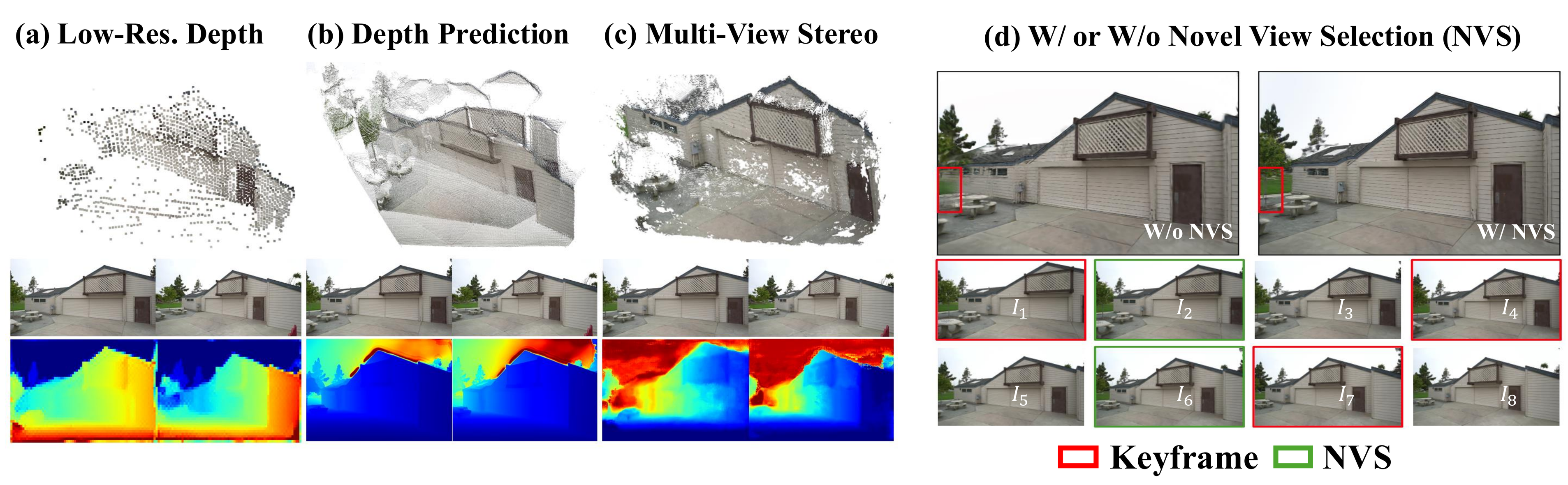
Online 3D Gaussian Splatting Modeling with Novel View Selection

Byeonggwon Lee, Junkyu Park, Soohwan Song Dongguk University, Republic of Korea

Khang Troung Giang 42dot, Republic of Korea

Summary

- **Goal**
  - Online 3D Gaussian Splatting (3DGS) modeling from RGB frames
- **Previous Approach: Monocular Dense SLAM**
  - Point clouds generated from low-resolution depth maps hinder dense reconstruction
  - Monocular depth prediction leads to depth inconsistencies
  - Reliance solely on keyframes is insufficient to capture the entire scene



