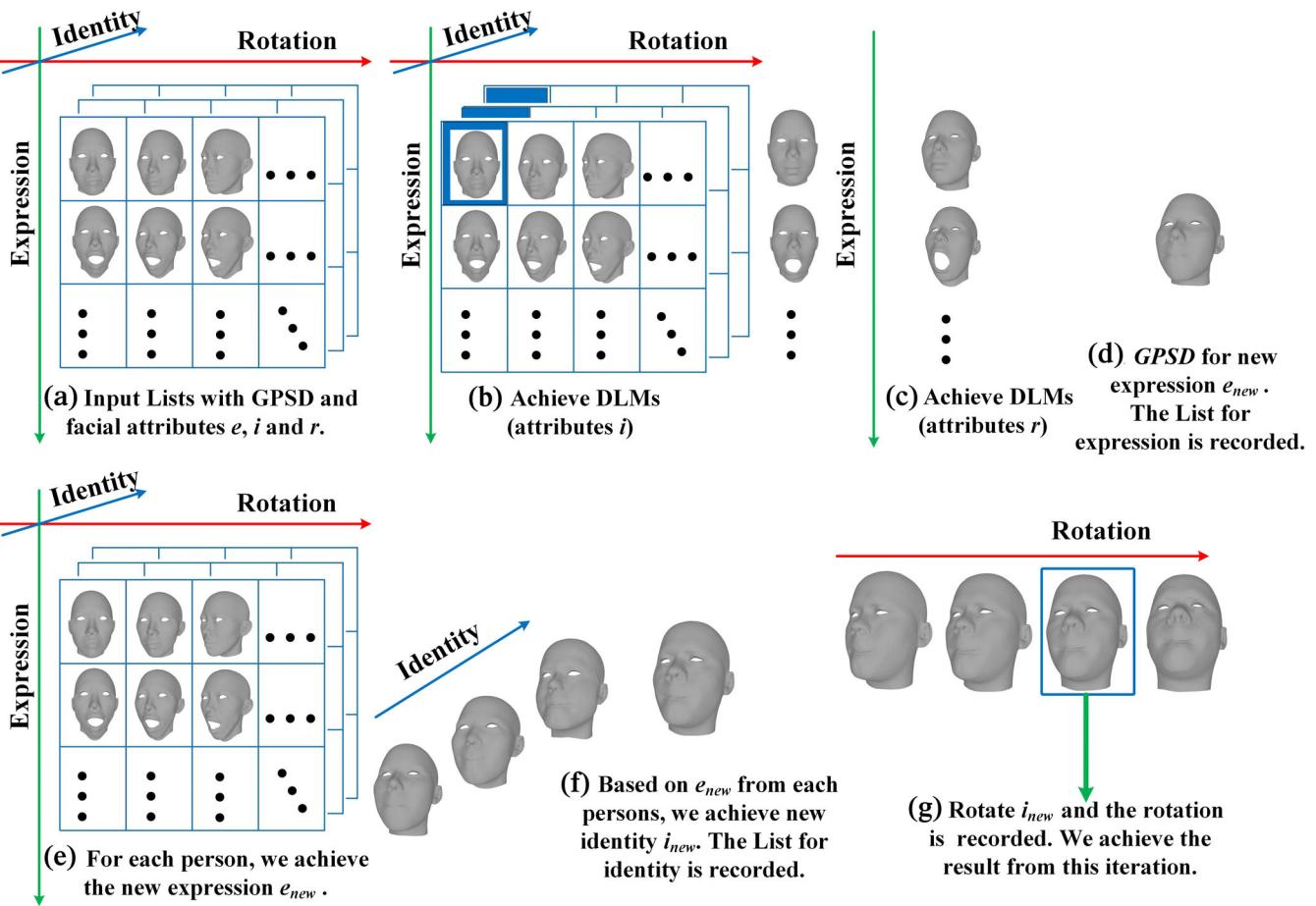
$$List_{rotation} = \{x_r, y_r, z_r\}. \quad (20)$$

The update process is described by Equation (21).

$$L_{sr}(T_{rotation}) = Rotate(List_{rotation}, L_{sr}(S_{image})). \quad (21)$$

The influence of contour landmarks should be considered in head pose estimation after rotation optimization. The reason is the positions of contour landmarks have been changed in facial image which are not consistent to the face model.<sup>27,28</sup> For accurate head pose estimation, the contour landmarks should be updated. We detect the parallel transfer points of contour landmarks and search the best match, which is equal to contour landmarks update in Reference 28. We update the new contour landmarks in DLM and achieve the final rotation parameters.

A complete iterative optimization of *GPSD* is comprised of a three searching process based on  $List_{expression}$ ,  $List_{identity}$ , and  $List_{rotation}$ . After one complete iteration of *GPSD*, the new DLM is constructed with different facial attributes and the *GPSD* updated lists are recorded. In next iteration of *GPSD*, the start DLM is obtained with the new facial attributes that are computed by a previous computation. When we extract the corresponding DLMs with certain facial attributes, different lists are used to reconstruct the data using Equations (17), (19), and (21). In Figure 6, we show an instance of complete iterative optimization of the *GPSD*. Through the iterative calculation of *GPSD*, the final 3D target  $L_{sr}(T_r)$  is achieved with a minimum distance energy  $E_{synthesis}$  to target  $L_{sr}(S_{image})$ .

