State-of-the-art in 3D face reconstruction from a single RGB image

Haibin Fu[1], Shaojun Bian[1,3], Ehtzaz Chaudhry[1], Andres Iglesias[2], Lihua You[1],

Jian Jun Zhang[1]

[1] Bournemouth University, Bournemouth BH12 5BB, UK
[2] University of Cantabria, Cantabria 39005, Spain
[3] Humain Ltd, Belfast BT1 2LA, UK

Abstract. Since diverse and complex emotions need to be expressed by different facial deformation and appearances, facial animation has become a serious and on-going challenge for computer animation industry. Face reconstruction techniques based on 3D morphable face model and deep learning provide one effective solution to reuse existing databases and create believable animation of new characters from images or videos in seconds, which greatly reduce heavy manual operations and a lot of time. In this paper, we review the databases and state-of-the-art methods of 3D face reconstruction from a single RGB image. First, we classify 3D reconstruction methods into three categories and review each of them. These three categories are: Shape-from-Shading (SFS), 3D Morphable Face Model (3DMM), and Deep Learning (DL) based 3D face reconstruction. Next, we introduce existing 2D and 3D facial databases. After that, we review 10 methods of deep learning-based 3D face reconstruction and evaluate four representative ones among them. Finally, we draw conclusions of this paper and discuss future research directions.

Keywords: Monocular RGB Image, 3D Face Reconstruction, 3D Morphable Model, Shape-from-shading , Deep Learning, 3D Face Database.

1   Introduction

Generating 3D face models from 2D images and videos is widely required in industry and academia. It greatly reduces time-consuming and repetitive manual operations of creating 3D shape sequences for every new character. Face reconstruction techniques based on 3D Morphable Face Model and deep learning provide one efficient solution to reuse existing databases and create believable facial animation of new characters just from images or video frames in seconds, which greatly reduce manual operations and loads of time.

3D face reconstruction is widely used in commerce and industry. Traditional 3D reconstruction methods can be divided into two categories: 1) active methods such as Structural light and Laser scanning, and 2) passive methods like Multi-view stereo, Shape-from-shading (SFS) and Data-driven reconstruction. Face reconstruction from a single image has been widely investigated in the fields of computer vision and com-

puter graphics. There are many methods tackling this problem. Such as 3DMM methods, SFS, and Deep learning based methods.

Apart from the above static 3D face reconstruction from a single image, reliable detection and recognition of faces has been an on-going research topic for decades. Face detection, face recognition and realistic face tracking remain a challenging task in dynamic environments, such as videos, where the problems of big data and low computational efficiency become much more serious. Reducing the data size but still keeping good realism and details is especially important for 3D face reconstruction in these dynamic environments.

In this paper, we will review the state-of-the-art methods of 3D face reconstruction from a single image. The review will include three categories of 3D reconstruction methods, facial databases, and deep learning-based 3D face reconstruction methods with open source code.

2   Single image-based 3D reconstruction methods

3D reconstruction is the process of capturing the geometry and appearance of real objects. Face reconstruction from a single image (analysis by synthesis) has been investigated extensively in computer vision and computer graphics communities. To restore the geometry as well as the skin color of the face needs to obtain geometry information and color information from the image. The first step is to choose the geometry representation of a face model. It could be point cloud, depth map/normal map or parametric model. Among them, a parametric model is the most widely used representation. Although faces are different, the geometric mechanism of faces is similar, which lays a foundation for expressing a complex face with few parameters and parameterized expression.

Main 3D face reconstruction methods can be classified into three categories: Shape-From-Shading (SFS), 3D Morphable Face Model (3DMM), and Deep Learning (DL) based 3D face reconstruction. In what follows, we will briefly review them. Since Deep Learning (DL) based 3D face reconstruction is becoming a hot topic, we will give more details about various deep learning based 3D face reconstruction methods with open source code and evaluate four representative ones in Section 4, after we make a survey on existing databases for face reconstruction in Section 3.

2.1 Shape-from-Shading (SFS) based 3D face reconstruction

Shape-from-Shading [1] is a very basic and classic 3D reconstruction method. The basic principle is to use the brightness information of grayscale images and the principle of brightness generation to obtain the normal vector of each pixel in a 3D space, and finally obtain the depth information according to the normal vector. We can get the geometric structure of an object by analyzing the changes of light and shade on the surface of the object. The reconstruction process is shown in Fig. 1. Shape from shading is known to be an ill-posed problem [2]. Therefore, SFS needs a reliable initial 3D shape as a reference to manually align the input facial image. With this prior

information, shape from shading is then used to restore the geometry. Features of facial symmetry have been used by many researchers to constrain this problem [3]. As an efficient and simple low dimensional representation, 3DMM to be reviewed below is widely used as an initial facial shape [4]. SFS can restore fine geometric details by optimizing the process given the light coefficient and reflection parameters. In the recent research study [5], the refinement exploits shading cues in the input image to augment the medium-scale model with fine-scale static and transient surface details. Compared with traditional 3DMM, SFS can get better details of a human face. It is a good way to add details such as acne skin and wrinkles to the 3D face model.

