


# An Efficient Three-Dimensional Reconstruction Approach for Pose-Invariant Face Recognition Based on a Single View

Minghua Zhao<sup>1</sup> , Ruiyang Mo<sup>1</sup>, Yonggang Zhao<sup>2</sup>,  
Zhenghao Shi<sup>1</sup>, and Feifei Zhang<sup>1</sup>

<sup>1</sup> Faculty of Computer Science and Engineering,  
Xi'an University of Technology, Xi'an 710048, China  
mh\_zhao@126.com

<sup>2</sup> Faculty of Earth Science and Engineering,  
Xi'an Shiyou University, Xi'an 710065, China

**Abstract.** A three-dimensional (3D) reconstruction approach based on a single view is proposed to solve the problem of lack of training samples while addressing multi-pose face recognition. First, a planar template is defined based on the geometric information of the segmented faces. Second, 3D faces are resampled according to the geometric relationship between the planar template and original 3D faces, and a normalized 3D face database is obtained. Third, a 3D sparse morphable model is established based on the normalized 3D face database, and a new 3D face can be reconstructed from a single face image. Lastly, virtual multi-pose face images can be obtained by texture mapping, rotation, and projection of the established 3D face, and training samples are enriched. Experimental results obtained using BJUT-3D and CAS-PEAL-R1 face databases show that recognition rate of the proposed method is 91%, which is better than other methods for pose-invariant face recognition based on a single view. This is primarily because the training samples are enriched using the proposed 3D sparse morphable model based on a new dense correspondence method.

**Keywords:** Pose-invariant face recognition · 3D face reconstruction · Single view · 3D morphable model · Dense correspondence

## 1 Introduction

Currently, face recognition technologies exhibit good performance under normal conditions [1–3]. However, face recognition with pose variations is still a major challenge because an object with different poses may appear considerably different owing to nonlinear deformation and self-occlusion [4]. Multi-pose face recognition based on a single view has high research value because it can be used in a wide range of practical applications. Researches on multi-pose face recognition technology based on a single view can be divided into two primary categories, i.e., pose correction methods and virtual multi-pose image synthesizing methods. Virtual multi-pose image synthesizing methods can improve the accuracy of face recognition because they generate appropriate virtual multi-pose face images to enrich training samples.

Among virtual multi-pose image synthesizing methods, three-dimensional (3D) face morphable models are an important area of research. The establishment of this model is based on a set of 3D faces [5]. However, original 3D face data cannot be used for linear calculation because the numbers of vertices and patches are different for each 3D face. It is necessary for all vertices and patches of the 3D faces to have one-to-one correspondence. A large number of methods have been proposed for dense correspondence of 3D faces. Blanz et al. proposed an optical flow algorithm and a bootstrapping algorithm to solve the dense correspondence problem of 3D faces [6, 7]. The method is valid when the 3D face is similar to the reference face. Yongli Hu et al. proposed a dense correspondence method based on a grid resampling method, and the result was more accurate [8]. However, the process was very complicated because considerable manual interaction was required. Gu et al. proposed a uniform grid resampling algorithm [9]. Gong Xun et al. proposed a grid resampling method based on a planar template [10]. The method located features based on a combination of two-dimensional (2D) and 3D texture information, and divided the surface of a 3D face using these features.

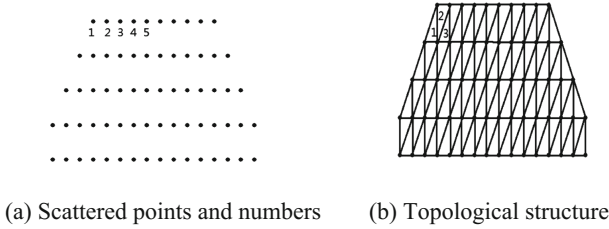
To simplify the construction of a 3D face model, an improved dense correspondence method based on planar template resampling with geometric information is proposed in this paper. A 3D face is reconstructed using a 3D sparse morphable model, and multi-pose face images are generated using texture mapping, model rotation, and projection. The rest of the paper is organized as follows: In Sect. 2, an improved dense correspondence method based on planar template resampling with geometric information is proposed. In Sect. 3, 3D face reconstruction based on the 3D face morphable model is presented. In Sect. 4, the experiments performed using BJUT-3D and CAS-PEAL-R1 databases are described, and the proposed method is compared to other methods. Finally, conclusions are stated in Sect. 5.