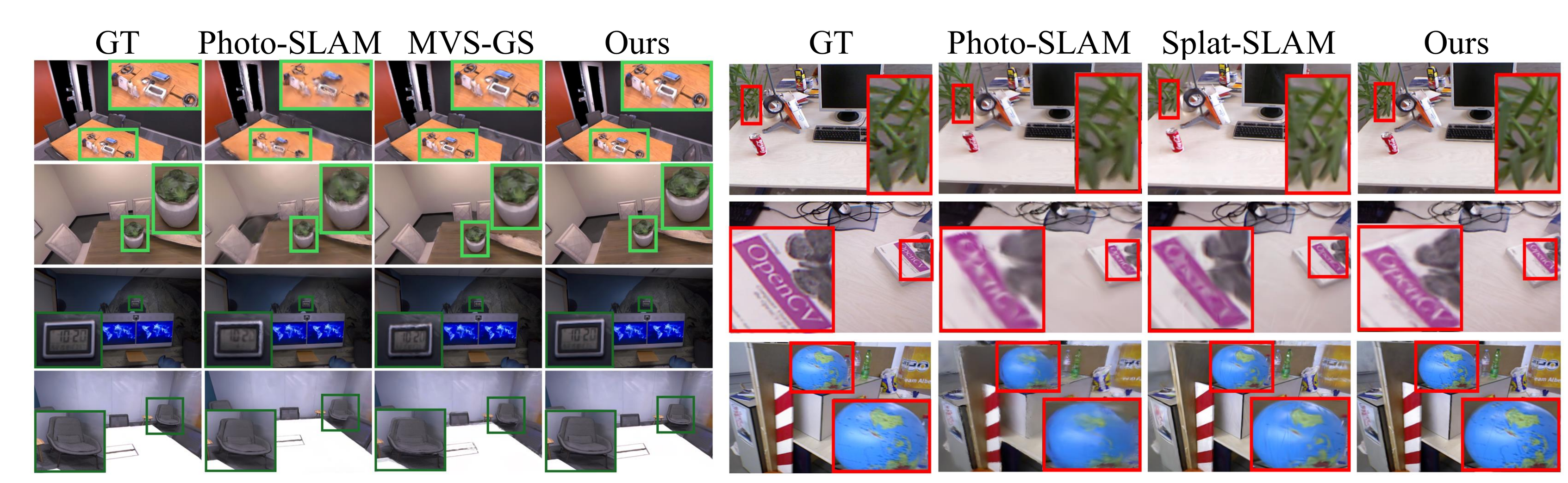
- **Key Idea**
  - Employ online MVS to estimate depth maps and initialize Gaussians
  - Select non-keyframes by estimating Gaussian uncertainty
  - Use non-keyframes to further train 3DGS and fill incomplete regions

- **NVS based on Gaussian uncertainty**
  - Select non-keyframes observing the most uncertain Gaussian
  - Gaussian uncertainty :