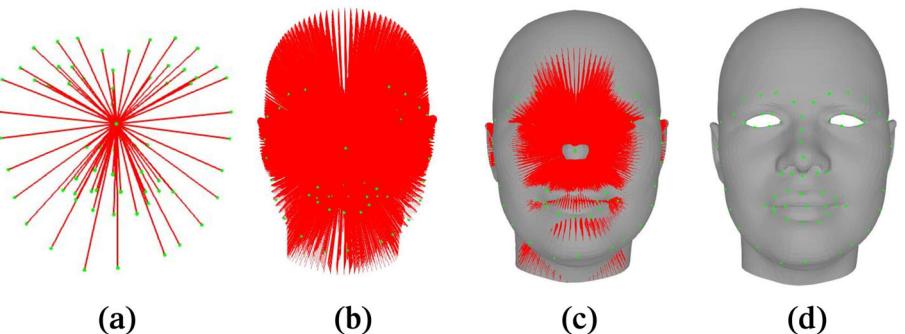


**FIGURE 6** Instance of complete optimization by *GPSD* (a–g). In pictures, the DLMs are represented by original 3D face for convenient view

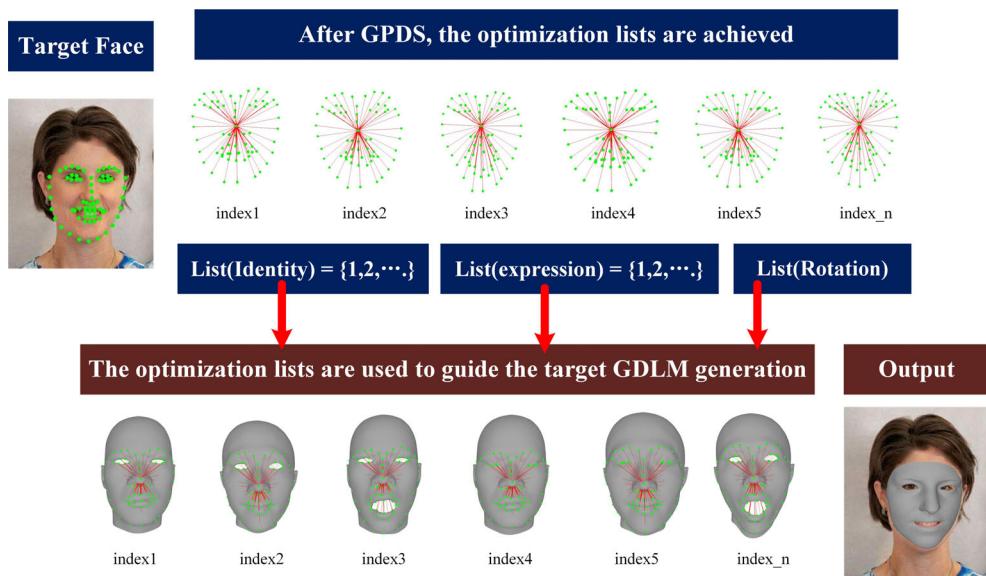
## 4.2 | 3D face reconstruction

Based on the *GPSD* result, a 3D face can be constructed. The DLMs in discrete shape space are constructed by the facial landmarks which can be regarded as the 3D face counterpart. Based on the DLM, we propose a global DLM (GDLM) that is constructed using all vertices.

In Figure 7, we show the comparison between DLM and GDLM. The construction method of the GDLM is the same as that of the DLM, all vertices in the original 3D face are used instead of the set of facial landmarks. The optimization lists guide the new DLM regenerated from the DLMs in discrete shape space. Each DLM in the space has a corresponding GDLM. The new DLM generating process based on the optimization lists can be applied to the GDLM generating process in parallel. Each vertex of the new 3D face is determined by the reconstruct GDLM, and the topological relation is inherited. The GDLM reconstruction process is shown in Figure 8.



**FIGURE 7** Instance of GDLM. (a) DLM, (b) GDLM, (c) GDLM with meshes, and (d) Final reconstruction result



**FIGURE 8** GDLM generation from the DLM optimization lists

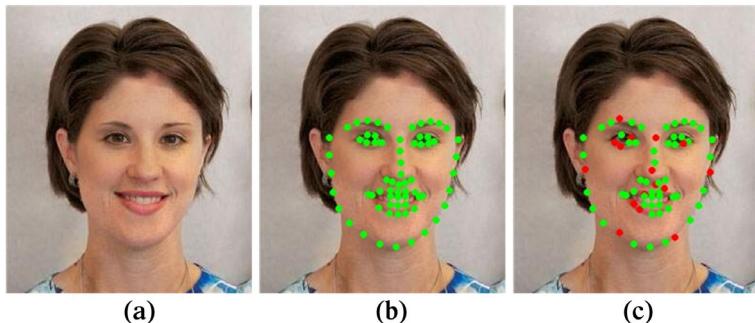
## 5 | EXPERIMENTS AND APPLICATIONS

We evaluate the performance of our method on platform VS2013, OpenGL4.6.0 and OpenCV2.4.10. The hardware platform is constructed by Intel Xeon E5620 2.4 GHz, 12G RAM, and TITAN XP. The 3D face database (FaceWareHouse) is used to construct the discrete shape space. The database has 150 persons and each with 20 expressions, we use a subset of the database to construct the discrete shape space ( $50 \times 20$ ). In our method, we choose a set of angles to rotate the 3D DLM in a subset of database (rotation around X:  $[-30, 30]$ , Y:  $[-30, 30]$ , Z:  $[-15, 15]$ , step = 5, includes  $13 \times 13 \times 7 = 1,183$ ). Finally, we obtain a 3D counterpart candidate set of the 3D DLM. The vertices of the face data in FaceWareHouse are aligned. We manually label the 3D facial landmarks in one face data. The landmarks' positions can be automatically assigned to other face objects. Using the 3D landmarks' positions, we construct 3D DLMs and discrete the shape space. For the 3D face modeling evaluation and related application exhibition from a single image, we select two face image databases (LFW and Helen) and one face sketch database (CUFS).

