Cross-Domain Synthetic Data Generation

From Avatars to Real-World Applications

Introduction

Modern avatar reconstruction increasingly relies on synthetic data to address the limitations of real-world capture. Synthetic human datasets offer full control over body shape, pose, and appearance, enabling large volumes of accurate, diverse, and richly annotated samples that are difficult, time-consuming, or costly to obtain through traditional scanning.

Measurement-Driven Avatars

We reconstruct metrically accurate 3D avatars from only frontal and lateral silhouettes [2]. Anthropometric measurements are estimated from the silhouettes and mapped to SMPL [6] shape parameters using regression trained on a large synthetic human dataset. This enables reliable body-shape reconstruction without 3D scans.

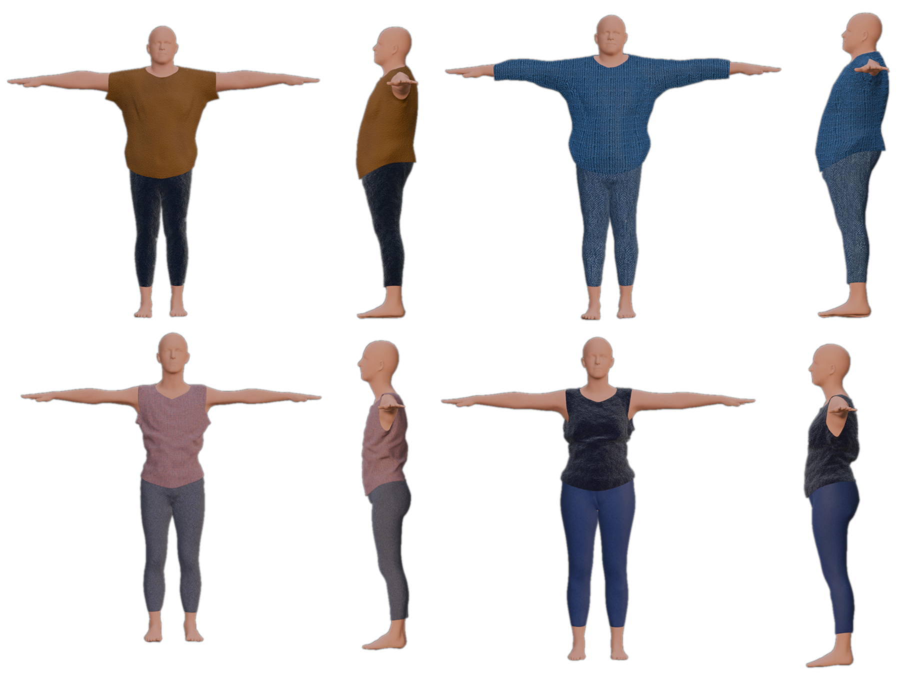


Figure 1. Anthropometrically correct avatars.

The pipeline follows a simple sequence: silhouettes are processed by a ResNet-based [3] model to extract key anthropometric cues, which are then translated into a parametric body model through a learned measurement-to-shape mapping. This structured flow ensures that each stage contributes directly to preserving realistic proportions and generating a coherent, metrically faithful avatar.

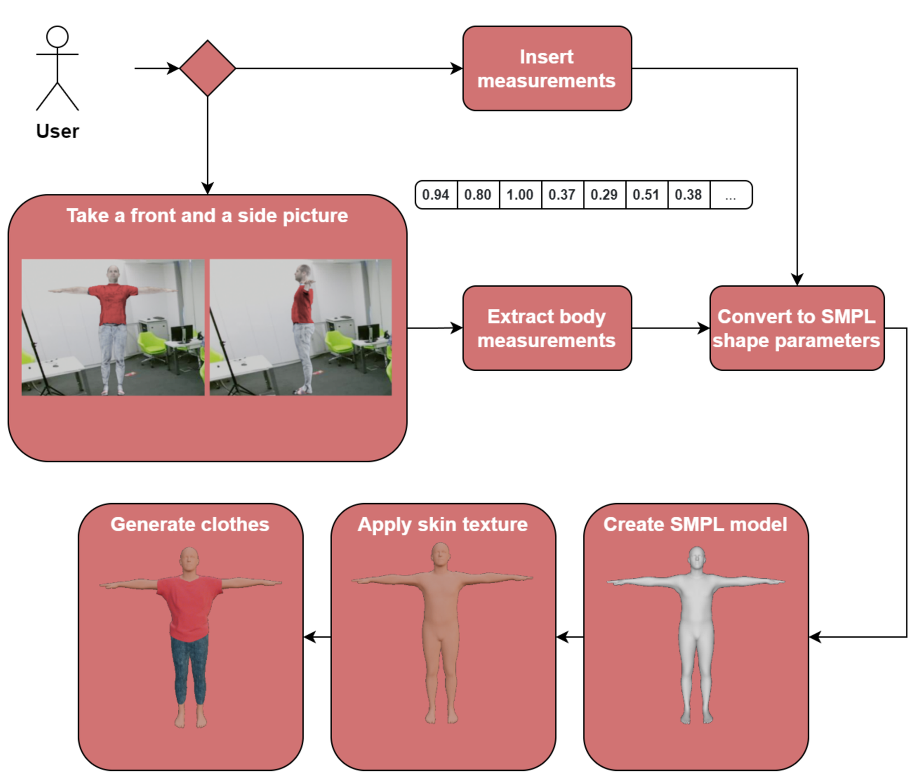
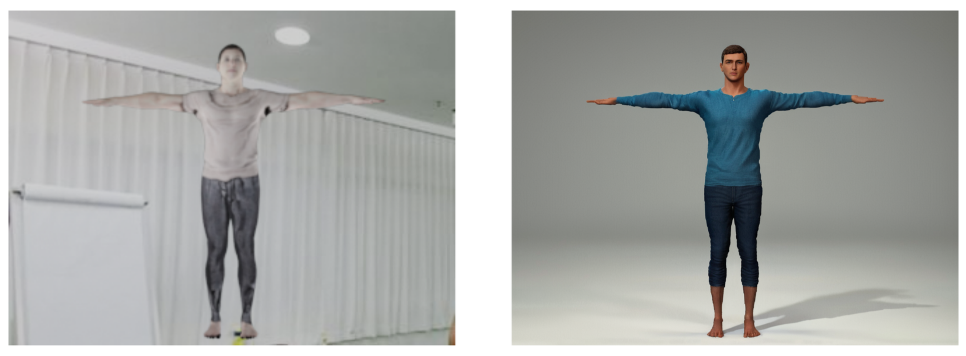