Fig. 1. Shape-from-shading based reconstruction: input image, estimated normal map, and rendered model (from left to right) [16].

2.2 3D Morphable Face Model (3DMM) based 3D face reconstruction

Reconstructing a 3D face shape from a single 2D photograph or from a video frame is an inherently ill-posed problem with many ambiguities. One way of solving some of the ambiguities is to use a 3D face model to aid the task. 3D Morphable Face Models (3DMMs) also called 3D model fitting have been developed for this aim. In 1999, Vetter and Blantz [6] introduced the 3DMM for the synthesis of 3D faces, a principal component analysis (PCA) basis for representing faces. It is one of the most commonly used methods for face reconstruction from a single image [7]. The advantage of using the 3DMM is that the solution space is constrained to represent only likely solutions, so that the problem could be simplified. In 2016, Zhu [8] proposed an automatic reconstruction process for 3DMM. Still, the automated initialization pipelines usually do not produce the same quality of reconstructions when only one image is used. In addition, the 3DMM solutions cannot extract fine details since they are not spanned by the principal components. The common 3DMMs used for facial tracking include Surrey Face Model (SFM), the 3D morphable face model created at the University of Surrey by Huber [9], and Basel Face Model (BFM) created at the University of Basel by Paysan et al. [10]. BFM is a generative 3D shape and texture model. It uses the principal component regression (PCR) model to carry out regression 2D face markers and reconstruct 3D faces with higher shape and texture accuracy (Fig. 2).

Although the 3DMM greatly reduces the data size and raises the computational efficiency, it cannot describe 3D models with high details and good realism due to the reduction of the computational accuracy.
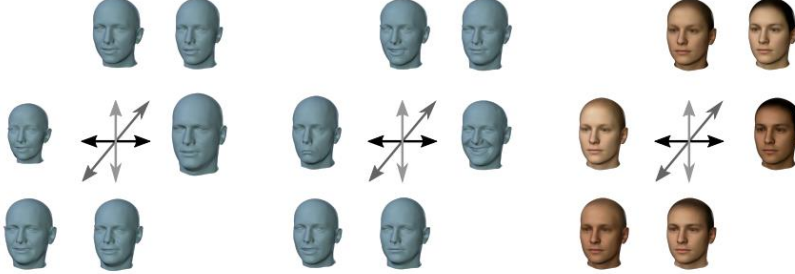
Fig. 2. BFM 2019 [12] shape, expression, and appearance variations.

2.3 Deep Learning (DL) based 3D face reconstruction

In the past few years, deep learning methods have revolutionized computer vision [12]. With the development of deep leaning, the classical 3D morphable face modelling and parameter estimation techniques are being replaced by or combined with deep learning [13]. Due to the speed and robustness of deep network, reliable performance is also achieved on large poses and in-the-wild images. Moreover, deep learning methods are adapted at extract fine details like wrinkles. Clearly, deep learning has made some achievements in 3D face reconstruction and face appearance modeling, especially in the following four aspects: frontal face to arbitrary poses, occlusion handling, separation of training and testing, synthetic data and unsupervised real image [14].

Deep learning-based face reconstruction methods normally contain 2 parts: a database and pipeline of 3D face reconstruction process (Fig. 3). One of the most important factors for the success of deep learning-based 3D face reconstruction is the availability of large amounts of training data [15]. We will discuss 2D and 3D face data-bases in Section 3.

The pipeline of 3D face reconstruction process can be classified as coarse face reconstruction and fine detail (wrinkles and pores) face reconstruction. For the coarse face reconstruction, face landmarks or 3DMM parameters are treated as supervision signal for training. Dense fine detail reconstruction methods predict dense shape variation rather than low-dimensional parameters.
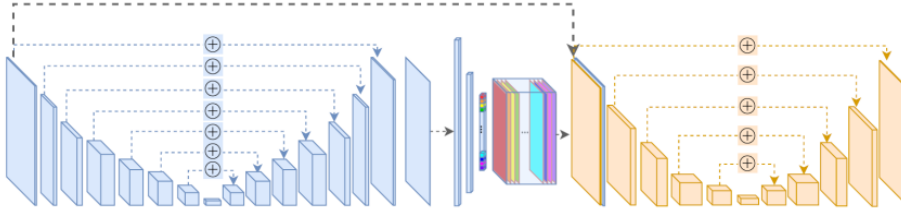

Fig. 3. Pipeline of deep learning based 3D face reconstruction Network architecture for DFDN [15], blue part for coarse model synthesis, and yellow part for fine detail synthesis.

