THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

ICLR 2025

Graph-Guided Scene Reconstruction from Images with 3D Gaussian Splatting

Chong Cheng[1]*, Gaochao Song[2]*, Yiyang Yao[3], Qinzheng Zhou[4], Gangjian Zhang[1], Hao Wang[1]

1 HKUST (GZ) 2 HKU 3 SCUT 4 UC Berkeley

Introduction

香港科技大学(广州)
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Motivation:

Currently, reconstructing open scenes and achieving high-quality novel view synthesis from uncalibrated and unordered images using NeRF and 3DGS still faces significant challenges:

- Reliance on precise camera poses, which are difficult to obtain due to GPS/IMU noise and computational challenges in real-world scenarios.
- Dependence on dense viewpoints, which are impractical to capture in many situations, particularly when using drones or smartphones for large-scale scenes.

Core Problem:

• How to efficiently reconstruct large-scale scenes from uncalibrated and sparsely captured images?

Method

香港科技大学(广州)
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Overall framework of GraphGS:

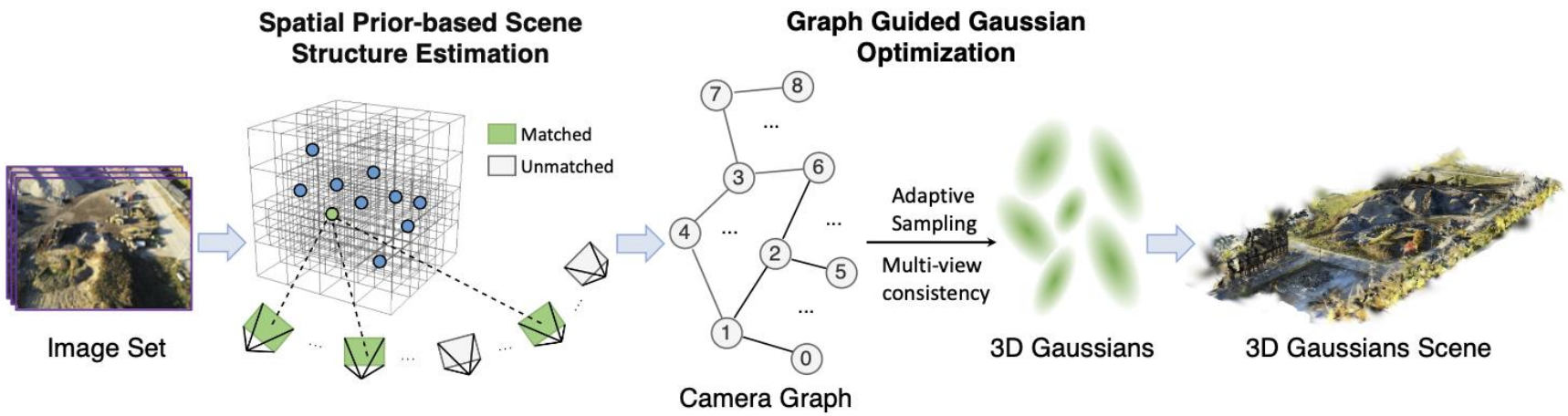


Figure 1: Framework of the GraphGS method for efficient large 3D scene reconstruction. The process begins with spatial prior-based structure estimation, followed by octree-based efficient or- ganization of initialization points. The camera graph is obtained at the end of structure estimation, which contains topology information of scene camera. The information in camera graph will be further us ed for the following gaussian optimization.

Method

香港科技大学(广州)
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Spatial Prior-Based Structure Estimation

