SUPPLEMENTARY MATERIAL

1. SUPPLEMENTARY.

1.1. Implementation Details

Our implementation is based on PyTorch and follows the optimization settings of CF-3DGS[1], unless stated otherwise. All experiments are conducted on an RTX A5000 GPU.

Camera Pose Estimation. We estimate monocular depth using Metric3DV2 [2], then lift it with camera intrinsics. To manage high-resolution inputs, we downsample the point cloud before fitting it to 3DGS. Before applying G-ICP for point cloud registration, we restrict the point cloud within a bounding box of $[-40, -40, 0]$ to $[40, 40, 60]$. Pose estimation is optimized 300 iterations, where the parameters of 3DGS remains fixed, and only the pose parameters are updated.

Scene Growing The global reconstruction begins with the first frame's depth estimation. Camera poses are estimated sequentially through the refined pose estimation process, while the scene expands progressively with new frames. On the KITTI-360 dataset, we set the voxel size to 1 and the voxel density threshold to 1.5, whereas on the Tanks and Temples dataset, we use a voxel size of 0.3 and a density threshold of 2.

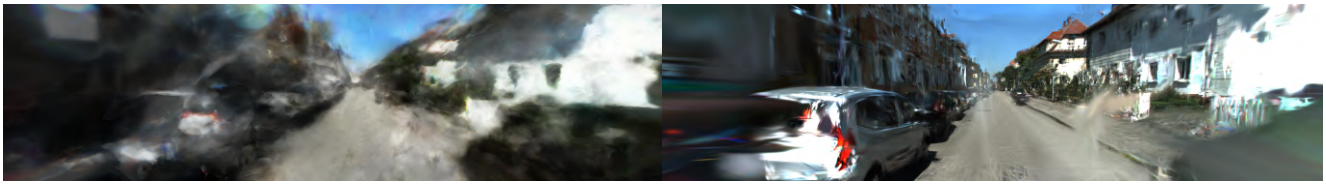
1.2. Additional Qualitative Results

We report qualitative comparison results for all five sequences in Kitti-360 here from Fig. 1 to Fig. 5.

2. REFERENCES

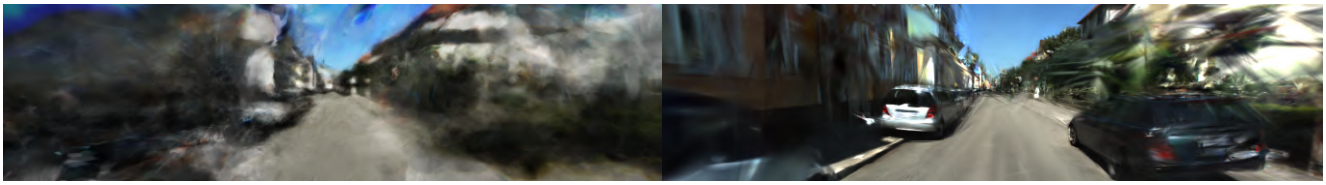
[1] Y. Fu, X. Wang, S. Liu, A. Kulkarni, J. Kautz, and A. A. Efros, "Colmap-free 3d gaussian splatting," *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20796–20805, 2024.

[2] M. Hu, W. Yin, C. Zhang, Z. Cai, X. Long, H. Chen, K. Wang, G. Yu, C. Shen, and S. Shen, "Metric3d v2: A versatile monocular geometric foundation model for zero-shot metric depth and surface normal estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 12, pp. 10579–10596, December 2024.

