



Robust3D: a robust 3D face reconstruction application

Zhihan Lv^{1,2}

Received: 28 May 2019 / Accepted: 19 July 2019 / Published online: 2 August 2019
© Springer-Verlag London Ltd., part of Springer Nature 2019

Abstract

In the process of reconstructing a historical event such as a rock concert only from video, the reconstruction of faces and expressions of the musicians is obviously important. However, in the process of rebuilding appearance, because of the low quality of the video of the recorded concert, the result of the reconstruction may be far from the real appearance. In this paper, a robust 3D face reconstruction application is described that can be applied to a video recording. The application first uses DeblurGAN program to run anti-ambiguity calculation and removes the ambiguity in the concert video. Then, the super-resolution program is used to enlarge every frame of the concert video by four times, thus making every frame of the video clearer. Finally, the 3D faces are obtained after 3D reconstruction of the processed video frames via the 3DMM_CNN program.

Keywords 3D face · Deblur · Super-resolution

1 Introduction

In recent years, three-dimensional face modeling and reconstruction technology have attracted more and more attention in the field of computer vision and computer graphics. Previously, many researchers have proposed how to reconstruct three-dimensional shapes from two-dimensional images. However, most algorithms require multiple images or videos to initialize the process of 3D face reconstruction. In many applications, only one image is available. Although some methods only use a single image for 3D face reconstruction, the resulting three-dimensional face is not realistic enough [1]. Therefore, we need an algorithm that can reconstruct a realistic three-dimensional face from a single image. Some researchers have proposed more accurate three-dimensional face algorithm, the most successful of which is based on the three-dimensional morphable model (3DMM) [2–10], and the algorithm of reconstructing three-dimensional face model by composite analysis and random optimization of multivariate cost functions. In order to improve the accuracy of facial feature

fitting, the whole face is first fitted, and then, the specific areas such as eyes, mouth and nose are fitted. The whole fitting process takes about half a minute. Obviously, because of this long time, the system is difficult to use in practice. 3D deformation model (3DMM) is only used for recognition under limited controlled observation conditions [11–15]. Based on the three-dimensional deformation model, an efficient, robust and accurate fitting algorithm, inverse compositional image alignment (ICIA), is proposed in document [16] to fit two-dimensional images. This method greatly improves the fitting efficiency. Previous work [17] presents a linear shape and texture fitting algorithm. This algorithm is similar to ICIA, and its speed is five times faster than stochastic optimization algorithm. In [18], an efficient method of face reconstruction is described, which combines 2D with 3D. A single front face photograph is used to reconstruct a three-dimensional face model. It requires only general facial expression and normal illumination. According to a personalized 3D face, a realistic virtual face can be obtained under different PIEs (posture, illumination and expression).

The contribution of this paper is to design a robust 3D face reconstruction application based on the characteristics of the live video. Firstly, the application uses DeblurGAN program to run the anti-ambiguity calculation and removes the ambiguity in the concert video. Then, the super-resolution program named VDSR is used to enlarge every

✉ Zhihan Lv
lyuzhihan@ub.edu

¹ Qingdao University, Qingdao, China

² University of Barcelona, Barcelona, Spain

frame of the concert video by four times, which makes every frame of the concert video clearer. Finally, the 3D face is obtained by 3D reconstructing the processed video frames via the 3DMM_CNN program.

1.1 Related research

Many previous attempts have been made to estimate the 3D surface of the face that appears in a single image. Before considering these, it is important to mention recent multi-image reconstruction methods using image sets (e.g., [19–23]). Recently, the generative countermeasure network (GAN) has achieved good results in image super-resolution reconstruction and in painting. GAN can retain the rich details of the image and create images that are very similar to the real image.

At present, there is no application of 3DMM to face recognition from a video of a concert. One reason is that the face images reconstructed by this method are unstable from an uncertain perspective. 3D simulation is either unstable, resulting in a large difference in the same individual’s 3D simulation, or it is too generalized, leading to most of the images being similar. It also explains why some people recently proposed using rough, simple 3D shape approximation as a proxy when rendering faces to new views rather than facial representations [24–27]. We adopted the method of 3DMM_CNN, which can generate robust 3D face models from arbitrary face images. And we used convolution neural network (CNN) to adjust the face shape and texture of 3DMM according to the input image.

At present, 3D face reconstruction based on multiple face images can generate 3D face models with high accuracy, but a large number of images are needed. However, 3D face reconstruction based on single view appears to be difficult, which can be divided into the following categories.

