3D Facial Reconstruction from a Single Image Using a Hybrid Model Based on 3DMM and Deep Learning

Isha Deshmukh(✉), Vikas Tripathi, and Durgaprasad Gangodkar

Graphic Era (Deemed to be) University, Dehradun, India
{21062011,vikastripathi.cse,dr.gangodkar}@geu.ac.in

Abstract. A fundamental challenge in computer vision is accurately modelling 3D faces from images. It facilitates the creation of immersive virtual experiences, realistic facial animations, and reliable identity verification. This research introduces an innovative approach aimed at reconstructing intricate facial attributes, encompassing shape, pose, and expression from a single input image. The proposed methodology employs a fusion of two potent techniques: 3D Morphable Models (3DMMs) and advanced Deep Learning (DL) methodologies. By integrating DL into tasks like face detection, expression analysis, and landmark extraction, the framework excels in reconstructing realistic facial attributes from single images even in diverse environments. The framework achieves compelling results in reconstructing "in-the-wild" faces, exhibiting notable fidelity while preserving essential facial characteristics. Experimental evaluations confirm the effectiveness and robustness of our approach, confirming its adaptability across various scenarios. Our research contributes to the advancement of 3D face modelling techniques, addressing the challenges of accurate reconstruction and holding promise for applications in virtual reality, facial animation, medical, security, and biometrics.

Keywords: 3D Facial Reconstruction · 3D Morphable Model · Face Detection · Facial Landmark Detection · Face Model Fitting

1 Introduction

Facial reconstruction from images has evolved into a critical challenge in computer vision. The accurate modelling and reconstruction of the 3D shape, pose, and expression of a face from an image has garnered significant attention and found crucial applications in domains such as virtual reality, facial animation, medical, security, and biometrics [1–4]. The advancements in accurate and realistic 3D face modelling have paved the way for immersive virtual experiences, lifelike facial animations in movies and games, and reliable identity verification systems [5–7].

Image-based methods for facial reconstruction have played a pivotal role in driving progress in this field. These methods leverage the abundance of visual information available in images and enable the reconstruction process without requiring expensive and intrusive hardware setups. The development of accurate and effective image-based

B. J. Choi et al. (Eds.): IHCI 2023, LNCS 14532, pp. 115–126, 2024.
https://doi.org/10.1007/978-3-031-53830-8_12

reconstruction techniques has been considerably aided by the abundance of readily accessible image data, the rapid progress of Deep Learning (DL) techniques, and the easy availability of large-scale datasets.

The domain of 3D face modelling has made significant advancements lately, propelled by breakthroughs in DL techniques, particularly, Convolutional Neural Networks (CNN) for image processing [8]. This progress has been fueled by the need to overcome the limitations of traditional approaches that relied on handcrafted features and labour-intensive manual annotation. To achieve accurate facial reconstruction, researchers have explored various techniques and representations, aiming to capture the intricate details of human faces.

An extensively employed approach for 3D facial modelling is the utilization of 3D Morphable Models (3DMMs) [9], which offers a versatile and parametric representation of facial geometry and appearance [10]. 3DMMs provide a condensed yet comprehensive representation that can be utilised to reconstruct and modify 3D faces by capturing the variations in facial form, texture, and expressions within a low-dimensional space.

This paper introduces an innovative approach to 3D face modelling that combines the strengths of 3DMMs and integrates DL-based techniques for face detection, landmark extraction and expressions. The primary focus is on achieving a realistic reconstruction of facial shape, pose, and expression from a single input image, particularly in complex "in-the-wild" conditions. Our method aims to overcome the limitations and deliver a robust and efficient solution for facial reconstruction.

This paper makes several significant contributions:

- A unique approach that combines the flexibility and parametric representation of 3DMMs by leveraging the precision of DL-based approaches. This integration enables realistic and detailed reconstruction of facial geometry and appearance from an input image.
- The proposed method uses a single image for reconstruction that can be implemented in real-time systems as it is computationally less expensive and has more processing speed.
- The proposed method holds promise for various applications such as virtual reality, facial animation, and biometrics, where 3D facial modelling is crucial.

