Toward Realistic 3D Avatar Generation with Dynamic 3D Gaussian Splatting for AR/VR Communication

Hail Song*
KAIST UVR Lab

ABSTRACT

Realistic avatars are fundamental for immersive experiences in Augmented Reality (AR) and Virtual Reality (VR) environments. In this work, we introduce a novel approach for avatar generation, combining 3D Gaussian Splatting with the parametric body model, SMPL. This methodology overcomes the inefficiencies of traditional image/video-based avatar creation, which is often slow and requires high computing resources. The integration of 3D Gaussian Splatting for representing human avatar offers realistic and real-time rendering for AR/VR applications. We also conducted preliminary tests to verify the quality of avatar representation using 3D Gaussian Splatting. These tests, displayed alongside outcomes from existing methods, demonstrate the potential of this research to significantly contribute to the creation of realistic avatars in the future. Additionally, several key discussions are presented, essential for developing and evaluating the system and providing valuable insights for future research.

Index Terms: Computing methodologies—Artificial intelligence—Computer vision—Reconstruction Computing methodologies—Computer graphics—Shape modeling—Point-based models;

1 INTRODUCTION AND RELATED WORK

In the field of Augmented Reality (AR) and Virtual Reality (VR), avatars with realistic styles are crucial for communication. Users in virtual environments utilize avatars to express their identity and communicate with remote users. Various research efforts have been conducted to create avatar models that reflect the actual appearance of users.

Among these studies, parametric body models have been widely used to create animatable bodies and visualize facial expressions [4, 6]. These models use statistical methods for 3D body representation and simplify the depiction of realistic body meshes by utilizing a set of shape and pose parameters. Their approach maintains the correspondence between the model's vertices and body parts and includes UV coordinates, which facilitate texture and normal map modifications and enhance usability in various applications.

Various academic methods that have adopted parametric body models as their basic 3D representation have enabled the reconstruction of 3D avatars based on the actual appearance of users, utilizing images [6], videos [1], or video streams [7]. However, 3D human models based on template avatar meshes have been unable to represent high-fidelity avatars or complex clothing topologies.

Recently, Neural Radiance Fields (NeRF) [5] have shown impressive performance in novel view synthesis and scene representation. NeRF has also enabled the representation of realistic and dynamic avatars by embedding dynamic features and SMPL mesh information [2]. However, despite their ability to render realistic avatars, the computational cost of calculating color and opacity through per-pixel ray marching has been significant. This process is particularly challenging for securing high FPS, especially when considering Head-Mounted Displays (HMDs) that require binocular rendering.

*e-mail: hail96@kaist.ac.kr

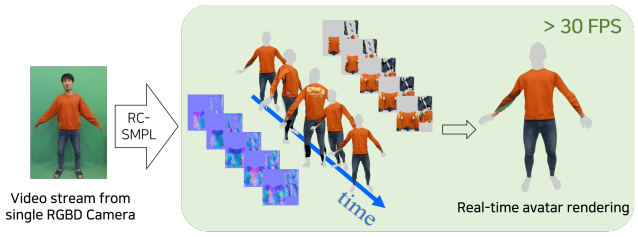

Figure 1: Concept diagram of RC-SMPL [7]. The system can reconstruct 3D human body models with texture and normal maps during real-time rendering.

To overcome the low rendering FPS, the technique of 3D Gaussian Splatting [3] has recently emerged and gained attention. This method handled the three-dimensional elements composing a scene as dense splats possessing 3D Gaussian distribution information, enabling representation from new views. Furthermore, due to its characteristic of following a Gaussian distribution, inputting a new view vector into the rendering process requires significantly fewer computational resources compared to traditional NeRF-based methods. With its ability to realistically represent scenes and achieve high FPS exceeding 100, Gaussian Splatting is expected to be effective in representing VR scenes and AR objects.

Against this background, we propose a dynamic 3D Gaussian Splatting technique for realistic 3D avatar generation. By embedding SMPL parameters along with the high FPS and realistic scene representation shown by 3D Gaussian Splatting, we aim to demonstrate real-time rendering on HMDs that accurately reflects body shape and pose.

2 PREVIOUS RESEARCH : RC-SMPL

Before developing an avatar generation system based on 3D Gaussian Splatting, we conducted research on real-time avatar generation [7] using the SMPL model [4]. While traditional image/video-based avatar creation methods could conveniently generate 3D human models, they were time-consuming in model reconstruction. Methods using NeRF [2] also showed high reconstruction quality but required substantial computing resources, making them unsuitable for the real-time demands of AR/VR environments. Therefore, we proposed a system that rapidly creates avatars by asynchronously accumulating the texture and normal maps of the SMPL model. This system reduced the time required for avatar creation and utilized a single RGBD camera, minimizing time and spatial costs while acquiring avatar models comparable to existing video-based method [1]. The concept figure is at Fig. 1. The rendering results of the system are at Fig. 2. However, due to the expressive limitations of the SMPL model, it was impossible to detail clothing, and the mesh-based avatar representation showed lower quality compared to studies based on NeRF.