## 5.1 | 3D face modeling quality evaluation

In the modeling process, the facial landmarks in a facial image directly affect the modeling result. We use the landmarks detection method<sup>26</sup> to extract 68 facial landmarks from different facial regions from the facial image. The *GPSD* is affected by two factors: facial landmarks for DLM construction and facial samples for discrete shape space construction. To evaluate the facial landmarks that are sensitive during the  $E_{synthesis}$  optimization, we add two sets of random displacements to the landmark subset; randomly selecting eight and 16 facial landmarks from different facial regions and adding the randomly displacements ( $\leq 5$  mm). One set of the landmarks' movement are signed in Figure 9.

To evaluate the modeling performance in different scale of face samples in the discrete shape space construction, we use three face sample sets with different numbers of people (10, 50, 100) from Facewarehouse to construct the discrete



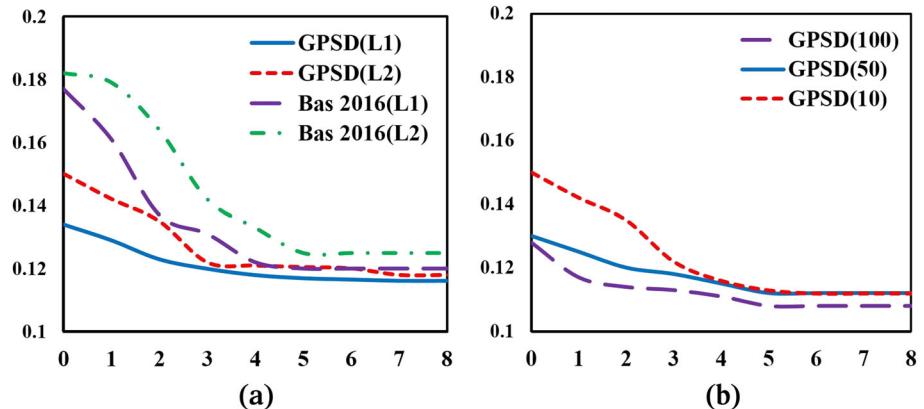
**FIGURE 9** Movement in landmarks subset. The green points are the original detection results. The red points are selected with a random movement.

shape space. In Figure 10, the  $E_{synthesis}$  reduction curve graph is shown for different iteration steps and different landmarks sets by the two methods and different scales of face samples in discrete shape space. The data in transverse axis represent the iteration step's number, in longitudinal axis represent the  $E_{synthesis}$  value. Figure 10a shows the  $E_{synthesis}$  values by two methods with different sets of facial landmarks' movement (L1: eight landmarks, L2: 16 landmarks). To evaluate the performance of the GPSD, we compare the gradient descent algorithm method<sup>9</sup> and GPSD for  $E_{synthesis}$  optimization. Figure 10b shows the  $E_{synthesis}$  values by our methods with different numbers of face samples (10, 50, 100). The target facial images have 300 samples which are selected from Helen and LFW randomly. The results show that the GPSD has better convergence property, robustness to local facial landmarks, and novelty performance in small face samples. Our discrete shape space is constructed using 50 persons' data (totally 50\*20 original face objects, index 1–20 and 100–124 in Facewarehouse) in applications. The 50 persons include facial data with different genders and ethnic groups to keep the basic diversity.

To evaluate the final modeling result, we compute the procrustes analysis error<sup>23</sup> between facial image and facial modeling result. Based on different landmarks' sets, we compare the procrustes analysis errors of our method in Table 1 (25 and 68 landmarks are selected based on method,<sup>26</sup> 83 and 106 landmarks are selected based on tools of face++: <https://api-cn.faceplusplus.com/facepp/v3/face/analyze>). The quality of our modeling result cannot be improved by adding landmarks when the landmarks have covered the global facial regions. The reason is that redundant landmarks may produce overfitting phenomenon.

We also compare the procrustes analysis errors from different methods.<sup>9,14,27</sup> Such modeling method are also based on facial landmarks. The 3D Morphable Model in Reference 9 is trained by Blanz.<sup>6</sup> The 3D Morphable Model in Reference 27 is trained by facial data from Reference 6 and Facewarehouse, which include 300 persons and 47 expressions for each person. The core tensor space in Reference 14 is trained from Facewarehouse (150\*47). In Table 2, we show the procrustes analysis error of different methods in LFW with different head poses.

In Table 3, we show the procrustes analysis error of different methods in Helen with different head poses. The face modeling speed of different methods should be evaluated. The modeling process of our method can be divided into



**FIGURE 10**  $E_{synthesis}$  reduction curve graph for different iterations and different conditions

**TABLE 1** Procrustes analysis error based on different facial landmarks' set in two image sets (LFW and Helen)

Database	Error				Facial landmarks' sets
	25 <sup>26</sup>	68 <sup>26</sup>	83 (Face++)	106(Face++)	
LFW	0.16	0.1	0.18	0.19	
Helen	0.15	0.13	0.16	0.18	

**TABLE 2** Procrustes analysis error by different methods for images in LFW with different face rotations (the rotation angle is absolute value)