## 2 An Improved Dense Correspondence Method Based on Planar Template Resampling with Geometric Information

### 2.1 Definition of Planar Template

After face segmentation on the cylindrical expansion of the original 3D faces, segmented 2D texture maps are aligned using the nose tip positions and overlapped. Among the overlapped points, the pixels that belong to more than half of the images are used to construct a planar template, which is shown in Fig. 1(a).

The corresponding vertices of the selected pixels are scattered, and meshes are built using these vertices to construct a normalized 3D face model. The topological structure is constructed through the following steps: First, vertices are connected according to the connection rules of vertices on the vertical and horizontal iso-lines. Then, each small rectangular area is divided into two triangles by the diagonal between the bottom left corner and the top right corner of this area. Lastly, the triangles are numbered from top to bottom and left to right, which is shown in Fig. 1(b). The template can be used to generate the normalized 3D face model.



**Fig. 1.** Definition of planar template

### 2.2 Resampling of Original 3D Faces

An original 3D human face is not consistent with the definition of the planar template, and it is necessary to resample it.

(a) Vertical sampling

For an original 3D face,  $f_i$  ( $1 \leq i \leq m$ ), the vertex on the  $k$ th line can be calculated using interpolation.  $pre_{ver}$  and  $fol_{ver}$  shown in Eqs. (1) and (2) are used to represent the line numbers before and after the  $k$ th line, respectively.

$$pre_{ver} = floor(ratio_{ver} \times k) \tag{1}$$

$$fol_{ver} = ceil(ratio_{ver} \times k) \tag{2}$$

In Eqs. (1) and (2),  $floor$  is a rounded down function,  $ceil$  is a rounded up function, and  $ratio_{ver}$  is the ratio between the height of the segmented 2D texture image,  $f_i(h)$ , and the height of the planar template,  $Tmpt(h)$ .

$$ratio_{ver} = f_i(h)/Tmpt(h) \tag{3}$$

The resampling data (including the shape and texture data) of the vertex,  $v_k(vert)$ , on the  $k$ th line can be calculated using interpolation between the vertex,  $v_{pre}(ori)$ , on the  $pre$  line and the vertex,  $v_{fol}(ori)$ , on the  $fol$  line.

$$\begin{aligned}
 v_k(vert) &= v_{pre}(ori) \times (1 - \Delta r) + v_{fol}(ori) \times \Delta r \\
 \Leftrightarrow \begin{cases} s_k(vert) = s_{pre}(ori) \times (1 - \Delta r) + s_{fol}(ori) \times \Delta r \\ t_k(vert) = t_{pre}(ori) \times (1 - \Delta r) + t_{fol}(ori) \times \Delta r \end{cases}
 \end{aligned} \tag{4}$$

where  $\Delta r = ratio \times k - pre$ .

(b) Horizontal sampling

The interpolation operation that was used in vertical sampling is used to calculate the  $l$ th data on the  $k$ th line.  $pre_{hor}$  and  $fol_{hor}$  shown in Eqs. (5) and (6) are used to represent the row numbers before and after the  $l$ th row, respectively.

$$pre_{hor} = \text{floor}(\text{ratio}_{hor} \times l) \quad (5)$$

$$fol_{hor} = \text{ceil}(\text{ratio}_{hor} \times l) \quad (6)$$

where  $\text{ratio}_{hor}$  is the ratio between the width of the segmented 2D texture image,  $f_i(w)$ , and the width of the planar template,  $Tmpt(w)$ .