$$U_{n,i} = \alpha_1 \lambda_{n,i} + \alpha_2 A_{n,i}$$

  - $\lambda_{n,i}$: Largest eigenvalue of Cov
  - $A_{n,i}$: Gaussian gradients

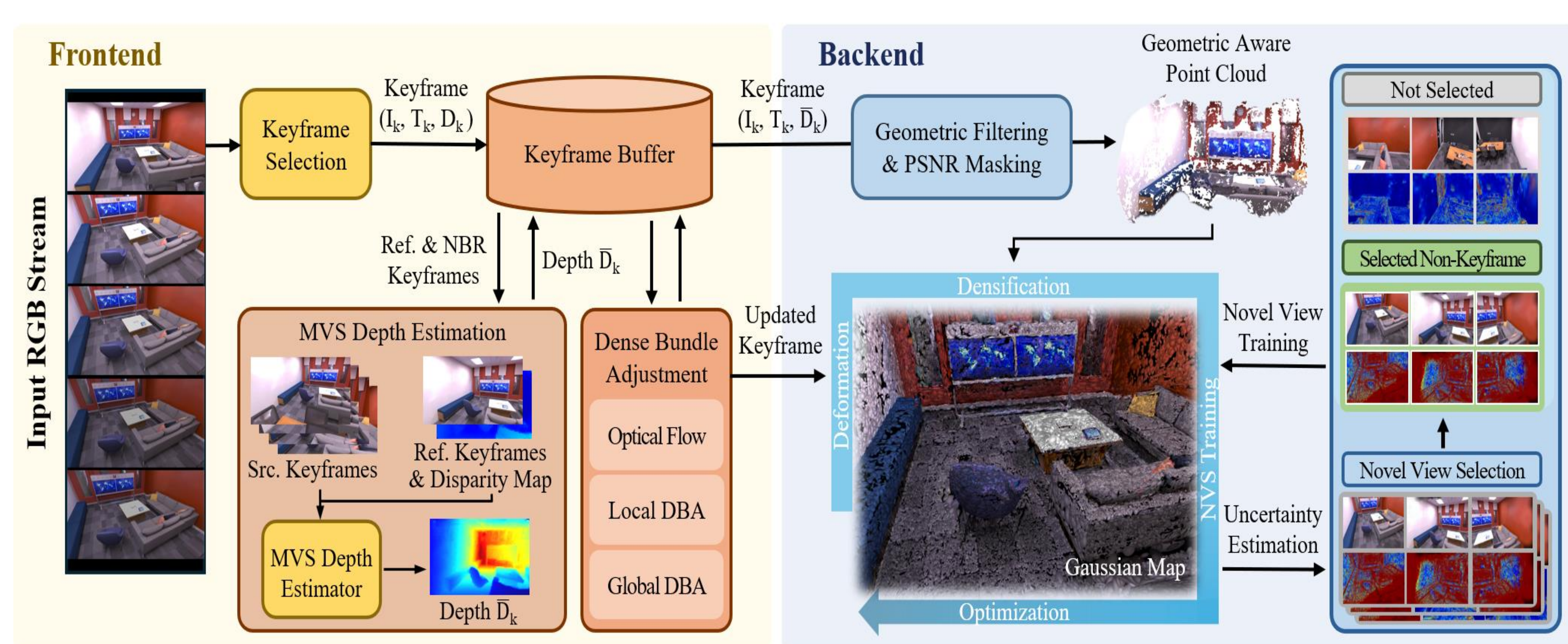
- **Results**



Method

- **Frontend**
  - Build keyframe pose graph and sparse depths via DROID-SLAM
  - Run GBA every 30 keyframes to maintain global frame consistency
  - Estimate high-resolution depth maps from keyframes within a local time window using the SOTA MVS method, MVSFormer
- **Backend**
  - Generate new Gaussians from the MVS depth maps for each keyframe and integrate them into the 3DGS model
  - Independently of the frontend, continuously train the 3DGS model until the next keyframe is received
  - Non-keyframes selected through NVS are also used for training