Firstly, statistical shape representations. For example, the widely used 3DMM method uses many aligned 3D face shapes for 3D face reconstruction. This method cannot generate faces with individual features. Recently, CNN has been used to adjust the face parameters of 3DMM [28]. However, they found that lack of sufficient training data is a major problem in face recognition. Unlike the algorithm presented in this paper, they generate training face images based on the sampling of 3DMM face models. Face images generated by this method are prone to over-fitting problem [29]. Therefore, they can only train a shallow residual network.

Secondly, scene assumption methods. In order to obtain the correct face reconstruction model, one kind of research estimates the scene and angle of the input picture. Some methods use information such as light source, facial

reflection and facial symmetry to estimate [30]. However, such estimates do not apply in reality [31].

Thirdly, example-based methods. It adjusts the template 3D face according to the input image [32–34]. This method can be used to generate the invisible side of a face in face recognition.

Fourthly, landmark fitting methods. This kind of reconstruction method first detects the facial recognition points [35, 36] and then compares the recognition points to the 3D model [37, 38].

2 Methodology

2.1 Implementation of DeblurGAN

Given a blurred image I_b , we expect to reconstruct a clear image I_s . To this end, we constructed a generative countermeasure network and trained a CNN as a generator G_{θ_G} and a discriminant network D_{θ_D} [39].

2.1.1 Network architecture

The overall structure is given in Fig. 1.

The structure of generator CNN is given in Fig. 2.

The network structure is similar to that proposed by Johnson in the task of style migration. The author added “ResOut” or “global skip connection.” What CNN learns

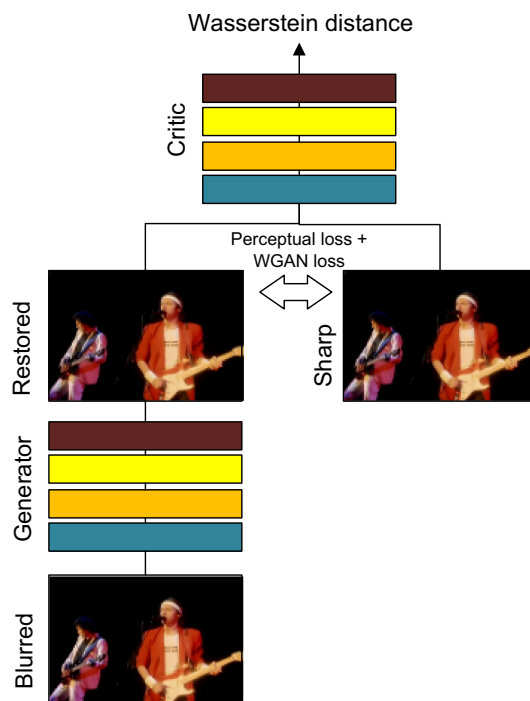


Fig. 1 Overall structure

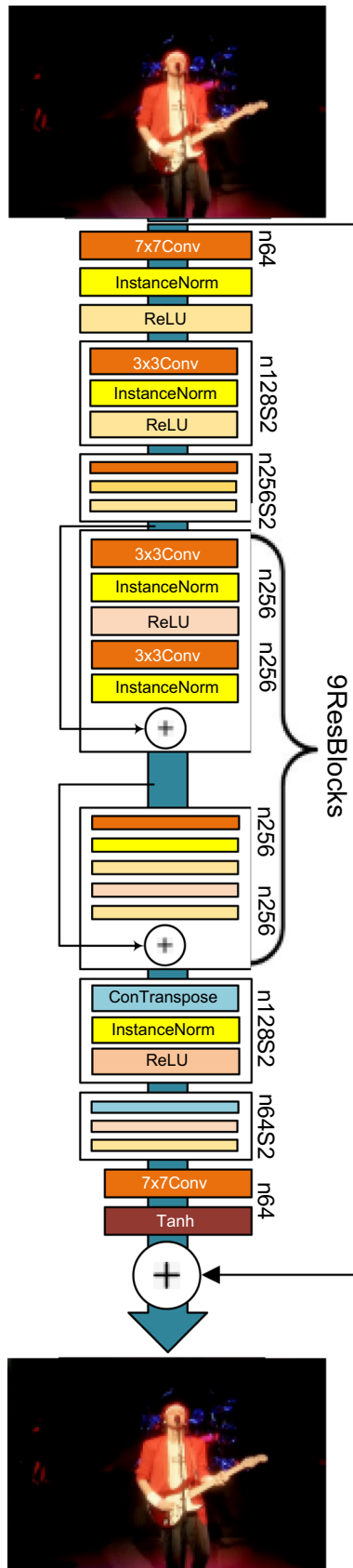


Fig. 2 Structure of generator CNN

is residual, namely $I_S = I_B + I_R$. It makes training faster and model generalization more powerful.

