Face Reconstruction

Ritik Chauhan[1], Mohit Singh[2], Rajesh Kumar Pathak[3]
[1,2,3]*Dept. of computer science and engineering, M.tech Integrated, Niet, Greater Noida, AKTU*

Abstract—3D face reconstruction is a challenging problem and also an important task in the field and within the domain of computer vision and graphics. In the present time, most bulk of facial image data of interest does not come solely from single images but instead from videos, which encompass and contain a wealth of rich dynamic information.3D face shape is widely used in many areas, such as medical care, medical cosmetology, entertainment, and security. And it contributes to other fields such as face alignment, face recognition, and face editing. This paper is a review of 3D face reconstruction from videos.

Keywords—Face reconstruction, 3D, 2D, face recognition, face alignments.

I. INTRODUCTION

During the last few years, Deep Learning frameworks have succeeded in significantly increasing the accuracy of monocular 3D face reconstruction, even in cases of unconstrained image data. The current state of the artwork is able to robustly reconstruct fine details of the 3D facial geometry as well as a reliable estimation of the captured subject's facial anatomy.

This is beneficial for applications, such as augmented reality, performance capture, visual effects, photo-realistic video synthesis, human-computer interaction, personalized avatars, etc.

A crucial factor for the limitations of some existing methods is the fact that most methods use weak 2D supervision from landmarks predicted by face alignment methods. While these landmarks can yield an estimation of the facial shape, they fail to provide an accurate representation of the expressive details of a highly deformable mouth region.

It is also important to note that the shapes of the human mouth are perceptually correlated with movements and the realism of a 3D talking head.

To overcome the limitations of the existing literature, this work tackles the problem of monocular 3D face reconstruction from a video[29], with a strong focus on the mouth area and its expressions and movements. We highlight and address the fact that an accurate 3D reconstruction of a human face in a video should retain those mouth expressions and movements.

In this work, we focus on landmark localization, in particular, on facial landmark localization, which is commonly posed in terms of locating relatively few landmarks, or keys.

points. We use the mesh topology comprised of 400+ points arranged in fixed quads. The points have been manually selected in accordance with the supposed applications, such as expressive AR effects, virtual accessories, and real-time applications.

II. METHODS

However, researchers find that 2D alignment has difficulties dealing with large poses or occlusion problems. With the development of deep learning, many computer vision problems have been well solved by utilizing Convolution Neural Networks (CNNs) [1,2,30,31].

Thus, some works start using CNNs to estimate the 3D Morphable Model (3DMM) coefficients [1,11] or 3D model warping functions to restore the corresponding 3D information from a single 2D facial image, providing dense face alignment and 3D face reconstruction results[31].

As described above, the main problems our method can solve are 3D face alignment and face reconstruction.

The proposed method consists of two main tasks-

- Face Detection and Tracking
- Reconstruction of the Face

The details of these methods are as follows:

1. 3DMM-based methods: the 3D morphable model was first proposed by Thomas Vetter et al. [5] in the article " A deformable model for the synthesis of 3D faces".

Up to now, many 3D face reconstruction methods have been developed on the basis of this model. How to get these fitting parameters is the main problem for

3DMM. With the development of deep learning, many methods have been proposed to provide more possibilities for solving the parameter problem.

2. Other methods: Shape from shading is a method for recovering 3D information from a single image proposed by Horn in 1980. SFS uses the change of the normal vector of the smooth object surface, which changes the brightness of the incident light on the object surface and then reflects the shape of the object. And recently, some researchers use some image processing methods such as UV maps, and Epipolar Plane Images to achieve single-image 3D face reconstruction.

Our method consists of different modules and is highly based on phrases like Human face Detection from Videos, Face Detection, and Tracking and Reconstruction of face[28].

At present-

1. 3DMM-based methods 3D mesh renderer using TensorFlow-

3DMM is a statistical model of 3D facial shape and texture. It has been widely used in 3D face reconstruction and face recognition [16,17,18,32,24].