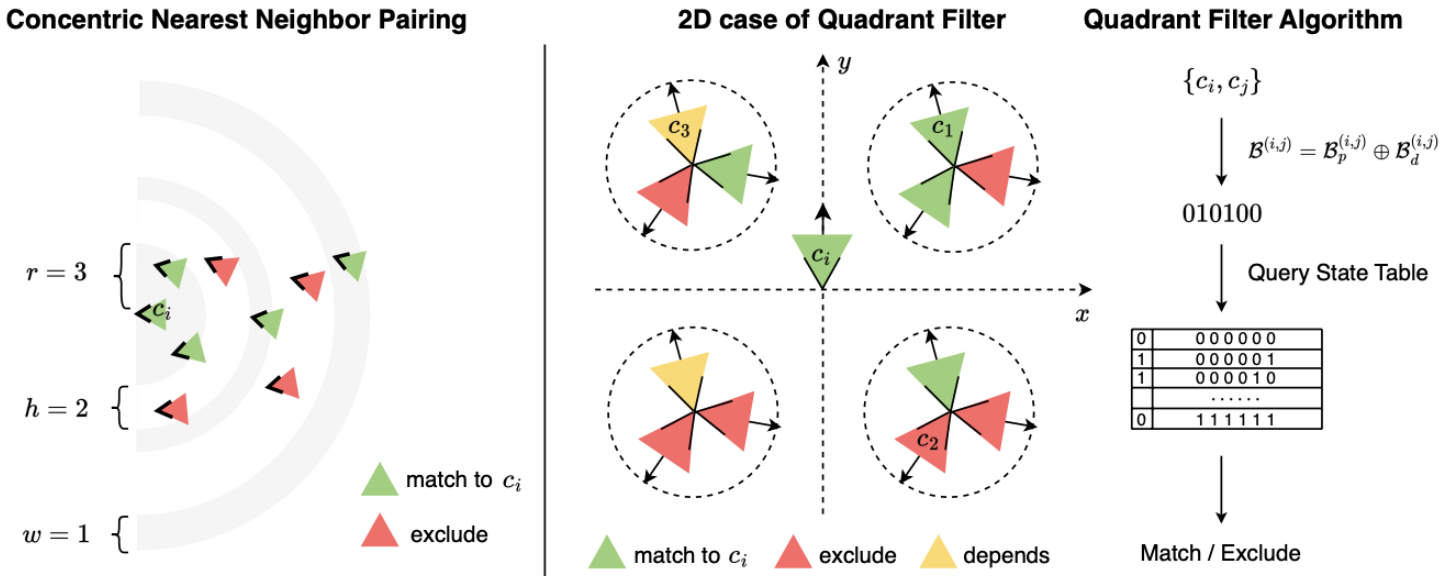


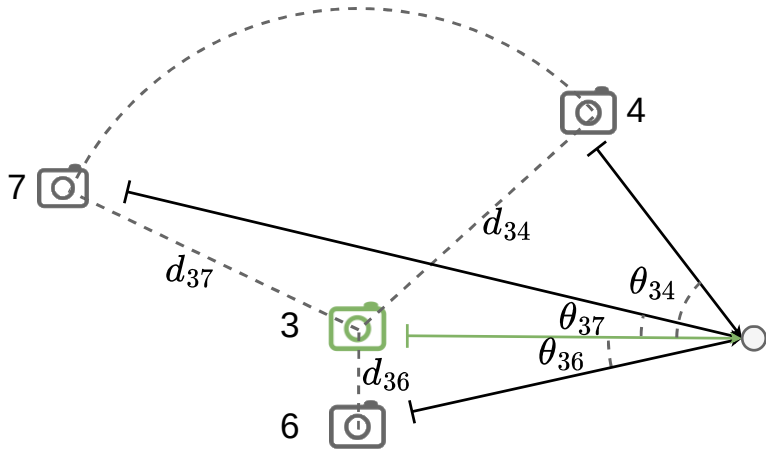
Figure 2: Illustration of spatial prior-based structure estimation, including two parts: Concentric Nearest Neighbor Pairing (left) and Quadrant Filter (right).

- Concentric Nearest Neighbor Pairing
- Quadrant Filter
- Octree Initialization

Method

香港科技大学(广州)
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Camera Graph-Guided Optimization



$$d_{36} < d_{37} = d_{34}$$
$$\theta_{37} = \theta_{36} < \theta_{34}$$
$$\Downarrow$$
$$w_e(3,4) < w_e(3,7) < w_e(3,6)$$

Figure 3: Example of edge weights.

Multi-View Consistency Loss:

$$\mathcal{L}_{\text{cons}} = \lambda \sum_{i,j} \sum_{p} \| I_i(p) - I_j(\mathbf{K}_j(R_{ji}\mathbf{K}_i^{-1}p + T_{ji})) \|$$

Node weight $w_n\,(i)$:

$$w_n(i) = \sum_{i \neq j \neq k} \frac{\sigma_{jk}(i)}{\sigma_{jk}}$$

Edge weight $w_e\,(i,j)$:

$$w_e(i,j) = \frac{e^{-k\left\|p^{(i)} - p^{(j)}\right\|_2}}{1 - e^{-d^{(i)} d^{T(j)}}}$$