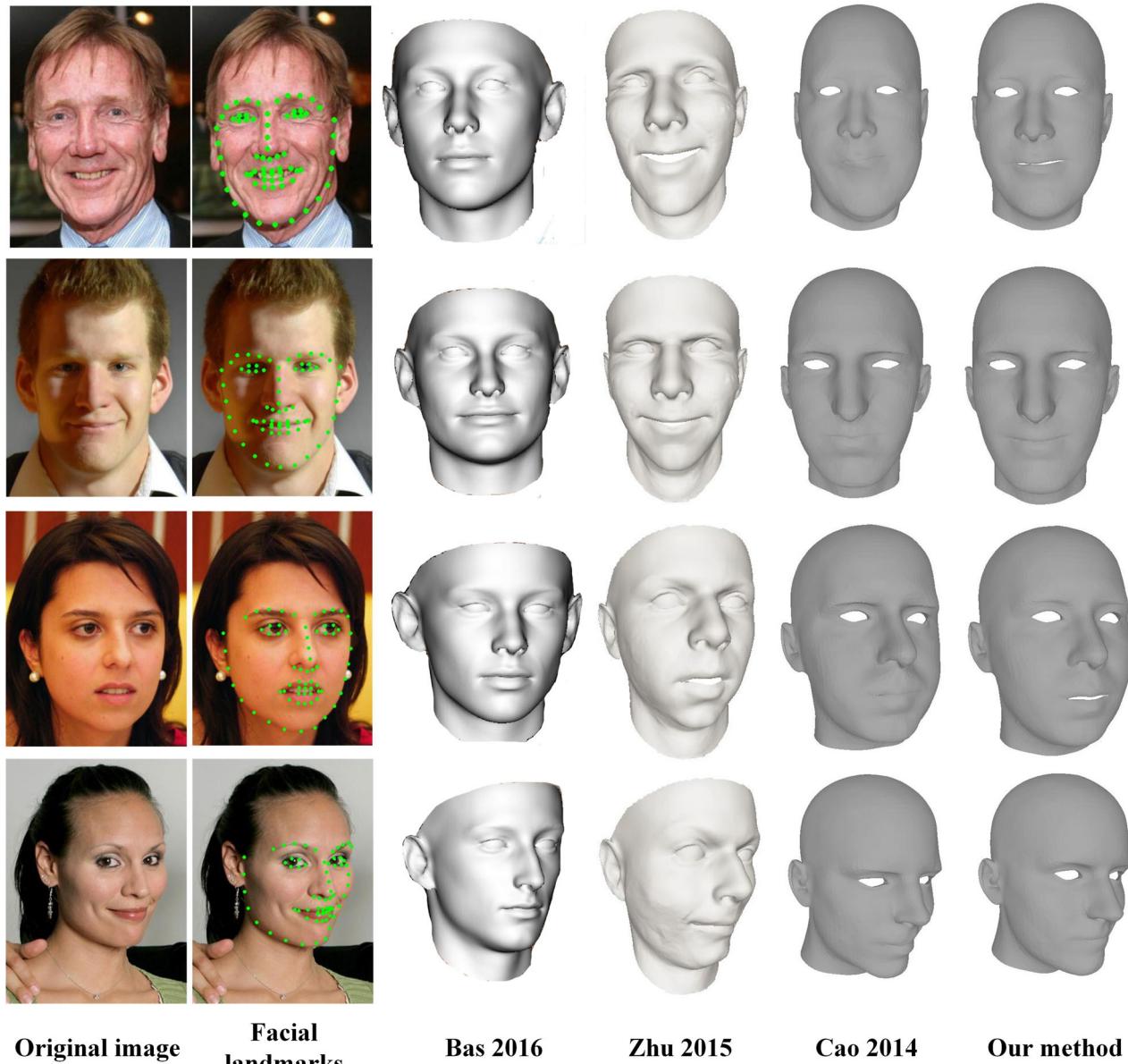
Method	Procrustes analysis error (LFW)					Mean value
	<10°	10°–25°	25°–35°	35°–45°	>45°	
Bas et al., 2016 <sup>9</sup>	0.13	0.15	0.16	0.16	0.18	0.16
Zhu et al., 2015 <sup>27</sup>	0.13	<b>0.13</b>	0.14	<b>0.14</b>	0.15	<b>0.13</b>
Cao et al., 2014 <sup>14</sup>	0.14	0.14	0.15	0.16	0.16	0.15
Our method	<b>0.12</b>	<b>0.13</b>	<b>0.13</b>	<b>0.14</b>	<b>0.14</b>	<b>0.13</b>

Procrustes analysis error (Helen)						
Method	Head rotation					Mean value
	<10°	10°–25°	25°–35°	35°–45°	>45°	
Bas et al., 2016 <sup>9</sup>	0.13	0.15	0.15	0.16	0.18	0.15
Zhu et al., 2015 <sup>27</sup>	<b>0.11</b>	0.13	<b>0.13</b>	<b>0.14</b>	0.16	0.13
Cao et al., 2014 <sup>14</sup>	0.12	0.14	0.14	0.15	0.15	0.14
Our method	<b>0.11</b>	<b>0.12</b>	<b>0.13</b>	<b>0.14</b>	<b>0.15</b>	<b>0.12</b>

Method	LFW1 (s)	LFW2 (s)	Helen1 (s)	Helen2 (s)
Bas et al., 2016 <sup>9</sup>	62	67	66	71
Zhu et al., 2015 <sup>27</sup>	11	14	13	16
Cao et al., 2014 <sup>14</sup>	45	48	52	57
Our method	<b>9</b>	<b>12</b>	<b>12</b>	<b>14</b>

**TABLE 3** Procrustes analysis error by different methods for images in Helen with different face rotations (the rotation angle is absolute value)

