ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/333007308

3D Facial Expression Modeling based on facial landmarks in single image

Article *in* Neurocomputing · May 2019

DOI: 10.1016/j.neucom.2019.04.050

CITATION
1

READS
55

4 authors, including:

Lv Chenlei
Nanyang Technological University

14 PUBLICATIONS **8** CITATIONS

SEE PROFILE

zk wu
Beijing Normal University

162 PUBLICATIONS **732** CITATIONS

SEE PROFILE

Xingce Wang
Beijing Normal University

58 PUBLICATIONS **199** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Computational Forensic View project

Project Image process View project

All content following this page was uploaded by Lv Chenlei on 08 November 2019.

The user has requested enhancement of the downloaded file.

Neurocomputing 355 (2019) 155–167

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

3D facial expression modeling based on facial landmarks in single image

Chenlei Lv[a,b], Zhongke Wu[a,b,*], Xingce Wang[a,b], Mingquan Zhou[a,b]

[a] College of Information Science and Technology, Beijing Normal University, Beijing, China
[b] Engineering Research Center of Virtual Reality and Applications, Ministry of Education, Beijing Key Laboratory of Digital Preservation and Virtual Reality for Cultural Heritage, Beijing Normal University, Beijing 100875, China

ARTICLE INFO

Article history:
Received 26 July 2018
Revised 15 March 2019
Accepted 30 April 2019
Available online 10 May 2019

Communicated by Su-Jing Wang

Keywords:
Facial expression modeling
Kendall shape space
Head poses

ABSTRACT

Facial expression modeling is important for many applications such as human emotional analysis and facial animation. Generally, facial expression modeling from single 2D facial image is difficult. Different head poses and scales of facial data in images affect the accuracy of the modeling results. We propose a new 3D facial expression modeling method which is based on facial landmarks from single image. Using the facial landmarks, expression modeling can be processed in Kendall shape space. The Kendall shape space is mathematic space, the facial expression modeling process in Kendall shape space can be regarded as a geodesic path search between different faces. The modeling result is more accurate. The 3D facial expression modeling result is convenient to obtain from 2D facial image with different head poses. In experiments, we show the 3D facial expression modeling performance by our method, which include expression editing and evaluation in public facial database: JAFFE, LFW, Helen and RAF-DB.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

Facial expression modeling has been researched for many years. The related technologies are used in many applications such as face recognition, human emotional analysis, face track and 3D facial animation. Facial expression modeling constructs a standard form for different facial expressions. It provides the necessary tools for facial expression quantitative analysis and synthesis. For 2D facial image, the facial expression modeling is difficult. The reason is the facial image have different head poses and scales in generally. In facial image, the geometric features are incomplete. To remove the influence of head poses and scales, a 3D facial expression counterpart should be reconstructed from 2D facial image. In 3D scene, the facial data can be aligned by facial landmarks and the facial expression modeling result can be achieved. In summary, the core problem of facial expression modeling is constructing the 3D facial expression counterpart from a 2D image.

The traditional methods of 3D facial expression counterpart reconstruction are based on facial landmarks fitting by energy function [1,2]. The head poses and scales are considered in the energy function. Using different energy optimization methods (least square, gradient descent), a 3D facial expression counterpart can be computed from a 2D facial image. However, such methods have significant drawbacks. Firstly, the computation cost of the methods is huge. The fitting process includes four degrees of freedom: identity, expression, head pose and scale. It increases the complexity of energy function. Secondly, the modeling result is limited by the local optimum. The complexity energy function increases probability of falling into local optimum.

We propose a 3D facial expression modeling method based on Kendall theory. The method includes three steps: 1.We extract the facial landmarks from a facial image to construct a 2D facial landmarks based model; 2.We build Kendall shape space based on 2D facial landmarks models with different poses of Facewarehouse (Facewarehouse is a classical 3D facial expression database); 3.We input the 2D facial landmarks model as source into Kendall shape space and search the target 3D facial landmarks model of the database. Using the target 3D facial landmarks model, we can recover the target 3D facial data. Compared to traditional methods, our method has two advantages. Firstly, the computation cost is lower. In Kendall shape space, the face data is represented by a standard form, the head poses and scale are removed to a certain extent. The fitting computation is reduced obviously. Secondly, the fitting process in Kendall shape space can be regarded as a

* Corresponding author at: College of Information Science and Technology, Beijing Normal University, Beijing, China.
E-mail addresses: chenleilv@mail.bnu.edu.cn (C. Lv), zwu@bnu.edu.cn, zhongkewu@gmail.com (Z. Wu), wangxingce@bnu.edu.cn (X. Wang), mqzhou@bnu.edu.cn (M. Zhou).

https://doi.org/10.1016/j.neucom.2019.04.050
0925-2312/© 2019 Elsevier B.V. All rights reserved.

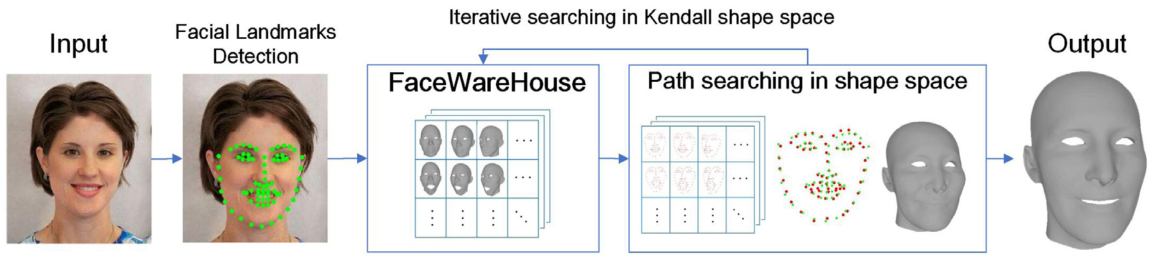

Fig. 1. The pipeline of our facial expression modeling method.

geodesic path search between different faces. It is a global optimal scheme with shape space restriction. The result is more accurate and stable. In Fig. 1, we show the pipeline of our method.

In summary, our main contributions are as follows:

(1) We propose a 3D facial expression modeling framework from 2D facial image to 3D model. The modeling result remains the original 2D facial information includes head pose and expression.
(2) We propose a facial expression synthesis method in Kendall shape space. The method is called path search in Kendall shape space (PSKS). Using the method, we can achieve an accurate and stable 3D fitting face from 2D facial image which is not limited by local optimum.
(3) We propose a facial expression representation which is based on facial landmarks. The representation is called discrete landmark model (DLM). It can be used to fitting 2D facial image to 3D facial model in Kendall shape space.

The other parts of our paper are organized as following. In Section 2, we discuss some related works of facial expression modeling. In Section 3, we introduce the fundamentals of shape space. In Section 4, we illustrate the representation of facial expression which is based on facial landmarks. In Section 5, we illustrate the facial expression modeling method in Kendall shape space. In Section 6, we show the facial expression modeling performance of our method.

2. Related works

Facial expression modeling has been researched for many years. The related works can be divided into five parts: 2D face modeling, 3D geometric modeling, 3D static modeling, 3D dynamic modeling and 3D facial expression modeling by deep learning and manifold learning. The methods based on 3D static modeling and 3D dynamic modeling have been used in many applications such as face synthesis and facial animation.

2D face modeling methods construct face model based on 2D image features. The features include face contour, facial feature landmarks and texture. Cootes proposed the classical face modeling methods: Active Shape Model (ASM) [3] and Active Appearance Models (AAM) [4]. The method was used to represent the facial features by facial landmarks. Based on the methods, some improved algorithms [5,6] were proposed for precise facial expression analysis. Huang and Huang [7] proposed a 2D facial modeling method for facial expression recognition. The facial features were represented by facial landmarks and facial contour. Some methods constructed face space to analysis facial images. The face space was based on principal component analysis (PCA). Penev and Sirovich [8] analyzed the performance of the face space and proposed several improvement solutions. Chen and Huang [9] used the similar method to facial expression clustering. Paul and Gavrilova [10] proposed a geometric facial modeling method for face detection. The modeling process was based on face space. Generally, the feature extraction of 2D face modeling methods was convenient to implement. However, the influence of different head poses was difficult to be removed.

3D geometric modeling methods extracted geometric features from multi-facial images or 3D facial data to construct facial model. Decarlo et al. [11] used anthropometric facial features to build facial model. Lee et al. [12] constructed facial model by multi-views of facial data. To combine the facial images, the facial contours were used to achieve the optimization result. Ansari and Mohamed [13] proposed an automatic facial modeling method by two orthogonal views of facial data. The method was employed for 3D face recognition. Wang et al. [14] proposed a facial expression analysis method by surface geometric features analysis. Soyel and Demirel [15] proposed 3D distance-vector features for facial expression recognition. The features were based on facial features measurement. Peng et al. [16] propose a face modeling method which is based on surface B-Spline reconstruction. Zhan et al. [17] propose a 3D face modeling method which is based on Sparse Iterative Closest Point. Zhenbo et al. [18] provided a facial expression recognition method which was based on multi-level appearance features. Wu et al. [19] proposed a multi-view facial expression feature coding method based on multi-layer feature model. Such methods constructed facial features from multi-facial images or 3D facial data directly. Although the influence of head pose was considered in facial modeling, it was difficult to achieve facial representation with accurate facial expression.