3 Face Databases

Face databases are fundamental for 3D face reconstruction and evaluation. In this section, we review existing face databases. In general, existing face databases could be divided into two categories: 2D face databases (Table 1), and 3D face databases (Table 2).

Table 1. Comparison of 2D face databases

Dataset	Subjects	landmarks	Images	Resolution	Male	Year	Ref
PUT	100	30	9,971	2048x1536	89%	2008	[18]
Multi-PIE	337	68	750k	3072x 204	-	2010	[19]
MUCT	276	76	3,375	480x640	49%	2010	[20]
SCface	130	21	4.16k	1200x1600	88%	2011	[21]
AFLW	-	21	25993	-	41%	2011	[22]
GBU	437	-	6.51k	128x128	58%	2012	[23]
Helen	-	194	20330	1200x900	-	2012	[24]
300-W	300	68	600	16.2k, 3.3M	-	2013	[25]
FaceScrub	530	-	10.6k	-	50%	2014	[26]
IJB-A	500	-	25k	-	-	2015	[27]
VGG-Face	2,622	-	2.6M	256x256	47%	2015	[28]
CelebA	10,177	5	202k	512x512	50%	2015	[29]
UMDFaces	8,277	21	8277	-	-	2016	[30]
MegaFace	690,572	49	1m	100x100	-	2016	[31]
LS3D-W	-	68	230k	-	-	2017	[32]
LFW	5,749	-	13k	250x250	77%	2018	[33]
FFHQ	-	-	70k	1024×1024		2019	[34]
Tufts	113	-	10k	-	34%	2020	[35]

Table 1 above shows some popular 2D face databases since 2008. It provides the information about the numbers of subjects, landmarks, and images as well as resolution, gender of the subjects, year of creation, and the publications for these 2D face databases.

Among the 2D face databases shown in Table 1, the most commonly used 2D face databases are Multi-PIE [19], 300W [25] and CelebA [29]. The CMU Multi-PIE face database contains more than 750,000 images of 337 people with a range of facial expressions [19]. 300W is well known because it was the first facial landmark localization Challenge [25]. CelebFaces Attributes Dataset (CelebA) is a large-scale face attributes dataset with more than 200K celebrity images, each with 40 attribute annotations [29].

Table 2 below shows existing 3D face databases since 2006. The methods used to create 3D databases can be either active or passive. Active methods are Structured

Light Systems (SLS), Laser Scanners (LS) and Time-of-Flight (TOF) sensors. Passive approach could be Multi View Stereo (MVS).

Table 2. Comparison of 3D face databases

Dataset	Subjects	Express	Vertex	Source[*]	Male	Texture	Year	Ref
ND2006	888	6	112k	LS	-	640×480	2006	[36]
BU-3DFE	100	25	10-20k	SLS	44%	1300×900	2006	[37]
BU-4DFE	101	6	30-50k	SLS	43%	1040×1329	2008	[38]
UoY	350	10	5-6k	KN	-	360 x 240	2008	[39]
Bosphorus	105	35	35k	SLS	-	1600×1200	2009	[40]
BJUT-3D	1200	1-3	200k	LS	50%	478 ×489	2009	[41]
D3DFACS	10	38AU	30k	MVS	40%	1024×1280	2011	[42]
Florence	53	1	30-40k	KN	75%	3341×2027	2011	[43]
FaceWarehouse	150	20	11k	KN	-	640×480	2014	[44]
BP4D-Spontan	41	27AU	37k	MVS	44%	1040×1392	2014	[45]
BP4D+	140	7AU	30-50k	MVS	41%	1040×1392	2016	[46]
UHDB31	77	1	25k	MVS	69%	2048×2448	2017	[47]
4DFAB	180	6	100k	KN	66%	1200×1600	2018	[48]
EB+	60	7AU	30-50k	MVS	41%	1040×1392	2019	[49]
Facescape	938	20	2m	MVS	-	4096×4096	2020	[50]
BFM	200	-	53490	SLS	50%	768 x 1024	2009	[9]
SFM	169	-	29587	SLS	-	768 x 1024	2016	[10]
LSFM	9663	-	53215	-	48%		2017	[51]
LYHM	1212	-			50%		2017	[52]

*MVS-Multi view stereo, KN-Kinect, SLS-Structure light system, LS-Laser scan.

3D face databases normally contain facial images aligned with their ground truth 3D shapes. These databases are essential for 3D face reconstruction and evaluation. For example, Basel face model (BFM) [9] are commonly used for 3D face reconstruction, and BU-3DFE (Binghamton University 3D Facial Expression), BU-4DFE and BP4D-Spontaneous are normally used for evaluations.