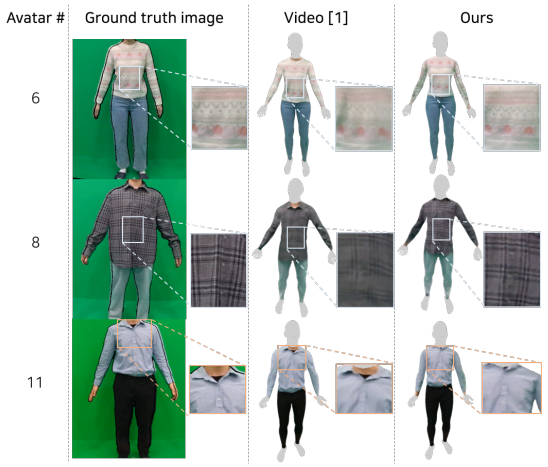

Figure 2: Rendering results of avatar generation methods. From left to right: Ground truth image, avatar generated by the video-based restoration method by Alldieck et al. [1] (Video), and avatar generated by RC-SMPL [7].

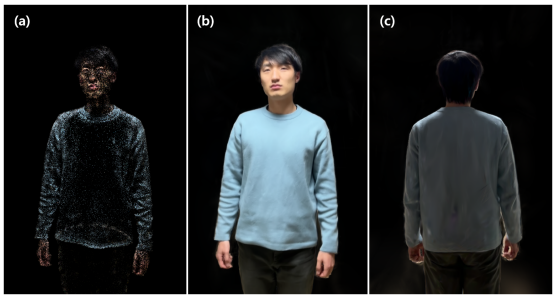

Figure 3: Rendering results of 3D human body from 3D Gaussian Splatting. (a) Point clouds from the results, (b) Front side, (c) Back side

3 Research Plan

Building on a fundamental understanding of avatar creation, we are in the process of developing a dynamic avatar based on 3D Gaussian Splatting and SMPL. 3D Gaussian Splatting is a rendering technique that represents 3D scenes as splats with a high-density Gaussian distribution.

3.1 Training 3D Gaussian Splatting for expression of avatar

Firstly, we conducted a test to check the rendering quality of 3D Gaussian Splatting for a static person. We recorded a video by circling the person positioned at the center for 20 seconds. From the captured video, we selected five frames per second and used the Structure-from-Motion (SfM) algorithm to determine camera positions, which were then utilized as input for training 3D Gaussian Splatting. The rendering results are in Fig. 3. We observed that the range of expression in the splats captured details such as clothing wrinkles and facial expressions much better than other methods presented in Fig. 2. However, since 3D Gaussian Splatting is fundamentally designed for static scene representation, it may not achieve high restoration effects when expressions or poses change. We plan to further develop this by embedding SMPL's body shape and pose information into the splats to enable dynamic body representation. An iPhone 13 was used for video recording, and the footage, with a resolution of 1920x1080 at 60FPS, served as the training data.

3.2 VR renderer for 3D Gaussian Splatting

To visualize the results on a VR HMD, rendering for both eyes is necessary, requiring parallel image rendering. To verify the real-time rendering capabilities of 3D Gaussian Splatting, we tested rendering speed using the 6 Degrees of Freedom (6 DoF) input from Meta Quest Pro[1], and assessed the FPS levels during the extraction of rendering results for the left and right eyes. Despite binocular rendering, we confirmed that a stable 70FPS was maintained.

4 Discussion

In this work, we have presented a plan to utilize 3D Gaussian Splatting and SMPL for realistic 3D avatar generation. Additionally, a simple comparison with existing avatar creation methods allowed us to observe potential improvements in avatar quality due to the introduction of this new technology. By completing the project currently under development, we aim to enable more realistic avatar representations in AR/VR environments, enhancing the sense of presence in metaverse interactions. Furthermore, we plan to expand our research to enable interactions using the point cloud information generated during 3D Gaussian Splatting representation.

Finally, we seek advice on the following agendas related to generating 3D avatars.

- Can we achieve satisfactory rendering results by animating 3D Gaussian Splats using Linear Blend Skinning (LBS)?
- What considerations are needed for combining Gaussian rendering techniques with traditional graphic pipelines for representing objects or avatars?
- What would be the appropriate experimental methods and metrics for testing interactions with the generated avatars?

Acknowledgments

This research was supported by National Research Council of Science and Technology(NST) funded by the Ministry of Science and ICT(MSIT), Republic of Korea(Grant No. CRC 21011) and Korea Institute for Advancement of Technology(KIAT) grant funded by the Korea Government(MOTIE) (P0012746, The Competency Development Program for Industry Specialist)

References

[1] T. Alldieck, M. Magnor, W. Xu, C. Theobalt, and G. Pons-Moll. Video based reconstruction of 3d people models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8387–8397, 2018.

[2] T. Jiang, X. Chen, J. Song, and O. Hilliges. Instantavatar: Learning avatars from monocular video in 60 seconds. *arXiv preprint arXiv:2212.10550*, 2022.

[3] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics (ToG)*, 42(4):1–14, 2023.

[4] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black. Smpl: A skinned multi-person linear model. *ACM transactions on graphics (TOG)*, 34(6):1–16, 2015.

[5] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021.

[6] G. Pavlakos, V. Choutas, N. Ghorbani, T. Bolkart, A. A. Osman, D. Tzionas, and M. J. Black. Expressive body capture: 3d hands, face, and body from a single image. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10975–10985, 2019.

[7] H. Song, B. Yoon, W. Cho, and W. Woo. Rc-smpl : Real-time cumulative smpl-based avatar body generation. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2023.

[1] https://www.meta.com/quest/quest-pro/