3D static modeling methods have been researched for many years. The basic idea of the method was constructing a static facial feature space. In the space, the face modeling process was transferred to a linear optimization problem. Blanz and Vetter [20] proposed the classical facial modeling method: 3D morphable model (3DMM). 3DMM was used to reconstruct 3D facial data with special characteristics such as gender and age. Blanz and Vetter [21] used the same method for 3D face recognition. Paysan et al. [22] improved the 3DMM for pose and illumination invariant face recognition. Mena-Chalco [23] proposed the similar method for face modeling with different expressions. Pighin et al. [24] proposed facial photographs regenerate method by 3D shape morphing. Bas et al. [1] proposed a 3DMM fitting method which was considering the edge information. The method improved the image fitting speed. Booth et al. [25] proposed a framework for 3DMM with large scale. Duan et al. [26] proposed a 3D face modeling method based on partial least squares regression. The static modeling methods constructed a face space by facial features analysis. In face space, different faces were transferred to regular representation. It reduced the complexity of face analysis. However, the performance of the methods were limited by the facial samples. For facial expression modeling, the method required more computation.

3D dynamic modeling methods constructed facial model based on various facial samples with different expressions. Lu and Jain [27] proposed a deformation modeling method for 3D face matching. Ichim et al. [28] proposed 3D facial modeling method in mobile phone. The reconstruction 3D facial data included texture and wrinkle. Jin et al. [29] used the facial frontal and side images to reconstruct the high-fidelity 3D facial model. Some methods proposed a core tensor space to reduce the computation of facial reconstruction. Vlasic et al. [30] proposed the multilinear models for facial expression modeling. Mpiperis et al. [31] proposed the facial expression recognition methods by the same framework. Cao et al. [2] proposed a facial expression database (Facewarehouse) which was constructed the multilinear models with facial surface fitting. The framework was used in face tracking and facial animation [32]. Gao and Tian [33] proposed a face modeling method for face recognition, which was based on tensor subspace analysis. Song et al. [34] developed a face recognition method based on local tensor model. Generally, the reconstruction process of such methods required optimize an energy equation. The local optimization also limited the quality of facial modeling.

Deep learning & Manifold learning for face expression modeling were researched in recent years. The methods based on deep learning framework used various facial features in facial modeling process. Elaiwat et al. [35] proposed a Restricted Boltzmann Machines (RBM) based method for facial expression analysis. Zhang et al. [36] proposed a multi-facial images Deep Neural Network (DNN) for facial expression recognition. Lopes et al. [37] proposed a facial expression samples extend method based on Convolutional Neural Networks (CNN). Huibin [38] constructed Deep Fusion Convolutional Neural Networks (CNN) for multi-model based facial expression analysis. The methods based on manifold framework were reconstructing facial model under the restriction of the manifold. Brunton et al. [39] introduced the shape space for facial data analysis. Kurtek and Drira [40] used the elastic measure in shape space to analysis different 3D faces. Alashkar et al. [41] mapped the 3D facial data flow into Grassmann manifold for facial recognition. Patel and Smith [42] combined static facial modeling and manifold learning framework for facial modeling. The deep learning frameworks could achieved better facial expression recognition results. However, for 3D facial expression modeling, the frameworks were too complicated. In our method, we construct the facial expression modeling in a Kendall shape space which is a manifold based on discrete points. In following parts, we will discuss the details of our method.

3. Fundamentals

Our method is based on the theory of Kendall shape space [43]. Shape is an important concept to describe the geometric information of object, especially for facial expression models which have complicated external form and contours. Comparing the similarity between different shapes is a challenging task. The shape should be invariant in some symmetric transformations such as scaling, translation and rotation. For mathematic descriptions of shape, there is a theory which is called shape space. Shape space is a quotient space of isometric Lie group actions. Different symmetric transformations such as scaling, translation and rotation are regarded as a Lie group acting smoothly on a manifold. Defining a metric in the manifold by geodesic distance, the similarity of shapes can be computed and the influence of symmetric transformations can be removed.

$$M/G = \{[p] | p \in M\} \tag{1}$$

$$d_{M/G}([p],[q]) = \inf_{g \in G} d_M(p, g.q) \tag{2}$$

$$G.D \mapsto (b, a, O; D) \mapsto aO.D + b.1_k \tag{3}$$

In Eq. (1), G is a Lie group acting smoothly on a manifold M. For p in M, the orbit of p is defined as $[p]$. In generally, the metric in quotient space is hardly achieved, it should be computed indirectly. In Eq. (2), the metric in quotient space is achieved by optimization searching from group action in M. In Eq. (3), we provide the group action G which includes translation b, scaling a, rotation O. $1k = (1, 1) \in R^{1\times k}.g \in G, q \in D$.

A classical method to construct a shape space through discrete points is called Kendall shape space. Kendall shape space provides statistical shape analysis tools to measure the similarity between different shapes. The discrete points' distribution can be regarded as a shape of the points' set. Using Kendall theory, the facial landmarks sets, as the discrete points' sets, can be compared.

$$M = R^{m\times k}\backslash\{0\}, D \in M, D = (x_{(1)}, \dots, x_{(k)}) \tag{4}$$

In Eq. (4), Kendall shape space is defined. The k is the number of points in a shape. M is the manifold and the dimension of M is $m \times k$. D is the discrete point sets. To construct the Kendall shape space, the three symmetric transformations should be removed. We define two discrete point sets $D(a)$ and $D(b)$. In Kendall shape space, the shape measurement of the $D(a)$ and $D(b)$ can be represented by Eq. (5).

$$d_{\text{Kendall}}(D(a), D(b)) = \inf_{g \in G} d_M(D(a), g.D(b)) \tag{5}$$

For discrete point set, removing scaling and translation just requires simple computation. The centroid of discrete point set is computed. The influence of scaling and transfer can be removed by centroid alignment and scaling normalization. In Eq. (6), we show the process.

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

Using the result of Eq. (6), we can rewrite the Eq. (5) to Eq. (7). O represents the rotation group. S_m^k represents the pre-shape space of Kendall shape space.

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

To remove the influence of rotation, there have several methods. In Kendall shape space, the process can be transferred to a singular value decomposition problem. The computation can be interpreted as the alignment of eigenvectors in matrix. In Eq. (8), we show the computation. Z is orthogonality of D_s.

$$\begin{aligned} & d_{S_m^k}(Z(a), O.Z(b)) = arccos\text{tr}(\text{R}\Lambda) \\ & \quad Z(b)Z(a)^t = U\Lambda V, \text{R} \in SO(m) \end{aligned} \tag{8}$$

In Kendall theory, the orthogonality is optimally registered for the rotation group O. Combining the different equations, we achieve the final computation for geodesic distance representation in Eq. (9).

$$d_{Kendall}(D(a), D(b)) = inf d_{S_m^k}(Z(a), O.Z(b)) = arccos tr(\Lambda) \tag{9}$$

Using the geodesic distance, we can compute the similarity of different shapes with discrete points. Following the geodesic path in the space, the new shape can be generated. In Eq. (10), we show the new shape in the geodesic path between two shapes.

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

In our method, we transfer the facial data to a discrete points based representation. Then we can use the tools of Kendall shape space to measure different faces and generate new face. The facial expression modeling result can be achieved from facial image by geodesic path search in the space.

4. Facial expression representation

In Kendall theory, the shape is represented by discrete points. To use the relevant tools of Kendall theory in facial expression modeling, we should find a facial expression representation with register facial landmarks. We propose a discrete landmarks model (DLM) to represent the facial features. The DLM is based on facial landmarks. Based on facial image, the DLM is constructed by 2D points. For 3D facial object, the DLM is constructed by the 3D points. The DLM provides generally facial representation of 2D facial image and 3D facial object. We use the classical facial landmarks detection method [44] to extract the landmarks positions from facial image. The construction process of DLM include three parts: 1. translation remove equation; 2. scaling remove equation; 3. rotation alignment equation. In Eq. (11) to Eq. (13), we show the construction of DLM.

$$\begin{aligned} &F = \{x_1, \ldots, x_k\}, B \in F \\ &L = \{v_1, \ldots, v_k\}, v_i = x_i - B \end{aligned} \tag{11}$$

In Eq. (11), F is the set of the facial landmarks, B is the benchmark of F. The benchmark is used to uniform location of the landmarks from different faces. We select the nasal tip from F to be the benchmark. L is the set of landmark vectors which are obtained by the landmarks minus the benchmark. Then we achieve the preliminary landmarks model.

$$\begin{aligned} &s(L) = \left(\sum_{i=1}^{k} \|v_i\|\right) \\ &L_s = \{v_1^{'}, \ldots, v_k^{'}\}, v_i^{'} = \frac{v_i}{s(L)}, \sum_{i=1}^{k} \left\|v_i^{'}\right\| = 1 \end{aligned} \tag{12}$$

In Eq. (12), s is the scale of the L. Different scale of L influences the analysis of facial features relative positions. The landmarks model should have a unique scale. To remove the scaling factor, we normalize L by s. The new landmarks model L_s is constructed. The representation of L_s is consist to the discrete points group in Eq. (6).

