STAG: Smart Tools and Applications in Graphics (2023)
G. Manfredi, G. Caggianese, N. Capece, and U. Erra (Editors)

Poster

AvatarizeMe: A Fast Software Tool for Transforming Selfies into Animatable Lifelike Avatars Using Machine Learning

G. Manfredi[1], N. Capece[1], and U. Erra[1]

[1] University of Basilicata, Department of Mathematics, Computer Science and Economics, Potenza 85100, Italy

Abstract

Creating realistic avatars that faithfully replicate facial features from single-input images is a challenging task in computer graphics, virtual communication, and interactive entertainment. These avatars have the potential to revolutionize virtual experiences by enhancing user engagement and personalization. However, existing methods, such as 3D facial capture systems, are costly and complex. Our approach adopts the 3D Morphable Face Model (3DMM) method to create avatars with remarkably realistic features in a bunch of seconds, using only a single input image. Our method extends beyond facial shape resemblance; it meticulously generates both facial and bodily textures, enhancing overall likeness. Within Unreal Engine 5, our avatars come to life with real-time body and facial animations. This is made possible through a versatile skeleton for body and head movements and a suite of 52 face blendshapes, enabling the avatar to convey emotions and expressions with fidelity. This poster presents our approach, bridging the gap between reality and virtual representation, and opening doors to immersive virtual experiences with lifelike avatars.

CCS Concepts

• ***Computing methodologies*** → ***Computer vision; Machine learning; Parametric curve and surface models; Texturing;***

1. Introduction

The intent of creating realistic avatars, mirroring the facial features of individuals captured in input images, holds huge promise for enhancing user engagement, immersion, and presence in virtual environments [LRG*17; WMC; LMG*20]. However, existing state-of-the-art approaches have some limitations. Conventional methods, such as sophisticated 3D facial capture systems [DHT*00; BHPS10; AWLB17; CWW*16], are effective but expensive and complex to deploy widely, as they rely on multi-perspective acquisitions for accurate texture mapping. There are also methods that generate a textured 3D face or head mesh from a single image [CRG*23; JZD*18]. However, it is challenging to develop a fast 3D modeling algorithm capable of seamlessly adapting these finished meshes to a 3D human body during the application's execution. In contrast, our approach aims to overcome these constraints by using the 3DMM [BV99; GPKZ19] to quickly create avatars with realistic features, given a single input image. Our method does not limit itself solely to the face shape resemblance; it extends its purview to the generation of facial and bodily textures, thereby enhancing the overall likeness of the avatar to the source image. Furthermore, within Unreal Engine 5 (UE5), our avatars gain life-like dynamism. They are endowed with the ability to respond in real time, moving their body and facial expressions convincingly. This is facilitated by the introduction of a versatile skeleton for body and head movements and a comprehensive suite of 52 face blendshapes, which enable the avatar to transmit expressions with fidelity.

2. 3D Morphable Face Model

Our methodology involves using the FLAME (Faces Learned with an Articulated Model and Expressions) model [LBB*17], a parameterized vertex-based generic head model learned from an extensive dataset of over 33,000 accurately aligned 3D scans. The model consists of a template mesh in the “zero pose,” a shape blendshape function for identity-related shape variation, corrective and expression blendshapes. For our specific application, we utilize the subset of blendshapes from the FLAME model related to identity shape variation. These selected blendshapes provide the foundation for aligning the 3DMM with the facial landmarks and contours present in the input image.

3. Preliminary Operation: Blender Integration

In the preliminary phase, we utilize Blender, a versatile 3D modeling and animation software, for three crucial tasks:

1. **Attachment of FLAME Head Mesh to the Body:** We used the MB-Lab [Bas] add-on for Blender to generate a generic avatar equipped with a pre-existing skeleton and a set of blendshapes designed for body deformation. To seamlessly integrate