2 Related Work

The growth in 3D facial modelling and reconstruction has been particularly driven by the need to accurately capture the pose, shape, and expression of human faces. Previous studies have explored various approaches, with a particular emphasis on 3DMMs and image-based methods. This section provides an overview of the related work in these areas, highlighting the key contributions and limitations of each approach.

2.1 3DMM-Based Methods

3DMMs are widely used for 3D facial modelling, providing a parametric framework to capture facial geometry and appearance variations. Blanz and Vetter [9] pioneered the concept and showcased their effectiveness in reconstructing faces from 3D scans. These

models encode shape, texture, and expression variations in a low-dimensional linear space, enabling efficient and compact representation.

Since then, researchers have made significant advancements in 3DMMs to enhance their accuracy and applicability. Booth et al. [11] introduced the Large-Scale Facial Model (LSFM), a comprehensive 3DMM that incorporates statistical information from a diverse human population. The model analyzes the high-dimensional facial manifold, revealing clustering patterns related to age and ethnicity. Although the LSFM shows promise for medical applications due to its sensitivity to subtle genetic variations, further research and validation in this domain are warranted.

Tran et al. [12] introduced a method that learns a nonlinear 3DMM model from a large set of unconstrained face images, eliminating the need for 3D scans. They employed weak supervision and leveraged a large collection of 2D images. Similarly, [13] utilized an encoder-decoder architecture to estimate projection, lighting, shape, and albedo parameters, resulting in a nonlinear 3DMM. However, the learned shape exhibits some noise, especially around the hair region. In [14], an approach to enhance the nonlinear 3D face morphable model by incorporating strong regularization and leveraging proxies for shape and albedo was presented. The method utilized a dual-pathway network architecture that balances global and local-based models. Nevertheless, the model may face challenges when dealing with extreme poses and lighting conditions.

Dai et al. [15] proposed the Liverpool-York Head Model (LYHM), a fully-automatic statistical approach for 3D shape modelling, enhancing correspondence accuracy and modelling ability. However, variations in lighting, expressions, or occlusions may impact texture mapping quality. Similarly, [16] introduced 3DMM-RF, a facial 3DMM combining deep generative networks and neural radiance fields for comprehensive rendering, yet challenges remain in accurately rendering occluded areas and flattened eye representation in the training data.

In [17], the authors introduced the SadTalker system to create stylized audio-driven animations of talking faces from single images. This approach involves generating 3D motion coefficients from audio and utilizing a unique 3D-aware face rendering method for animation. However, the emphasis of this method is primarily on lip movement and eye blinking, leading to generated videos having fixed emotions.

2.2 Image-Based Methods

These methods leverage the abundance of visual information available in images and enable the reconstruction process without requiring expensive and intrusive hardware setups. Recent advancements in DL techniques have revolutionized image-based facial reconstruction.

Jiang et al. [18] employed a coarse-to-fine optimization strategy for 3D face reconstruction, refining a bilinear face model with local corrective deformation fields. However, it is sensitive to face deviations from the training datasets, ambiguities in albedo and lighting estimation, and reliance on the quality of detected landmarks. In [19], the incorporation of expression analysis and supervised/unsupervised learning for proxy face geometry generation and facial detail synthesis was proposed. Their method excels in handling surface noise and preserving skin details, but it has limitations in accounting for occlusions, hard shadows, and low-resolution images.

Afzal et al. [3] utilized feature extraction and depth-based 3D reconstruction method. However, the method does not consider facial expressions, which limits its applicability in dynamic scenarios or facial recognition applications. On the other hand, [20] focused on high-fidelity facial texture reconstruction using GANs and DCNNs for single-image reconstruction. Their approach achieves impressive results but may face challenges with extreme expressions, challenging lighting conditions, limited data availability, and computational complexity, impacting its real-time performance and scalability.