**TABLE 4** Face modeling average time in different facial data sets by different methods

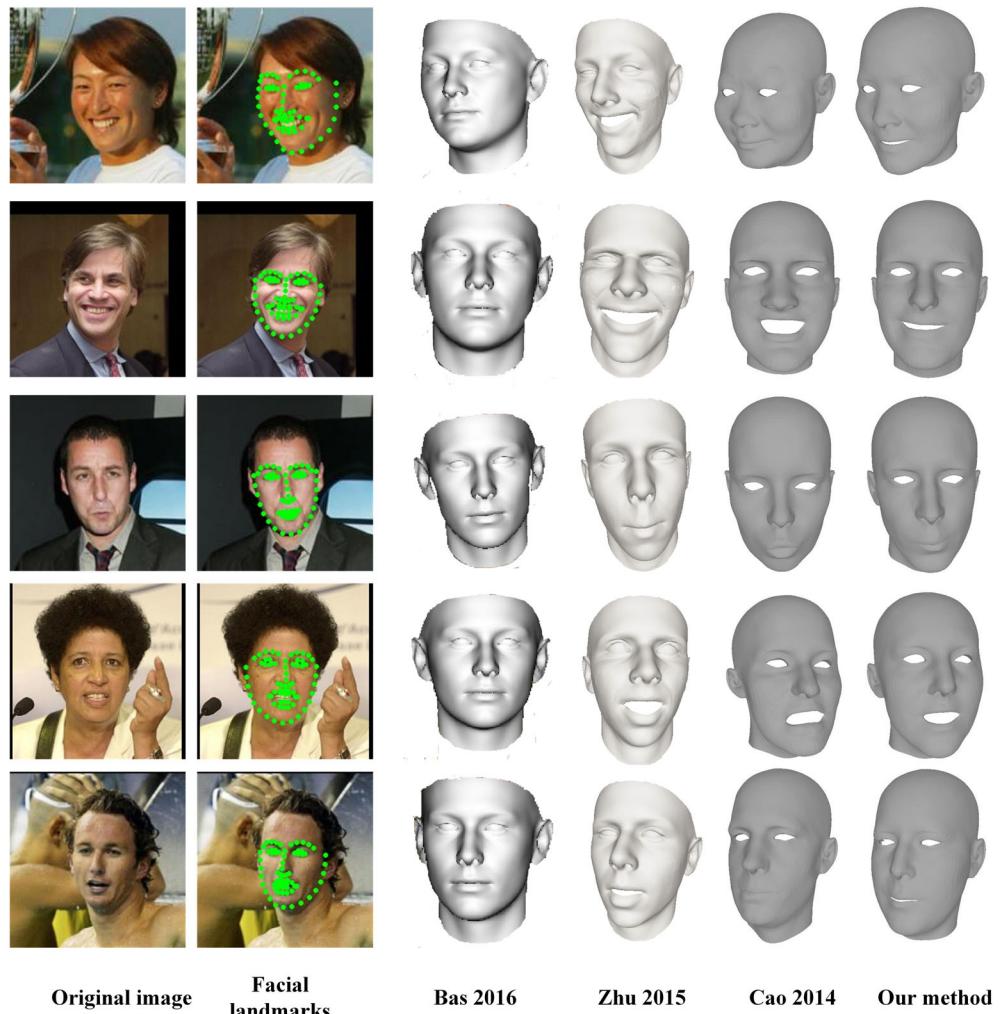


**FIGURE 11** 3D face modeling results by different methods in Helen

independent steps in each iterative steps. For *GPDS* optimization, each middle DLM can be computed independently. Such property of our method provide sufficient conditions to employ a parallel computing structure to accelerate the modeling speed.

In Table 4, we show the face modeling average time in different facial data sets by different methods. The facial data set LFW1 includes face images from LFW, which have small head poses ( $<35^\circ$ ). The LFW2 includes face images with large head poses ( $\geq 35^\circ$ ). The construction conditions of data sets, Helen1 and Helen2, are same as to LFW1 and LFW2. The head pose estimations in LFW and Helen are processed by face detection tools (<https://github.com/ShiqiYu/libfacedetection>).

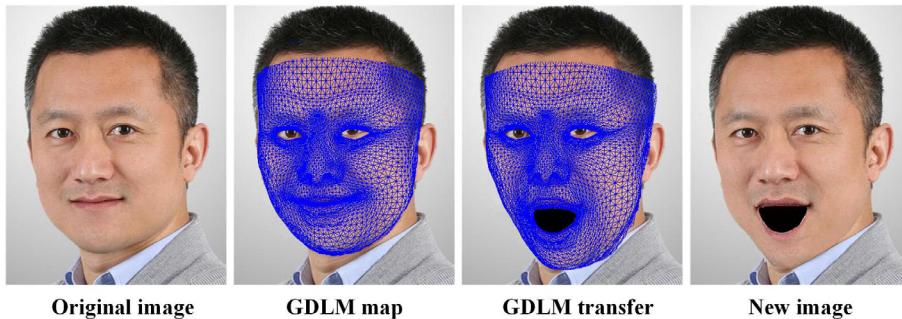
In Figures 11 and 12, we show some face modeling instances from facial images in Helen and LFW. In summary, the results show that our modeling method can achieve similar modeling result from a small facial data set (50\*20) to other classical modeling method. In Figure 13, we show some instances of 3D face modeling results which map the 3D model into facial image.