The network structure of the discriminator is the same as that of PatchGAN.

2.1.2 Loss function

The loss function uses the sum of “content loss” and “adversarial loss”:

$$L = \underbrace{L_{GAN}}_{\text{adv loss}} + \underbrace{\lambda \cdot L_X}_{\text{content loss}} = \text{total loss}$$

In this experiment, $\lambda = 100$.

2.1.3 Adversarial loss

When training the original GAN (vanilla GAN), the problems of gradient disappearance and mode collapse may be encountered, for which it is very difficult to train. The Wasserstein GAN (WGAN) proposed later uses the Wasserstein-1 distance to make training less difficult. Later, Gulrajani and others proposed adding “gradient penalty” item, which further improved the stability of training. WGAN-GP achieves stable training on various GAN structures and hardly needs to adjust the super-parameters. This paper uses WGAN-GP, adversarial loss formula as follows:

$$L_{GAN} = \sum_{n=1}^N -D_{\theta_D}(G_{\theta_G}(I^B))$$

2.1.4 Content loss

Content loss is to evaluate the difference between the generated clear image and ground truth. Two commonly used options are L1 (also known as MAE, mean absolute error) loss and L2 (also known as MSE) loss. Recently, “perceptual loss” has been proposed, which is essentially a L2 loss, but it calculates the distance between feature map generated by CNN and feature map of ground truth. Definitions are as follows:

$$L_X = \frac{1}{W_{ij}H_{ij}} \sum_{x=1}^{W_{ij}} \sum_{y=1}^{H_{ij}} (\phi_{ij}(I^S)_{x,y} - \phi_{ij}(G_{\theta_G}(I^B))_{x,y})^2$$

where ϕ_{ij} represents the feature map of the output of the j th convolution layer before the first max pooling layer (after activation) after the image is input into VGG19 (pre-trained on ImageNet). $W_{ij}H_{ij}$ represents the dimension of feature map.

2.1.5 Motion blur generation

Compared with other image-to-image translation tasks, such as super-resolution and stylization, it is difficult to obtain clear-blurred image pairs for training. A common method is to use high-speed camera to shoot the video, get clear images from video frames and synthesize blurred images. Another method is to use a variety of “blur kernels” on clear image convolution to obtain synthetic blurred images. DeblurGAN is further expanded on the basis of the existing second method. The proposed method can simulate a more complex “blur kernel.”

Firstly, DeblurGAN adopts Boracchi and Foi’s method of random motion trajectory generation [40], which is generated by Markov random process, and then generates blur kernel by “sub-pixel interpolation” of trajectory.

2.2 VDSR

VDSR is based on the residual network ResNet [41] proposed by He Kaiming in 2015. ResNet has solved the problem that can not be trained when the network structure is deep, and its performance has been improved. Residual network structure has been applied in a lot of work [42].

As was mentioned by the author in the VDSR paper, the input low-resolution image and the output high-resolution image are similar to each other to a great extent. That is to say, the low-frequency information carried by the low-resolution image is similar to the low-frequency information of the high-resolution image. It takes a lot of time to carry this part in training. In fact, we only need to learn high-resolution image and low-resolution image. The idea of residual network structure is particularly suitable for solving super-resolution problems, which can be said to affect the subsequent in-depth learning of super-resolution methods. VDSR is the most direct and obvious structure of learning residuals. Its network structure is shown in Fig. 3.

VDSR takes the low-resolution image which becomes the target size after interpolation as the input of the network and then adds the image and the residual learned by the network to get the final network output. VDSR has four main contributions: 1. deepening the network structure (20 layers), so that the deeper the network layer, the greater the field of feeling. This paper chooses a convolution core of 3×3 , and the network with the depth of D has $(2D + 1) \times (2D + 1)$ field of perception. 2. With residual learning, the residual image is sparse and most of the values are 0 or smaller, so the convergence speed is fast. VDSR also applies adaptive gradient clipping, which limits the gradient to a certain range, and can speed up the convergence process. 3. VDSR completes 0 operations on the image before each convolution, which ensures that all the feature maps and the final output image are consistent in size, and

solves the problem that the image will become smaller and smaller through gradual convolution. Experiments show that the predicted results of the complement 0 operation for boundary pixels can also be improved. 4. VDSR trains images of different multiples together, so that a model trained can solve the problem of super-resolution of different multiples.