The traditional 3DMM uses mean shape and a linear combination of a set of shape bases to generate a personalized 3D face shape.

The shape bases are typically extracted from a training set of 3D face scans by Principal Component Analysis (PCA). Through searching for the optimal linear fitting parameters, a reconstructed face rendering from a 3D face will be approximate to the original image [8]. Modeling highly variable 3D face shapes requires a quantity of high-quality 3D face scans[32].

3DMM (3D Morphable Model) is a widely used approach in face reconstruction that models the variations in shape and texture of human faces. It provides a statistical representation of the 3D geometry and appearance of faces, allowing for the reconstruction of realistic and detailed 3D face models from 2D images or other input data.

2. Face alignments:

For the past few years, alignments on the face were aimed at locating the set of 2D facial landmarks. Past works on such types of models include Active Shape Model(ASM) and Active Appearance Model(AMM) [3,4,5,6] and then with the advancement of deep learning, CNNs used to position the landmarks for increasing and for the improvements of performance[7, 8, 9,10].

Now, most problems are shifting on 3D face problems. There are many strategies for solving these problems and widely using at present are:

1. Fitting certain 3D face models [13] to present the full structure in 3D to guide the 3D landmark localization.
2. Do separately detection of 2D landmarks and their depth[10,11] .
3. Deep dense alignments [12] by combining a set of tens of thousands of points, are solved by registering and fitting all the points.

3. For 3d plane coordinates and it’s working-

We have adopted the policy that the x- and y-coordinates of the vertices correspond to the point locations in the 2D plane as given by the image pixel coordinates. The z-coordinates are interpreted as the depth relative to a reference plane passing through the mesh’s center of mass. They are re-scaled so that a fixed aspect ratio is maintained between the span of x-coordinates and the span of z-coordinates, i.e. a face that is scaled to half its size has its depth range (nearest to farthest) scaled down by the same multiplier[27].

When used on video input in the face tracking mode, a good facial crop is available from the previous frame prediction and the usage of the face detector is redundant. In this scenario, it is only used on the first frame and in the rare events of re-acquisition. It should be noted that with this setup, the network receives the inputs with faces reasonably centered and aligned. We argue that this allows to save some model representational capacity that could otherwise be spent on handling the cases with substantial rotation and translation[27].

In particular, we could reduce the number of related augmentations while gaining prediction quality.

III. THEORETICAL DISCUSSION

Uses of pytorch :

- PyTorch is a popular open-source machine learning library that utilizes the Torch library. It is widely used for various applications, including computer vision and natural language processing, and is primarily developed by Facebook's AI Research lab[26].

- PyTorch is released as free and open-source software, making it accessible to the community. It is licensed under the Modified BSD license, allowing users to modify and distribute the library according to their needs. While the Python interface of PyTorch is highly refined and receives the majority of development attention, there is also a C++ interface available [21,26].
- The features of PyTorch include:
  PyTorch offers tensor computation capabilities, similar to NumPy. It leverages the power of graphics processing units (GPUs) to accelerate computations, resulting in enhanced performance for demanding tasks.
  Pytorch provides a powerful framework for building deep neural networks. It incorporates a tape-based automatic differentiation system, enabling efficient computation of gradients. This feature is essential for training complex neural network models effectively[26].

IV. RESULTS

Firstly, we evaluate our method on a sparse set of 400+ facial landmarks and compare our result with all segments. As shown in Figures 1,2 3,4, our result slightly outperforms the state of the art method when calculating per distance with 2D coordinates. When considering the depth value, the performance of our method increases. Notice that, our method needs no other network to predict the z coordinate of landmarks, and the depth value can be obtained directly in our method.

The speed estimations are based on the TensorFlow Lite GPU framework. Also, the face uses limitation can be maximized up to 4 faces through our method, but we keep it on 2 faces for experiments.