4   3D face reconstruction methods based on deep learning

In this section, we review the recent development of open source 3D face reconstruction methods based on deep learning. First, we identify 10 methods with open source code, which were published between 2018 and 2020. Then, we evaluate four representative ones using their open source code.
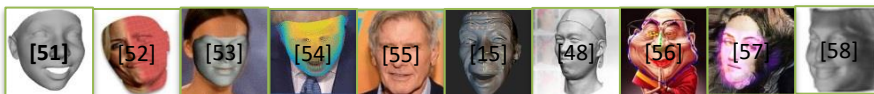
Fig. 4. An overview of state of the art 3D face reconstruction methods

Table 3. List of 2018 to 2020 open source 3D face reconstruction methods.

Method	Dataset	Vertices	Coarse	Fine	Year	Ref
ExpNet	CK+, EmotiW-17	-	√		2018	[53]
ITW	300W, BFM	5k	√		2018	[54]
MGCNet	300W-LP, Multi-PIE	36k	√	√	2018	[55]
PRNet	300W-LP, BFM	43,867	√		2018	[56]
DF2Net	CACD, BU- 3DFE	7-13k		√	2019	[57]
DFDN	AffectNet	53,149	√	√	2019	[15]
Facescape	Facescape	12,483	√	√	2020	[50]
Caricature	WebCaricature	6144	√		2020	[58]
3FabRec	300W	-		√	2020	[59]
UnsupNet	CelebA, BFM	-		√	2020	[60]

The identified 10 methods with open source code are shown in Table 3 and Fig. 4. Among them, UnsupNet [60], PRNet [56], DF2Net [57], and DFDN [17] will be evaluated in Subsections 4.1, 4.2, 4.3, and 4.4, respectively. In what follows, we briefly review other methods listed in the table 3.

ExpNet [53] produces expression coefficients, which better discriminate between facial emotions than those obtained using the state-of-the-art facial landmark detection techniques. 3D Face Morphable Models In-the-Wild (ITW) proposed in [54] provide a new fast algorithm for fitting the 3DMM to arbitrary images. The work captures the first 3D facial database with relatively unconstrained conditions. Multi-view Geometry Consistency (MGCNet) [55] first proposes an occlusion-aware view synthesis method to apply multi-view geometry consistency to self-supervised learning. Then, it designs three novel loss functions for multi-view consistency, which are the pixel consistency loss, the depth consistency loss, and the facial landmark-based epipolar loss. Facescape [50] presents a 3D Face Dataset containing 938 subjects and provides a fine 3D face construction method. Based on the constructed dataset and a nonlinear parametric model, Caricature [58] proposes a neural network based method to regress the 3D face shape and orientation from the input 2D caricature image. Fast Few-shot Face alignment by Reconstruction (3FabRec) [58] first trains an adversarial autoencoder to reconstruct faces via a low-dimensional face embedding. Then, it interleaves the decoder with transfer layers to retask the generation of color images to the prediction of landmark heat maps.

Through the investigation, we selected 4 representative methods: UnsupNet, PRNet, DF2Net and DFDN, and run their open source code with our selected image databases. The obtained results are given and discussed in the following subsections.

4.1 UnsupNet

The 3D facial models reconstructed from UnsupNet [60] are shown in Fig. 5. Unsup-Net learns 3D deformable object categories from single RGB images without external supervision. The method is based on an automatic encoder that decomposes each input image into depth, albedo, viewpoint, and illumination. For separating these components without oversight, it works well in symmetric structures.

Compared to other methods, UnsupNet could reconstruct the 3D models of human faces, cat faces, cars and other symmetric objects from single RGB images without any supervision or prior shape model such as 3DMM. However, UnsupNet has poor performance in reconstructing 3D faces with large poses and occlusion.

Fig. 5. The result of UnsupNet

4.2 PRNet

3D face reconstruction and face alignment are closely linked problems. Although a lot of work has been done to achieve both goals by obtaining 3DMM coefficients, it is limited by the quality of the model space. PRNet [56] can simultaneously reconstruct the 3D facial structure and provide dense alignment. A UV position map is created to record 3D position under UV coordinates. By using encoder-Decoder network mode, the joint task of 3D face reconstruction and dense face alignment from single RGB face images can be realized end-to-end. By using the 300W-LP database consisting of 300 indoor and 300 outdoor images downloaded from the web [25] as a training set, it can also get better results for the faces with large variations in poses.

Fig. 6. The result of PRNet

The 3D facial models reconstructed from PRNet [56] are shown in Fig. 6. In general, it can be divided into three parts: 3D face model, UV graph representation of 3D point cloud, and network structure design. The 3D model can be built on the Basel Face Model (BFM) [9] by 3DMM coefficient, while the expression can be built by the FW Model. The sum of the two is then projected to the camera coordinate system through the Pose (attitude Angle, translation and scale) parameters.

The advantages of PRNet [56] are model-free, reference-free, and not Voxel-based. The reconstruction and alignment work fine, including dense alignment of both visible and non-visible points (including 68 key points). The disadvantages are lack of details and low texture resolution. In addition, the distinct stripes could be seen on the model.

