Supplementary Materials for BiNAR: A Bi-Modal Framework for Non-Aligned RGB-IR 3D Reconstruction via Gaussian Splatting

Zhongwen Wang[1], Han Ling [1]*, Weihao Zhang [1], Yinghui Sun [2], Quansen Sun [1]†

[1]Nanjing University of Science and Technology
[2]Southeast University

{jankinwang, 321106010190, zhangweihao, sunquansen}@njust.edu.cn sunyh@seu.edu.cn

1. Summary

In this supplementary material, we provide more information. Specifically, we supplement the principle of RGB scene pre-reconstruction. We added the details of bi-modal rendering and optimization in Section 3.2.2. We show more experimental results, including the results of reprojection error in Section 5.2, full experimental results on the RGBT-Scenes dataset in Section 5.3, pixel-level aligned rendering results for all scenes in Section 5.4 and ablation experimental results for all scenes in Section 5.5. We also analyzed the limitations of BiNAR and looked ahead to future work.

2. Pre-reconstruction of RGB scenes

Thanks to the rich texture information and high resolution of RGB scenes, after completing the joint calibration of RGB and infrared cameras, we first use multiple images taken by the RGB camera to pre-reconstruct the scene to obtain a more reliable sparse point cloud and RGB camera pose. The Structure-from-Motion (SFM) [3] method has played an important role in many 3D scene reconstruction works [1, 2]. Specifically, SFM performs global or incremental pose optimization on all images and estimates the 3D coordinates of scene points to form a sparse point cloud. The process of optimizing the coordinates of the 3D points outside the camera can be written as the following Bundle Adjustment (BA) problem:

$$\min_{\{R_k,t_k\},\{X_p\}} \sum_{k=1}^{N} \sum_{p\in\Omega_k} ||\pi\left(K\ , R_k, t_k, X_p\right) - x_{k,p}||^2 \quad (1)$$

Where R_k, t_k represents the external parameters of the k-th camera image, X_p represents the spatial coordinates of the p-th 3D feature point, $x_{k,p}$ is the 2D pixel coordinate observed on the k-th image, and Ω_k represents the index of all feature points that are visibly associated with the k-th image. $\pi(\cdot)$ represents the function that projects the 3D points to the image coordinate system after the camera internal parameters. Function $\pi(\cdot)$ involves two main transformations: first, the 3D point $X_p \in \mathbb{R}^3$ is transformed from the world coordinate system to the camera coordinate system using the extrinsic parameters $R_k \in \mathbb{R}^{3\times3}$ and $t_k \in \mathbb{R}^3$:

$$X_c = R_k X_p + t_k \quad (2)$$

where $X_c = [X_c, Y_c, Z_c]^T$ denotes the coordinates of the point in the camera coordinate system. Next, the transformed point is projected onto the 2D image plane using the intrinsic matrix $K \in \mathbb{R}^{3\times3}$ of the camera:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{Z_c} K X_c, \quad \text{where} \quad K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (3)$$

where $[f_x, f_y]$ denote the focal lengths in pixel units, and $[c_x, c_y]$ are the coordinates of the principal point. Consequently, the explicit form of the projection function π can be written as:

$$\pi(K, R_k, t_k, X_p) = \begin{bmatrix} f_x \cdot \dfrac{X_c}{Z_c} + c_x \\ f_y \cdot \dfrac{Y_c}{Z_c} + c_y \end{bmatrix} \quad (4)$$

This projection accurately models the image formation process and serves as the foundation for reprojection error computation in the Bundle Adjustment (BA) optimization.