$$\text{ratio}_{hor} = f_i(w)/Tmpt(w) \quad (7)$$

The  $l$ th resampling data on the  $k$ th line (including the shape data and texture data) of the vertex,  $v_{k,l}(\text{final})$ , can be obtained using interpolation between the vertex,  $v_{k,pre}(\text{vert})$ , on the  $pre$  row and the vertex,  $v_{k,fol}(\text{vert})$ , on the  $fol$  row.

$$\begin{aligned} v_{k,l}(\text{final}) &= v_{k,pre}(\text{vert}) \times (1 - \Delta r) + v_{k,fol}(\text{vert}) \times \Delta r \\ &\Leftrightarrow \begin{cases} s_{k,l}(\text{final}) = s_{k,pre}(\text{vert}) \times (1 - \Delta r) + s_{k,fol}(\text{vert}) \times \Delta r \\ t_{k,l}(\text{final}) = t_{k,pre}(\text{vert}) \times (1 - \Delta r) + t_{k,fol}(\text{vert}) \times \Delta r \end{cases} \end{aligned} \quad (8)$$

where  $\Delta r = \text{ratio} \times l - pre$ .

After arranging the resampled data according to the topology structure of the planar template, each 3D face can be represented as two vectors.

$$\begin{cases} s_i = (x_1, y_1, z_1, \dots, x_j, y_j, z_j, \dots, x_n, y_n, z_n)^T \in R^{3n} \\ t_i = (r_1, g_1, b_1, \dots, r_j, g_j, b_j, \dots, r_n, g_n, b_n)^T \in R^{3n}, 1 \leq i \leq N \end{cases} \quad (9)$$

In Eq. (9),  $s_i$  is the shape vector composed of three coordinates of the  $i$ th 3D face,  $t_i$  is its corresponding texture vector composed of R value, G value, and B value,  $N$  is the number of 3D faces, and  $n$  is the number of facial points on the 3D face. Values with the same subscript represent the same facial feature points on different face vectors.

### 3 3D Face Reconstruction Based on Sparse Morphable Model

3D face reconstruction based on a sparse morphable model consists of two steps, i.e., model construction and face shape reconstruction. The first step involves acquisition and normalization of prototype 3D face data and establishment of a 3D morphable model. The second step is to match a target face image with the morphable model and accomplish the reconstruction of the face.

#### 3.1 Construction of 3D Morphable Model

Normalized 3D faces have an equal number of vertices and patches. They can be regarded as a linear space, where each element can be expressed linearly using other elements. The shape of a 3D face can be expressed as Eq. (10).

$$s_i = (x_1, y_1, z_1, \dots, x_k, y_k, z_k, \dots, x_n, y_n, z_n)^T \in R^{3n}, i = 1, \dots, m \quad (10)$$

where  $(x_k, y_k, z_k)$  is the coordinate of the  $k$  th vertex,  $v_k, n$  is the number of vertices, and  $m$  is the number of 3D faces. The linear space constructed using  $m$  3D faces can be expressed as Eq. (11).

$$S = (s_1, \dots, s_m) \in R^{3n \times m} \tag{11}$$

The shape of a new face can be expressed as Eq. (12).

$$s_{new} = S \cdot a, a = (a_1, \dots, a_i, \dots, a_m) \tag{12}$$

where  $a_i \in [0, 1]$  and  $\sum_{i=1}^m a_i = 1$ .

The  $m'(m' \leq m - 1)$  vectors corresponding to the  $m'(m' \leq m - 1)$  largest eigenvalues can be used to construct the feature matrix,  $Q = (q_1, \dots, q_{m'})$ , and Eq. (12) can be rewritten as Eq. (13).

$$s_{new} = \bar{s} + Q \cdot \beta = \bar{s} + \Delta s \tag{13}$$

In Eq. (13),  $\bar{s} = \frac{1}{m} \sum_{i=1}^m s_i$  and  $\beta = (\beta_1, \dots, \beta_i, \dots, \beta_{m'})^T \in R^{m'}$ . This implies that a specific face can be obtained using a deformation on the average face.