DOI: 10.2312/stag.20231304

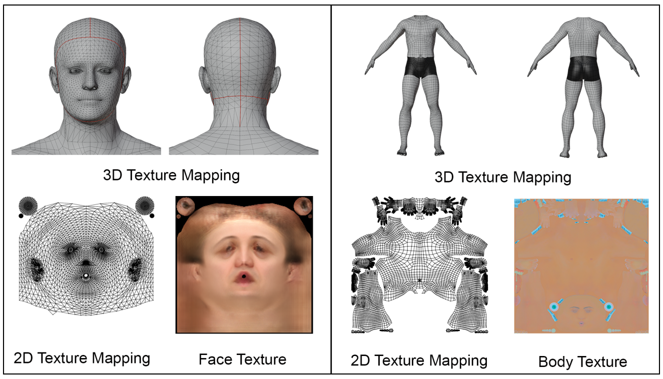

Figure 1: *Texture Mapping for the head and the body of the avatar.*

the FLAME head model into this generic avatar, we replaced the avatar's original head with the FLAME head mesh, which inherently includes the necessary blendshapes for head deformation. Subsequently, we adjusted the body blendshapes to harmonize with the new head structure. Finally, we connected the newly integrated head to the neck and hand skeleton bones, ensuring a natural fusion of the FLAME head with the avatar's body.

2. **Texture Mapping:** In this phase, we used the Blender texture mapping capabilities to map textures onto both the head and body components of the avatar individually (see Figure 1). This separation ensures that the facial texture is accurately derived from the input image, enhancing the overall likeness of the avatar to the source picture (see Figure 2).
3. **Creation of 52 Blendshapes for Facial Expressions:** Blender plays a pivotal role in crafting the 52 ARKit face blendshapes [ARK] necessary for animating the avatar's facial expressions in real-time. These blendshapes provide the foundation for conveying a wide range of emotions.

4. Runtime Operation: Image Processing, Blendshapes' Values Extraction, and Texture Creation

During the runtime operation, we proceed to transform a single input image into a realistic avatar (see Figure 3). This phase includes:

1. **Image Capturing and Preprocessing:** In the runtime phase, the application starts by capturing a single user's facial image. A face detection model [ZZL*17] identifies the bounding box coordinates for the detected face and crops the input image accordingly. Next, a facial landmark detection model [BT17] pinpoints key facial landmarks such as the eyes, nose, and mouth. To further refine the 3D face model fitting, we employ a semantic segmentation neural network [NMT*18] to separate the face from the background, enhancing the precision of facial deformation.
2. **Blendshapes' Values Detection for Facial Deformation:** This phase involves optimizing various facial parameters to align the 3D FLAME model with the detected face in the input image.
3. **Creation of Head and Body Textures:** To generate the head texture, our approach initially applies a default texture to the 3D avatar's face. Subsequently, employing a photometric loss, the algorithm compares the avatar's face to the input image, finely adjusting the texture to align with the reference image. The photometric loss function is the Frobenius norm of the difference between the Gaussian-filtered input image and the Gaussian-filtered texture [BBLR15]. For the body texture, the algorithm utilizes a color transfer technique to harmonize the colors between a default body texture and the avatar's face texture, ensuring a consistent and realistic appearance across the entire avatar. To do so, it calculates color statistics for both images, analyzing the L, a, and b channels in the *Lab* color space. Then it performs a color adjustment process, which involves shifting and scaling the L, a, and b channels of the body texture to match the corresponding statistics of the face texture.

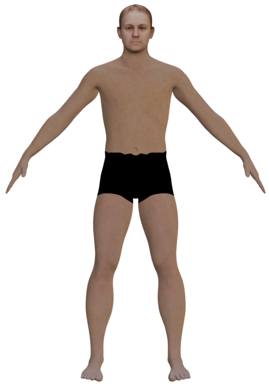

Figure 2: *Textured avatar with FLAME head.*

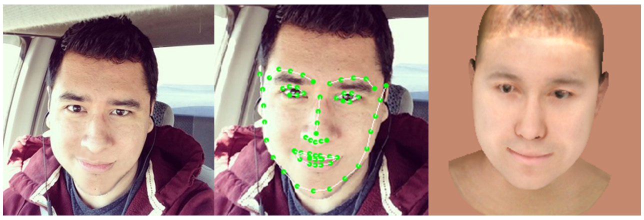

Figure 3: *Runtime Operation. The selfie image was sourced from the Selfie-Image-Detection-Dataset [Bha].*

5. Real-Time Operation

In the final phase, we used a MediaPipe plugin for UE5 for the real-time face and body animation, allowing the avatar to convey a wide range of facial expressions and body movements with striking fidelity.

6. Conclusions

In this poster, we have introduced a fast method for generating highly realistic avatars from a single input image. Our approach goes beyond mere facial shape resemblance as it can craft both facial and bodily textures, enhancing the avatar realism. Furthermore, our avatars, when integrated into UE5, gain the capability to respond in real time, with fluid body and facial movements, thanks to the incorporation of a skeleton for body and head movements, complemented by a comprehensive suite of 52 face blendshapes for expressions. Our ongoing work could focus on refining facial textures and potentially extending our method to include the creation of realistic hair, all deriving from a single selfie.

