High-Quality Static Head Avatars with Gaussian Splatting

Jie Peng (Frank) Chen
Affiliations, University of Toronto

Motivation

- 3D Gaussian Splatting is effective for rendering high-quality 3D scenes, making it ideal for static head avatars in AR/VR and gaming. However, static head avatars for novel view synthesis remain underexplored, with recent research focusing primarily on animated head avatars for capturing facial expressions. Moreover, 3D Gaussian Splatting faces two main issues: reliance on SfM for point cloud initialization, which struggles with smooth geometries in human head avatars, and its discrete representation, which leads to artifacts like popping and floating.

- To address these challenges, we explore random point initialization with iterative addition and pruning within the head avatar mask. Leveraging a NeRF-inspired MLP, we aim to predict color and opacity for smoother rendering, reduced artifacts, and improved visual quality.

Related Work

- **3D Gaussian Splatting** [1] is a highly efficient method for real-time rendering with high quality, offering fast rendering performance and accurate synthesis for unseen view directions, making it ideal for static head avatars.

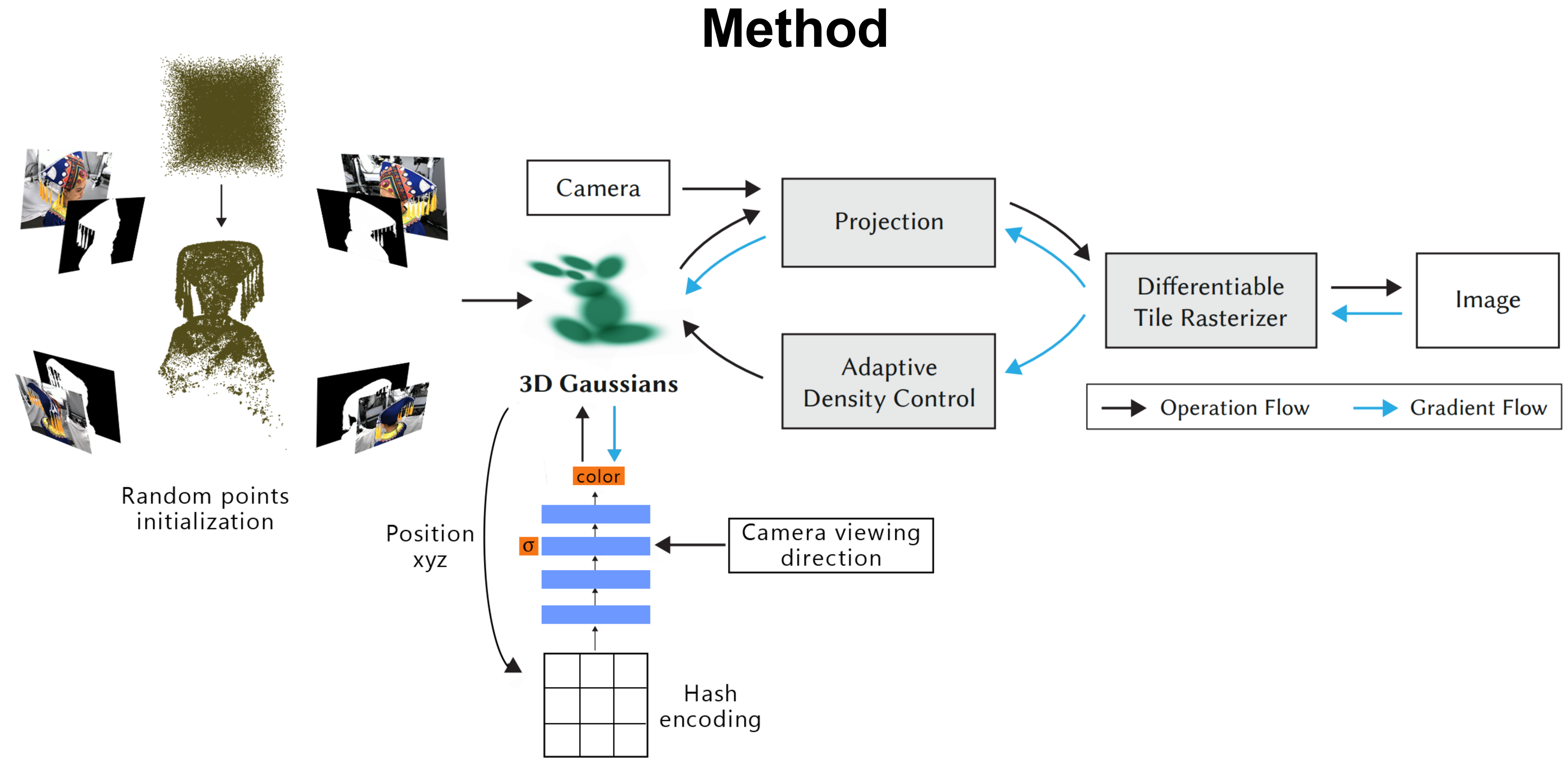
- **MonoGaussianAvatar** [2] introduced random point initialization techniques for monocular portrait videos, demonstrating their effectiveness and inspiring our approach to initializing points in static head avatars.