4.3 DF2Net

A deep Dense-Fine-Finer Network (DF2Net) [57] could decompose the reconstruction process into three stages: D-Net, F-Net, and Fr-Net. Training data are generated by 3DMM. D-Net uses U-Net architecture proposed in [61] to map input images to dense depth images. F-Net refines the D-Net output by fusing depth and RGB domain characteristics, and its output is further enhanced by Fr-Net using a novel multi-resolution super column structure. In addition, three types of data: 3D model synthesis data, 2D image reconstruction data, and fine facial images, were introduced to train these networks.

Fig. 7. The result of DF2Net

The 3D facial models reconstructed from DF2Net [56] are shown in Fig. 7. The qualitative evaluation shows that DF2Net can effectively reconstruct facial details, such as small crow's feet and wrinkles. It can perform qualitative and quantitative analysis of real world images and BU3DFE datasets [37] with superior or comparable performance to the most advanced methods.

Apart from that DF2Net can reconstruct good details, especially for wrinkle and crow's feet, the reduction degree of the image is also very good. The disadvantages of DF2Net are: 1) when the mesh is incomplete, only partially detected face is reconstructed, 2) the eyes are sculptured concave, and 3) large poses, half face, and occlusions are not recognized and could not be reconstructed sometimes. Compared with

PRNet, DF2Net is less robust but can create more personalized with individual details.

4.4 DFDN

The Deep Facial Detail Net (DFDN) method [17] is adept at recovering fine geometric details like wrinkles. For facial detail synthesis, it combines with supervised and unsupervised learning. The method is based on Conditional Generative Adversarial Net (CGAN) that adopts both appearance and geometry loss functions. The 3D facial models reconstructed from DFDN [15] are shown in Fig. 8. Compared with other methods, DFDN takes the longest time to reconstruct the 3D models. One of the reasons for this is DFDN conducts emotion prediction to determine a new expression-informed proxy, which is quite useful in animation. On the other side, emotion prediction could also lead to under-regularized global shape and inaccurate prediction for single image input. For example, a neutral face pose with lip drop might lead to open mouth surprise prediction as shown in Fig. 8.
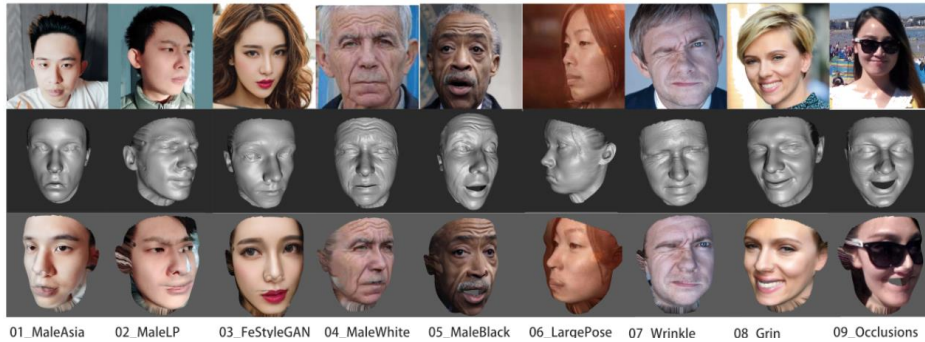

Fig. 8. The result of DFDN

4.5 Evaluations

This subsection compares the visual results of UnsupNet, PRNet, DF2Net and DFDN shown in Fig. 9. The UnsupNet method shows inaccurate global model and limited details. The global reconstruction by 3DMM and PRNet is normally over-regularized and misses details like winkles. In contrast, DF2Net and DFDN could capture details but the reconstructed 3D model is under-regularized.
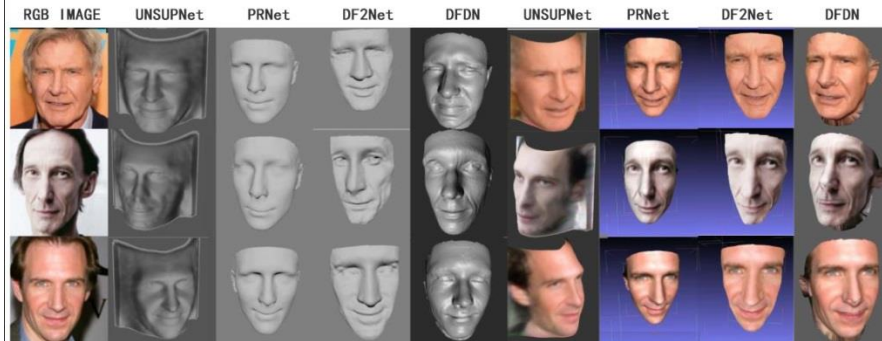
Fig. 9. The mesh and textured results obtained from UnsupNet, PRNet, DF2Net, and DFDN.