References

[ARK] ARKit. *ARKit Face Blendshapes*. https://developer.apple.com/documentation/arkit/arfaceanchor/blendshapelocation. Accessed: 2023-09-11 2.

[AWLB17] Achenbach, Jascha, Waltemate, Thomas, Latoschik, Marc Erich, and Botsch, Mario. "Fast generation of realistic virtual humans". *Proceedings of the 23rd ACM symposium on virtual reality software and technology*. 2017, 1–10 1.

[Bas] Bastioni, Manuel. *MB-Lab*. https://mb-lab-community.github.io/MB-Lab.github.io/. Accessed: 2023-09-11 1.

[BBLR15] Bogo, Federica, Black, Michael J, Loper, Matthew, and Romero, Javier. "Detailed full-body reconstructions of moving people from monocular RGB-D sequences". *Proceedings of the IEEE international conference on computer vision*. 2015, 2300–2308 2.

[Bha] Bhatt, Jigar. *Selfie-Image-Detection-Dataset*. https://www.kaggle.com/datasets/jigrubhatt/selfieimagedetectiondataset. Accessed: 2023-09-11 2.

[BHPS10] Bradley, Derek, Heidrich, Wolfgang, Popa, Tiberiu, and Sheffer, Alla. "High resolution passive facial performance capture". *ACM SIGGRAPH 2010 papers*. 2010, 1–10 1.

[BT17] Bulat, Adrian and Tzimiropoulos, Georgios. "How far are we from solving the 2d & 3d face alignment problem? (and a dataset of 230,000 3d facial landmarks)". *Proceedings of the IEEE international conference on computer vision*. 2017, 1021–1030 2.

[BV99] Blanz, V and Vetter, T. "A Morphable Model for the Synthesis of 3D Faces". *26th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 1999)*. ACM Press. 1999, 187–194 1.

[CRG*23] Caselles, Pol, Ramon, Eduard, Garcia, Jaime, et al. "Sira: Relightable avatars from a single image". *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2023, 775–784 1.

[CWW*16] Cao, Chen, Wu, Hongzhi, Weng, Yanlin, et al. "Real-time facial animation with image-based dynamic avatars". *ACM Transactions on Graphics* 35.4 (2016) 1.

[DHT*00] Debevec, Paul, Hawkins, Tim, Tchou, Chris, et al. "Acquiring the reflectance field of a human face". *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*. 2000, 145–156 1.

[GPKZ19] Gecer, Baris, Ploumpis, Stylianos, Kotsia, Irene, and Zafeiriou, Stefanos. "Ganfit: Generative adversarial network fitting for high fidelity 3d face reconstruction". *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019, 1155–1164 1.

[JZD*18] Jiang, Luo, Zhang, Juyong, Deng, Bailin, et al. "3D face reconstruction with geometry details from a single image". *IEEE Transactions on Image Processing* 27.10 (2018), 4756–4770 1.

[LBB*17] Li, Tianye, Bolkart, Timo, Black, Michael J, et al. "Learning a model of facial shape and expression from 4D scans." *ACM Trans. Graph.* 36.6 (2017), 194–1 1.

[LMG*20] Lattas, Alexandros, Moschoglou, Stylianos, Gecer, Baris, et al. "AvatarMe: Realistically Renderable 3D Facial Reconstruction" in-the-wild"". *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020, 760–769 1.

[LRG*17] Latoschik, Marc Erich, Roth, Daniel, Gall, Dominik, et al. "The effect of avatar realism in immersive social virtual realities". *Proceedings of the 23rd ACM symposium on virtual reality software and technology*. 2017, 1–10 1.

[NMT*18] Nirkin, Yuval, Masi, Iacopo, Tran, Anh Tuan, et al. "On Face Segmentation, Face Swapping, and Face Perception". *IEEE Conference on Automatic Face and Gesture Recognition*. 2018 2.

[WMC] Wilser, Nicola, Maillot, Yvan, and Cordier, Frédéric. "A survey of 3D human body reconstruction from single and multiple camera views". *Available at SSRN 4333245* () 1.

[ZZL*17] Zhang, Shifeng, Zhu, Xiangyu, Lei, Zhen, et al. "S3fd: Single shot scale-invariant face detector". *Proceedings of the IEEE international conference on computer vision*. 2017, 192–201 2.