- **System Framework with Parallel Frontend–Backend Operation**

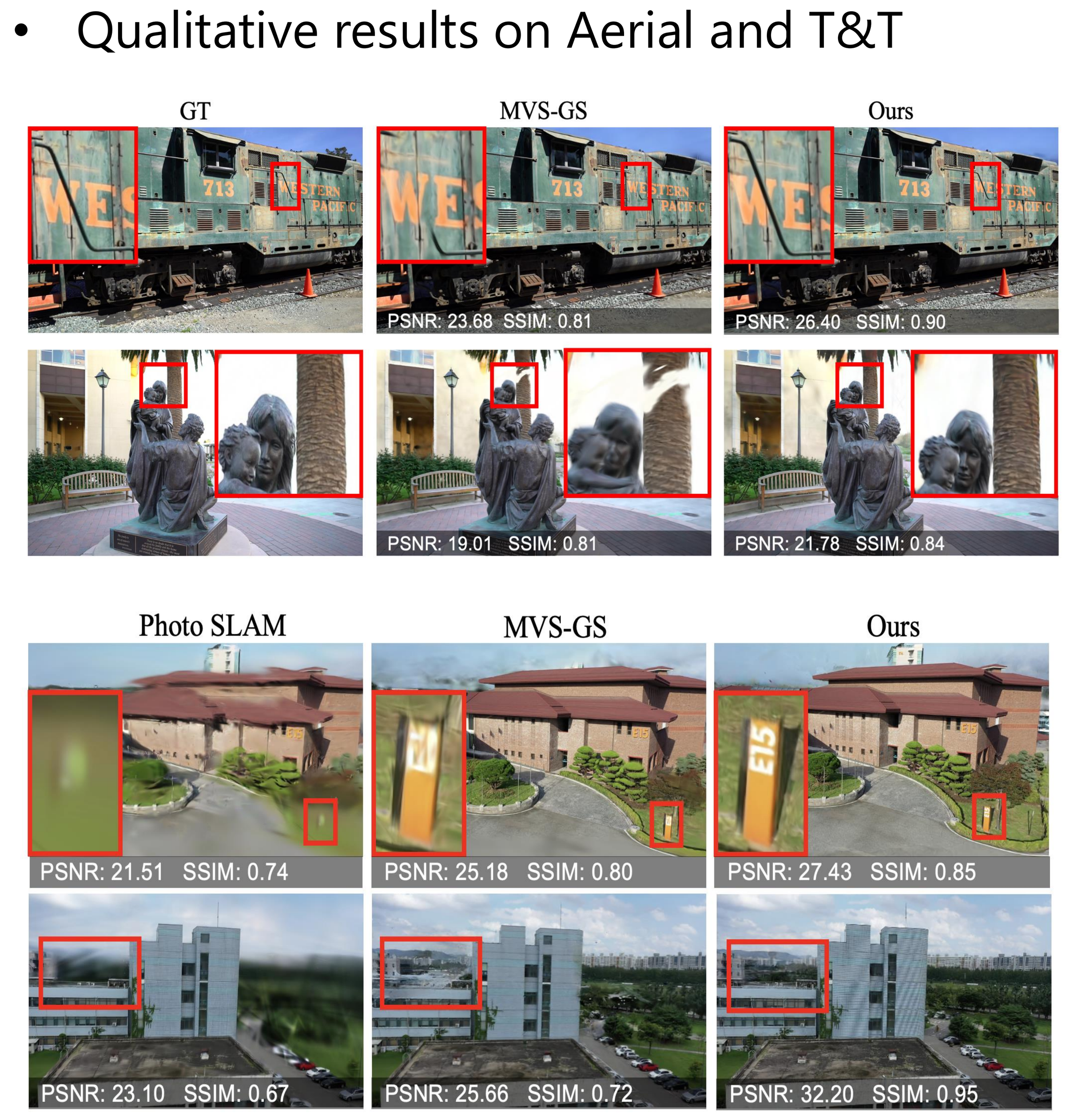


Results

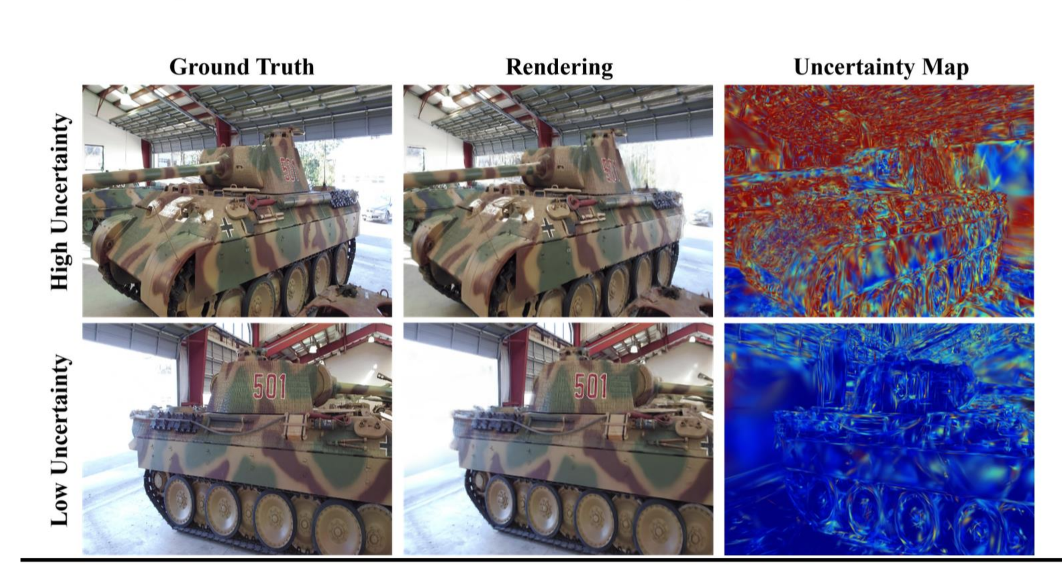
- **Results for Replica dataset**
  - Quantitative results on Replica evaluation set