In [21], AvatarMe, a method for reconstructing high-resolution realistic 3D faces through single "in-the-wild" images was proposed. The approach includes facial mesh reconstruction and head topology inference that allows for complete head models with textures. However, the training dataset contains insufficient cases of individuals from various ethnicities, potentially resulting in lower performance in reconstructing faces. In [22], a model utilizing a generative prior of 3D GAN and an encoder-decoder network was proposed that can be generalized to new identities efficiently. This addresses the limitation of personalized methods and expands practicality.

Approaches for the reconstruction from multi-view images were explored by [23, 24], and [25]. The approach of [23] combined traditional multi-view geometry with DL techniques, but it relies on high-quality 3D scans, limiting performance. A fast and accurate spatial-temporal stereo matching scheme using speckle pattern projection was proposed by [24], while [25] introduced a method for high-quality 3D head model recovery from a few multi-view portrait images. However, results depend on input image quality and computational demands may restrict real-time or resource-constrained applications. Obtaining sufficient high-quality images from different viewpoints can be challenging in practical or real-world settings.

3DMMs have offered a parametric representation for capturing facial variations, while image-based methods have leveraged DL techniques to extract information directly from images. However, several challenges remain, including the robustness to varying illumination and occlusion, handling large pose variations, and preserving fine-scale details in "in-the-wild" scenarios. The proposed method aims to address these challenges by leveraging the advantages of both 3DMM-based and image-based approaches, providing a more accurate and robust framework for 3D facial reconstruction.

3 Methodology

This section outlines the methodology employed for the proposed approach. The process involves several key steps, including initialization, face detection and landmark extraction, fitting process, and output generation. Figure 1 provides a visual representation of the methodology proposed in our research.

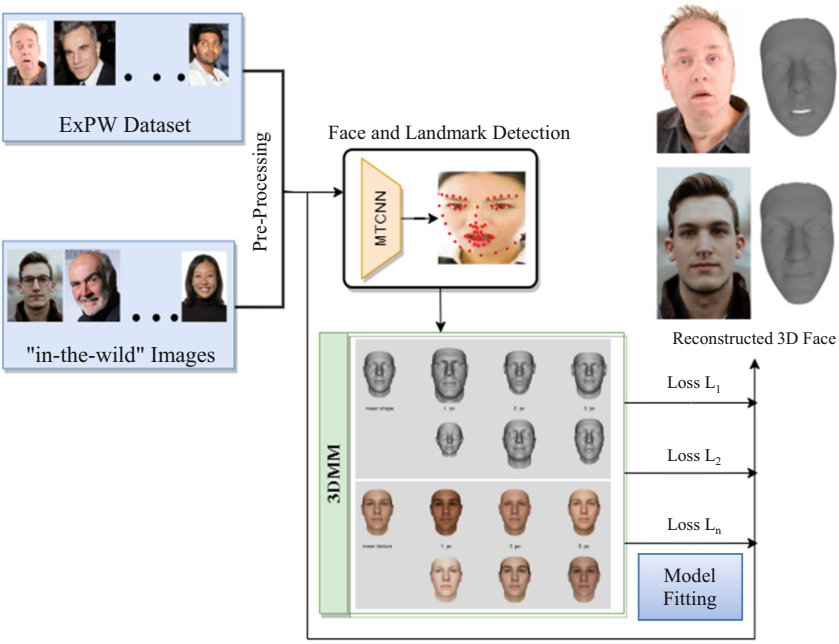

Fig. 1. Our facial reconstruction process includes face detection, landmark detection, and refinement of a 3D face model based on the input image. By optimizing the model to minimize discrepancies between projected and extracted landmarks, we achieve a realistic reconstruction. The final result is a composed image obtained by rendering and compositing the reconstructed face with the original image, enabling further analysis.

3.1 3D Morphable Model

For reconstruction using 3DMM, we utilize the popular Basel Face Model (BFM) 2009 [26]. The parameterization of each face involves angular meshes with 53,490 vertices.

$$\mathrm{S} = \mathrm{S}(\alpha, \beta) = \overline{\mathrm{S}} + B_{\mathrm{id}}\alpha + B_{\mathrm{exp}}\beta \tag{1}$$

$$\mathrm{T} = T(\gamma) = \overline{T} + B_t\gamma \tag{2}$$

In Eqs. (1) and (2), $\overline{\mathrm{S}}$ and $\overline{T}$ represent the average face shape and texture, respectively. B_{id}, B_{exp}, and B_t are the PCA bases of identity, expression, and texture respectively. These bases are scaled with standard deviations. The coefficient vectors α, β, and γ are used to generate a 3D face.

