Detail Reconstruction of 3D Face Model from Single Image

Hongxin Xu¹, Ruoming Lan*, Tianping Li*
¹School of Physics and Electronic Science, Shandong Normal University, Jinan, Shandong, China
*School of Physics and Electronic Science, Shandong Normal University, Jinan, Shandong, China
Corresponding author’s e-mail: sdsdltp@sdnu.edu.cn; lanrm0616@163.com

Abstract. With the development of computer vision technology, 3D face model reconstruction technology has a great breakthrough. In view of the current popular deep learning method, we use an algorithm based on neural network to restore the details of 3D face model. We design a high frequency detail generation network to generate displacement map based on the texture extracted from the original image. The displacement map contains a large number of high frequency detail features of the target face. By embedding the displacement map into the face model, the details of the model can be restored.

1. Introduction
After years of development, image processing has been relatively mature. Face recognition is also one of the most important applications in the field of computer vision. In most environments, the current face recognition technology can recognize the target face efficiently. 2D image information is not complete enough to meet the needs of research, so 3D reconstruction method has become one of the hot research methods. With the application of deep learning method in image processing [1], artificial reconstruction method is replaced, and efficient 3D face reconstruction algorithm is more suitable for research and application. At present, most of the known 3D face reconstruction algorithms only reconstruct the rough model or a small amount of detail model. As these models contain most of the information, but the accuracy is not ideal. We propose an algorithm based on neural network to reconstruct 3D models from single face images and restore high frequency details.

2. Non-front face alignment
Our proposed algorithm mainly deals with the problem of face pose and face details. Bazal et al. proposed a Basel face model (BFM) based on principal component analysis (PCA) [2]. BFM is a parametric model, which can be obtained by inputting shape parameters, texture parameters and other data. Although this model has a high degree of fit to the target face, it contains most of the face information, but it is difficult to deal with high frequency details such as facial wrinkles. Our method uses a neural network to generate the displacement map based on the real texture map, which contains most of the details [3].

2.1. Camera model
In the field of 3D reconstruction, a camera model is often needed to simulate the image acquisition process of real camera equipment. The camera model takes the normal of the face image as the positive direction, considering the pose angle, the zoom and displacement of the image.
Generally speaking, the camera model uses orthogonal perspective projection to achieve imaging. Orthographic perspective projection cannot scale images effectively. So we use weak perspective projection similar to orthogonal perspective projection to solve this problem [5]. Suppose that the real model shape vector \( S \) of the non-frontal target face, the rotation matrix relative to the positive direction of the camera model is \( R \in \mathbb{R}^{3 \times 3} \), and the weak perspective projection function is:

\[
P = FS + t, \tag{1}
\]

where \( F = f * \Lambda * R \), \( f \) is the focal length parameter of the camera model, \( \Lambda \) is the orthogonal matrix, and \( t \) is the displacement parameter.

2.2. Feature point detection

When processing non frontal face images, some information of face pose self occlusion is lost seriously, so it is difficult to follow-up operation. Aiming at the problem of face pose, a face alignment method is proposed. Referring to PRN, we try to apply weight to 3D dense face model in UV vector space, so as to solve the problem that it is difficult to obtain partial information of self occlusion [6]. Face feature points are distributed in the main components of the face, we put a large weight on these feature points to highlight the recognition degree of the face.

In the aspect of feature point detection, we used dlib library based on regression tree set cascade method [7]. The feature points are detected, and then the dictionary is indexed to get the face deflection angle, and then the rotation matrix is calculated. Figure 1 is an example of feature point detection, and the image is from 300W dataset [8].

![Feature point detection example](image)

Figure 1. Example of feature point detection.

3. 3D face detail reconstruction

Our method divides 3D face reconstruction into two steps: smooth model reconstruction and model detail generation. Smooth model lacks high frequency details, and model details generation can make up for the details of smooth model.

3.1. Reconstruction of smooth model using BFM

The BFM based on PCA is a parametric model. In this paper, we need to input parameters to get the shape model with smooth surface. The shape vector of the reconstructed model is \( S' = (x_i, y_i, z_i) \) by inputting the shape parameter \( \alpha_i \) to influence the vector:

\[
S_{\text{smooth}} = \sum_{i=1}^{n} \alpha_i S_i \tag{2}
\]

The detected feature points are used to calculate PCA parameters, but they can not represent all the information of 2D face. Compared with the real face model, the model reconstructed by BFM will inevitably lose some information.

\[
S - S_{\text{smooth}} = \sum_{k} \omega_k \| L_k - P(x_k, y_k) \|_2, \tag{3}
\]

where \( L_k \) is the kth feature point of plane face, \( P(x_k, y_k) \) is the projection coordinate of the kth vertex of the 3D model on the 2D plane, \( \omega_k \) is the weight of the kth feature point.
3.2. High frequency detail generation

In the field of 3D model reconstruction, high frequency details are essential for a high precision model. The number of facial details is large and complex, so it is unrealistic to analyze the details by only one face image. For 3D detail model, the manual reconstruction method is time-consuming and labor-consuming, so we propose a detail generator based on CGAN to restore the high frequency details of the reconstructed model.