Method	Off0	Off1	Off2	Off3	Off4	Rm0	Rm1	Rm2	Avg.
PSNR ↑									
Q-SLAM	36.31	37.22	30.68	30.21	31.96	29.58	32.74	31.25	32.49
MonoGS	32.00	31.21	23.26	25.77	23.85	23.53	25.00	22.42	25.88
Splat-SLAM	40.81	40.64	35.19	35.03	37.40	32.25	34.31	35.95	36.45
Photo-SLAM	36.99	37.52	31.79	31.62	34.17	29.77	31.30	33.18	33.29
MGS-SLAM	35.51	34.25	30.83	31.86	34.38	29.91	31.06	31.49	32.41
MVS-GS	41.02	42.04	34.00	34.65	33.33	32.20	31.54	35.84	35.58
Ours	**43.93**	**43.98**	**37.98**	**36.31**	**39.59**	**34.88**	**37.99**	**39.60**	**39.28**
SSIM ↑									
Q-SLAM	0.94	0.94	0.90	0.88	0.89	0.83	0.91	0.87	0.89
MonoGS	0.90	0.88	0.82	0.84	0.86	0.75	0.79	0.81	0.83
Splat-SLAM	0.97	**0.99**	0.97	**0.97**	0.97	**0.96**	**0.97**	0.96	0.97
Photo-SLAM	0.96	0.95	0.93	0.92	0.94	0.87	0.91	0.93	0.93
MGS-SLAM	0.94	0.93	0.90	0.92	0.95	0.89	0.90	0.91	0.92
MVS-GS	0.98	0.98	0.95	0.96	0.95	0.95	0.92	0.96	0.96
Ours	**0.99**	**0.99**	**0.98**	**0.97**	**0.98**	**0.96**	**0.97**	**0.98**	**0.98**
LPIPS ↓									
Q-SLAM	0.13	0.15	0.20	0.19	0.18	0.18	0.16	0.15	0.17
MonoGS	0.23	0.22	0.30	0.24	0.34	0.33	0.35	0.39	0.30
Splat-SLAM	0.05	0.07	0.06	0.04	0.10	0.09	0.06	0.05	0.06
Photo-SLAM	0.06	0.06	0.09	0.09	0.07	0.10	0.08	0.07	0.08
MGS-SLAM	0.07	0.11	0.12	0.07	0.08	0.08	0.09	0.09	0.09
MVS-GS	0.05	0.05	0.09	0.07	0.10	0.10	0.13	0.07	0.08
Ours	**0.02**	**0.02**	**0.03**	**0.03**	**0.03**	**0.04**	**0.04**	**0.03**	**0.03**

- Qualitative results on Aerial and T&T



- High vs. Low Uncertainty Visualization



	PSNR ↑	SSIM ↑	LPIPS ↓	Uncertainty Value ↓
High Uncertainty	19.609	0.632	0.193	1236.9
Low Uncertainty	28.787	0.899	0.132	446.22

[1] 3D Gaussian splatting for real-time radiance field rendering, ACM Transactions on Graphics, 2023.

[2] MVSFormer: Multi-View Stereo by Learning Robust Image Features and Temperature-based Depth, TMLR, 2023.

[3] Photo-SLAM: Real-time Simultaneous Localization and Photorealistic Mapping …, CVPR, 2024

[4] MVS-GS: High-Quality 3D Gaussian Splatting Mapping via Online Multi-View Stereo, IEEE Access, 2025