Sparse view 3D reconstruction with Gaussian splatting

Sri Siddarth Chakravathy
sp313@gatech.edu

Jinchu Li
jinchu.li@gatech.edu

Kunal Aneja
kaneja6@gatech.edu

Sohan Malladi
smalladi32@gatech.edu

Introduction

Novel-view synthesis has garnered significant attention in recent years due to its wide-ranging applications in various fields, including gaming, filmmaking, and augmented reality (AR). One of the leading techniques in this domain is Gaussian Splatting, known for its ability to generate high-quality 3D reconstructions efficiently.

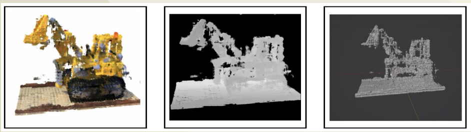


Figure 1: Sparse-view renderings of Nerf-Synthetic LEGO dataset from left to right rendered novel view, depth map of the view and mesh extracted from the renderings

Despite its advantages, Gaussian Splatting faces significant challenges in **sparse-view scenarios** where limited input images are available. In real-world applications, collecting extensive multi-view datasets can be impractical or time-consuming, which poses a substantial barrier to the widespread adoption of these techniques. This limitation is particularly critical in contexts where rapid 3D content generation is required. Addressing the sparse-view issue is essential to enhance the efficiency and accessibility of 3D reconstruction processes, enabling broader applicability across various industries and research fields. To overcome these challenges, we propose a novel three-stage architecture that enhances Gaussian Splatting in sparse-view scenarios:

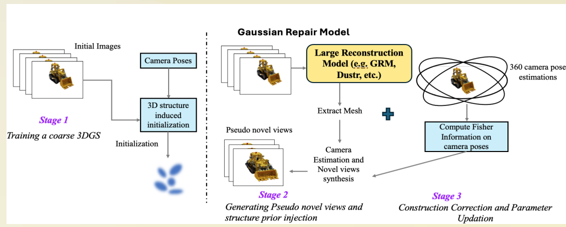
- **Initialization with Visual Hulls**: We train a coarse 3D Gaussian Splatting model using object structure priors derived from visual hulls to capture the basic geometry of the scene from limited views.
- **Gaussian Repair Model**: We employ a large reconstruction model to generate pseudo-views, enhancing structural details and improving overall accuracy.
- **Fisher Information for View Selection**: We utilize Fisher Information to identify and correct errors in novel view poses, enabling targeted parameter updates and further refinement.

Project Goals

Our primary aim is to produce high-quality 3D reconstructions from sparse-view inputs, thereby expanding the applicability of Gaussian Splatting in practical scenarios. Specifically, our goals are:

- **Develop a Sparse-View Reconstruction Framework**: Create a method capable of reconstructing detailed 3D objects from as few as 3–4 input images.
- **Enhance Scene Structure and Multi-View Consistency**: Leverage physical information and object structure priors to improve the geometric accuracy of reconstructions.
- **Optimize View Selection Using Fisher Information**: Select the most informative views to maximize the utility of sparse input data and reduce model uncertainty.
- **Outperform Existing Baselines**: Demonstrate superior performance in key metrics such as PSNR, SSIM, and LPIPS compared to state-of-the-art methods.

Methodology



Overview:

Our framework leverages physical information to enhance scene structure in sparse-view 3D reconstruction. By utilizing object structure priors and improving multi-view consistency, we reconstruct high-quality 3D objects from as few as 10-15 input images. The methodology consists of four main components:

1. Initialization with Visual Hull

- Create a Visual hull that acts a geometric scaffold for creating 3D gaussians using *Segment Anything Model* (SAM)
- Sample 3D points within the visual hill and project them onto image planes to keep the valid points.
- Assign colors to points by getting pixel values from the reference images and converting them into 3D Gaussians.