The expression bases, B_{exp}, utilized in our method, as described in [27], consist of 53,215 vertices. To reduce dimensionality, a subset of these bases is selected, resulting in coefficient vectors $\alpha \in \mathbb{F}^{80}$, $\beta \in \mathbb{F}^{64}$ and $\gamma \in \mathbb{F}^{80}$ where $\mathbb{F}$ represents the field of real numbers. It is important to note that the cropped model used in our approach contains 35,709 vertices.

3.2 Camera Model

A perspective camera model is employed to record the 3D-2D projection geometry of the face. The camera model incorporates a focal length determined through empirical observations, enabling us to precisely represent the connection between the 3D face and its 2D projection.

The 3D pose of the face, denoted as θ, is expressed using a rotation matrix $R \in SO(3)$ (Special Orthogonal group in three dimensions) and a translation vector $t \in \mathbb{F}^3$ (three-dimensional space). These parameters, R and t, define the camera's orientation and position relative to the face. By applying this camera model, we can project the 3D facial information onto a 2D image plane, facilitating further analysis and processing of the face data.

3.3 Illumination Model

The illumination model used is based on the concept of spherical harmonics (SH) [28, 29] basis functions H_b: $\mathbb{F}^3 \rightarrow \mathbb{F}$. The colour C at a vertex with normal vector n and tangent vector t, parameterized by the coefficients γ, can be expressed as the dot product between t and the linear combination of spherical harmonic basis functions:

$$C(n, t|\gamma) = t \cdot (\gamma 1 * \Phi 1(n) + \gamma 2 * \Phi 2(n) + ... + \gamma B * \Phi B(n)) \tag{3}$$

In Eq. (3), $\Phi 1(n)$, $\Phi 2(n)$,..., $\Phi B(n)$ represent the spherical harmonic basis functions evaluated at the normal vector n. The coefficients $\gamma 1, \gamma 2, \ldots, \gamma B$ are the weights associated with each basis function.

3.4 Model Fitting

Model fitting is a crucial stage in the reconstruction process, as it seeks to optimize the parameters of the 3D face model for precise alignment with the face in the input image and detected landmarks.

Face and Landmark Detection. Before initiating the model fitting process, the input image undergoes a series of preprocessing steps. Initially, the face region is detected using multi-task Cascaded Convolutional Networks (MTCNN) [30]. Subsequently, 68 facial landmarks are extracted using the landmark detection model presented by [31].

Loss Functions. These functions are used to measure the discrepancy between the expected values and the actual data during the model-fitting process.

Photometric Loss. The resemblance between the rendered image created by the 3D model and the input image is determined by comparing their colour and texture. This comparison is performed using a skin-aware photometric loss, as described by [32] given by the Eq. (4):

$$\mathcal{L}_p(x) = \frac{\sum_{i \in \mathcal{M}} A_i \cdot \|I_i - I_i'\|_2}{\sum_{i \in \mathcal{M}} A_i} \tag{4}$$

In this equation, i represents the pixel index, $\mathcal{M}$ represents the projected face region, and A_i is the skin colour-based attention mask.

Reflectance Loss. We use the naive Bayes classifier of mixture models [33] to compute the skin-colour probability P_i for each pixel i in order to handle difficult and complicated facial appearances, such as occlusions like beards and makeup. This is shown in the Eq. (5) and (6):

$$A_i = \begin{cases} 1, if P_i > 0.5 \\ P_i, otherwise \end{cases} \tag{5}$$

Therefore, predicted reflectance loss is calculated by

$$\mathcal{L}_R(x) = \frac{1}{|S|} \cdot \sum\nolimits_{i \in S} R_i'^2 \tag{6}$$

where, $|S|$ is the number of skin pixels and $R_i'^2$ is the difference between the predicted reflectance and the mean reflectance for pixel I.