### 3.2 Face Shape Reconstruction Based on Sparse Morphable Model

Face shape reconstruction based on the sparse morphable model can be described as the analysis of global shape deformation,  $\Delta s$ , according to the shape deformation,  $\Delta s^f$ .

$$\Delta s^f = L(s - \bar{s}) = L(Q \cdot \beta) = Q^f \cdot \beta \tag{14}$$

where  $L$  represents the selected feature points.

The computation of  $\beta$  can be transformed into determining the optimal solution of the objective function, as shown in Eq. (15).

$$E(\beta) = \|Q^f \cdot \beta - \Delta s^f\|^2 + \eta \cdot \|\beta\|^2 \tag{15}$$

where the first term represents reconstruction error, the second term represents random fluctuation, which improves the robustness of the model to noise, and  $\eta \geq 0$  is an adjusting parameter.

According to singular value decomposition, Eq. (16) is obtained by taking the derivative of  $\beta$  with respect to the objective function.

$$\arg \min \|E(\beta)\| = V \cdot \left( \frac{\lambda_i}{\lambda_i^2 + \eta} \right) \cdot U^T \cdot \Delta s^L \tag{16}$$

where  $U \in R^{l \times l}$  and  $V \in R^{m' \times m'}$ .

Combining Eq. (16) with Eq. (13), the final reconstruction result of the given face can be described as Eq. (17).

$$s_{new} = \bar{s} + Q \cdot V \cdot \left( \frac{\lambda_i}{\lambda_i^2 + \eta} \right) \cdot U^T \cdot \Delta s \quad (17)$$

## 4 Experiments and Analysis

### 4.1 3D Face Database and the Segmentation Performance

The BJUT-3D database [11] is used to evaluate the performance of the proposed method. There are approximately 200000 vertices and 400000 triangular faces for one 3D face model. An example of original 3D face data is shown in Fig. 2. Face segmentation results based on geometric information are shown in Fig. 3.

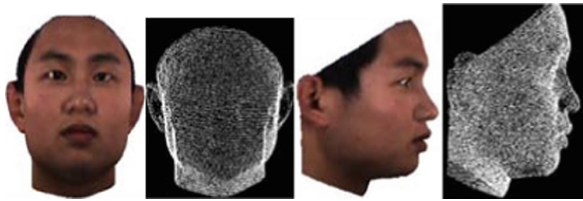


Fig. 2. Geometry and texture information of one person in BJUT-3D database

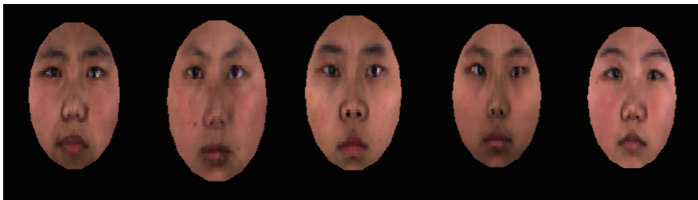


Fig. 3. Face segmentation results based on geometric information

### 4.2 Dense Correspondence Performance of the Proposed Method

The data of 100 3D faces from the BJUT-3D database, consisting of 50 females and 50 males, are used to evaluate the dense correspondence performance of the proposed method. A comparison between the original 3D point cloud data and the dense correspondence results is shown in Fig. 4. It can be observed from Fig. 4 that the 3D face database can be normalized using the proposed method and is suitable for linear operations to be performed later.

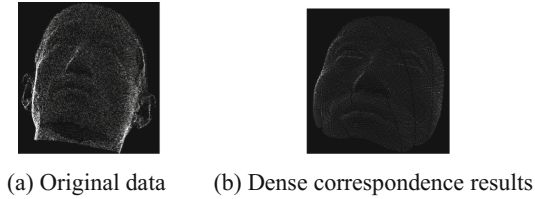


Fig. 4. Dense correspondence results of the proposed method

### 4.3 3D Reconstruction Results of the Proposed Method

A hundred face images with a natural expression and no light variation are used from the CAS-PEAL-R1 database to evaluate the reconstruction performance of the proposed method. Two original face images are shown in Fig. 5(a). The images are normalized to a size of  $164 \times 146$  pixels. The 3D reconstruction results obtained using the proposed method are shown in Fig. 5(b). The texture mapping results are shown in Fig. 5(c). Figure 5(d)–(n) show the multi-pose results obtained by rotation and