In Kendall theory, rotation is removed by the computation in Eqs. (8) and (9). In our method, we use a simple method to instead the process. The DLM is constructed by facial landmarks. The landmarks include semantic information. We can construct a local coordinate system by certain landmarks to align the facial landmarks.

$$L_{sr} = \{T(v_1^{'}), \ldots, T(v_k^{'})\} \tag{13}$$

$$d_{\text{Kendall}}(L_{sr}(a), L_{sr}(b)) = arccos(L_{sr}(a) \cdot L_{sr}(b)) \tag{14}$$

In Eq. (13), we achieve the L_{sr}, which is the final representation of DLM. For the vectors in L_s, the rotation factors are still existing. The influence comes from the different head poses of 3D facial data. We introduce transform function T to remove the rotation from the landmarks model. The function T is a coordinate transformation based on a local coordinate system. 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 computed in Eq. (14), which is transferred from Eq. (9). In Fig. 2, we show the instance of local coordinate system. For 2D facial image, the DLM representation is regarded as a 2D reflection of 3D facial data. The task of 3D facial expression modeling from 2D facial image can be transferred to DLM reconstruction from the 2D reflection. Using the DLM, we achieve the fitting method between 3D facial model and 2D facial image. For coordinate system construction in 3D face object, the facial landmarks detection results can be achieved by [45].

5. Facial expression modeling

Facial expression modeling is constructing the best fitting result from template facial database to source facial image. Based on the DLM representation, the modeling process can be processed by geodesic path searching in Kendall shape space. The similarity of different DLMs can be computed by Eq. (14), which is employed to represent the energy function to show the modeling process.

The function represents the geodesic distance of the DLMs in Kendall shape space. S_{image} represents the facial landmarks from facial image, $L_{sr}(S_{image})$ represents the DLM of S_{image}. T_r represents the target DLM which is constructed from the DLM tensor of template facial database to fit the source DLM in facial image. To construct DLM tensor from the database, we sign the 68 facial landmarks in the face object manually. The face samples in the database are aligned and the landmarks can be mapped into different samples. Based on the DLM tensor, the modeling process minimizes the distance energy $E_{synthesis}$ to achieve target DLM. In Eq. (15), we show the energy function.

$$E_{synthesis} = d_{\text{Kendall}}(L_{sr}(S_{image}), L_{sr}(T_r)) \tag{15}$$

To achieve the target DLM $L_{sr}(T_r)$ with minimum $E_{synthesis}$, we propose the synthesis algorithm: Path Search in Kendall Shape Space (PSKS). PSKS searches the geodesic path in Kendall shape space to reduce the $E_{synthesis}$ between generate DLM from database and source DLM from facial image. Searching process includes three directions: rotation, expression and identity. The PSKS is following the directions to recover the target DLM to fit source DLM. The target DLM can be generated after the optimization process. The facial model can be reconstructed by the target DLM.

5.1. Path search in Kendall shape space (PSKS)

We introduce the process of PSKS. In Fig. 4, the blue point represents the source DLM. The path search in Kendall shape space(PSKS) is employed for DLM construction. For PSKS, an initialization DLM from DLM tensor should be computed at first. It is used to be start point in Kendall shape space. The DlM from template facial database is a 3D DLM. However, the DLM in facial image is 2D DLM. To compare the two kinds of DLM in Kendall shape space, we rotate each 3D DLM from database to achieve 2D reflection of DLM. Combining the new samples by rotation, the 3D DLM are organized to a 2D DLM tensor, which includes rotation, expression and identity). The initialization DLM is the one with the minimum $E_{synthesis}$ from the tensor. In Fig. 3, we show the tensor with the direction indexes.

Starting from the initialization DLM, we synthesis the new DLMs by PSKS. The basic idea of PSKS is updating the DLM in a certain direction. Through changing the updating direction iteratively, the new DLM is achieved which reduce the distance energy. The Kendall shape space is not a linear space. In Kendall shape space, the new DLM can't be achieved by a linear method. We can only use the tools of Kendall theory to synthesis the target DLM. Following the limit of the searching process in Kendall shape space, we explain the implementation of PSKS which is based on Eqs. (9) and (10). In Fig. 4, we show an instance of PSKS in a certain direction (expression).

The DLM includes three attributes: angle, expression and identity, which are consistent with directions of PSKS. From the DLM tensor, we achieve the initialization DLM which has minimum distance energy. We extract the attributes of initialization DLM. The DLM construction can be divided into different attributes optimization independently. For expression optimization, we extract the candidate expression DLM set from the DLM tensor which have same identity and rotation information of initialization DLM. We call the set to expression DLM set. The DLM in the set have

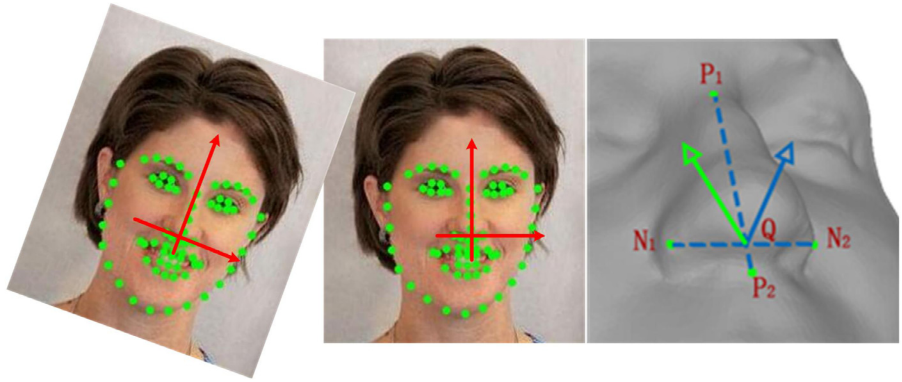

Fig. 2. The instances of local coordinate system by serval facial landmarks. In 3D scene, the points of P_1, P_2, N_1, N_2 are landmarks around the nose area. The point Q is pedal of vector P_1P_2 and Q which is computed by $P_1P_2 \times N_1N_2$. In 2D scene, the local coordinate system can be achieved by the P_1P_2.

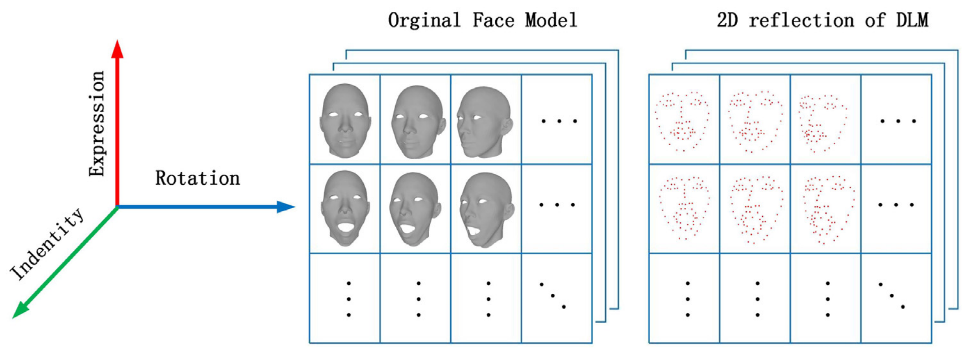

Fig. 3. The tensor of DLMs from Facewarehouse.