Landmark Loss. It calculates the distance between the projected landmarks of the 3D model and the corresponding detected landmarks in the input image to ensure precise alignment. For landmark loss during the detection, we use Eq. (7):

$$\mathcal{L}_l(x) = \frac{1}{N} \sum_{n=1}^{N} \omega_n \| q_n - q_n'(x) \|^2 \tag{7}$$

Here, ω_n represents the manually assigned landmark weight for specific landmarks such as mouth and nose points.

Gamma Loss. The gamma loss encourages consistent gamma correction by measuring the deviation of gamma correction parameters from their mean value, as shown in Eq. (8):

$$\mathcal{L}_g(x) = \| \triangle \lambda \|^2 \tag{8}$$

where $\triangle \lambda$ is the difference between the gamma correction parameters and their mean value.

4 Results Analysis

Our experimental setup involved the ExPW dataset [34], which consists of approximately 91, 793 "in-the-wild" images with seven fundamental expression categories assigned to each face image; as well as other images found on the internet. The experimental setup included an Intel Core i7 processor, an NVIDIA RTX 3050 Ti graphics card, and 16GB of RAM. The implemented method combines DL and computer vision techniques for face fitting and 3D reconstruction. We utilized Python programming language and leveraged the open-source libraries OpenCV [35], Pytorch3D [36], and NumPy [37] for implementation.

We employed the MTCNN algorithm [30] to detect faces in images, resizing them to 224x224 pixels. For landmark detection, the face-alignment method [31] was utilized. The fitting process began with refining the BFM's pose (rotation and translation) to align with the detected face. Subsequently, the BFM was deformed to capture shape and expression details. Fitting was optimized using the Adam optimizer [38] to minimize the discrepancy between projected 3D landmarks of the BFM and extracted 3D landmarks from the image. The optimizer also minimized a combination of various loss terms, iteratively refining BFM parameters for minimizing overall loss. After fitting, optimized BFM parameters rendered a deformed face image. This image was composited with the original input, replacing the face region. The composed image, BFM coefficients, and mesh could be saved as output for diverse applications.

To assess the performance of our approach, we conducted comparisons with state-of-the-art approaches and relevant baseline methods. The evaluation encompasses both qualitative visual comparisons and quantitative analysis utilizing a variety of loss metrics. Figures 2 and 3 illustrate the outcomes of our approach juxtaposed with those of prominent state-of-the-art techniques. The visual comparisons vividly underscore the strengths of our method in capturing intricate facial details, expressions, and lifelike texture mapping. Across various test images, our method consistently generates more accurate and realistic 3D facial reconstructions, effectively preserving the nuances of individual appearances. Table 1 showcases a comprehensive summary of the computed losses across different types. This table presents representative values for each loss category, complemented by their corresponding mean and standard deviation. These metrics not only offer a concise snapshot of the experimental results but also provide insights into the dispersion and trends of the loss values.

While direct comparisons are limited by dataset variations and evaluation metrics, our method exhibits promising performance in terms of facial reconstruction and expression preservation. Our method offers several significant advantages. Firstly, it eliminates the need for manual marking of landmarks, which is required by many other methods. Additionally, it is computationally efficient as it only requires a single image for efficient 3D reconstruction. This efficiency is achieved through less expensive computations and faster processing speeds, enabling real-time implementation.

One aspect to consider is that the method may encounter challenges when dealing with occlusions, such as individuals wearing sunglasses, despite its ability to handle faces with spectacles. In such cases, the reconstructed 3D face may exhibit dark areas under the eyes, reflecting the colour of the sunglasses. Addressing these occlusion challenges and improving the generation of realistic facial features in such scenarios would be a valuable avenue for future enhancements. Furthermore, refining the model's ability to reproduce finer details, including wrinkles and eye tracking can contribute to achieving even greater realism in the reconstructed 3D faces.