Results

香港科技大学(广州)
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Waymo & KITTI

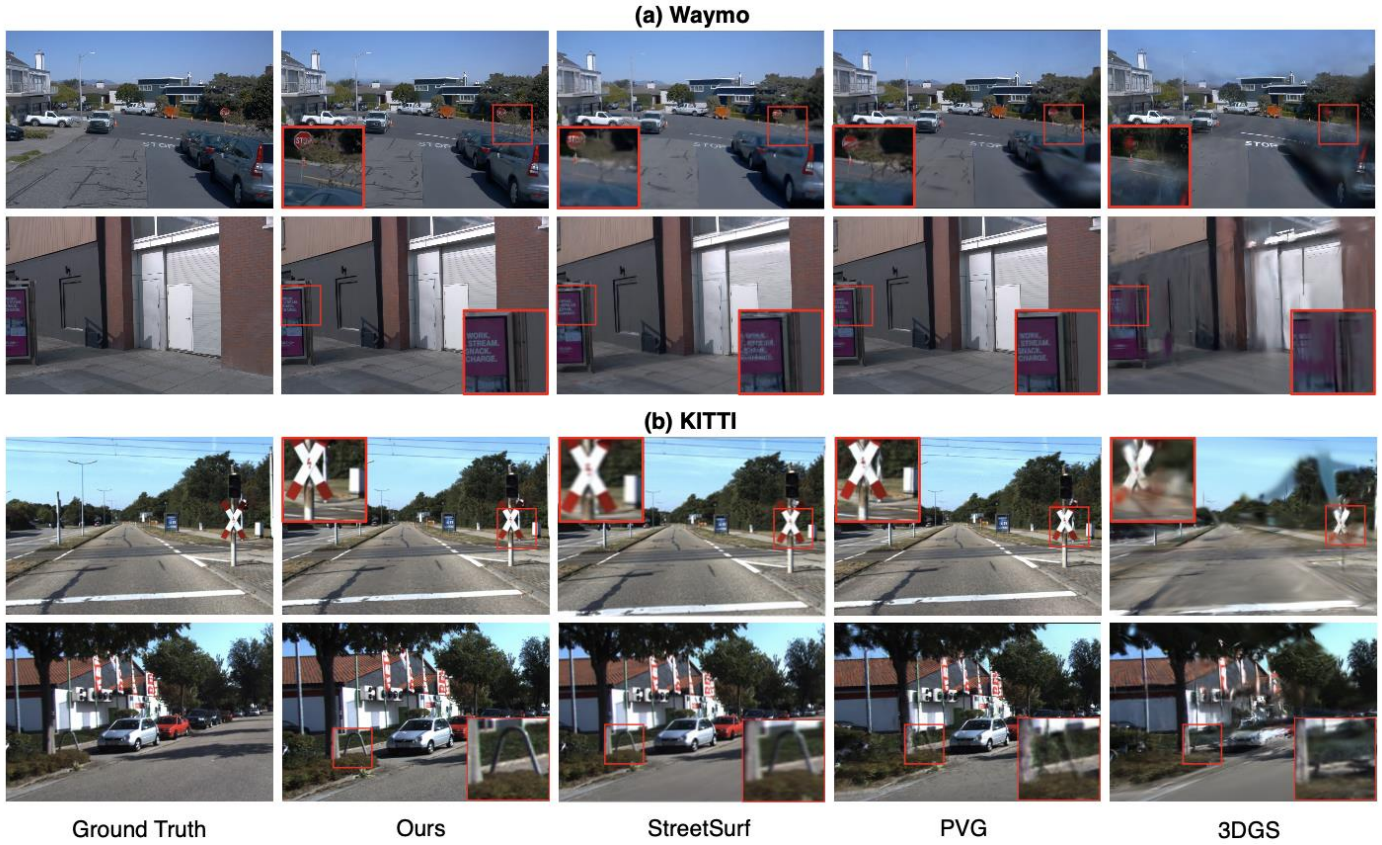


Figure 4: Qualitative comparison of novel view synthesis results on thea Waymo and KITTI.

Method	Waymo				KITTI			
	FPS ↑	PSNR ↑	SSIM ↑	LPIPS ↓	FPS ↑	PSNR ↑	SSIM ↑	LPIPS ↓
Mip-NeRF 360 (Barron et al., 2022)	0.042	22.42	0.698	0.471	0.053	20.68	0.650	0.480
S-NeRF (Xie et al., 2023)	0.001	19.22	0.515	0.400	0.008	18.71	0.606	0.352
StreetSurf (Guo et al., 2023)	0.097	23.78	0.822	0.401	0.037	22.48	0.763	0.304
Zip-NeRF (Barron et al., 2023)	0.500	26.21	0.815	0.389	0.610	21.41	0.665	0.470
UC-NeRF (Cheng et al., 2023)	0.032	26.72	0.800	0.375	0.051	24.05	0.721	0.400
BARF (Lin et al., 2021)	0.041	9.07	0.235	1.021	0.071	10.68	0.250	0.990
SPARF (et al., 2023b)	-	-	-	-	-	-	-	-
UP-NeRF (et al., 2023a)	0.120	26.16	0.876	0.375	-	-	-	-
EmerNeRF (Yang et al., 2023a)	0.043	25.92	0.763	0.384	0.28	25.24	0.801	0.237
3DGS (Kerbl et al., 2023)	**63**	25.08	0.822	0.319	**125**	19.54	0.776	0.224
PVG (Chen et al., 2024)	50	28.11	0.849	0.279	59	26.63	0.885	**0.127**
Ours	52	**29.43**	**0.899**	**0.217**	79	**26.98**	**0.887**	0.157

Table 1: Quantitative comparison of novel view synthesis results on the Waymo and KITTI.