![Diagram of high frequency detail generator](image)

The network is divided into two parts: compression channel and expansion channel. Each layer consists of two convolution layers and one maximum pooling layer. Image as input, compression channel through dimension reduction, each layer generates half dimension feature map. The generated feature map is fused with the dimension reduced image in the expansion channel to generate a new image, and the last layer outputs the generated image.

The generated image is grayed and input to the PatchGAN for threshold judgment. If the judgment is successful, the displacement map will be output; if the judgment fails, it will be re input into the detail generator. Loss function of this network [9]:

\[
\begin{align*}
L_G &= \mathcal{V}_{\text{GAN}}(D, G) + \lambda_1 L_1(G), \\
L_D &= -\mathcal{V}_{\text{GAN}}(D, G),
\end{align*}
\]

where \(\lambda_1\) is set to 100.

![Comparison of smooth model and detail model](image)

Figure 3. Comparison of smooth model and detail model. (a) Original image. (b) Smooth model. (c) Detail model.
4. Experimental evaluation

We use the original image reconstruction model in Facescape dataset [10], and compare our method with the advanced method PRN. Facescape dataset is a collection of face information by deep camera equipment, which includes the original image and the corresponding real model. We calculate the point cloud error between the reconstructed model and the real model, and evaluate the accuracy of the model.

Figure 4. Error heat map.

Figure 4 shows the error heat distribution between the algorithm reconstruction model and the real database model. The four images in the first line are the original images that can be published in Facescape dataset. For this evaluation, we use ICP algorithm to compare the error values of 50K point clouds of the model, and calculate the root mean square error (RMSE). The detailed data are shown in Table 1. As can be seen from the data in Table 1, the accuracy of our method is higher than that of PRN in most cases.

| Method   | (a)  | (b)  | (c)  | (d)  |
|----------|------|------|------|------|
| Our method | 5.82 | 6.11 | 5.01 | 5.72 |
| PRN      | 5.97 | 6.57 | 4.94 | 6.48 |

5. Conclusion

The proposed method uses the displacement map to enhance the high frequency details of the smooth model. Considering that the pose of the target face will interfere with the reconstruction process, our method uses a face alignment method to identify the positive direction of the face, and obtains the face feature points through the detection of dlib library. The feature points are used to calculate the BFM parameters. After the experimental evaluation, our method in terms of accuracy has achieved relatively ideal results.
Acknowledgments
The authors would like to thank the authors of Facescape dataset and 300W dataset for sharing their experimental data with us. This work is supported by Shandong Provincial Natural Science Foundation (ZR2020MF119).

References
[1] Zhang K, Zuo W, Chen Y, et al. Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising[J]. IEEE Transactions on Image Processing, 2017, 26(7): 3142-3155.
[2] Paysan P, Knothe R, Amberg B, et al. A 3D face model for pose and illumination invariant face recognition[C]. 2009 sixth IEEE international conference on advanced video and signal based surveillance, 2009: 296-301.
[3] Garrido P, Zollhöfer M, Casas D, et al. Reconstruction of personalized 3D face rigs from monocular video[J]. ACM Transactions on Graphics (TOG), 2016, 35(3): 1-15.
[4] Park H S, Shiratori T, Matthews I, et al. 3D trajectory reconstruction under perspective projection[J]. International Journal of Computer Vision, 2015, 115(2): 115-135.
[5] Bruckstein A M, Holt R J, Huang T S, et al. Optimum fiducials under weak perspective projection[J]. International Journal of Computer Vision, 1999, 35(3): 223-244.
[6] Feng Y, Wu F, Shao X, et al. Joint 3d face reconstruction and dense alignment with position map regression network[C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 534-551.
[7] Kazemi V, Sullivan J. One millisecond face alignment with an ensemble of regression trees[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2014: 1867-1874.
[8] Zhu X, Lei Z, Liu X, et al. Face alignment across large poses: A 3d solution[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 146-155.
[9] Kuang H, Ding Y, Ma X, et al. 3D Face Reconstruction with Texture Details from a Single Image Based on GAN[C]. 2019 11th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA). IEEE, 2019: 385-388.
[10] Yang H, Zhu H, Wang Y, et al. FaceScape: a large-scale high quality 3D face Dataset and detailed riggable 3D face prediction[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 601-610.