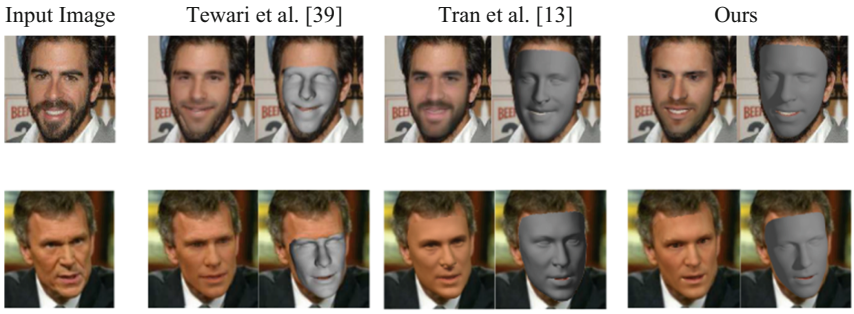

Fig. 2. The visual comparison of our outcomes with other innovative methods.

Table 1. Summary of the calculated losses for different loss types. It includes three representative values for each loss type to provide a concise representation of the results along with their mean and standard deviation. These values provide insights into the variations and distribution of the losses, offering a concise overview of the experimental results.

Loss Type	Measure 1	Measure 2	Measure 3	Mean	Standard Deviation
Landmark Loss $\mathcal{L}_l(x)$	0.000062	0.000098	0.000152	0.000104	0.000045
Photometric loss $\mathcal{L}_p(x)$	0.053055	0.095443	0.093362	0.080953	0.020273
Texture Loss $\mathcal{L}_T(x)$	0.006199	0.009445	0.007481	0.007375	0.001303

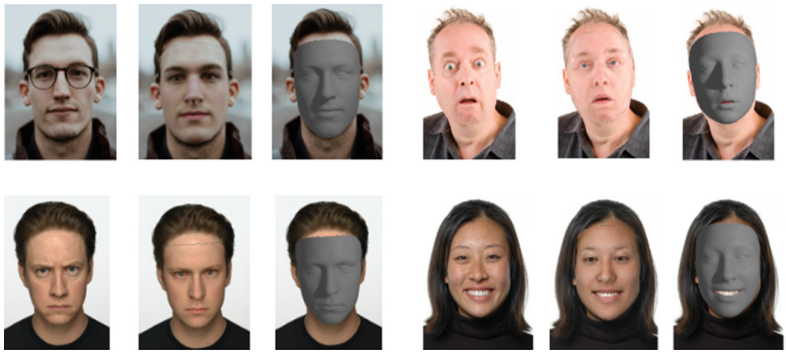

Fig. 3. The figure presents a compilation of reconstructed 3D faces utilizing our proposed method, highlighting its ability to generate realistic facial reconstructions.

5 Conclusion

In this research, we demonstrated a technique for generating a 3D facial model from a single input image. It incorporates face and landmarks detection models to precisely locate and extract the facial region. The approach enhances the quality of the reconstructed face by using photometric consistency constraints, local corrective deformation fields, and coarse-to-fine optimization. The use of a single image eliminates the need for multiple images or complex scanning setups, making the reconstruction process more practical and cost-effective. It further utilizes fitting process optimizations to minimize various loss functions, resulting in a refined and realistic 3D reconstruction, enabling further analysis and application possibilities.

Overall, the paper presents a powerful approach for 3D face reconstruction, with potential applications in computer graphics, virtual reality, facial animation, and biometrics. While the effectiveness relies on the performance of the detection models and input image quality, fine-tuning the optimization parameters can further enhance the accuracy and fidelity of the reconstruction. By providing a comprehensive solution for reconstructing 3D faces from a single image, this paper opens doors for advancements in facial modelling and realistic virtual representations.