All the tests and comparisons were conducted on a desktop PC with an Intel CPU i7-8650U at 2.11 GHz, 16GB of RAM and NVIDIA GeForce GTX1060. As for the running time for each 3D reconstruction, UnsupNet takes about 12 seconds to obtain 3D mesh. PRNet takes about 90 milliseconds to obtain both 3D mesh and 68 2D landmarks. The number of vertices of reconstructed mesh is 43,867. DF2Net takes about 20 seconds to obtain point cloud and takes 420ms to obtain 3D mesh with 77,767 vertices. DFDN takes about 1 minute and 47 seconds to obtain 3D mesh, texture, and the displacement map with 53,149 vertices.

5 Conclusions and future research directions

In this paper, we have provided a detailed survey about state-of-the-art methods in 3D face reconstruction from a single RGB image. We reviewed 3D morphable model, shape from shading, and deep learning based 3D face reconstruction methods, introduced commonly used 2D and 3D face databases, identified 10 deep learning-based 3D face reconstruction methods with open source code, and evaluated and compared four representative ones. How to optimize the coarse-to-fine framework and keep the balance of global shape and fine facial details are still unsolved challenges.

In the future, the challenges are how to improve the comparability of 3DMM, find the balance of global shape and fine facial details, add the other parts of parametric models, such as eyes, teeth, hair and skin details, and develop an efficient method to improve the processing time for future scenario use. Moreover, the solution should be robustness and be able to deal with a variety of light conditions; sometimes hard shadows can lead to incorrect normal map or displacement estimates. Furthermore, the solution also needs to take into account the effects of large pose and occlusion (such as hair or glasses). Another limitation of existing techniques is that they cannot handle blurry images and low-resolution images. The geometric detail prediction scheme normally relies heavily on the resolution of pixels. Therefore, the preprocess should be able to sharp the RGB image or correct the perspective to achieve distortion-free wide-angle portrait.

Overall, although existing research has been carried out in 3D face reconstruction, the research in this field is still incomplete and there are still many problems to be studied. Realizing an automatic and real-time system capable of handling unconstrained environmental conditions will open up huge potential for new applications in the gaming, safety and health industries. We hope that this survey will provide a useful guideline for current and future researchers in this field.

Acknowledgements

References

1. Zhao, W.-Y., Chellappa, R.: Symmetric shape-from-shading using self-ratio image. International Journal of Computer Vision 45 (1), 55-75. (2001)
2. Zhang, Ruo., Tsai, P., Cryer, J.-E., Shah, M.: Shape-from-shading: a survey. IEEE transactions on pattern analysis and machine intelligence 21(8), 690-706. (1999)
3. Zhu, X., Lei, Z., Yan, J., Yi, D., Li, S.-Z.: High-fidelity pose and expression normalization for face recognition in the wild. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 787-796. (2015)
4. Li, Y., Ma, L., Fan, H., Mitchell, K.: Feature-preserving detailed 3D face reconstruction from a single image. In Proceedings of the 15th ACM SIGGRAPH European Conference on Visual Media Production, pp. 1-9. (2018)
5. Garrido, P., Zollhöfer, M., Casas, D., Valgaerts, L., Varanasi, K., Pérez, P., Theobalt, C.: Reconstruction of personalized 3D face rigs from monocular video. ACM Transactions on Graphics (TOG) 35(3), 1-15. (2016)
6. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3D faces. In Proceedings of the 26th annual conference on Computer graphics and interactive techniques, (1999)
7. Tran, L., Liu, X.: Nonlinear 3D face morphable model. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7346-7355. (2018)
8. Zhu, X., Lei, Z., Liu, X.: Face alignment across large poses: A 3D solution. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 146-155. (2016)
9. Huber, P.: Real-time 3D morphable shape model fitting to monocular in-the-wild videos. University of Surrey (2017)
10. Paysan, P., Knothe, R., Amberg, B., Romdhani, S., Vetter, T.: A 3D face model for pose and illumination invariant face recognition. In 2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance, pp. 296-301. IEEE, (2009)
11. Liu, P., Yu, H. Cang, S.: Adaptive neural network tracking control for underactuated systems with matched and mismatched disturbances. Nonlinear Dynamics, 98(2), pp.1447-1464. (2019)
12. Gerig, T., Morel-Forster, A., Blumer, C., Egger, B., Luthi, M., Schönborn, S., Vetter, T.: Morphable face models-an open framework. In 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), pp. 75-82. IEEE, (2018)
13. Egger, B., Smith, W., Tewari, A., Wuhrer, A., Zollhoefer, M.: 3D morphable face models—past, present, and future. ACM Transactions on Graphics (TOG) 39(5), 1-38 (2020)

14. Zollhöfer, M., Thies, J., Garrido, P.: State of the art on monocular 3D face reconstruction, tracking, and applications. In Computer Graphics Forum, 37(2), 523-550 (2018)

15.