2.3 Regressing 3DMM parameters with a CNN

We use CNN to adjust the 3DMM face shape parameter [43] according to the input face image. At present unconstrained face and 3D ground truth data sets are too small for training depth neural networks. However, we find three advantages.

Firstly, 3D face can be accurately estimated by multiple images of the same face.

Secondly, there are many data sets of multiple pictures of a single individual at present.

Thirdly, there are currently very effective depth neural networks for face recognition.

2.3.1 Acquisition of training data

We adopted the recently published method of generating multi-image 3DMM [44]. We use this method to generate 3DMM on CASIA WebFace data set. These 3D face models serve as ground truth for training our CNN. Multi-image 3DMM reconstruction consists of two steps: First, 500 K images are selected from CASIA data set to estimate the parameters of 3DMM. Second, the 3DMM generated by different photographs of the same individual is aggregated to obtain a single individual’s 3DMM (about 10 K individuals).

2.3.1.1 The 3DMM representation Our system uses the popular Basel face model (BFM), which is the best single view 3D model currently open. The generation model of a face includes two parts: face type and texture. The generating function is:

$$S' = \hat{s} + W_S \alpha, T' = \hat{t} + W_T \beta.$$

2.3.1.2 Single-image 3DMM fitting Two different methods were used to match each training picture with 3D MM. For image I , we estimate that α^* and β^* represent images similar to image I . The best face feature point detector (CLNF) is used to detect $K = 68$ face feature points $P_k \in R^2, k \in 1 \dots K$ and confidence values ω . Face feature points are used to initialize the angle of the input face in the 3DMM coordinate system. The angle is expressed as six degrees of freedom: angle $r = [r_\alpha, r_{\beta a}, r_{\gamma a}]$ and translation

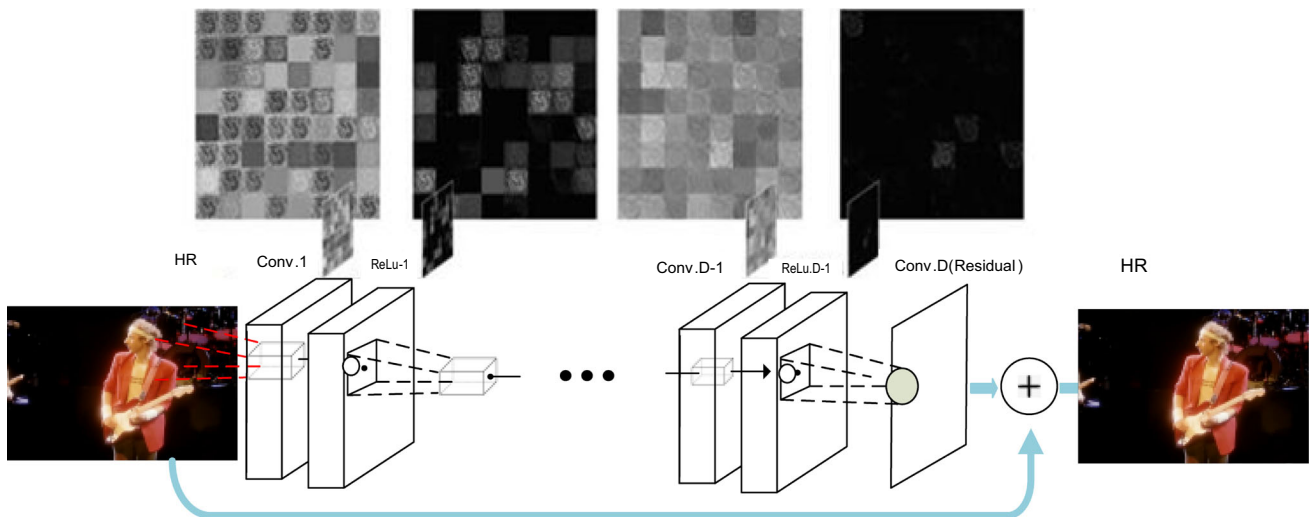


Fig. 3 VDSR network structure

$t = [t_X, t_Y, t_Z]$. Then, face shape, texture, angle, light and color are processed.

2.3.1.3 Multi-image 3DMM fitting Multi-image 3D MM generation is realized by the facial and texture parameters of 3D MM generated by different images of individual pool.

$\bar{\gamma} = \sum_{i=1}^N \omega_i \cdot \gamma_i$, where $\sum_{i=1}^N \omega_i = 1$ is the confidence values generated for CLNF face feature detection.