References

1. Widanagamaachchi, W.N., Dharmaratne, A.T.: 3D Face Reconstruction from 2D Images. In: 2008 Digital Image Computing: Techniques and Applications, pp. 365–371. IEEE (2008). https://doi.org/10.1109/DICTA.2008.83
2. Zollhöfer, M., et al.: State of the Art on Monocular 3D Face Reconstruction, Tracking, and Applications. Computer Graphics Forum. **37**, 523–550 (2018). https://doi.org/10.1111/cgf.13382
3. Afzal, H.M.R., Luo, S., Afzal, M.K., Chaudhary, G., Khari, M., Kumar, S.A.P.: 3D Face Reconstruction From Single 2D Image Using Distinctive Features. IEEE Access. **8**, 180681–180689 (2020). https://doi.org/10.1109/ACCESS.2020.3028106
4. Diwakar, M., Kumar, P.: 3-D Shape Reconstruction Based CT Image Enhancement. In: Handbook of Multimedia Information Security: Techniques and Applications, pp. 413–419. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-030-15887-3_19
5. Uddin, M., Manickam, S., Ullah, H., Obaidat, M., Dandoush, A.: Unveiling the Metaverse: Exploring Emerging Trends, Multifaceted Perspectives, and Future Challenges. IEEE Access. 1–1 (2023). https://doi.org/10.1109/ACCESS.2023.3281303
6. Jha, J., et al.: Artificial intelligence and applications. In: 2023 1st International Conference on Intelligent Computing and Research Trends (ICRT), pp. 1–4. IEEE (2023). https://doi.org/10.1109/ICRT57042.2023.10146698
7. Sharma, H., Kumar, H., Gupta, A., Shah, M.A.: Computer Vision in Manufacturing: A Bibliometric Analysis and future research propositions. Presented at the (2023)
8. Khari, M., Garg, A.K., Gonzalez-Crespo, R., Verdú, E.: Gesture Recognition of RGB and RGB-D Static Images Using Convolutional Neural Networks. Int. J. Interact. Multi. Artifi. Intell. **5**, 22 (2019)
9. 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, pp. 187–194 (1999)

10. Kittler, J., Huber, P., Feng, Z.-H., Hu, G., Christmas, W.: 3D Morphable Face Models and Their Applications. Presented at the (2016). https://doi.org/10.1007/978-3-319-41778-3_19
11. Booth, J., Roussos, A., Ponniah, A., Dunaway, D., Zafeiriou, S.: Large Scale 3D Morphable Models. Int. J. Comput. Vis. **126**, 233–254 (2018)
12. Tran, L., Liu, X.: Nonlinear 3D Face Morphable Model. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7346–7355. IEEE (2018)
13. Tran, L., Liu, X.: On Learning 3D Face Morphable Model from In-the-wild Images. IEEE Trans. Pattern Anal. Mach. Intell. **43**, 157–171 (2019). https://doi.org/10.1109/TPAMI.2019.2927975
14. Tran, L., Liu, F., Liu, X.: Towards High-Fidelity Nonlinear 3D Face Morphable Model. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1126–1135. IEEE (2019). https://doi.org/10.1109/CVPR.2019.00122
15. Dai, H., Pears, N., Smith, W., Duncan, C.: Statistical modeling of craniofacial shape and texture. Int. J. Comput. Vis. **128**, 547–571 (2020). https://doi.org/10.1007/s11263-019-01260-7
16. Galanakis, S., Gecer, B., Lattas, A., Zafeiriou, S.: 3DMM-RF: convolutional radiance fields for 3D face modeling. In: 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 3525–3536. IEEE (2023)
17. Zhang, W., et al.: SadTalker: Learning Realistic 3D Motion Coefficients for Stylized Audio-Driven Single Image Talking Face Animation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8652–8661 (2023)
18. Jiang, L., Zhang, J., Deng, B., Li, H., Liu, L.: 3D face reconstruction with geometry details from a single image. IEEE Trans. Image Process. **27**, 4756–4770 (2018). https://doi.org/10.1109/TIP.2018.2845697
19. Chen, A., Chen, Z., Zhang, G., Mitchell, K., Yu, J.: Photo-realistic facial details synthesis from single image. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 9428–9438. IEEE (2019). https://doi.org/10.1109/ICCV.2019.00952
20. Gecer, B., Ploumpis, S., Kotsia, I., Zafeiriou, S.: GANFIT: generative adversarial network fitting for high fidelity 3D face reconstruction. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1155–1164. IEEE (2019). https://doi.org/10.1109/CVPR.2019.00125
21. Lattas, A., et al.: AvatarMe: Realistically Renderable 3D Facial Reconstruction "In-the-Wild." In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 757–766. IEEE (2020). https://doi.org/10.1109/CVPR42600.2020.00084
22. Yu, W., et al.: NOFA: NeRF-based One-shot Facial Avatar Reconstruction. In: Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Proceedings, pp. 1–12. ACM, New York, NY, USA (2023)
23. Bai, Z., Cui, Z., Rahim, J.A., Liu, X., Tan, P.: Deep facial non-rigid multi-view stereo. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5849–5859. IEEE (2020). https://doi.org/10.1109/CVPR42600.2020.00589
24. Fu, K., Xie, Y., Jing, H., Zhu, J.: Fast spatial–temporal stereo matching for 3D face reconstruction under speckle pattern projection. Image Vis. Comput. **85**, 36–45 (2019). https://doi.org/10.1016/j.imavis.2019.02.007
25. Wang, X., Guo, Y., Yang, Z., Zhang, J.: Prior-Guided Multi-View 3D Head Reconstruction. IEEE Trans. Multimedia **24**, 4028–4040 (2022)
26. 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). https://doi.org/10.1109/AVSS.2009.58