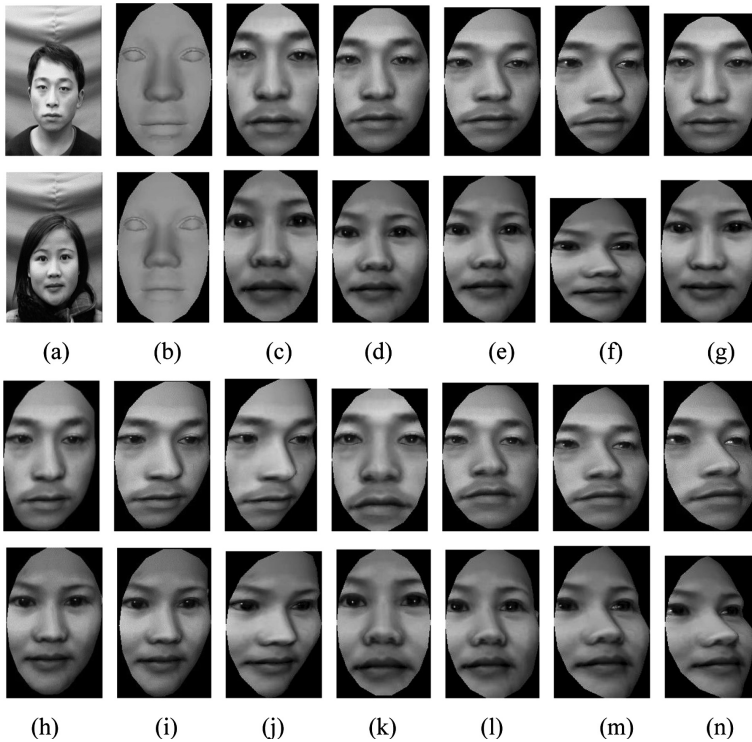


Fig. 5. 3D face reconstruction results (a) Input images (b) Reconstructed results (c) Texture mapping results (d)–(n) Multi-pose mapping results of the 3D face

projection of the 3D faces. Figures 5(d), (e), and (f) show the projected results of 15°, 30°, and 45° rotation, for the pose of looking in front horizontally. Figures 5(g), (h), (i), and (j) show the projected results of 0°, 15°, 30°, and 45° rotation, for the pose of looking down. Figures 5(k), (l), (m), and (n) show the projected results of 0°, 15°, 30°, and 45° rotation, for the pose of looking up.

#### 4.4 Recognition Results of the Proposed Method

A hundred subjects with a natural expression and no light variation are used from the CAS-PEAL-R1 database to evaluate the performance of the proposed method for face recognition. Twelve multi-pose images for each subject obtained by rotation and projection of the 3D sparse morphable model, as described in Sect. 4.3, are used as training samples. Five real multi-pose face images for each subject are used as testing samples. The performance of the proposed method is compared to that of conventional methods, and the results are shown in Table 1.

**Table 1.** Recognition performance comparison of the proposed method and other methods

Methods	Recognition rate
Enhanced (PC) <sup>2</sup> A [12]	0.71
PAFE [13]	0.88
TPSM [14]	0.90
Proposed method	0.91

For the enhanced projection-combined principal component analysis ((PC)<sup>2</sup>A) method, the recognition rate is relatively low. This is primarily because the new images are dependent and are highly correlated with the original image. The recognition rates of the pose adaptive feature extraction (PAFE) method and the triplet pose sparse matrix (TPSM) method are relatively high because a priori information about 3D face data is used. For the proposed method, training images are enriched by rotation and projection of the reconstructed 3D face. It can be observed from Table 1 that the recognition rate of the proposed method is 91%, which is superior to state-of-the-art methods for pose-invariant face recognition.