2.3.2 Learning to regress pooled 3DMM

For each individual in the data set, there are multiple images and a single pool’s 3DMM. We use this data to train the model so that the model can generate similar 3DMM feature vectors according to different pictures of the same individual.

As shown in Figs. 1, 2 and 3, we use 101-layer deep ResNet network for face recognition. The output layer of the neural network is a 198-dimensional 3DMM eigenvector. Then, the pooled 3DMM generated by CASIA image is used as the target value to fine-tune the neural

network. We also tried using the VGG-16 structure, which turned out to be slightly worse than the ResNet structure.

2.3.2.1 The asymmetric Euclidean loss In our experiments, we found that using Euclidean loss would result in a lack of detail in the output of 3D faces. Therefore, we introduce asymmetric Euclidean loss.

$$L(\gamma_p, \gamma) = \lambda_1 \cdot \underbrace{\|\gamma^+ - \gamma_{\max}\|_2^2}_{\text{over-estimate}} + \lambda_2 \cdot \underbrace{\|\gamma_p^+ - \gamma_{\max}\|_2^2}_{\text{under-estimate}},$$

using the element-wise operators:

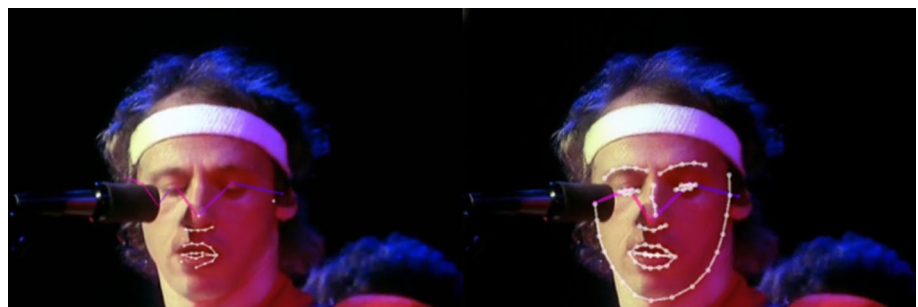
$$\gamma^+ \doteq \text{abs}(\gamma) \doteq \text{sign}(\gamma) \cdot \gamma$$

$$\gamma_p^+ \doteq \text{sign}(\gamma) \cdot \gamma_p$$

$$\gamma_{\max} \doteq \max(\gamma^+, \gamma_p^+).$$

Among them, γ is the target pooled 3DMM value, γ_p is the input value and $\gamma_{1,2}$ is the balance value of over- and under-estimation errors. In practice, we set $\lambda_1 = 1, \lambda_2 = 3$ to encourage the model to learn more details.

Fig. 4 (Left) Face detection on the original image; (right) face detection on the deblur image



2.3.2.2 Neural network super-parameters We use SGD optimizer with mini-batch of 144, momentum of 0.9 and L2 weight decay of 0.0005 to train the model. The learning rate is 0.01. When the verification set loss is saturated, we reduce the learning rate until the verification set loss stops decreasing.

2.3.3 Parameter based on 3D–3D recognition

After training, the CNN neural network can convert the input image into 3DMM parameter γ_p , which is $f : I \rightarrow \gamma_p$. We use this parameter γ_p as face feature for face recognition. The similarity formula of two faces is as follows:

