Three-view 3D Face Reconstruction Using Facial Landmarks and Pose Estimation

Robert Hakobyan

National Polytechnic University of Armenia
Yerevan, Armenia
e-mail: hakobyan.rob@gmail.com

Abstract— This paper presents an innovative approach for creating a 3D model of a person's face using three images captured from different angles. The proposed method involves detecting and extracting facial landmarks from each image, as well as estimating the pose of the face in each image. By aligning the registered images and estimated poses, a 3D model of the face can be reconstructed. This model is then refined using texture information from the input images and prior knowledge of facial structure. The approach demonstrates potential applications in areas such as cyber and physical security, computer graphics, biometrics, and forensics. It is important to consider ethical concerns related to facial image usage when applying this method.

Keywords— 3D face reconstruction, facial landmarks, pose estimation, computer vision, image registration, texture mapping, stereo reconstruction.

I. INTRODUCTION

Creating precise 3D facial models has become increasingly crucial in modern applications such as biometrics, forensics, graphics, and computer cyber and physical security. These models help to generate realistic facial animations, assist in forensic investigations by visualizing suspects' faces from various angles, and enable facial recognition systems for security purposes. Nonetheless, producing a 3D model of a face from 2D images is a difficult task that requires careful consideration of variables like lighting, pose, and facial expression [1].

Recent developments in computer vision techniques have shown promising outcomes for 3D facial reconstruction from multiple images. The use of facial landmarks and pose estimation has been particularly effective in creating precise and realistic 3D face models. Despite these favorable outcomes, however, there is presently no open-source software accessible for three-view 3D face reconstruction using three images that is both accurate and easy to customize.

This paper presents a unique approach for 3D facial reconstruction using three images of the same person's face from various angles, which aims to bridge this gap. The proposed technique employs facial landmarks and pose estimation to register the input images to a common coordinate system and carry out 3D reconstruction of the face. The resulting 3D model is subsequently refined using texture information from the input images and prior knowledge of facial structure. The proposed method is accurate, customizable, and user-friendly, making it a valuable contribution to the open-source community for 3D face reconstruction from three images.

II. IMAGE PROCESSING

Image processing is a method utilized to alter or extract pertinent data from images. This procedure involves inputting an image and obtaining a result that could encompass elements within the image, color schemes, or other desired information. The main objectives of image processing are image enhancement and image analysis.

Figure 1. Deblurring of an image

The goal of image enhancement is to augment specific characteristics of an image, such as reducing blur or boosting the visual clarity, etc. It's important to note that enhancement techniques suitable for cosmic images may not be applicable to indoor images [2].

Image analysis involves the use of various techniques by a computer to extract information about an image. This includes 2D and 3D object recognition including face recognition, image segmentation, and other techniques [3].

Image analysis is the process of obtaining significant information from images, primarily digital images, using digital image processing techniques. Image processing is a rapidly expanding field and is finding increasing applications in areas such as engineering and computer science. The image processing procedure involves three key steps:

- Inputting the image for analysis.
- Processing and modifying the image using various techniques.

https://doi.org/10.51408/csit2023_66

- Outputting the modified image or the essential information obtained during the image processing procedure.

These three steps are critical in extracting relevant information from images and making them suitable for various applications, such as computer vision, medical imaging, remote sensing, and many others.

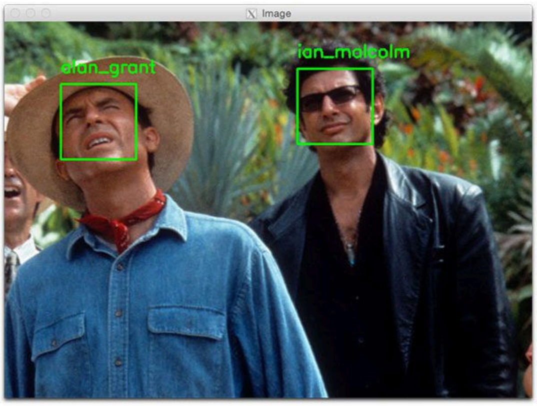

Figure 2. Face recognition

Facial recognition is a well-known application of image analysis that involves utilizing image processing techniques to recognize and authenticate individuals based on their facial characteristics. This technology has a broad range of uses, including security systems, mobile devices, and biometric identification [4].