27. Guo, Y., Zhang, J., Cai, J., Jiang, B., Zheng, J.: CNN-based real-time dense face reconstruction with inverse-rendered photo-realistic face images. IEEE Trans. Pattern Anal. Mach. Intell. **41**, 1294–1307 (2019). https://doi.org/10.1109/TPAMI.2018.2837742
28. Ramamoorthi, R., Hanrahan, P.: A signal-processing framework for inverse rendering. In: Proceedings of the 28th annual conference on Computer graphics and interactive techniques, pp. 117–128. ACM, New York, NY, USA (2001). https://doi.org/10.1145/383259.383271
29. Ramamoorthi, R., Hanrahan, P.: An efficient representation for irradiance environment maps. In: Proceedings of the 28th annual conference on Computer graphics and interactive techniques, pp. 497–500. ACM, New York, NY, USA (2001)
30. Schroff, F., Kalenichenko, D., Philbin, J.: FaceNet: a unified embedding for face recognition and clustering. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 815–823. IEEE (2015). https://doi.org/10.1109/CVPR.2015.7298682
31. Bulat, A., Tzimiropoulos, G.: How far are we from solving the 2D & 3D face alignment problem? (and a Dataset of 230,000 3D Facial Landmarks). In: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 1021–1030. IEEE (2017)
32. Deng, Y., et al.: Accurate 3D face reconstruction with weakly-supervised learning: from single image to image set. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 285–295. IEEE (2019). https://doi.org/10.1109/CVPRW.2019.00038
33. Jones, M.J., Rehg, J.M.: Statistical color models with application to skin detection. Int. J. Comput. Vis. **46**, 81–96 (2002). https://doi.org/10.1023/A:1013200319198
34. Hou, Z.-D., Kim, K.-H., Lee, D.-J., Zhang, G.-H.: Real-time markerless facial motion capture of personalized 3D real human research. Int. J. Inter. Broadcas. Comm. **14**, 129–135 (2022)
35. OpenCV: Open Source Computer Vision Library (2015)
36. Johnson, J., et al.: Accelerating 3D deep learning with PyTorch3D. In: SIGGRAPH Asia 2020 Courses, p. 1. ACM, New York, NY, USA (2020). https://doi.org/10.1145/3415263.3419160
37. Harris, C.R., et al.: Array programming with NumPy. Nature **585**, 357–362 (2020)
38. Kingma, D.P., Ba, J.: Adam: a method for Stochastic Optimization. CoRR. abs/1412.6980 (2014)
39. Tewari, A., et al.: MoFA: Model-Based Deep Convolutional Face Autoencoder for Unsupervised Monocular Reconstruction. In: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 3735–3744. IEEE (2017)