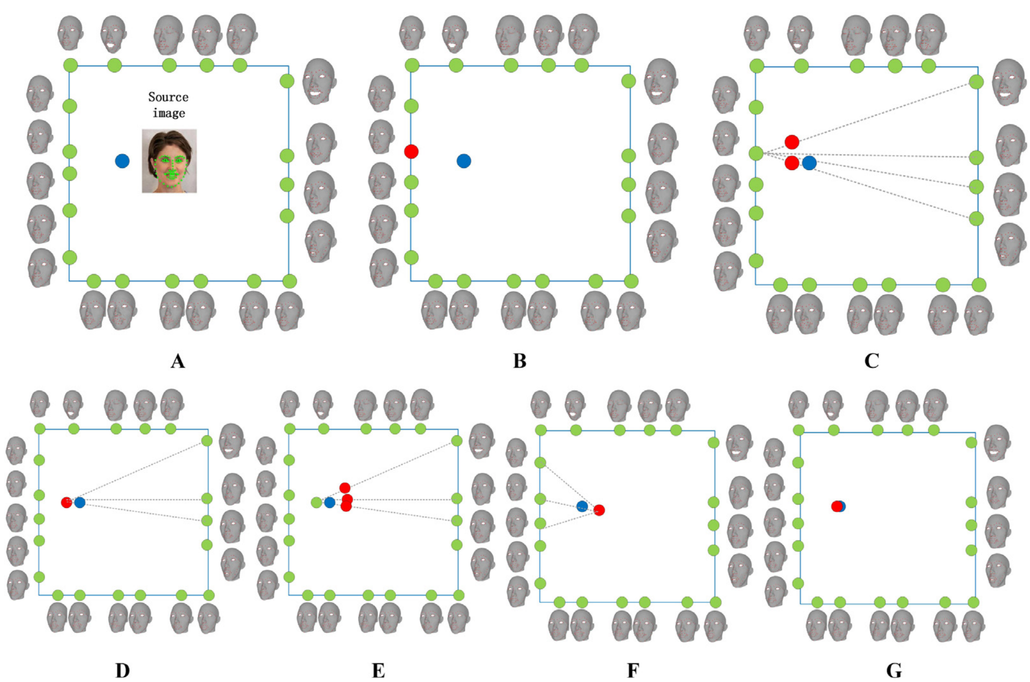

Fig. 4. The instance of expression searching by PSKS in Kendall shape space.

same angle and identity attributes to the initialization DLM. Firstly, we compute the start DLM from the expression DLM set with the minimum distance energy to the source DLM. In Fig. 4B, the red point represents the start DLM. Secondly, we compute a set of new DLMs by geodesic path searching in Kendall shape space. We call the DLM set to temporary expression DLM set. In Fig. 4C, the red points represent temporary expression DLM set. In temporary expression DLM set, each new DLM is generated by Eq. (10). The start DLM represents the $D(a)$, the other DLMs of the set represent the $D(b)$. Then we update the new start DLM from temporary expression DLM set which has minimum distance energy to source DLM. In Fig. 4D, the red point represents the new start DLM. According to the step of k in Eq. (10) and iterative update start DLM, the final DLM is computed (red point in Fig. 4G). For identity searching, the process of PSKS is the same as expression searching. For angle searching, we rotate the angle of the start DLM to fit the source DLM. Through searching the geodesic path by the three directions iteratively, we achieve the final fitting DLM when the distance energy is converged. Algorithm 1 outlines the process of searching expression direction by PSKS.

Algorithm 1 Expression searching by PSKS.

Require: Initialization DLM $L_{sr}(T_{init})$and source DLM $L_{sr}(S_{image})$.
Extract the attributes angle and identity of initialization DLM.
Extract the candidate expression DLM set.
Set searching parameter step $k = 0.1$.
Set $E_{synthesis} = Equation15(L_{sr}(S_{image}) and L_{sr}(T_{init}))$.
while Search geodesic path from start DLM to DLM set. **do**
  Temporary expression DLM set = Equation9 (start DLM, candidate expression DLM set, k=0.1).
  for do i in Temporary expression DLM set
    $E_i = Equation15(L_{sr}(S_{image}), L_{sr}(T_i))$.
    Select the minimum E_i.
    Select the corresponding DLM $L_{sr}(T_i)$.
    $List_{identity}$ = GPSD(identity DLMs);
    if (**then**$E_i < E_{synthesis}$)
      $E_{synthesis} = E_i$.
      Start DLM = $L_{sr}(T_i)$.
      Record $L_{sr}(T_i)$ into path searching list
    else
      Out
    end if
  end for
end while
Ensure: $L_{sr}(T_{init})$ with the path searching list

5.2. Model regenerate by DLM

Using the PSKS to reduce the distance energy, we achieve the target DLM with path searching list of 3 directions. However, the DLM is a discrete landmarks model which cannot be used to represent the global face model. The face model should be reconstructed by target DLM and path searching list in Kendall shape space.

The path searching list includes three directions. The process of model regenerate is following different directions. Firstly, we construct the facial model with identity attribute. The path searching list records the searching paths of identity. We select the corresponding facial model in facial database. Then we achieve an facial expression set of face models with the same identity and different expressions by Eq. (10). Secondly, we follow the expression direction of path searching list to generate the facial model from the expression set. The process is similar to PSKS for expression in Fig. 4. Finally, we rotate the facial model by the angle record from the path searching list. Algorithm 2 shows the regenerate process. Combining the PSKS and model regenerate, the final fitting result of facial expression modeling is achieved. The whole process is shown in Algorithm 3. In Fig. 5, we show the process of face model regenerate. We also show the model regenerate result by different iterative steps in Fig. 6.

6. Experiments and application

We evaluate the performance of our facial expression modeling method in four public facial database: JAFFE, LFW, Helen and RAF-DB[46,47]. JAFFE is a facial expression database. It includes 213 Japanese facial images with different expressions. LFW is a classical facial image database. It includes 13233 facial images from 5749 persons. It is widely used in facial landmarks detection and facial recognition. Helen is another facial database which includes 2330 facial images. RAF-DB is a wild facial database which has accurate facial expressions labels. We use the Facewarehouse [2] to be the template facial database. Around the different coordinate axes(X, Y, Z), we achieve 125 DLM reflections for each face model in template facial database. The rotation step is 15 degrees. Combining the new samples by rotation, the face models are organized to a tensor (125*20*150, rotation, expression and identity). The facial expression modeling process of our framework is based on the DLM tensor.

Algorithm 2 Facial model regenerate.

Require: Path searching list.
Extract the list of identity searching record I and first.
Extract the list of expression searching record E.
Extract the angle of rotation.
Set facial expression model set FL
for do$i$ to 20//20 is the number of expressions for single person in tensor.
  Extract the first model in I to init target face model F.
  The expression index is i
  for do$j$ in I
    Extract the model F_j by the expression i, identity j.
    Generate new model F_{new} by Equation $10(F, F_j)$.
    $F = F_{new}$.
  end for
  Put the F into FL.
end for
for do$i$ in E
  Extract the model F_i by index i in FL.
  Generate new model F_{new} by Equation $10(F2, F_i)$.
  $F2 = F_{new}$.
end for
Rotate the $F2$ by the angle
Ensure: $F2$

Algorithm 3 Facial expression modeling process.

Require: facial image S.
Extract the facial landmarks S_{image}.
Construct source DLM $L_{sr}(S_{image})$.
Construct DLM tensor from facewarehouse.
Searching the DLM tensor to achieve the initialization DLM $L_{sr}(T_{init})$.
Initialize the $E_{synthesis}$ by $L_{sr}(S_{image})$ and $L_{sr}(T_{init})$.
while
  do$L_{sr}(T_{expression})$ = Expression Searching($L_{sr}(T_{init}), L_{sr}(S_{image})$).
  $L_{sr}(T_{identity})$ = Identity Searching($L_{sr}(T_{expression}), L_{sr}(S_{image})$).
  $L_{sr}(T_{rotation})$ = Rotation Searching($L_{sr}(T_{expression}), L_{sr}(S_{image})$).
  $L_{sr}(T_{init}) = L_{sr}(T_{rotation})$.
  $E_{temporary}$ = Equation 15($L_{sr}(S_{image})$ and $L_{sr}(T_{init})$).
  if (**then**$E_{temporary} < E_{synthesis}$)
    $E_{synthesis} = E_{temporary}$.
  else
    $L_{sr}(T_r) = L_{sr}(T_{init})$.
    Jump out
  end if
end while
Regenerate (path searching list).
Ensure: the facial expression model.

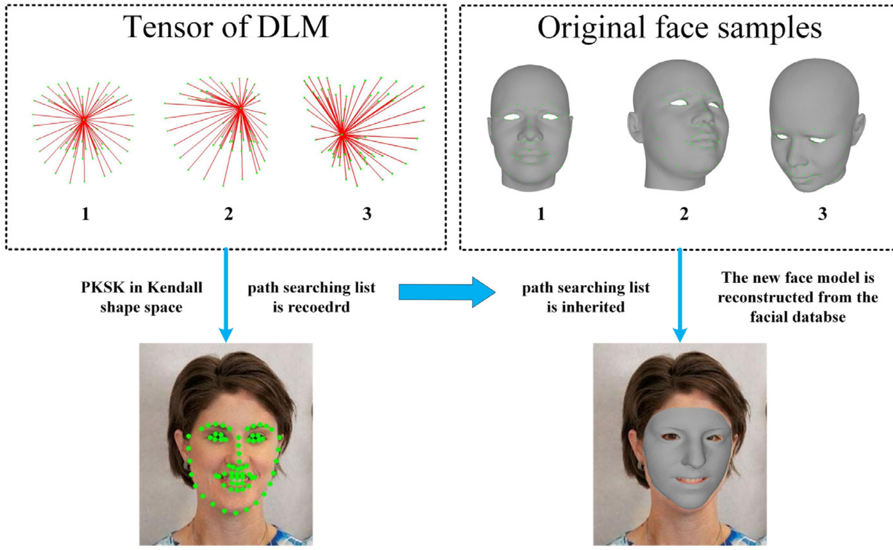

Fig. 5. face model regenerate from target DLM and path searching list.