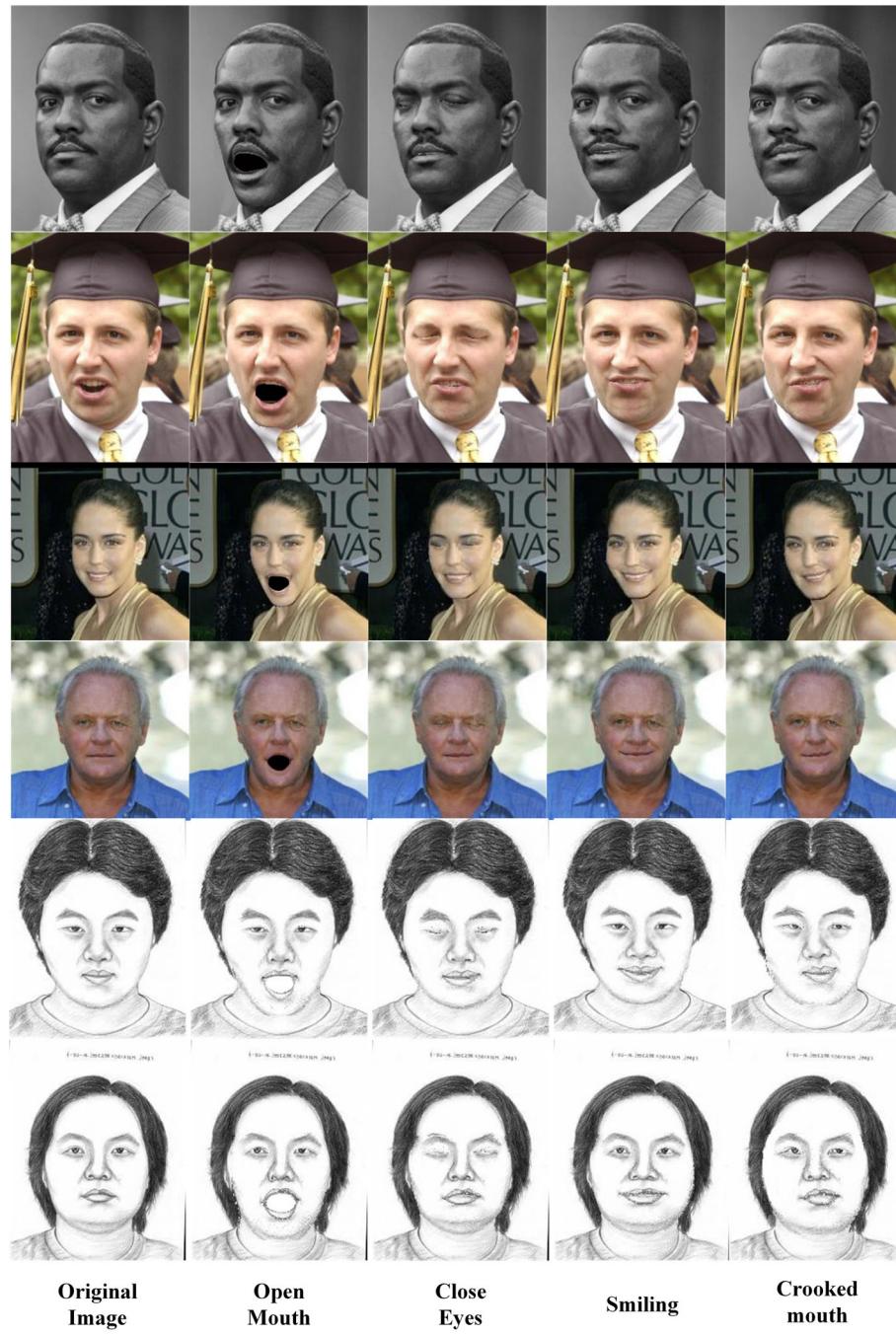
**FIGURE 12** 3D face modeling results by different methods in LFW



**FIGURE 13** Instances of 3D face model mapping results by our method



**FIGURE 14** Instance for facial expression editing by GDLM mapping



**FIGURE 15** Facial expression editing result in facial images

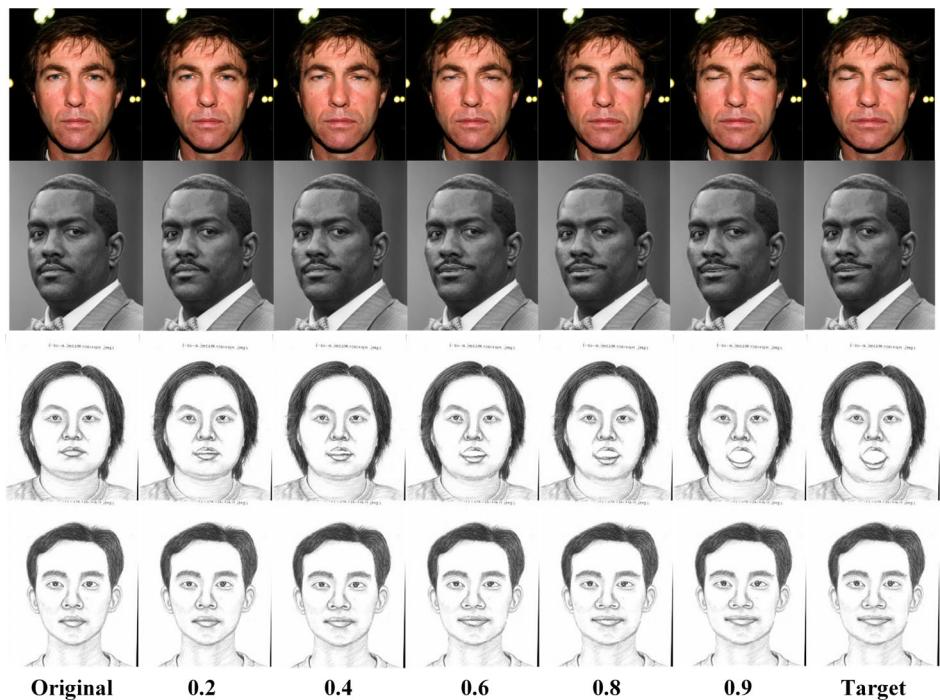
## 5.2 | Application to face expression editing

Based on our face modeling method, we propose related applications for use in face expression editing. After the modeling process, we obtain a GDLM that matches the target facial image. In Figure 14, we map the GDLM onto the facial image.