## 5 Conclusions and Discussions

To synthesize virtual multi-pose faces, a 3D face reconstruction method based on a single view is proposed in this paper. This method includes three aspects. The first is planar template establishment based on geometric information. The second is vertical and horizontal resampling of 3D face data based on the geometric relationship between the planar template and original 3D face data, based on which normalized 3D face data is obtained. The third is that a 3D sparse morphable model is constructed based on the normalized 3D face data, and multi-pose face images are generated by texture mapping, rotation, and projection of the established 3D face. The proposed method has two



contributions. First, the dense correspondence process is performed automatically and the model construction process does not involve manual interaction, which makes it superior to other methods. Second, virtual multi-pose face images are synthesized using the proposed method and training samples are enriched, because of which recognition performance improves.

A pose-invariant face recognition problem without illumination variations is considered in this study. Face recognition with pose and illumination variations is more difficult. Future research will focus on a pose-invariant face recognition problem with illumination variations based on a single view.

**Acknowledgments.** This work was partially supported by a grant from the National Natural Science Foundation of China (No. 61401355), a grant from the Key Laboratory Foundation of Shaanxi Education Department, China (No. 14JS072) and a grant from Science and Technology Project Foundation of Beilin District, Xi'an City, China (No. GX1621). The authors also thank anonymous reviewers for their valuable comments.

## References

1. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.: Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **31**(2), 210–227 (2009)
2. Ding, C., Choi, J., Tao, D., Davis, L.S.: Multi-directional multi-level dual-cross patterns for robust face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **38**(3), 518–531 (2016)
3. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: application to face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(12), 2037–2041 (2006)
4. Sharma, R., Patterh, M.S.: A new pose invariant face recognition system using PCA and ANFIS. *Optik-Int. J. Light Electron Opt.* **126**(23), 3483–3487 (2015)
5. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3D faces. In: *Computer Graphics Proceedings SINGRAPH 1999*, pp. 187–194 (1999)
6. Blanz, V., Vetter, T.: Face recognition based on fitting a 3D morphable model. *IEEE Trans. Pattern Anal. Mach. Intell.* **25**(9), 1063–1074 (2003)
7. Blanz, V.: Face recognition based on a 3D morphable model. In: *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*, vol. 25(9), pp. 617–624. IEEE Computer Society (2006)
8. Hu, Y., Yin, B., Gu, C., Cheng, S.: 3D face reconstruction based on the improved morphable model. *Chin. J. Comput.* **28**(10), 1671–1679 (2005)
9. Gu, C., Yin, B., Hu, Y., Cheng, S.: Resampling based method for pixel-wise correspondence between 3D faces. In: *The Proceedings of the International Conference on Information Technology: Coding and Computing*, pp. 614–619 (2004)
10. Gong, X., Wang, G.: 3D face deformable model based on feature points. *J. Softw.* **20**(3), 724–733 (2009)
11. Hu, Y., Zhang, Z., Xu, X., Fu, Y., Huang, T.S.: Building large scale 3D face database for face analysis. In: Sebe, N., Liu, Y., Zhuang, Y., Huang, T.S. (eds.) *MCAM 2007*. LNCS, vol. 4577, pp. 343–350. Springer, Heidelberg (2007). doi:[10.1007/978-3-540-73417-8\\_42](https://doi.org/10.1007/978-3-540-73417-8_42)
12. Songcan, C., Daoqiang, Z., Zhihua, Z.: Enhanced (PC)<sup>2</sup>A for face recognition with one training image per person. *Pattern Recognit. Lett.* **25**(10), 1173–1181 (2004)

13. Yi, D., Lei, Z., Li, S.Z.: Towards pose robust face recognition. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, vol. 9(4), pp. 3539–3545. IEEE Computer Society (2013)
14. Moeini, A., Moeini, H., Faez, K.: Real-time pose-invariant face recognition by triplet pose sparse matrix from only a single image. In: 2014 Proceedings of 22nd International Conference on Pattern Recognition, pp. 465–470 (2014)