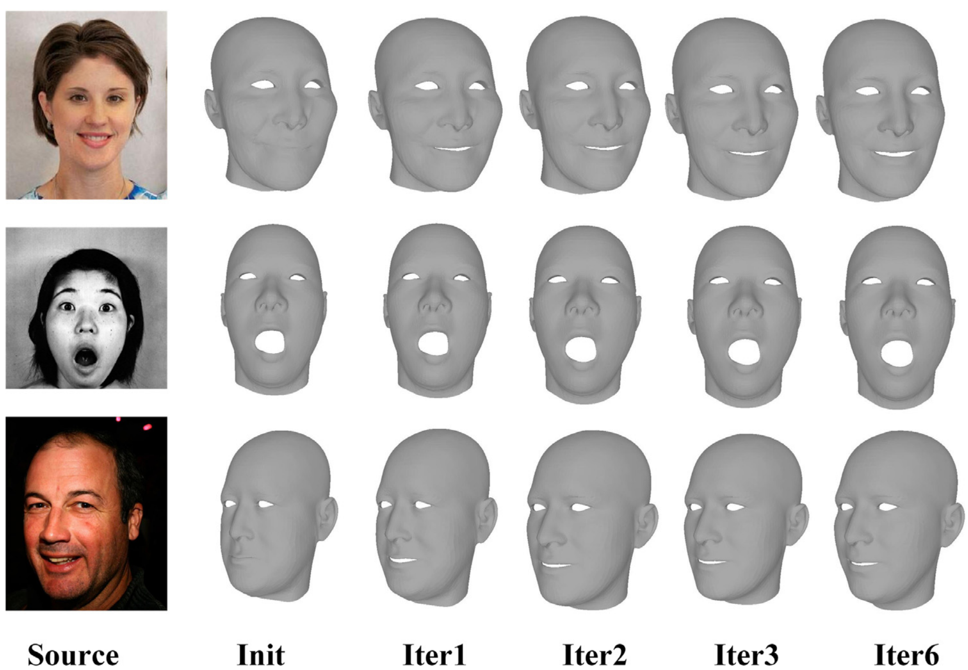

Fig. 6. Facial expressions modeling result by different iterative steps of our method.

6.1. Evaluation for facial landmarks' distortion

In our framework, positions of facial landmarks affect the DLM construction. It cannot guarantee the accuracy of each facial landmark is detected using a detection method. Therefore the evaluation of our face modeling method with some facial landmarks' distortion is important. To evaluate the sensitive of our method for $E_{synthesis}$ optimization, we add two sets (L1 and L2) of random displacements to the landmark subset in LFW; randomly selecting 8 and 16 facial landmarks from different facial regions and adding the randomly displacements($\leq$5mm). One set of the landmarks' movement is signed in Fig. 7A. In Fig. 7B, the $E_{synthesis}$ reduction curve graph is provided for different iteration steps and different landmarks sets. We employ the method [1] to be a competitor which is also based on the facial landmarks to achieve face modeling result.

6.2. Evaluation in JAFFE

The facial images in JAFFE have clearly expression labels. The labels include angry, disgust, fear, neutral, sadness, happiness and surprise. Some expressions with different labels are difficult to distinguish. We reclassify the facial images into 4 new expression labels: Annoy (angry, disgust, fear, and sadness), Happy (happiness), Surprise and Normal (neutral). We build the modeling test set for each label. To evaluate the modeling accuracy, we propose the Kendall error which is computed by the facial landmarks in Eq. (15) which has been discussed in [43]. The Kendall error is the geodesic distance between 2D facial source image DLM and 3D facial DLM. It reflects the shape similarity of the facial expressions. We compare the Kendall error of different modeling method: 3D morph model (3DMM)[1], bilinear model[2] and our modeling method. In Fig. 8, we show some modeling instances by different

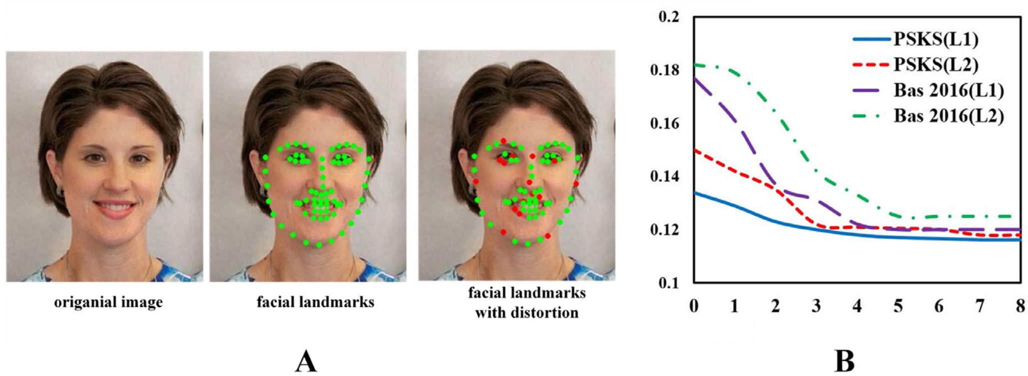

Fig. 7. $E_{synthesis}$ reduction curve graph for two methods with two set of facial landmarks distortions. In B, the longitudinal axis represent the $E_{synthesis}$ value,the transverse axis represent iterative step in optimization process.

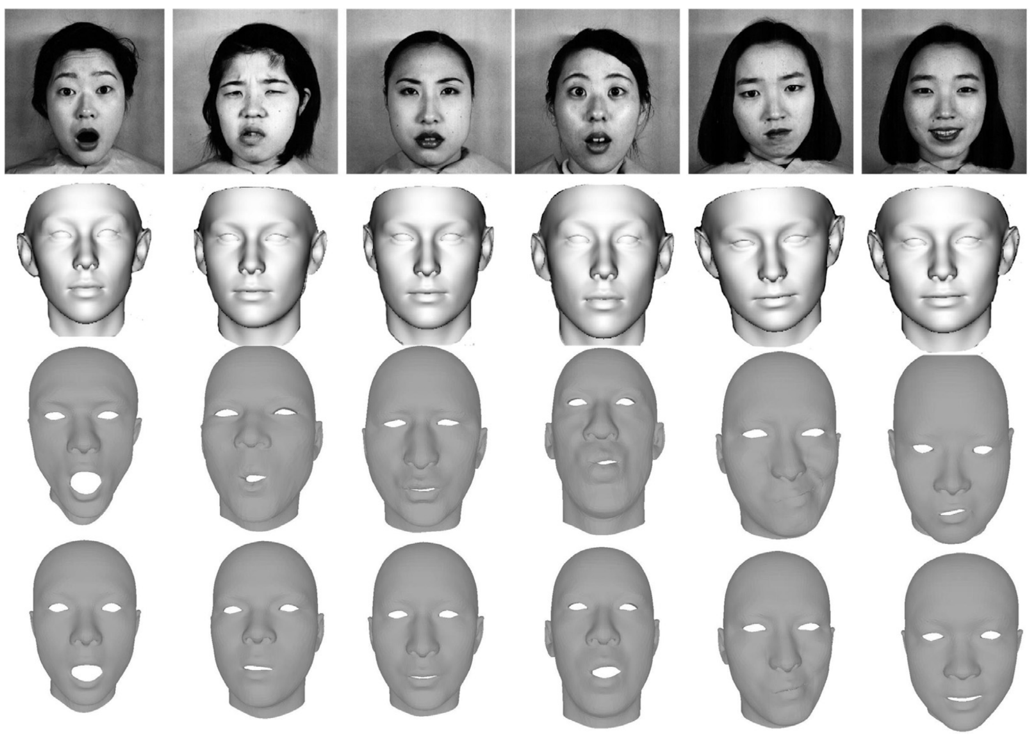

Fig. 8. Facial expression modeling results by different methods in JAFFE. Source facial images are shown in first row. The modeling results by different method are shown in other rows (Row 2: 3D morph model, Row 3: Bilinear model, Row 4: Our method.