16. Sun, L., Zhao, C., Yan, Z., Liu, P., Duckett, T. and Stolkin, R.: A novel weakly-supervised approach for RGB-D-based nuclear waste object detection. IEEE Sensors Journal, 19(9), pp.3487-3500. (2018) Sengupta, S., Kanazawa, A., Castillo, C.-D., Jacobs, D.-W.: SfSNet: Learning Shape, Reflectance and Illuminance of Faces in the Wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6296-6305 (2018)

17. Chen, A., Chen, Z., Zhang, G., Mitchell, K., Yu, J.: Photo-realistic facial details synthesis from single image. In Proceedings of the IEEE International Conference on Computer Vision, pp. 9429-9439 (2019)

18. Kasinski, A., Florek, A., Schmidt, A.: The PUT face database. Image Processing and Communications, 13(3-4), 59-64 (2008)

19. Gross, R., Matthews, I., Cohn, J., Kanade, T., Baker, S.: Multi-pie. Image and Vision Computing, 28(5), 807-813 (2010)

20. Milborrow, S., Morkel, J., Nicolls, F.: The MUCT landmarked face database. Pattern Recognition Association of South Africa 201(0) (2010)

21. Grgic, M., Delac, K., Grgic, S.: SCface–surveillance cameras face database. Multimedia tools and applications, 51(3), 863-879 (2011)

22. Koestinger, M., Wohlhart, P., Roth, P., Bischof, H.: Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization. In 2011 IEEE international conference on computer vision workshops, pp. 2144-2151. IEEE, (2011)

23. Lui, Y.-M., Bolme, D., Phillips, P.-J., Beveridge, J.-R.: Preliminary studies on the good, the bad, and the ugly face recognition challenge problem. In 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 9-16. IEEE, (2012)

24. Zhu, X., Ramanan, D.: Face detection, pose estimation, and landmark localization in the wild. In conference on computer vision and pattern recognition, IEEE, (2012)

25. Sagonas, C., Antonakos, E., Tzimiropoulos, G., Zafeiriou, S., Pantic, M.: 300 faces in-the-wild challenge: Database and results. Image and vision computing 47, 3-18 (2016)

26. Ng, H.-W., Winkler, S.: A data-driven approach to cleaning large face datasets. In 2014 IEEE international conference on image processing (ICIP), pp. 343-347. IEEE, (2014)

27. Klare, B.-F., Klein, B., Taborsky, E., Blanton, A.: Pushing the frontiers of unconstrained face detection and recognition: IARPA Janus Benchmark A. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1931-1939. (2015)

28. Parkhi, O.-M., Vedaldi, A., Zisserman, A.: Deep face recognition. (2015)

29. Liu, Z., Luo, P., Wang, X., Tang. X.: Deep learning face attributes in the wild. In Proceedings of the IEEE international conference on computer vision, pp. 3730-3738. (2015)

30. Bansal, A., Nanduri, A., Castillo, C.-D., Ranjan, R., Chellappa, R.: UMDfaces: An annotated face dataset for training deep networks. In 2017 IEEE International Joint Conference on Biometrics (IJCB), pp. 464-473. IEEE, (2017)

31. Shlizerman, I.-K., Seitz, S.-M., Miller, D., Brossard, E.: The megaface benchmark: 1 million faces for recognition at scale. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4873-4882.(2016)

32. Bulat, A., Tzimiropoulos, G.: How far are we from solving the 2D & 3D face alignment problem. In Proceedings of the IEEE International Conference on Computer Vision, pp. 1021-1030. (2017)

33. Huang, G.-B., Mattar, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. (2008)

34. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4401-4410. (2019)