Fig.1

Fig.2

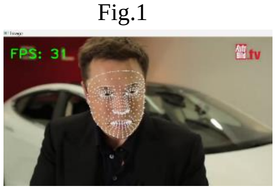

Fig.3

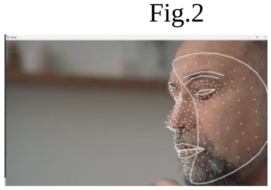
Fig.4

The process of reconstructing the frontal face from the non-frontal face is based on facial features. From the non-frontal face by estimating the roll angle and from the salient landmark points the frontal face is reconstructed with stretching and mirroring operations. The angle correction by tilting the image[28] and it can view with any 3d library is shown in Figure 5,6,7.

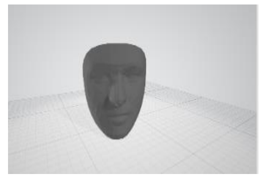
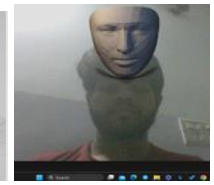
Fig.5 Fig.7


Fig.6

V. ACKNOWLEDGMENTS

1. We have presented the method for the reconstruction of 3D heads. This work proposes a method to identify face detection and its alignment in a video sequence and take the snapshot of the human face in different poses and reconstructing the original image by fusing the images. Our extensive subjective and objective evaluations have verified that the results of 3D reconstruction are significantly preferred to counterpart methods that rely only on geometric losses for the mouth movements and methods that use direct 3D supervision.
2. However, this task becomes more difficult in the presence of different variations in brightness, lighting, contrast levels, poses, and backgrounds. This work proposes a method to identify reconstructing the original image by fusing the images. We demonstrated significant improvements in terms of speed, accuracy, and perceptual quality compared to single-image methods. Then we further improved our approach by implementing a custom dynamic reconstruction algorithm.