- **Instant Neural Graphics Primitives** [3] developed a method to train and render MLPs using multiresolution hash input encoding, which influenced our adoption of hash encoding for predicting color and opacity in static head avatar rendering.

References

[1] Kerbl, Bernhard, Kopanas, Georgios, Leimkühler, Thomas, Drettakis, George, "3D Gaussian Splatting for Real-Time Radiance Field Rendering," ACM Transactions on Graphics, vol. 42, no. 4, July 2023.
[2] Chen, Yufan, Wang, Lizhen, Li, Qijing, Xiao, Hongjiang, Zhang, Shengping, Yao, Hongxun, Liu, Yebin, "MonoGaussianAvatar: Monocular Gaussian Point-based Head Avatar," arXiv preprint, 2023.
[3] Müller, Thomas, Evans, Alex, Schied, Christoph, Keller, Alexander, "Instant Neural Graphics Primitives with a Multiresolution Hash Encoding," ACM Transactions on Graphics, vol. 41, no. 4, July 2022, pp. 102:1–102:15.

Method



Initialization

- Randomly generate points within a cube, centered and sized based on camera positions. Project points onto the image plane, pruning those outside the mask. Repeat until no points are pruned or a maximum number of iterations is reached.

Color and Opacity Prediction

- Hash encode (x, y, z) to predict opacity (σ) and a 128-dimensional feature vector, encoding spatial and geometric details.
- We explore different prediction strategies, including predicting modifiers for opacity and spherical harmonics (SH) coefficients or directly predicting opacity and RGB color.