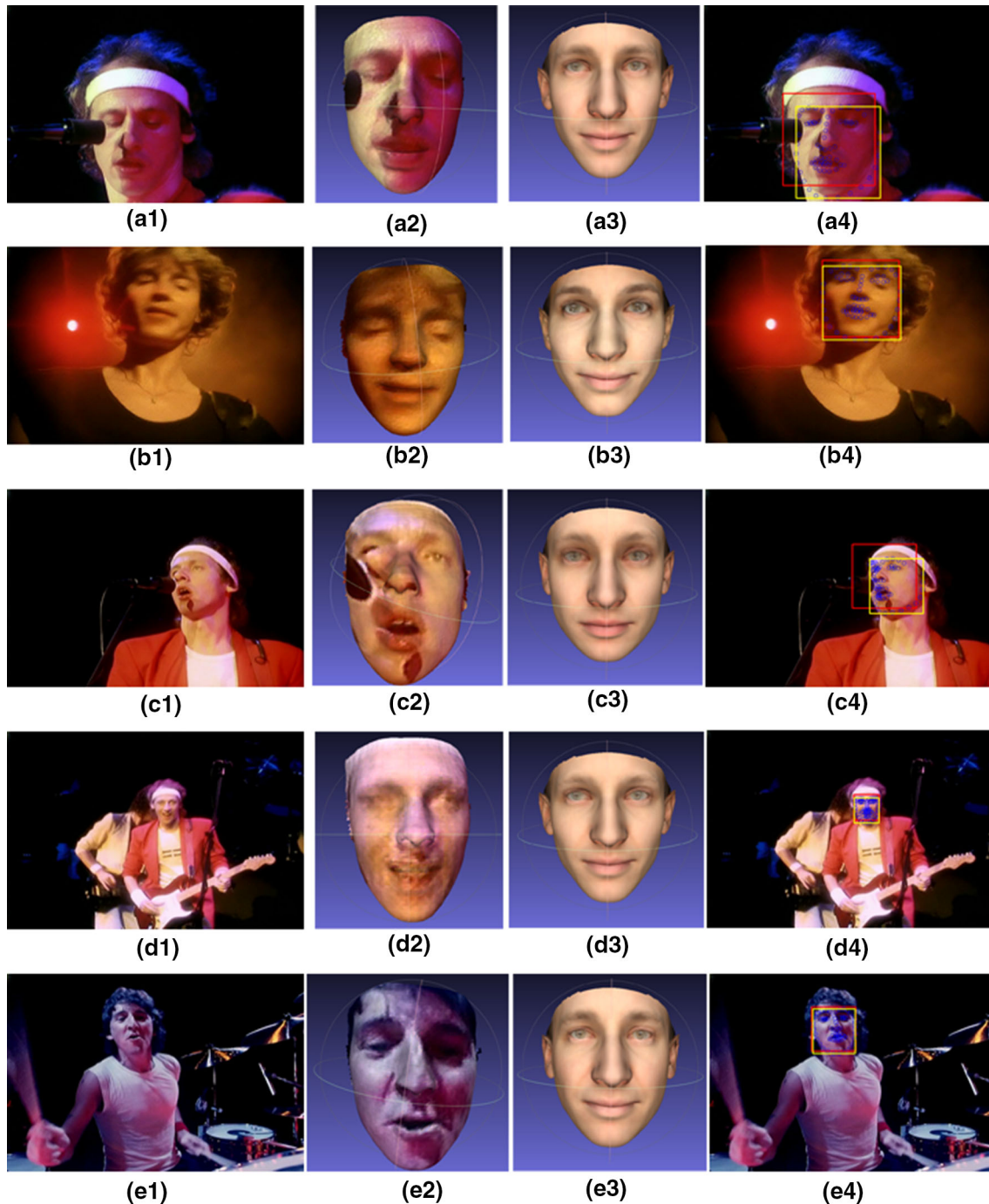


Fig. 5 Comparison of effects of 3DMM_CNN and PRNET

$$s(\gamma_1, \gamma_2) = \frac{\gamma_{p1} \cdot \gamma_{p2}^T}{\|\gamma_{p1}\| \cdot \|\gamma_{p2}\|}.$$

In some cases, a single individual has a set of pictures. For example, in the YTF data set, video contains multiple frames of a single individual; in the IJB-A data set, we use multiple data sources (pictures, videos) to simulate the 3DMM of each face frame in each video to get an average 3DMM parameter.

3 Experiment

Figure 4 is a comparison of face detection between the original image and the deblur image. As can be seen, in the left image, the effect of face detection is not ideal. Only the nose and mouth were detected, but the outlines of eyes, eyebrows and faces were not detected. In the right image, the outline of the face, eyebrows, eyes, mouth, and nose are well detected. This shows that the Deblur program improves the clarity of the blurred image and the recognition rate of image detection. The rendering of the face contour is using OpenPose [45].

Figure 5 is a comparison between the effects of Robust3D and PRNET. Among them, (a1)–(e1) are screenshots of concert videos after Deblur and Super-resolution; (a2)–(e2) are three-dimensional face models generated after PRNET processing; (a3)–(e3) are three-dimensional face models generated after the treatment of Robust3D; (a4)–(e4) are the results of facial feature detection with Robust3D. As shown in the figure, the image generated by PRNET is more realistic and similar to the original image, but it is greatly affected by the environment and does not filter out the environmental noise. For example, the right face of the three-dimensional face generated in (a2) has a black spot on it, which is the microphone. In (a3), there is no microphone noise. This is because (a3) using the method of 3DMM, only the features of the face and other effective factors are collected, so the environmental noise is filtered out. Similarly, in (c2) there are environmental noises such as microphones, and in (c3), there are no environmental noises such as microphones. In addition, in (d2), the features of human face such as human eyes are not clear and there is a phenomenon of human eyes missing, which is due to the small face in the original image (d1). In (d3), human eyes and other features are very clear, which is also the advantage of using the 3DMM method.