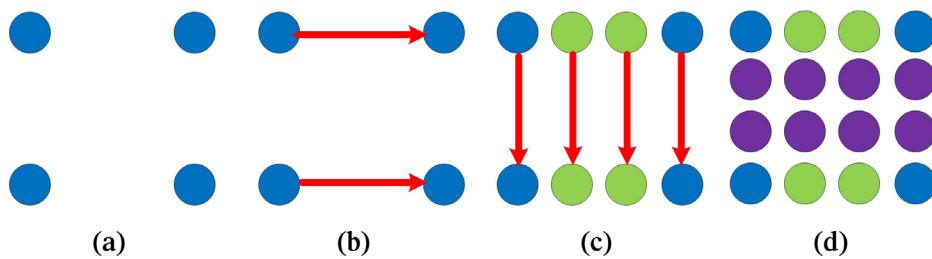
Each pixel in the facial image can be mapped onto a point in GDLM, which can be regarded as the texture information. If we change the expression attribute in the GDLM from the 3D facial database, and map the texture back to the image, the new facial image with a new expression is achieved. To change the expression attribute, we just change the record of  $List_{expression}$  in the final iteration of  $GPSD$ , and the expression is changed to reconstruct the GDLM. In Figure 15, we show some facial expression editing results of facial images and sketches from LFW, Helen, and CUFS.

## 5.3 | Application to face expression interpolation

Based on facial expression editing, we can construct the facial expression interpolation between different facial expressions. We set GDLM  $g_1$  to be the start and GDLM  $g_2$  to be the end. We input  $g_1$  and  $g_2$  into Equation (10), instead of  $D(a)$  and  $D(b)$ . Using a different parameter  $\tau$ , the new GDLMs between  $g_1$  and  $g_2$  are reconstructed, and new facial images with continuous expression changes are achieved. In Figure 16, we show some instances of continuous expression changes from facial images and sketches. We can use a start GDLM ( $gm_0$ ) and several direction GDLMs ( $gm_1, gm_2$ , and  $gm_3$ ) to construct the face expression interpolation in a matrix. The construction method of the face expression interpolations matrix is shown in Figure 17, and two instances are shown in Figure 18.



**FIGURE 16** Instances of continuous expression changes from facial images; from the original image to the target image with different expressions. The facial expression interpolations are generated using parameter  $\tau$  in equation (10).



**FIGURE 17** Construction of facial expression transfer matrix. (a) The blue points represent four GDLMs with different expressions. (b) Two continuous expressions are reconstructed and follow two directions. The red line represents the expression transfer from one to another. (c) The green points represent new GDLMs. (d) The expression for the matrix is finally obtained



**FIGURE 18** Instances of facial expression transfer matrix. The facial images with blue borders are corresponding to the blue points in Figure 17

## 6 | CONCLUSIONS

We propose a novel 3D face modeling method from a single image based on discrete shape space. In our modeling method, various 3D faces with new attributes (such as a facial expression transfer matrix) can be constructed from a small face sample set. The *GPSD* supports many new interpolations generated between different faces with different attributes in high precision. The *GPSD* is different to traditional energy optimization methods as, it independently reduces the distance energy following different facial attributes. In each iteration of *GPSD*, the start DLM is updated from the global DLMs in the discrete shape space, and the different attribute information is updated for each step from the global samples. The probability of *GPSD* falling into a local optimum is lower. The synthesis DLM can be used to construct the GDLM. The synthesis process for GDLM does not lose the facial feature details from the original face samples. In the experiments and applications, our modeling results are more accurate and can provide natural facial images for facial expression editing and expression interpolation. The limitation of our method includes two parts: the method depends on facial landmarks' positions and it cannot achieve accurate local feature reconstruction such as whiskers, eyeballs, hair, and wrinkles. In future work, we aim to use deep learning framework to reconstruct the facial details from 2D facial image in modeling process.

## ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their constructive comments. This research was partially supported by the National Key Cooperation between the BRICS of China (No. 2017YFE0100500), National Key R&D Program of China (No. 2017YFB1002604), and Beijing Natural Science Foundation of China (No.4172033). The authors thank the provider (Zhejiang University in China, Kun Zhou) of face database Facewarehouse and the provider (University of Basel) of 3DMM. They also thank the face detection tools provider (Shenzhen University in China, Shiqing Yu).