14

35. Panetta, K., Samani, A., Yuan, X., Wan, Q., Agaian, S., Rajeev, S., Kamath, S.: A Comprehensive Database for Benchmarking Imaging Systems. IEEE Transactions on Pattern Analysis and Machine Intelligence 42(3). (2020)
36. Faltemier, T.-C., Bowyer, K.-W., Flynn, P.: Using a multi-instance enrollment representation to improve 3D face recognition. In 2007 First IEEE International Conference on Biometrics: Theory, Applications, and Systems, pp. 1-6. IEEE, (2007)
37. Yin, L., Wei, X., Sun, Y., Wang, J., Rosato, M.-J.: A 3D facial expression database for facial behavior research. In 7th international conference on automatic face and gesture recognition pp. 211-216. IEEE, (2006)
38. Yin, L., X. Chen, Y. Sun, T. Worm, M.: 3D Dynamic Facial Expression Database, IEEE Inter." In Conf. on Automatic Face and Gesture Recognition, Amsterdam (2008)
39. Heseltine, T., Pears, N.: Three-dimensional face recognition using combinations of surface feature map subspace components. Image and Vision Computing 26(3), 382-396 (2008)
40. Savran, A., Alyüz, N., Dibeklioğlu, H., Çeliktutan, O., Gökberk, B., Sankur, B., Akarun, L.: Bosphorus database for 3D face analysis. In European workshop on biometrics and identity management, pp. 47-56. Springer, Heidelberg, (2008)
41. Yin, B., Sun Y., Wang C., Ge Y.: BJUT-3D large scale 3D face database and information processing. Journal of Computer Research and Development 46(6), 1009 (2009)
42. Cosker, D., Krumhuber, E., Hilton, A.: A FACS valid 3D dynamic action unit database with applications to 3D dynamic morphable facial modeling. In 2011 international conference on computer vision, pp. 2296-2303. IEEE, (2011)
43. Bagdanov, A.-D., Bimbo, A.-D.: The florence 2D/3D hybrid face dataset. In Proceedings of ACM workshop on Human gesture and behavior understanding, pp. 79-80. (2011)
44. Cao, C., Weng, Y., Zhou, S., Tong, Y., Zhou., K.: Facewarehouse: A 3D facial expression database for visual computing. IEEE Transactions on Visualization and Computer Graphics 20(3), 413-425 (2013)
45. Zhang, X., Yin, L., Cohn, J.-F.: BP4D-spontaneous: a high-resolution spontaneous 3d dynamic facial expression database. Image and Vision Computing 32(10), 1-6 (2014)
46. Zhang, Z., Girard, J.-M., Wu, Y., Zhang, X., Liu, P., Ciftci, U., Canavan, S.: Multimodal spontaneous emotion corpus for human behavior analysis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3438-3446. (2016)
47. Le, H.-A., Kakadiaris, I.-A.: UHDB31: A dataset for better understanding face recognition across pose and illumination variation. In Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 2555-2563. (2017)
48. Cheng, S., Kotsia, I., Pantic, M., Zafeiriou, S.: 4DFAB: A large scale 4D database for facial expression analysis and biometric applications. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 5117-5126. (2018)
49. Ertugrul, I.-O., Cohn, J.-F., Jeni, L.-A., Zhang, A., Yin, L. Ji, Q.: Cross-domain au detection: Domains, learning approaches, and measures. In 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition pp. 1-8. IEEE, (2019)
50. Yang, H., Zhu, H., Wang, Y., Huang, M., Shen, Q.: FaceScape: a Large-scale High Quality 3D Face Dataset and Detailed Riggable 3D Face Prediction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 601-610. (2020)
51. Booth, J., Roussos, A., Ponniah, A., Dunaway, D., Zafeiriou, S.: Large scale 3D morphable models. International Journal of Computer Vision 126 (2-4), 233-254 (2018)
52. Dai, H., Pears, N., Smith, W.-A., Duncan, C.: A 3D morphable model of craniofacial shape and texture variation. In Proceedings of the IEEE International Conference on Computer Vision, pp. 3085-3093. (2017)
53. Chang, F., Tran, A.-T., Hassner, T., Masi, I., Nevatia, R., Medioni, G.: ExpNet: Landmark-free, deep, 3D facial expressions. In 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition , pp. 122-129. IEEE, (2018)

54. Booth, J., Roussos, A., Ververas, E., Antonakos, E., Ploumpis, S., Panagakis, Y.: 3D reconstruction of "in-the-wild" faces in images and videos. IEEE transactions on pattern analysis and machine intelligence 40(11), 2638-2652 (2018)

55. Shang, J., Shen, T., Li, S., Zhou, L., Zhen, M., Self-supervised monocular 3D face reconstruction by occlusion-aware multi-view geometry consistency. (2020)

56. Feng, Y., Wu, F., Shao, X., Wang, Y., Zhou, X.: Joint 3D face reconstruction and dense alignment with position map regression network. In Proceedings of the European Conference on Computer Vision , pp. 534-551. (2018)

57. Zeng, Xiaoxing, Xiaojiang Peng, Yu Qiao.: DF2Net: A Dense-Fine-Finer Network for Detailed 3D Face Reconstruction. In Proceedings of the IEEE International Conference on Computer Vision, pp. 2315-2324. (2019)

58. Zhang, J., Cai, H., Guo, Y., Peng, Z.: Landmark Detection and 3D Face Reconstruction for Caricature using a Nonlinear Parametric Model. (2020)

59. Browatzki, B., Wallraven, C.: 3FabRec: Fast Few-shot Face alignment by Reconstruction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6110-6120. (2020)

60. Wu, S., Rupprecht, C., Vedaldi, A.: Unsupervised Learning of Probably Symmetric Deformable 3D Objects from Images in the Wild. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1-10. (2020).

61. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N., Hornegger J., Wells W., Frangi A. (eds) Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015, LNCS, vol. 9351, pp. 234-241. Springer, Cham (2015)