Once the face is recognized, the next step is to extract landmarks from it. There are several algorithms available for this purpose, with DLIB and MTCNN being the two primary libraries used for facial landmark detection.

A. DLIB

DLIB is an open-source modern C++ toolkit containing machine learning algorithms and tools for creating complex software in C++ to solve real world problems. It is used in both industry and academia in a wide range of domains including robotics, embedded devices, mobile phones, and large high performance computing environments[5].

Figure 3. Facial landmarks extracted using DLIB

B. MTCNN

MTCNN is a facial detection and recognition algorithm used to detect and extract facial landmarks from images. It is an abbreviation for "Multi-Task Cascaded Convolutional Neural Network," and is widely used in computer vision applications such as face detection and recognition [6].

Figure 4. Facial landmarks extracted using MTCNN

C. Comparison

Both MTCNN and DLIB are effective tools for landmark detection, but the choice between the two depends on the application's specific requirements. MTCNN is more appropriate for applications demanding high accuracy and detection of multiple faces, while DLIB is suitable for real-time applications.

The facial recognition system in our application utilizes MTCNN due to its outstanding performance and accuracy in our evaluations. MTCNN's cascade approach to face detection and landmark localization has proven to be highly accurate in recognizing multiple faces within an image and extracting their facial landmarks.

In contrast to DLIB's reliance on HOG features and a linear classifier, MTCNN's deep learning-based technique makes it more appropriate for tasks requiring high accuracy and detection of multiple faces.

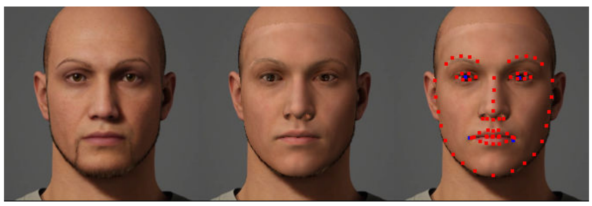

Figure 5. MTCNN usage example

Figure 6. DLIB usage example

The provided examples show that in our case MTCNN produces significantly better results than the DLIB algorithm.

III. Existing 3D model reconstruction systems

The field of 3D modeling has witnessed remarkable advancements over the years. These advancements have allowed us to convert a set of 2D photographs into detailed and immersive 3D models. Despite the progress made, many of the existing methods have their limitations, preventing their widespread adoption.

One common drawback is the reliance on a single image for 3D model reconstruction. Although this approach offers convenience by requiring only one input, working with just one image can limit the system's ability to capture intricate details and complexity of a subject, resulting in incomplete or distorted representations which lack accuracy and fidelity, restricting their usability in facial recognition.

On the other end of the spectrum, some 3D modeling systems demand a significant number of images for the reconstruction process. While this approach can potentially produce highly detailed and precise models, it brings challenges in terms of cost, control, and practicality. Acquiring and processing many images can be time-consuming and expensive, requiring substantial resources and expertise. Moreover, managing the vast amount of data generated by these systems can be overwhelming, making it difficult to effectively control considering the speed that is required when authenticating a person not to keep them waiting or to apprehend them promptly.

For instance, the implementation created by S. Xu et al. provides an intriguing illustration of a system that builds a 3D model from a single photograph. According to their study they use Python and a weakly-supervised learning strategy to create a 3D model from a single image. For CNN-based 3D face reconstruction, the method used by Xu et al. requires a hybrid-level training with lax supervision. Their technology demonstrates quick, accurate, and reliable reactivity to a variety of locations and occlusions [7].