Results

香港科技大学(广州)
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Mill 19

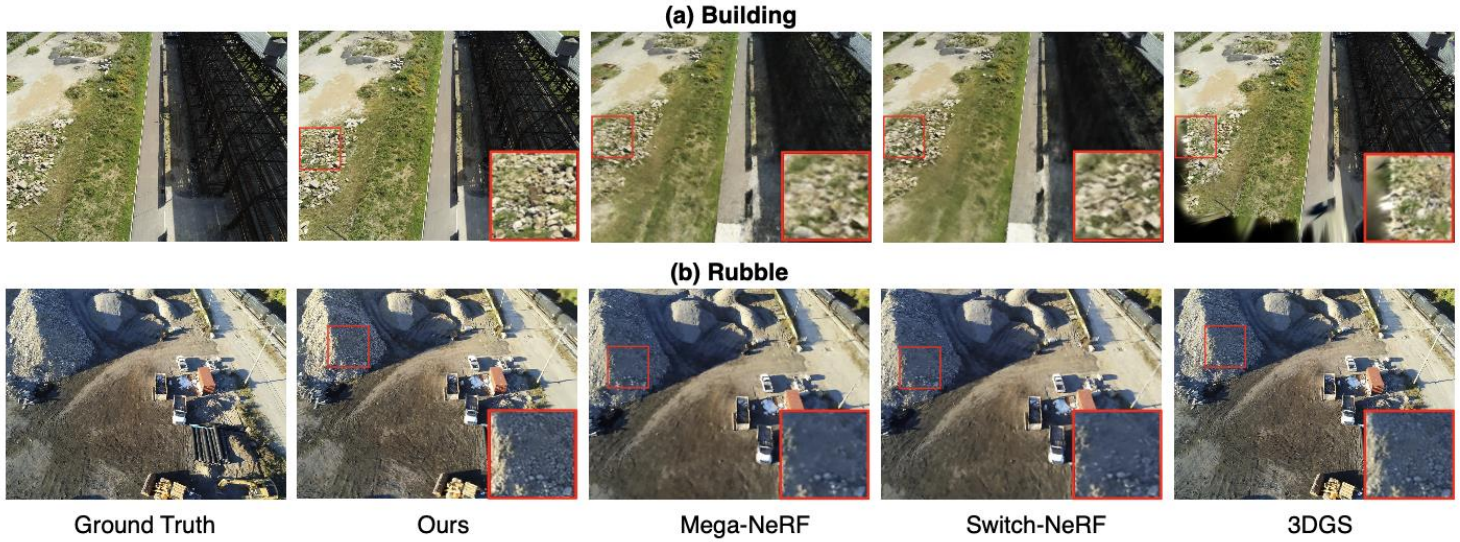


Figure 5: Qualitative comparison of novel view synthesis results in the Mill 19 large scene dataset.

Method	Building			Rubble		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
Mega-NeRF (Turki et al., 2022)	20.93	0.547	0.349	24.05	0.553	0.373
Switch-NeRF (Mi & Xu, 2023)	21.54	0.579	0.294	24.31	0.562	0.329
3DGS (Kerbl et al., 2023)	23.01	0.769	0.164	26.78	0.800	0.161
VastGaussian (Lin et al., 2024)	23.50	0.804	**0.130**	26.92	0.823	**0.132**
Ours	**26.60**	**0.854**	0.163	**27.03**	**0.869**	0.185

Table 2: Quantitative comparison of novel view synthesis results in the Mill 19.

Results

香港科技大学(广州)
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Real-world data

GT

Ours

GT

Ours

Conclusion

香港科技大学(广州)
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

GraphGS enables high-quality, fast reconstruction of large scenes from images without ground truth poses, benefiting applications like VR, gaming, and the metaverse.

Contributions:

- Introduced spatial prior-based structure estimation that bootstraps scene reconstruction from uncalibrated images without requiring precise input poses.
- Developed camera graph-guided 3D Gaussian Splatting that enforces multi-view consistency, improving quality and achieving faster convergence.
- Achieved SOTA performance on multiple benchmarks without using ground-truth poses, effectively handling challenging real-world scenes.

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Thank You