4 Conclusion

By training CNN, the input image can be converted to 3DMM parameters and 3D face reconstruction can be performed. The accuracy of the Robust3D method is higher than that of the existing methods. At the same time, the performance is excellent on the live video data set of the concert. The disadvantage of using Robust3D is that there are some differences between the reconstructed face and the real face. In future, we will improve our work by Big data technology [46, 47].

Acknowledgements This work was funded by the European Research Council (ERC) Advanced Grant to M. Slater, MoTIVE, ERC-2016-ADG 742989.

Compliance with ethical standards

Conflict of interest The author declares that there is no conflict of interest.

References

- Zhang R, Tsai PS, James EC et al (1999) Shape from shading: a survey. *IEEE Trans Pattern Anal Mach Intell* 21(8):690–706
- Blanz V, Vetter T (1999) A morphable model for the synthesis of 3D faces. In: *Proceeding of the SIGGRAPH'99*. Los Angeles, USA, ACM Press, New York
- Blanz V, Vetter T (2003) Face recognition based on fitting a 3D morphable model. *IEEE Trans Pattern Anal Mach Intell* 25(9):1063–1074
- Blanz V, Scherbaum K, Vetter T, Seidel H (2004) Exchanging faces in images. *Comput Gr Forum* 23(3):669–676
- Blanz V, Vetter T (1999) Morphable model for the synthesis of 3D faces. In: *Proceedings of the ACM SIGGRAPH conference on computer graphics*
- Chu B, Romdhani S, Chen L (2014) 3D-aided face recognition robust to expression and pose variations. In: *Proceedings of the conference on computer vision pattern recognition*
- Paysan P, Knothe R, Amberg B, Romhani S, Vetter T (2009) A 3D face model for pose and illumination invariant face recognition. In: *International conference on advanced video and signal based surveillance*
- Romdhani S, Vetter T (2003) Efficient, robust and accurate fitting of a 3D morphable model. In: *Proceedings of the international conference on computer vision*
- Tang H, Hu Y, Fu Y, Hasegawa-Johnson M, Huang TS (2008) Real-time conversion from a single 2d face image to a 3D text-driven emotive audio-visual avatar. In: *International conference on multimedia and expo*, pp 1205–1208, IEEE
- Yang F, Wang J, Shechtman E, Bourdev L, Metaxas D (2011) Expression flow for 3D-aware face component transfer. *ACM Trans Gr* 30(4):60
- Blanz V, Romdhani S, Vetter T (2002) Face identification across different poses and illuminations with a 3d morphable model. In: *International conference on automatic face and gesture recognition*, pp 192–197