Figure 2. Avatar generation pipeline.

By relying on synthetic training data, the system learns stable measurement relationships that generalize well to real silhouettes with varying quality.

Avatar Refinement

Synthetic humans in T-pose are generated together with clean segmentation masks, and their visual realism is enhanced using ControlNet-guided Stable Diffusion [1].



(a) original (b) enhanced

Figure 3. Synthetic data enhanced by stable diffusion.

The enhanced outputs introduce richer textures, more natural shading, and subtle visual details that are difficult to achieve with purely procedural rendering. This refinement step helps reduce the synthetic–to–real domain gap.

Neural Enhancement

In addition, the corresponding 3D models can be further refined using neural representations such as NeRF [7] or Gaussian Splatting [4]. These methods capture subtle surface detail and realistic lighting effects that traditional mesh-based techniques often miss. By combining neural fields with classical geometry, the resulting models achieve higher visual fidelity and more natural appearance.

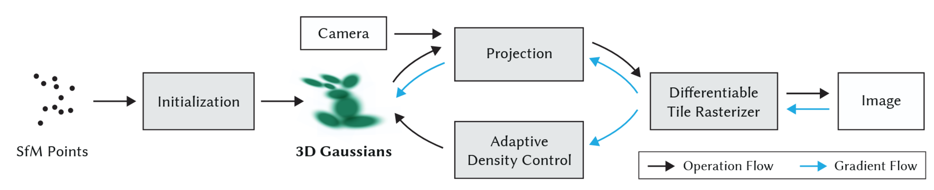


Figure 4. Gaussian Splatting optimization [4].

Gaussian-Based Avatar Model

FLAME [5] parameters are first extracted from the input video using VHAP [8], providing an initial coarse head model. This geometry is then refined with GaussianAvatars [9], where Gaussian Splatting recovers detailed appearance and realistic surface structure. The final representation captures high-fidelity textures and geometry beyond what traditional meshes can achieve.

Figure 5. From FLAME estimate to Gaussian avatar.

The entire reconstruction in the Figure 5 was generated from a single monocular video, demonstrating that high-fidelity head avatars can be obtained without multi-view capture or specialized hardware.

Conclusion

Synthetic data provides a powerful foundation for both human-centered and industrial 3D vision tasks. By combining procedurally generated datasets with modern neural techniques such as diffusion models, NeRFs, and Gaussian Splatting, it is possible to achieve high levels of visual fidelity and metric accuracy even from minimal or imperfect inputs. Our results demonstrate that silhouettes, T-pose renders, and monocular videos can be transformed into detailed 3D avatars through hybrid pipelines that merge classical geometry with neural representations. These approaches significantly reduce the need for costly data capture while narrowing the synthetic-to-real gap, enabling scalable and reliable 3D vision solutions. As synthetic generation and neural rendering continue to evolve, their integration will further broaden the applicability of 3D models.

References

[1] Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, Dustin Podell, Tim Dockhorn, Zion English, and Robin Rombach.
Scaling rectified flow transformers for high-resolution image synthesis.
In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24. JMLR.org, 2024.

[2] M. Halaj, D. Škorvánková, and M. Madaras.
Metrically Accurate 3D Human Avatars from Silhouette Images.
Journal of WSCG, 33(1):33–42, 2025.

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun.
Deep residual learning for image recognition.
In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[4] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis.
3D Gaussian Splatting for Real-Time Radiance Field Rendering.
ACM Transactions on Graphics, 42(4), July 2023.

[5] Tianye Li, Timo Bolkart, Michael. J. Black, Hao Li, and Javier Romero.
Learning a model of facial shape and expression from 4D scans.
ACM Transactions on Graphics, (Proc. SIGGRAPH Asia), 36(6):194:1–194:17, 2017.

[6] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black.
SMPL: A Skinned Multi-Person Linear Model.
ACM Trans. Graphics (Proc. SIGGRAPH Asia), 34(6):248:1–248:16, October 2015.

[7] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng.
Nerf: Representing scenes as neural radiance fields for view synthesis.
In *ECCV*, 2020.

[8] Shenhan Qian.
VHAP: Versatile Head Alignment with Adaptive Appearance Priors, sep 2024.

[9] Shenhan Qian, Tobias Kirschstein, Liam Schoneveld, Davide Davoli, Simon Giebenhain, and Matthias Nießner.
GaussianAvatars: Photorealistic Head Avatars with Rigged 3D Gaussians.
In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20299–20309, 2024.

FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY
Univerzita Komenského v Bratislave

Skeletex RESEARCH

MATFYZ CONNECTIONS

Martin Halaj, Martin Madaras
KAI FMFI UK, Skeletex Research s.r.o.