## ORCID

Dan Zhang  <https://orcid.org/0000-0001-5676-0656>

Na Liu  <https://orcid.org/0000-0002-4572-4155>

## REFERENCES

1. Decarlo D, Metaxas D, Stone M. An anthropometric face model using variational techniques. Proceedings of the Conference on Computer Graphics and Interactive Techniques, Orlando, Florida; 1998. p. 67–74.
2. Lee KS, Wong KH, Or SH, Fung YF. 3d face modeling from perspective-views and contour-based generic-model. Real-Time Imag. 2001;7(2):173–182.
3. Ansari AN, Abdel-Mottaleb M. Automatic facial feature extraction and 3d face modeling using two orthogonal views with application to 3d face recognition. Pattern Recogn. 2005;38(12):2549–2563.
4. Kurtek S, Drira H. A comprehensive statistical framework for elastic shape analysis of 3d faces. Comput Graph. 2015;51:52–59.
5. Alashkar T, Amor BB, Daoudi M, Berretti S. A grassmann framework for 4d facial shape analysis. Pattern Recogn. 2016;57(C):21–30.
6. Blanz V. A morphable model for the synthesis of 3d faces. Proceedings of the Conference on Computer Graphics and Interactive Techniques, Los Angeles; 1999. p. 187–194.
7. Pighin F. Synthesizing realistic facial expressions from photographs. ACM SIGGRAPH. New York, NY: ACM SIGGRAPH 2005; 2005; p. 9.
8. Paysan P, Knothe R, Amberg B, Romdhani S, Vetter T. A 3d face model for pose and illumination invariant face recognition. Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance, Genova, Italy; 2009. p. 296–301.
9. Bas A, Smith WAP, Bolkart T, Wuhrer S. Fitting a 3d morphable model to edges: A comparison between hard and soft correspondences. Proceedings of the Asian Conference on Computer Vision. Cham: Springer; 2016. p. 377–391.
10. Lüthi M, Jud C, Gerig T, Vetter T. Gaussian process morphable models. IEEE Trans Pattern Anal Mach Intell. 2016;40(8):1860–1873.
11. Patel A, Smith WAP. 3d morphable face models revisited. Proceedings of the 2009 CVPR 2009. IEEE Conference on Computer Vision and Pattern Recognition, Miami, Florida; 2009. p. 1327–1334.
12. Lu X, Jain AK. Deformation modeling for robust 3d face matching. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York; 2006. p. 1377–1383.
13. Ichim AE, Bouaziz S, Pauly M. Dynamic 3d avatar creation from hand-held video input. ACM Trans Graph. 2015;34(4):1–14.
14. Cao C, Weng Y, Zhou S, Tong Y, Zhou K. Facewarehouse: A 3d facial expression database for visual computing. IEEE Trans Vis Comput Graph. 2014a;20(3):1.
15. Jin H, Wang X, Zhong Z, Hua J. Robust 3d face modeling and reconstruction from frontal and side images. Comput Aid Geometr Des. 2017;50:1–13.
16. Guo Y, Zhang J, Cai J, Jiang B, Zheng J. Cnn-based real-time dense face reconstruction with inverse-rendered photo-realistic face images. IEEE Trans Pattern Anal Mach Intell. arXiv:1708.00980. 2019;41(6):1294–1307.

17. Han X, Gao C, Yu Y. Deepsketch2face: A deep learning based sketching system for 3d face and caricature modeling. *ACM Trans Graph.* 2017;36(4):126:1–126:12.
18. Schönborn S, Egger B, Morel-Forster A, Vetter T. Markov chain Monte Carlo for automated face image analysis. *Int J Comput Vis.* 2017; 123:160–183.
19. Joshi P, Wen CT, Desbrun M. Learning controls for blend shape based realistic facial animation. *Proceedings of the SIGGRAPH 2003 Conference on Sketches and Applications: In Conjunction with the Conference on Computer Graphics and Interactive Techniques, 2003, San Diego, CA; 2003, p. 187–192.*
20. Tena JR, Fernando DLT, Matthews I. Interactive region-based linear 3d face models. *ACM SIGGRAPH.* New York: ACM Trans Graph. 2011;30(4):p. 76.
21. Neumann T, Varanasi K, Wenger S, Wacker M, Magnor M, Theobalt C. Sparse localized deformation components. *ACM Trans Graph.* 2013;32(6):1–10.
22. Bradley D, Beeler T. An anatomically-constrained local deformation model for monocular face capture. *ACM Trans Graph.* 2016;35(4):115.
23. Kendall DG. Shape manifolds, procrustean metrics, and complex projective spaces. *Bull Lond Math Soc.* 1984;16(2):81–121.
24. Lee JM. Riemannian manifolds: An introduction to curvature. Berlin, Germany: Springer Science and Business Media, 2006.
25. Kahler M. Shape spaces and shape modelling analysis of planar shapes in a Riemannian framework [Ph.D. thesis]. Durham County, England: Durham University; 2012.
26. Kazemi V, Sullivan J. One millisecond face alignment with an ensemble of regression trees. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus; 2014.* p. 1867–1874.
27. Zhu X, Lei Z, Yan J, Dong Y, Li SZ. High-fidelity pose and expression normalization for face recognition in the wild. *Comput Vis Pattern Recogn.* 2015;1:787–796.
28. Jiang L, Zhang J, Deng B, Li H, Liu L. 3d face reconstruction with geometry details from a single image. *IEEE Trans Image Process.* 2017;27(99):1–1.

## AUTHOR BIOGRAPHIES



**Dan Zhang** is studying for PhD degree in School of Artificial Intelligence, Beijing Normal University (BNU). Her research interests include computer vision, craniofacial morphology, computer graphic and discrete differential geometry.



**Chenlei Lv** is studying for PhD degree in School of Artificial Intelligence, Beijing Normal University (BNU). His research interests include computer vision, 3D biometrics, computer graphic, discrete differential geometry and conformal geometric.



**Na Liu** is studying for PhD degree in School of Artificial Intelligence, Beijing Normal University (BNU). Her current research interests include computer graphic, artificial intelligence and machine learning.



**Zhongke Wu** is Full Professor in School of Artificial Intelligence, Beijing Normal University (BNU), China. Currently he is the member of Steering Committee for Professional Teaching of Animation, digital media in Colleges and universities of Ministry of Education, China and member of CCF CAD and Graphics and CCF Human Computer Interaction. He led and took part in various research and development projects in computer graphics and related areas. Prof. WUs current research interests include computer graphics, animation virtual reality, geometric modeling, volume graphics and medical imaging.



**Xingce Wang** is a Full Professor in School of Artificial Intelligence, Beijing Normal University, PR China. She is major in the 3D modeling and 3D visualization. Her current research interests include computer graphic, medical imaging, artificial intelligence and Machine learning.

**How to cite this article:** Zhang D, Lv C, Liu N, Wu Z, Wang X. 3D face modeling from single image based on discrete shape space. *Comput Anim Virtual Worlds*. 2020;31:e1943. <https://doi.org/10.1002/cav.1943>