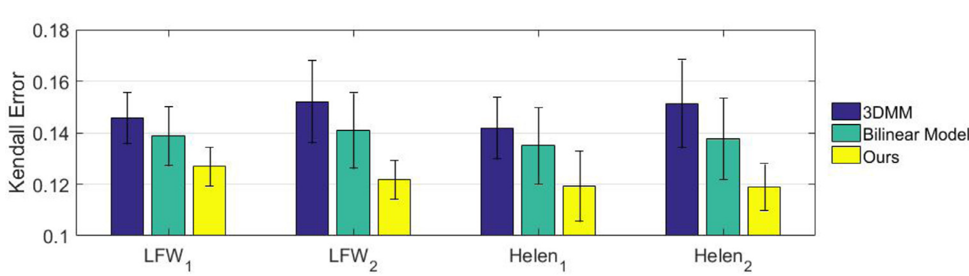

Fig. 9. Kendall error comparisons for different expressions in JAFFE.

method. In Fig. 9, we show the Kendall error bar graph(the Kendall error is introduced by Kendall [43]) by different methods in different expression sets.

6.3. Evaluation in LFW & Helen

The facial images in LFW and Helen are closer to application scene. The databases include different facial expressions and head poses. It takes a big challenge for expression modeling. To evaluate the accuracy of our expression modeling methods in different head poses, we use our modeling methods in LFW and Helen to compute the Kendall error. In order to evaluate the poses robustness of our method further, we build two subsets from LFW and Helen. The facial data in the subsets have obviously head poses (1000 in LFW and 300 in Helen, the rotation of the head pose is bigger than 15 degrees). In Figs. 10 and 11 we show some instances of

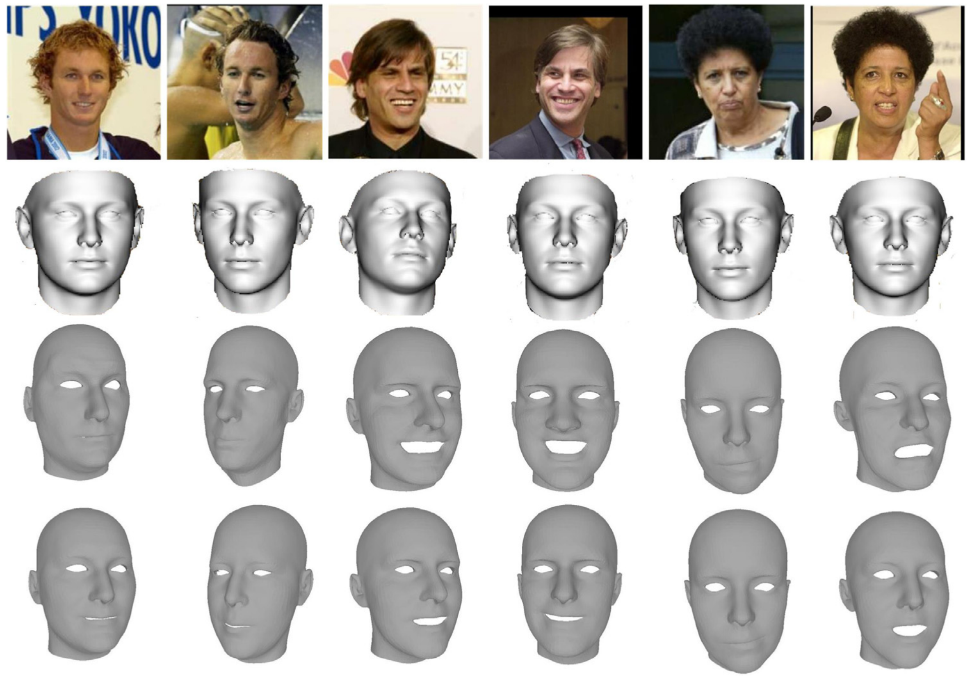

Fig. 10. Facial expression modeling results by different methods in LFW. (Row 2: 3D morph model, Row 3: Bilinear model, Row 4: Our method).

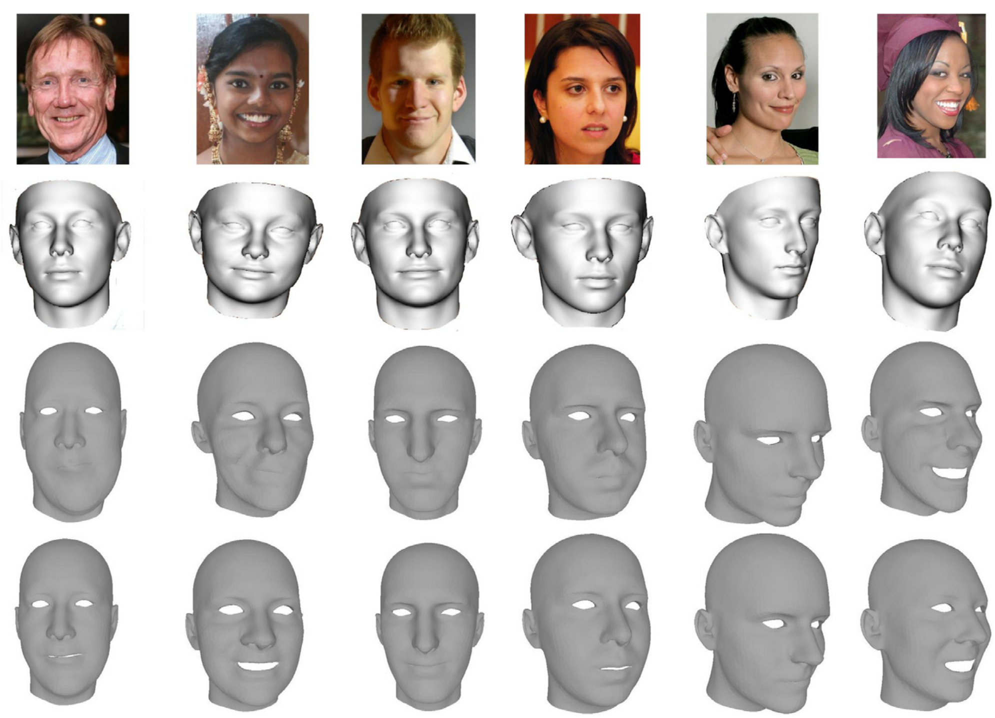

Fig. 11. Facial expression modeling results by different methods in Helen. (Row 2: 3D morph model, Row 3: Bilinear model, Row 4: Our method).

modeling result by different methods. In Fig. 12, we show the Kendall error bar graph by different methods in different facial image sets.

6.4. Evaluation in RAF-DB

Based on facial expression modeling method, the facial expression can be measured to a certain extent. We provide a expression classification evaluation based on test set from RAF-DB. We select a subset from the RAF-DB to be the reference group. According to the modeling result of PSKS, the geodesic path parameters can be recorded. We map the records from different models into the same DLM identity model, which can be regarded as the identity attribute removing. Then we achieve the DLM expression model with same identity attribute from different faces. We select 300 facial images (100 samples * 3 expression labels) to be the reference group. The DLM expression models are computed from the reference group. To achieve the final expression recognition result, we compute the distance between the DLM expression model from input facial image and the models in reference group. The DLM with

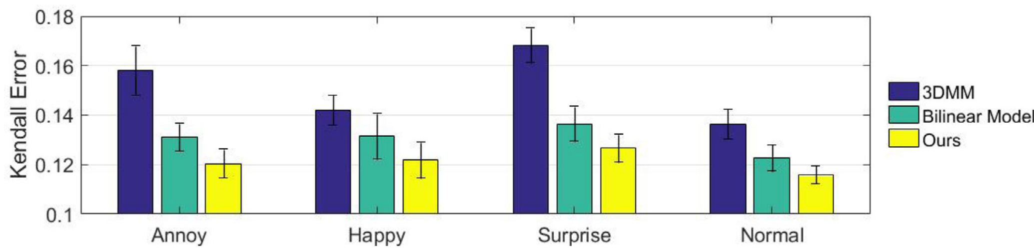

Fig. 12. Kendall error comparisons for different poses in LFW & Helen. (LFW1 and Helen1 are original LFW and Helen databases. LFW2 and Helen2 are the facial data sets with obviously rotation).

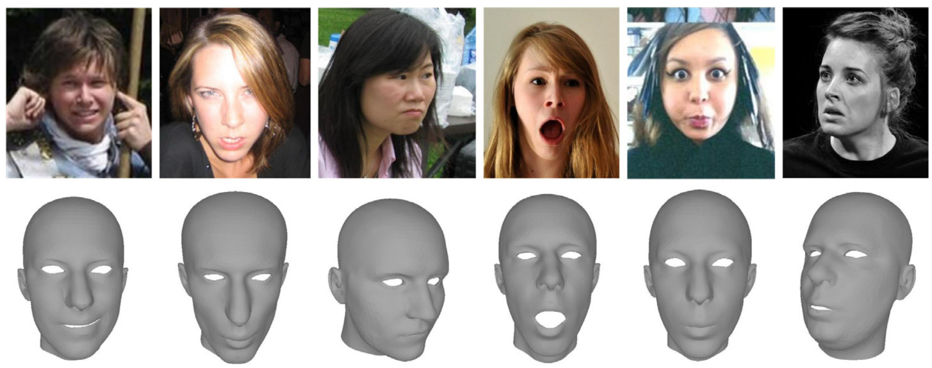

Fig. 13. Facial expression modeling results in RAF-DB.