2. Floater Elimination

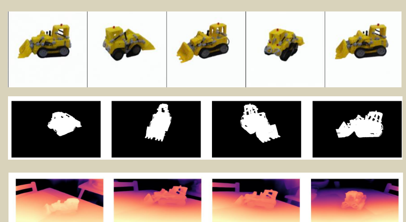
- Find irrelevant Gaussians using KNN to calculate the distance between nearby Gaussians
- Employ adaptive thresholding to purge the Gaussians that reside outside the object's expected geometry
- Incrementally reduce the threshold during optimization in order to refine the scene representation

3. Gaussian Repair Model

- Apply diffusion models like ControlNet to repair images that are corrupted into useful outputs
- Align the model with missing information by adding 3D noise to the Gaussian attributes
- Optimize the repaired views to improve the 3D representation quality

4. Fisher Information for Reconstruction Gain

- Use Fisher information to determine which views are the most informative views in order to reduce model uncertainty.
- Improve the scene reconstruction by selecting the views that maximize the sparse input data.
- Refine certain regions and improve overall performance using this process



Results

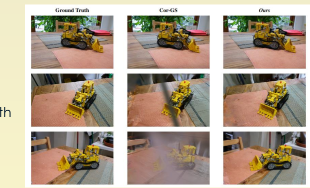
Mip-NeRF 360 (Kitchen Scene)

Model	PSNR (dB)	SSIM	LPIPS
Vanills 3DGS	20.3	0.69	0.09
CoR-GS	23.1	0.73	0.12
Ours	**23.6**	**0.74**	**0.12**

NeRF Synthetic Dataset (Lego Scene)

Model	PSNR (dB)	SSIM	LPIPS
Vanills 3DGS	15.3	0.49	0.06
CoR-GS	21.1	0.53	0.11
Ours	**21.4**	**0.56**	**0.12**

Our method outperforms baselines in sparse-view 3D reconstruction, achieving higher PSNR and SSIM for fidelity and structure, with comparable LPIPS for perceptual quality. Visuals from Mip-NeRF 360 and NeRF Synthetic show sharper details, fewer artifacts, and accurate depth maps. Key innovations like visual hull initialization, floater elimination, and Fisher Information ensure optimal data use and superior reconstructions, surpassing Vanilla 3DGS and CoR-GS across datasets.



Conclusions

We proposed a framework that significantly improves Gaussian Splatting for sparse-view 3D reconstruction. By integrating visual hull initialization with SAM, floater elimination for refining scenes, and a Gaussian Repair Model for filling gaps, we achieved robust geometry and enhanced visual fidelity. Leveraging Fisher Information for active view selection further optimized data usage, reducing uncertainty and improving reconstruction accuracy.

Our approach outperformed baselines like 3DGS and CoR-GS in PSNR, SSIM, and LPIPS across datasets like Mip-NeRF 360 and NeRF Synthetic. With high-quality reconstructions from just 3–4 views, our framework addresses the limitations of existing methods, enabling efficient 3D content generation for applications in gaming, AR, and filmmaking.

Future Work

- Extend to dynamic scenes using motion estimation.
- Incorporate semantic priors for ambiguous geometries.
- Optimize for real-time reconstruction and rendering.
- Improve robustness against noisy inputs for real-world scenarios.

References

[1] Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2020). **NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis.** *European Conference on Computer Vision (ECCV).*

[2] Kerbl, A., Bako, S., Leimkühler, T., Hanika, J., & Nießner, M. (2023). **3D Gaussian Splatting for Real-Time Radiance Field Rendering.** *ACM Transactions on Graphics (SIGGRAPH).*

[3] Xiong, J., Huang, S., Xu, J., & Theobalt, C. (2023). **SparseGS: Real-Time 360° Sparse View Synthesis Using Gaussian Splatting.** *arXiv preprint arXiv:2309.16175.*

[4] Wang, C., Mallya, A., Pal, D., Yang, R., & Liu, S. (2023). **SparseNeRF: Distilling Depth Ranking for Few-shot Novel View Synthesis.** *IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*

Datasets

- Mip-NeRF 360 Dataset
- NeRF Synthetic Dataset