REFERENCE

[1] Jourabloo, A., Liu, X.: Large-pose face alignment via cnn-based dense 3d model fitting. In: Computer Vision and Pattern Recognition. (2016)
[2] A. Jourabloo and X. Liu. Large-pose face alignment via cnnbased dense 3d model fitting. In CVPR, 2016.
[3] Zeyu Ruan, Changqing Zou, Member, IEEE, Longhai Wu, Gangshan Wu, Member, IEEE, Limin Wang, Member, IEEE, SADRNet: Self-Aligned Dual Face Regression Networks for Robust 3D Dense Face Alignment and Reconstruction[2021]
[4] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models—their training and application," Comput. Vis. Image Underst., 1995.
[5] P. Sauer, T. Cootes, and C. Taylor, "Accurate regression procedures for active appearance models," Proceedings of the British Machine Vision Conference (BMVC), 2011.
[6] Z. Feng, J. Kittler, M. Awais, P. Huber, and X. Wu, "Wing loss for robust facial landmark localisation with convolutional neural networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.
[7] L. Liu, G. Li, Y. Xie, Y. Yu, Q.Wang, and L. Lin, "Facial landmark machines: A backbone-branches architecture with progressive representation learning," IEEE Transactions on Multimedia,www.hindawi.com, 2019.
[8] X. Peng, X. Feris, Rogerio S.and Wang, and D. N. Metaxas, "A recurrent encoder-decoder network for sequential face alignment," in Proceedings of the European Conference on Computer Vision (ECCV), 2016.
[9] H. Zhang, Q. Li, and Z. Sun, "Joint voxel and coordinate regression for accurate 3d facial landmark localization," in International Conference on Pattern Recognition (ICPR), 2018.
[10] Yao-Kun Li, Yue-Zhao Yu, Yu-Liang Liu, Chao Gou. "MS-GCN: Multi-Stream Graph Convolution Network for Driver Head Pose Estimation" , 2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC), 2022
[11] A. Bulat and Y. Tzimiropoulos, "Convolutional aggregation of local evidence for large pose face alignment," in Proceedings of the British Machine Vision Conference (BMVC), 2016.
[12] Y. Liu, A. Jourabloo, W. Ren, and X. Liu, "Dense face alignment," in Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops, 2017.
[13] Zhu, X., Lei, Z., Liu, X., Shi, H., Li, S.Z.: Face alignment across large poses: A 3d solution. computer vision and pattern recognition (2016) 146–155
[14] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3d faces," in Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques, 1999.
[15] Y. Guo, L. Cai, and J. Zhang, "3d face from x: Learning face shape from diverse sources," IEEE Transactions on Image Processing, 2021.
[16] Brian Amberg, Reinhard Knothe, and Thomas Vetter. Expression invariant 3d face recognition with a morphable model. In 2008 8th IEEE International Conference on Automatic Face & Gesture Recognition, pages 1–6. IEEE, 2008.
[17] Volker Blanz and Thomas Vetter. Face recognition based on fitting a 3d morphable model. IEEE Transactions on pattern analysis and machine intelligence, 25(9):1063–1074, 2003.
[18] James Booth, Epameinondas Antonakos, Stylianos Ploumpis, George Trigeorgis, Yannis Panagakis, and Stefanos Zafeiriou. 3d face morphable models" in the wild". In Proceedings of the IEEE Conference on computer vision and pattern recognition, pages 48–57, 2017.
[19] James Booth, Anastasios Roussos, Stefanos Zafeiriou, Allan Ponniah, and David Dunaway. A 3d morphable model learnt from 10,000 faces. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 5543–5552, 2016.
[20] Hanxin Wang. A Review of 3D Face Reconstruction From a Single Image, pages 9-10,2021.
[21] https://pytorch.org/
[22] Liu, F., Zeng, D., Zhao, Q., Liu, X.: Joint face alignment and 3d face reconstruction. In: European Conference on Computer Vision, Springer (2016) 545–560
[23] Zhou, E., Fan, H., Cao, Z., Jiang, Y., Yin, Q.: Extensive facial landmark localization with coarse-to-fine convolutional network cascade. In:

Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on, IEEE (2013) 386– 391
[24] Richardson, E., Sela, M., Or-El, R., Kimmel, R.: Learning detailed face reconstruction from a single image. (2016) 15. Liu, Y., Jourabloo, A., Ren, W., Liu, X.: Dense face alignment. arXiv preprint arXiv:1709.01442 (2017)
[25] G. Trigeorgis, P. Snape, M. A. Nicolaou, E. Antonakos, and S. Zafeiriou. Mnemonic descent method: A recurrent process applied for end-to-end face alignment. In CVPR, 2016.
[26] wiki.hpcuser.uni-oldenburg.de
[27] Jiahao Wang, Zeyu Xiong, Yicun Duan, Junyu Liu, Zilin Song, Xiangjun Peng. "The Importance Distribution of Drivers' Facial Expressions Varies over Time!" , 13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, 2021.
[28] d.researchbib.com
[29] Yue Li, Liqian Ma, Haoqiang Fan, Kenny Mitchell. "Feature-preserving detailed 3D face reconstruction from a single image", Proceedings of the 15th ACM SIGGRAPH European Conference on Visual Media Production, 2018.
[30] Zhuo Chen, Yuesong Wang, Tao Guan, Yawei Luo, Luoyuan Xu, Wenkai Liu. "Learning 3D Face Shape From Diverse Sources With CrossDomain Face Synthesis", IEEE MultiMedia, 2022.
[31] Submitted to SVKM International School.
[32] Luan Tran, Xiaoming Liu. "Chapter 4 Learning 3D Face Morphable Model from In-the-Wild Images", Springer Science and Business Media LLC, 2021.
[33] scholar.archive.org
[34] Submitted to British University in Egypt.
[35] Feng Liu, Luan Tran, Xiaoming Liu. "3D Face Modeling From Diverse Raw Scan Data" , 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019.