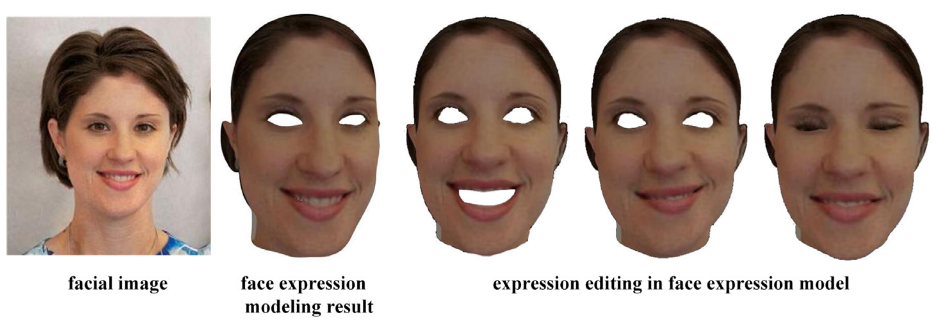

Fig. 14. Expression editing in 3D face expression model.

Table 1
Facial expression classification rate in RAF-DB. The expression label of Negative includes serval labels include: Disgusted and Sad. The expression label of Exaggerated includes serval labels include: Surprised, Fearful and Angry.

Expression Label	Happy	Negative	Exaggerated
RAF-DB	91%	85%	93%

Table 2
Facial expression modeling speed by different methods (Average time cost by second).

Method	JAFFE	LFW	Helen
3DMM	89 s	85 s	103 s
Bilinear Model	45 s	52 s	72 s
Our method	21 s	27 s	33 s

minimum distance is selected and the corresponding expression label is output. In Fig. 13, we show some instances of expression modeling. In Table 1, we show the expression classification rate.

6.5. Application for facial expression editing in image

Based on our modeling method, the facial expression in the image can be edited. In model regenerate process, we have introduced that the facial expression model can be reconstructed by target DLM and path searching list. From the facial image to 3D facial expression model, the pixels can be mapped into the 3D facial surface. Changing the expression record in path searching list, we can achieve new 3D face object with a new expression. In Fig. 14, we show the expression editing result in a 3D face model. The pixels can be updated according to the facial expression editing in 3D face model. Finally, facial expression editing for facial image can be processed. In Fig. 15, we show some facial expression editing results from facial images.

6.6. Summary of results

We compare different methods for facial expression modeling. The 3DMM reconstructs 3D facial model with accurate head poses. However, the method can't reconstruct the accuracy facial expressions. The face space by 3DMM hasn't accurate expression features. Therefore the Kendall error of 3DMM is much higher than other methods. The bilinear model can achieve the better facial expression result which is based on the regular facial expression database. However, the accuracy of the expression modeling is affected by the variety of facial samples in the face database. The linear optimization process of expression regenerate in the database can't achieve the accurate expression reconstruction model. The 3DMM and bilinear model are core spaces which neglect some

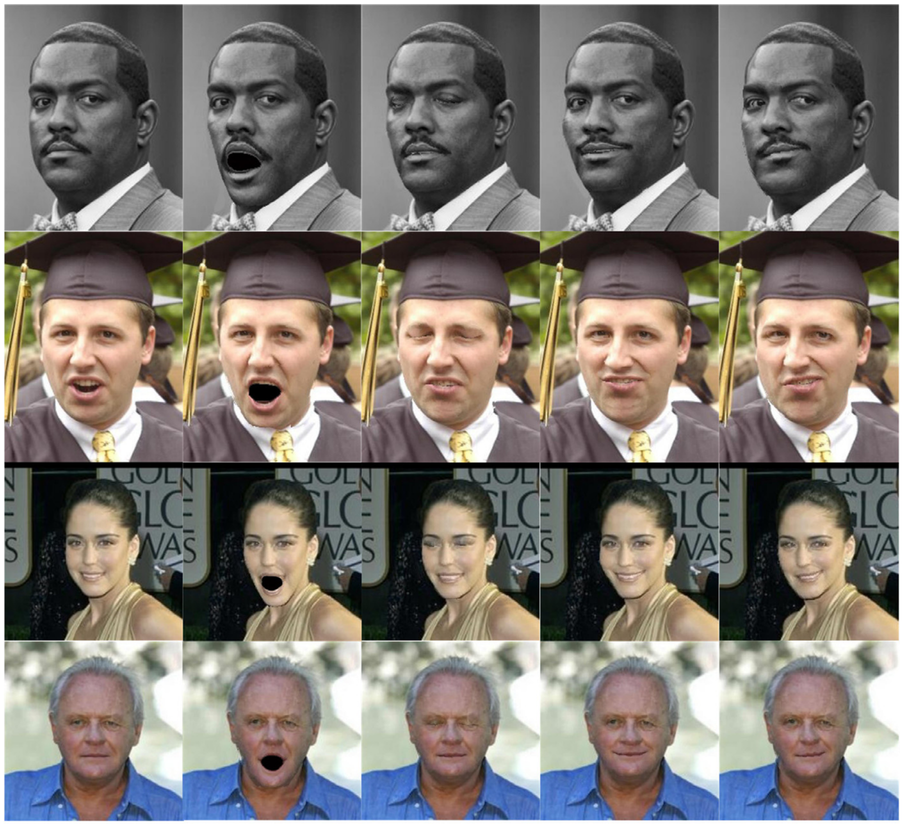

Fig. 15. Expression editing in facial images.

Table 3
Kendall error of different modeling methods.

Face database	JAFFE				LFW		Helen	
Method	Annoy	Happy	Surprise	Normal	LFW1	LFW2	Helen1	Helen2
3DMM	0.1581	0.1421	0.1682	0.1365	0.1456	0.1521	0.1418	0.1513
Bilinear Model	0.1311	0.1315	0.1365	0.1228	0.1387	0.1408	0.1350	0.1377
Our method	0.1205	0.1218	0.1266	0.1160	0.1269	0.1218	0.1193	0.1188

geometric features from the face model. Some details of face models have been removed in the modeling process. In our method, we don't lost the geometric details of the original face model. The model regenerate is based on the original facial database. The modeling process is driven by DLMs and PSKS. Using our method can achieve more accurate facial expression information. The fitting speed of our method is faster. In Table 2, we compare the average fitting speed of different methods. In Table 3, we compare the average Kendall error of different methods. Based on our method, the simple expression classification task can be processed without complex pre-process and data training process.

7. Conclusion

We propose a 3D facial expression modeling method based on Kendall theory. The method reconstructs the 3D facial model from 2D facial image while contains the accurate expression and head pose. We use facial expression representation (DLM) to fitting the 2D facial image from 3D facial data. The DLM fitting is processed by PSKS in Kendall shape space. Using the DLM fitting result, the 3D facial expression model can be regenerated. Our modeling method is robust to rotation, scaling and translation in facial images. The modeling result can reconstruct the accuracy 3D facial expression model by Kendall shape constraint. The optimization process is following the certain directions and the fitting energy is not limited by the local peak. The fitting speed is faster than other modeling methods. In future work, we will consider to add more facial features in the reconstruction process and provide accurate facial expression recognition method. We will extend the method in different applications such as facial animation and face recognition.

Conflict of interest

The authors declare that they have no competing interests.

Acknowledgments

This research was partially supported by the National Key Cooperation between the BRICS Program of China (No. 2017YFE0100500); National Key R&D Program of China (No. 2017YFB1002604, No. 2017YFB1402105, No. 2017YFB1002804); Beijing Natural Science Foundation of China (No. 4172033). We thank the provider (Zhejiang University in China, Kun Zhou) of

face database Facewarehouse and the provider (University of Basel) of 3DMM.