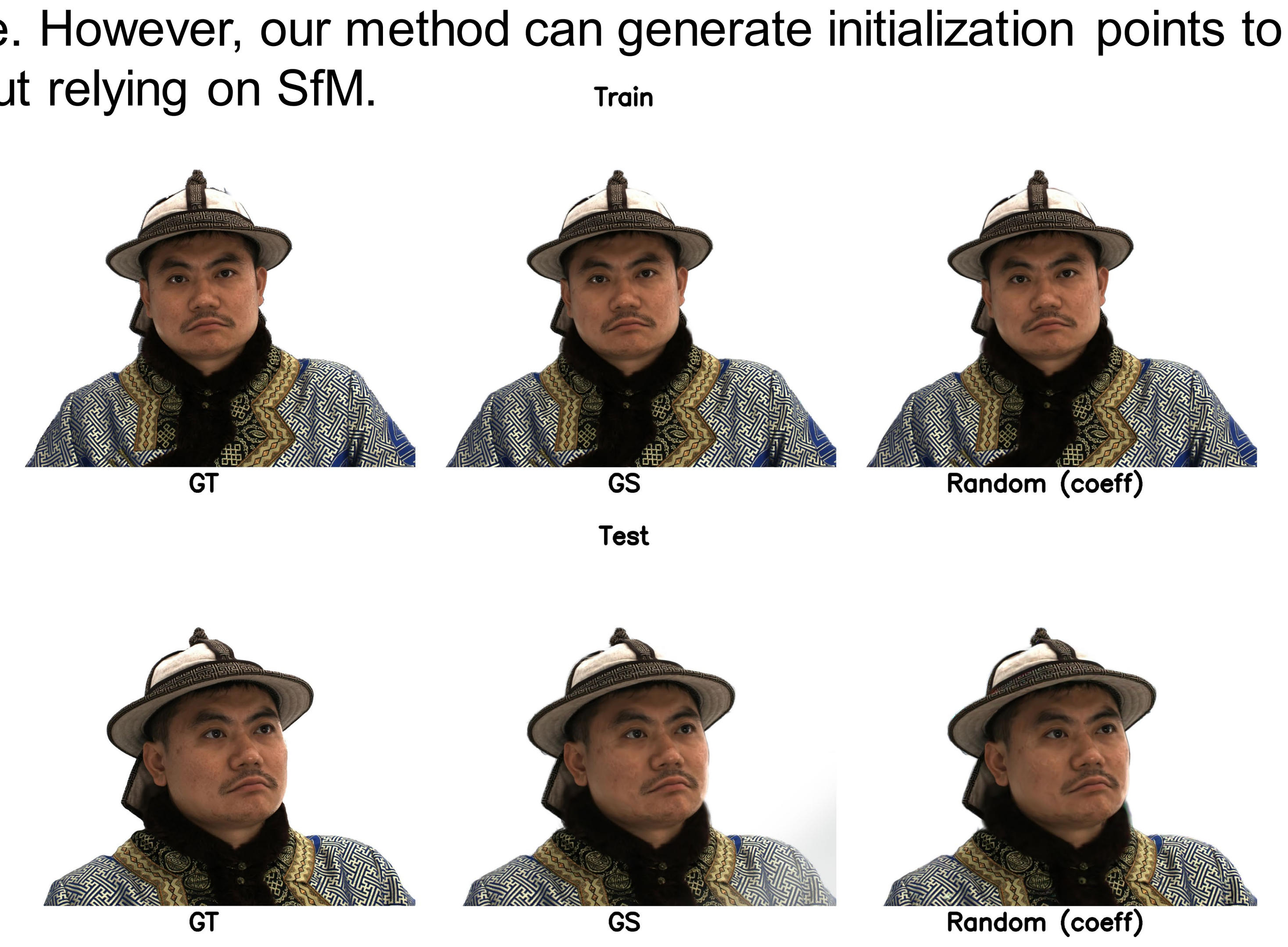
Experimental Results

- All experiments are conducted on Colab T4 GPU with 30,000 iterations, using 60 images from 60 cameras, split into 52 for training and 8 for testing.

	SSIM ↑	PSNR ↑	LPIPS ↓	# Gaussians	FPS
GS	0.918	22.604	0.103	455k	164.73
GS (rgb)	0.917	22.684	0.108	716k	22.9
GS (coeff)	0.916	22.545	0.109	436k	29.12
Random	0.915	22.128	0.114	246k	263.9
Random (rgb)	0.909	21.973	0.117	243k	53.38
Random (coeff)	0.913	21.999	0.115	218k	55.73

The comparisons are based on one subject from the RenderMe-360 dataset. "Random" refers to using random initialization. "RGB" indicates direct prediction of color RGB and opacity. "Coeff" represents predicting a multiplier modifier for SH coefficients and opacity.

- During experiments, the lack of key points extracted by SfM caused initialization failures for some subjects, leading to a complete generation failure. However, our method can generate initialization points to fit the avatar without relying on SfM.



Scan the QR code to view novel view synthesis (NVS).