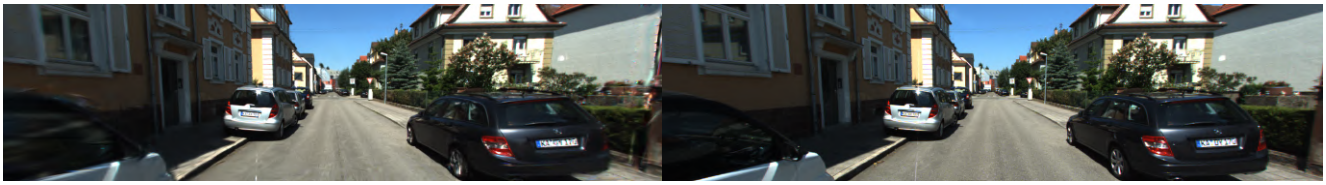


(a) Nope-NeRF (b) CF-3DGS

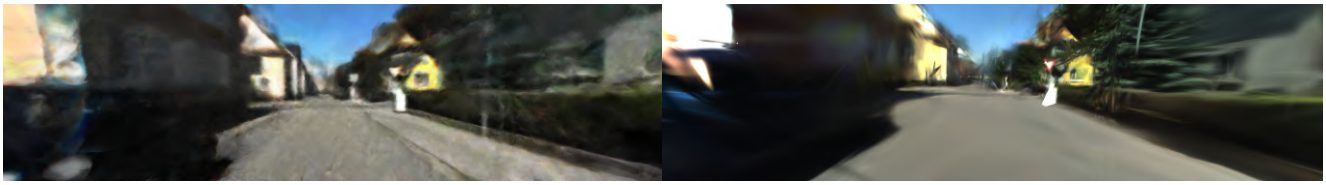
(c) Ours (d) Ground Truth



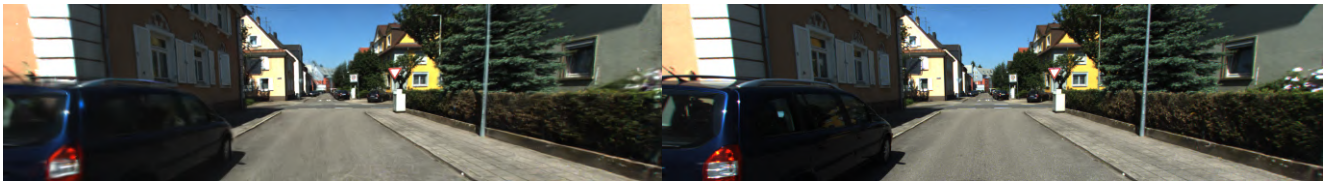
(a) Nope-NeRF (b) CF-3DGS



(c) Ours (d) Ground Truth



(a) Nope-NeRF (b) CF-3DGS



(c) Ours (d) Ground Truth

Fig. 1: Qualitative comparison for novel view synthesis on Seq_1.

(a) Nope-NeRF (b) CF-3DGS

(c) Ours (d) Ground Truth

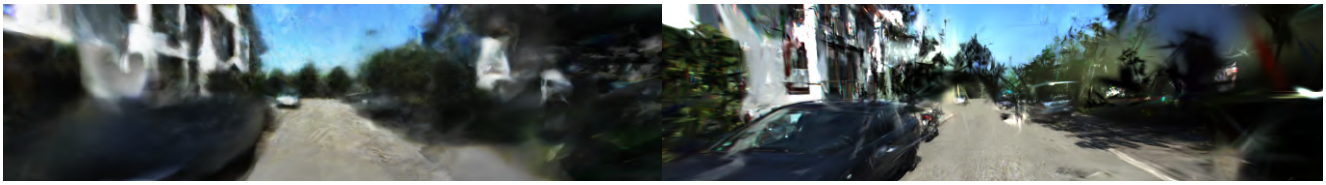
(a) Nope-NeRF (b) CF-3DGS

(c) Ours (d) Ground Truth

(a) Nope-NeRF (b) CF-3DGS

(c) Ours (d) Ground Truth

Fig. 2: Qualitative comparison for novel view synthesis on Seq_1.



(a) Nope-NeRF (b) CF-3DGS

(c) Ours (d) Ground Truth

(a) Nope-NeRF (b) CF-3DGS

(c) Ours (d) Ground Truth

(a) Nope-NeRF (b) CF-3DGS

(c) Ours (d) Ground Truth

Fig. 3: Qualitative comparison for novel view synthesis on Seq_3.

(a) Nope-NeRF (b) CF-3DGS

(c) Ours (d) Ground Truth

(a) Nope-NeRF (b) CF-3DGS

(c) Ours (d) Ground Truth

(a) Nope-NeRF (b) CF-3DGS

(c) Ours (d) Ground Truth

Fig. 4: Qualitative comparison for novel view synthesis on Seq_4.

(a) Nope-NeRF (b) CF-3DGS

(c) Ours (d) Ground Truth

(a) Nope-NeRF (b) CF-3DGS

(c) Ours (d) Ground Truth

(a) Nope-NeRF (b) CF-3DGS

(c) Ours (d) Ground Truth

Fig. 5: Qualitative comparison for novel view synthesis on Seq_5.