References

[1] A. Bas, W.A.P. Smith, T. Bolkart, S. Wuhrer, Fitting a 3D morphable model to edges: a comparison between hard and soft correspondences, in: Proceedings of the Asian Conference on Computer Vision, 2016, pp. 377–391.
[2] C. Cao, Y. Weng, S. Zhou, Y. Tong, K. Zhou, Facewarehouse: a 3D facial expression database for visual computing, IEEE Trans. Vis. Comput. Graph. 20 (3) (2014) 413–425.
[3] T.F. Cootes, C.J. Taylor, D.H. Cooper, J. Graham, Active shape models-their training and application, Comput. Vis. Image Underst. 61 (1) (1995) 38–59.
[4] G.J. Edwards, T.F. Cootes, C.J. Taylor, Face recognition using active appearance models, in: Proceedings of the IEEE Signal Processing and Communications Applications Conference, 1998, pp. 940–943.
[5] T. Hong, Y.B. Lee, Y.G. Kim, H. Kim, Facial expression recognition using active appearance model, in: Proceedings of the International Symposium on Neural Networks, 2006, pp. 69–76.
[6] C.C. Hsieh, M.K. Jiang, A facial expression classification system based on active shape model and support vector machine, in: Proceedings of the International Symposium on Computer Science and Society, 2011, pp. 311–314.
[7] C.L. Huang, Y.M. Huang, Facial expression recognition using model-based feature extraction and action parameters classification, J. Vis. Commun. Image Represent. 8 (3) (1997) 278–290.
[8] P.S. Penev, L. Sirovich, The global dimensionality of face space, in: Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, 2000, p. 264.
[9] X.W. Chen, T. Huang, Facial expression recognition: a clustering-based approach, Pattern Recognit. Lett. 24 (9) (2003) 1295–1302.
[10] P.P. Paul, M. Gavrilova, PCA based geometric modeling for automatic face detection, in: Proceedings of the International Conference on Computational Science and ITS Applications, 2011, pp. 33–38.
[11] D. Decarlo, D. Metaxas, M. Stone, An anthropometric face model using variational techniques, in: Proceedings of the Conference on Computer Graphics and Interactive Techniques, 1998, pp. 67–74.
[12] K.S. Lee, K.H. Wong, S.H. Or, Y.F. Fung, 3D face modeling from perspective-views and contour-based generic-model, Real-Time Imaging 7 (2) (2001) 173–182.
[13] N.A. A, A.M. Mohamed, Automatic facial feature extraction and 3d face modeling using two orthogonal views with application to 3D face recognition, Pattern Recognit. 12 (38) (2005) 2549–2563.
[14] J. Wang, L. Yin, X. Wei, Y. Sun, 3d facial expression recognition based on primitive surface feature distribution, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006, pp. 1399–1406.
[15] H. Soyel, H. Demirel, Facial expression recognition using 3d facial feature distances, in: Proceedings of the International Conference Image Analysis and Recognition, 2007, pp. 831–838.
[16] W. Peng, C. Xu, Z. Feng, 3d face modeling based on structure optimization and surface reconstruction with b-spline, Neurocomputing 179 (12) (2016) 228–237.
[17] S. Zhan, L. Chang, J. Zhao, T. Kurihara, H. Du, Y. Tang, J. Cheng, Real-time 3d face modeling based on 3d face imaging, Neurocomputing 252 (2017) 42–48.
[18] Y. Zhenbo, L. Guangcan, L. Qingshan, D. Jiankang, Spatio-temporal convolutional features with nested LSTM for facial expression recognition, Neurocomputing 317 (2018) 50–57.
[19] J. Wu, Z. Lin, W. Zheng, H. Zha, Locality-constrained linear coding based bi-layer model for multi-view facial expression recognition, Neurocomputing 239 (C) (2017) 143–152.
[20] V. Blanz, T. Vetter, A morphable model for the synthesis of 3D faces, in: Proceedings of the Conference on Computer Graphics and Interactive Techniques, 1999, pp. 187–194.
[21] V. Blanz, T. Vetter, Face recognition based on fitting a 3D morphable model, IEEE Trans. Pattern Anal. Mach. Intell. 25 (9) (2003) 1063–1074.
[22] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, T. Vetter, A 3d face model for pose and illumination invariant face recognition, in: Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance, 2009, pp. 296–301.
[23] J.P. Mena-Chalco, I. MacÃªdo, L. Velho, R.M.C. Jr, 3D face computational photography using PCA spaces., Vis. Comput. 25 (10) (2009) 899–909.
[24] F. Pighin, J. Hecker, D. Lischinski, R. Szeliski, Synthesizing realistic facial expressions from photographs, in: Proceedings of the ACM SIGGRAPH, 2005, p. 9.
[25] J. Booth, A. Roussos, A. Ponniah, D. Dunaway, S. Zafeiriou, Large scale 3D morphable models, Int. J. Comput. Vis. 126 (2–4) (2018) 1–22.
[26] F. Duan, D. Huang, T. Yun, L. Ke, Z. Wu, M. Zhou, 3d face reconstruction from skull by regression modeling in shape parameter spaces, Neurocomputing 151 (2015) 674–682.
[27] X. Lu, A.K. Jain, Deformation modeling for robust 3d face matching, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006, pp. 1377–1383.
[28] A.E. Ichim, S. Bouaziz, M. Pauly, Dynamic 3D avatar creation from hand-held video input, ACM Trans. Graph. 34 (4) (2015) 1–14.
[29] H. Jin, X. Wang, Z. Zhong, J. Hua, Robust 3d face modeling and reconstruction from frontal and side images, Comput. Aided Geom. Des. 50 (2017) 1–13.
[30] D. Vlasic, M. Brand, H. Pfister, Face transfer with multilinear models, in: Proceedings of the ACM SIGGRAPH, 2005, pp. 426–433.
[31] I. Mpiperis, S. Malassiotis, M.G. Strintzis, Bilinear models for 3-D face and facial expression recognition, IEEE Trans. Inf. Forensics Secur. 3 (3) (2008) 498–511.
[32] C. Cao, Q. Hou, K. Zhou, Displaced dynamic expression regression for real-time facial tracking and animation, ACM Trans. Graph. 33 (4) (2014) 1–10.
[33] X. Gao, C. Tian, Multi-view face recognition based on tensor subspace analysis and view manifold modeling, Neurocomputing 72 (16–18) (2009) 3742–3750.
[34] X. Song, Z.H. Feng, X. Yang, X. Wu, J. Yang, Towards multi-scale fuzzy sparse discriminant analysis using local third-order tensor model of face images, Neurocomputing 185 (C) (2016) 53–63.
[35] S. Elaiwat, M. Bennamoun, F. Boussaid, A spatio-temporal RBM-based model for facial expression recognition, Pattern Recognit. 49 (C) (2015) 152–161.
[36] T. Zhang, W. Zheng, Z. Cui, Y. Zong, J. Yan, K. Yan, A deep neural network–driven feature learning method for multi-view facial expression recognition, IEEE Trans. Multimed. 18 (12) (2016) 2528–2536.
[37] A.T. Lopes, E.D. Aguiar, A.F.D. Souza, T. Oliveira-Santos, Facial expression recognition with convolutional neural networks: coping with few data and the training sample order, Pattern Recognit. 61 (2016) 610–628.
[38] H. Li, J. Sun, Z. Xu, L. Chen, Multimodal 2d+3d facial expression recognition with deep fusion convolutional neural network, IEEE Trans. Multimed. 19 (12) (2017) 2816–2831.
[39] A. Brunton, A. Salazar, T. Bolkart, S. Wuhrer, Review of statistical shape spaces for 3d data with comparative analysis for human faces, Comput. Vis. Image Underst. 128 (11) (2014) 1–17.
[40] S. Kurtek, H. Drira, A comprehensive statistical framework for elastic shape analysis of 3d faces, Comput. Graph. 51 (2015) 52–59.
[41] T. Alashkar, B.B. Amor, M. Daoudi, S. Berretti, A Grassmann framework for 4d facial shape analysis, Pattern Recognit. 57 (C) (2016) 21–30.
[42] A. Patel, W.A.P. Smith, Manifold-based constraints for operations in face space, Pattern Recognit. 52 (2016) 206–217.
[43] D.G. Kendall, Shape manifolds, procrustean metrics, and complex projective spaces, Bull. Lond. Math. Soc. 16 (2) (1985) 81–121.
[44] V. Kazemi, J. Sullivan, One millisecond face alignment with an ensemble of regression trees, in: Proceedings of the Computer Vision and Pattern Recognition, 2014, pp. 1867–1874.
[45] F.M. Sukno, J.L. Waddington, P.F. Whelan, 3-d facial landmark localization with asymmetry patterns and shape regression from incomplete local features., IEEE Trans. Cybern. 45 (9) (2014) 1717–1730.
[46] S. Li, W. Deng, J. Du, Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2017, pp. 2584–2593.
[47] S. Li, W. Deng, Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition, IEEE Trans. Image Process. 28 (1) (2019) 356–370.

Chenlei Lv is studying for Ph.D. degree in College of information science and technology, Beijing Normal University (BNU). His research interests include computer vision, 3D biometrics, computer graphic, discrete differential geometry and conformal geometric.

Zhongke Wu is Full Professor in College of information science and technology, 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. His current research interests include computer graphics, animation virtual reality, geometric modeling, volume graphics and medical imaging.

Xingce Wang is Full Professor in College of Information Science and Technology, 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.

Mingquan Zhou is the super visor of doctor candidates and the dean in College of Information Science and Technology, Beijing Normal University, PR China. His current research interests include computer graphics, and 3D visualization.

View publication stats