12. Blanz V, Vetter T (2003) Face recognition based on fitting a 3d morphable model. *Trans Pattern Anal Mach Intell* 25(9):1063–1074
13. Chu B, Romdhani S, Chen L (2014) 3D-aided face recognition robust to expression and pose variations. In: *Proceedings of the conference on computer vision and pattern recognition*
14. Hu G, Yan F, Chan C-H, Deng W, Christmas W, Kittler J, Robertson NM (2016) Face recognition using a unified 3D morphable model. In: *European conference on computer vision*, Springer, Berlin, pp 73–89
15. Paysan P, Knothe R, Amberg B, Romdhani S, Vetter T (2009) A 3D face model for pose and illumination invariant face recognition. In: *International conference on advanced video and signal based surveillance*
16. Romdhani S, Vetter T (2003) Efficient, robust and accurate fitting of a 3D morphable model. In: *Proceedings of the 9th IEEE international conference on computer vision*. Nice, France, IEEE Computer Society
17. Romdhani S, Blanz V, Vetter T (2002) Face identification by fitting a 3D morphable model using linear shape and texture error functions. In: *Proceedings of the ECCV'02*. London, UK, Spring, Berlin, 2002, pp 3–19
18. Jiang Dalong H, Yuxiao YS et al (2005) Efficient 3D reconstruction for face recognition. *Pattern Recognit* 38(6):787–798
19. Liang S, Shapiro LG, Kemelmacher-Shlizerman I (2016) Head reconstruction from internet photos. In: *European conference on computer vision*, Springer, Berlin, pp 360–374
20. Piotraschke M, Blanz V (2016) Automated 3D face reconstruction from multiple images using quality measures. In: *Proceedings of the conference computer vision pattern recognition*
21. Roth J, Tong Y, Liu X (2015) Unconstrained 3D face reconstruction. In: *Proceedings of the conference on computer vision pattern recognition*, June
22. Roth J, Tong Y, Liu X (2016) Adaptive 3D face reconstruction from unconstrained photo collections. In: *Proceedings of the conference on computer vision pattern recognition*
23. Suwajanakorn S, Kemelmacher-Shlizerman I, Seitz SM (2014) Total moving face reconstruction. In: *European conference on computer vision*, Springer, Berlin, pp 796–812
24. Hassner T (2013) Viewing real-world faces in 3D. In: *Proceedings of the IEEE international conference on computer vision*, IEEE, pp 3607–3614
25. Hassner T, Harel S, Paz E, Enbar R (2015) Effective face frontalization in unconstrained images. In: *Proceedings of the conference computer vision pattern recognition*
26. Masi I, Tran A, Hassner T, Leksut JT, Medioni G (2016) Do we really need to collect millions of faces for effective face recognition? In: *European conference on computer vision*
27. Taigman Y, Yang M, Ranzato M, Wolf L (2014) Deepface: closing the gap to human-level performance in face verification. In: *Proceedings of the conference computer vision pattern recognition*, IEEE
28. Richardson E, Sela M, Kimmel R (2016) 3d face reconstruction by learning from synthetic data. In: *International conference on 3D vision*
29. Masi I, Tran A, Hassner T, Leksut JT, Medioni G (2016) Do we really need to collect millions of faces for effective face recognition? In: *European conference on computer vision*
30. Kemelmacher-Shlizerman I, Basri R (2011) 3D face reconstruction from a single image using a single reference face shape. *Trans Pattern Anal Mach Intell* 33(2):394–405
31. Dovgand R, Basri R (2004) Statistical symmetric shape from shading for 3D structure recovery of faces. In: *European conference on computer vision*, pp 99–113
32. Hassner T (2013) Viewing real-world faces in 3D. In: *Proceedings of the international conference on computer vision*, IEEE, pp 3607–3614
33. Hassner T, Basri R (2006) Example based 3D reconstruction from single 2D images. In: *Proceedings of the conference on computer vision pattern recognition workshops*, IEEE
34. Taigman Y, Yang M, Ranzato M, Wolf L (2014) Deepface: closing the gap to human-level performance in face verification. In: *Proceedings of the conference on computer vision pattern recognition*, IEEE
35. Huber P, Hu G, Tena R, Mortazavian P, Koppen W, Christmas W, Rtsch M, Kittler J (2016) A multiresolution 3D morphable face model and fitting framework. In: *International conference on computer vision theory and applications*
36. Jourabloo A, Liu X (2016) Large-pose face alignment via CNN-based dense 3D model fitting. In: *Proceedings of the conference on computer vision pattern recognition*
37. Saito S, Li T, Li H (2016) Real-time facial segmentation and performance capture from RGB input. In: *European conference on computer vision*
38. Zhu X, Lei Z, Liu X, Shi H, Li SZ (2016) Face alignment across large poses: a 3D solution. In: *Proceedings of the conference on computer vision pattern recognition*
39. Kupyn O, Budzan V, Mykhailych M, Mishkin D, Matas J (2018) Deblurgan: blind motion deblurring using conditional adversarial networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 8183–8192
40. Boracchi G, Foi A (2012) Modeling the performance of image restoration from motion blur. *IEEE Trans Image Process* 21(8):3502–3517
41. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 770–778
42. Kim J, Kwon Lee J, Mu Lee K (2016) Accurate image super-resolution using very deep convolutional networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 1646–1654
43. Tuan Tran A, Hassner T, Masi I, Medioni G (2017) Regressing robust and discriminative 3D morphable models with a very deep neural network. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 5163–5172
44. Piotraschke M, Blanz V (2016) Automated 3d face reconstruction from multiple images using quality measures. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 3418–3427
45. Cao Z, Hidalgo G, Simon T, Wei SE, Sheikh Y (2018) OpenPose: realtime multi-person 2D pose estimation using part affinity fields. *arXiv preprint arXiv:1812.08008*
46. Shi R, Gan Y, Wang Y (2018) Evaluating scalability bottlenecks by workload extrapolation. In: *2018 IEEE 26th international symposium on modeling, analysis, and simulation of computer and telecommunication systems (MASCOTS) 2018 Sep 25*, IEEE, pp 333–347
47. Shi R, Wang Y (2016) Cheap and available state machine replication. In: *2016 {USENIX} annual technical conference ({USENIX}{ATC} 16) 2016*, pp 265–279

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.