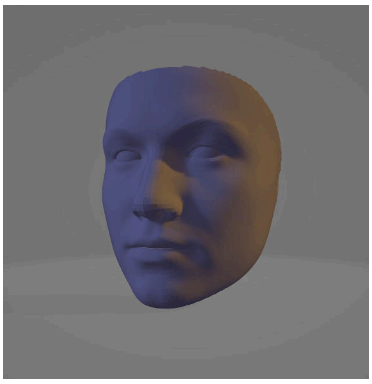

Figure 7. Single image 3D reconstruction example

Figure 7 shows an example generated by the model shown in Figure 4 using the MTCNN technique for landmark extraction to provide visual context. However, it is significant to highlight that one drawback of this specific 3D model is the lack of ears, which are essential for authenticating a person.

The implementation by Xu et al. shows the capability of constructing 3D models from a single image by using weakly-supervised learning methods. Although it displays outstanding responsiveness and resilience abilities, the absence of ears in the reconstructed model raises questions about its suitability for use in authenticating procedures. For 3D models to be accurate and reliable, particularly in industries where precise identification is essential, these restrictions must be addressed.

However, for now, the absence of ears in the 3D model generated from a single image remains a challenge that needs to be overcome. Further research and development efforts are required to refine the existing techniques and algorithms, enabling the creation of more comprehensive and accurate representations that encompass essential features such as ears. By doing so, the applications of single-image-based 3D modeling can be expanded to various domains, including biometrics, virtual reality, and entertainment, offering a wide range of immersive experiences and practical functionalities.

IV. Improving results

To enhance these results, two crucial modifications need to be implemented. First, the system should pass from using a single image to creating a 3D model based on three images. Additionally, the inclusion of ears in the reconstructed model must be prioritized.

Accomplishing these improvements requires careful consideration of the angles at which the pictures should be captured.

A. Camera angles

For the system to effectively utilize multiple images, specifically three in this case, it is crucial that the provided images contain sufficient information to enable accurate and robust reconstruction. This necessitates careful positioning of the cameras, with particular emphasis placed on the angles from which the cameras capture the individual being analyzed.

In determining the optimal camera angles for three-view 3D face reconstruction, it is essential to consider several factors. First, the angles should provide a comprehensive coverage of facial features, capturing a variety of perspectives to capture the true three-dimensional nature of the face. Second, the camera angles should be carefully chosen to minimize occlusions and shadows that could hinder accurate landmark detection and pose estimation. Additionally, the angles should be selected to avoid any potential distortions or biases that may arise from specific viewing perspectives. Striking a balance between capturing sufficient detail and avoiding potential pitfalls is the key to achieving reliable and precise three-view 3D face reconstruction.

Figure 8. Right and left profile views

Undoubtedly, the central camera should provide a direct view of the person. The other two cameras should be symmetrical to each other.
The best angles were determined to be around 35°, as they captured the most information about the ears, without capturing too much of the back of the head.

B. Ear reconstruction

To achieve realistic and dependable representations, precise ear reconstruction must be included in 3D facial models. Ears contribute to overall face identification and play an important part in authentication procedures. However, many existing 3D model reconstruction techniques frequently miss the reconstruction of ears, restricting their usefulness in a variety of disciplines.
When rebuilding ears in 3D models, several difficulties arise. The partial concealment of ears in photographs by hair or other facial structures makes proper recognition and reconstruction difficult. Furthermore, the ear's uneven shape and detailed characteristics necessitate specific procedures to achieve precise and realistic reconstruction.

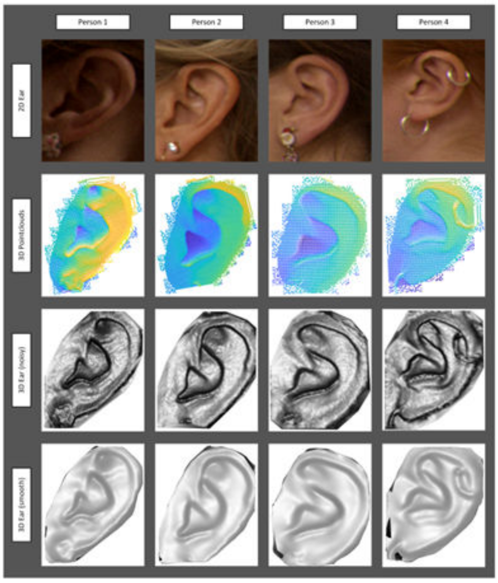

Figure 9. Examples of ear 3D reconstruction

One such example can be the following work by Md. Mursalin, M. Ahmed, and P. Haskell-Dowland “Biometric Security: A Novel Ear Recognition Approach Using a 3D Morphable Ear Model” where they reconstruct the ear using 3D morphable ear model (3DMEM) which is then used human recognition. Their proposed method can work in real time, which is crucial for real-time authentication. Examples are shown in Figure 9 [8].

V. Conclusion

Overall, the article explored various methodologies for 3D face reconstruction, focusing on the development of robust and reliable systems capable of authenticating individuals through image collection and reconstruction of their 3D facial models. While several approaches were discussed, it is important to note that no single system encompasses all the necessary features and achieves extensive accuracy.
One limitation observed in some systems is the reliance on a single image for 3D reconstruction. This approach restricts the system's accuracy as it may not capture the full range of facial expressions and variations that occur naturally. To address this, future research should focus on incorporating multiple images or exploring alternative data sources to improve the fidelity of the reconstructed 3D models.

Additionally, a noteworthy aspect missing in certain systems is the inclusion of the ear in the 3D facial model. The ear is an important anatomical feature that can provide valuable cues for accurate authentication. To enhance the reliability of these systems, efforts should be made to incorporate the ear into the reconstruction process, ensuring that the resulting 3D model captures all essential facial characteristics.
In conclusion, it is evident that further research and development are necessary to achieve a more advanced and dependable 3D face reconstruction system. Future efforts should aim to address the limitations discussed, such as incorporating multiple images to improve accuracy and ensuring the inclusion of the ear in the reconstructed 3D model. By advancing these techniques, we can pave the way for more trustworthy and effective systems that enhance security and authentication processes in various domains.

References

[1] T. Akimoto, Y. Suenaga, and R. S. Wallace, "Automatic creation of 3D facial models," *IEEE Computer Graphics and Applications*, vol. 13, no. 5, pp. 16-22, Sept. 1993.

[2] X. Guo, Y. Li, and H. Ling, "LIME: Low-Light Image Enhancement via Illumination Map Estimation," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 982-993, Feb. 2017.

[3] D. Huang, C. Shan, M. Ardabilian, Y. Wang, and L. Chen, "Local Binary Patterns and Its Application to Facial Image Analysis: A Survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 41, no. 6, pp. 765-781, Nov. 2011.

[4] Adrian Rosebrock, "Face recognition with OpenCV, Python, and deep learning" available at https://pyimagesearch.com/2018/06/18/face-recognition-with-opencv-python-and-deep-learning/, last visited May 2023.

[5] Davis E. King, Available at https://dlib.net, last visited May 2023.

[6] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499-1503, Oct. 2016.

[7] Y. Deng, J. Yang, S. Xu, D. Chen, Y. Jia, and X. Tong, “Accurate 3D Face Reconstruction with Weakly-Supervised Learning: From Single Image to Image Set”, *IEEE Computer Vision and Pattern Recognition Workshop (CVPRW) on Analysis and Modeling of Faces and Gestures (AMFG)*, 2019.

[8] Md. Mursalin, M. Ahmed, and P. Haskell-Dowland. "Biometric Security: A Novel Ear Recognition Approach Using a 3D Morphable Ear Model" *Sensors 22*, no. 22, 2022.