3. Bi-Modal Rendering and Optimization

Rendering from Gaussians requires accurate camera poses. However, due to the low resolution and lack of texture in IR images, directly estimating IR camera poses using traditional SFM methods is often unreliable or even infeasible. We use the rigid transformation matrix obtained from the aforementioned joint calibration to transform the RGB

*Corresponding authors.
†Corresponding authors.

camera pose into the IR camera pose, ensuring geometric consistency between the two modalities. Specifically, the pose of the RGB camera is represented by:

$$P_{RGB} = [R_{RGB}|t_{RGB}] \tag{5}$$

Then the IR camera pose is computed via:

$$P_{IR} = RT_{colmap} \cdot P_{RGB} \tag{6}$$

Given accurate camera poses, we render the scene using the shared geometry attributes (p_i, r_i, s_i) and modal-specific appearance attributes a_i^m from the unified Gaussian Field. Specifically, for a modality m, the rendered color at pixel u is defined as:

$$C_m(u) = \sum_{i=1}^{M} T_i(u)\alpha_i^m(u)c_i^m(u) \tag{7}$$

Where M is the total number of Gaussians; $T_i(u)$ denotes the transmittance accumulated from previous Gaussians; $\alpha_i^m(u)$ denotes the opacity and $c_i^m(u)$ denotes the color represented by spherical harmonics under modality m.

To improve rendering quality and stability, we adopt the rendering strategy of Mip-Splatting [4], which mitigates high-frequency aliasing and detail loss by adapting the scale and sampling density of Gaussians based on their distance to the camera and projected area on the image plane. This is especially critical for RGB-IR modalities with large resolution differences, as it allows for preserving detail in rendering without introducing aliasing. The updated rendering formula becomes:

$$C_m(u) = \sum_{i=1}^{M} T_i(u)\alpha_i^m\left(u, \sigma_i(u)\right) c_i^m\left(u, \sigma_i(u)\right) \tag{8}$$

Where $\sigma_i(u)$ represents the scale parameter.

Each Gaussian in the Unified Gaussian Field contains shared geometry and modal-specific appearance attributes. All parameters are jointly optimized by computing rendering losses from both RGB and IR images and backpropagating the gradients. The rendering loss of RGB and IR is defined by $L1$ and $DSSIM$ loss:

$$\begin{aligned} L_m &= (1-\lambda_{dssim})\, L1\left(\hat{I}_m, I_m\right) \\ &+\lambda_{dssim}\left[1 - SSIM\left(\hat{I}_m, I_m\right)\right] \end{aligned} \tag{9}$$

where $\hat{I}_m$ denotes the rendered image and I_m is the ground truth. A balancing factor $\lambda_{dssim} = 0.2$ is used in all experiments. The total loss becomes:

$$L_{Total} = \alpha L_{RGB} + L_{IR} \tag{10}$$

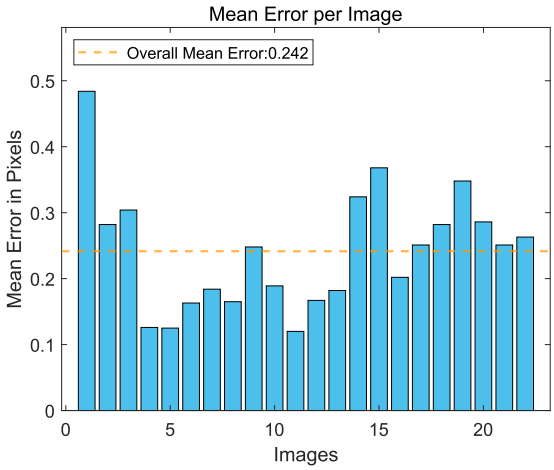

Figure 1. Cross-modal reprojection error statistics. The reprojection error of all images is 0.242 pixels, which has reached sub-pixel accuracy and provides reliable geometric consistency for subsequent bi-modal 3D scene reconstruction.

In practice, we find that setting $\alpha = 1$ effectively balances the contribution of both modalities. Alternating optimization of the two modalities ensures that the geometry attributes are consistently updated across modalities, while appearance attributes are optimized independently, enabling spectral-specific visual representations under unified geometry.

4. More Experimental Results

4.1. Calibration Reprojection Verification

We collected 22 pairs of RGB-IR images, and Fig. 1 shows the average reprojection error of 22 pairs of images. The blue bar represents the error of each image, and the orange dashed line gives the overall average error of 0.242px. Except for individual samples, the error of all images is less than 0.35px. This result shows that the proposed joint calibration achieves sub-pixel accuracy, laying a solid foundation for the subsequent bi-modal 3D scene reconstruction.

4.2. Rendering quality on the RGBT-Scenes dataset

As shown in the Fig. 2, we rendered the RGB and IR channels in ten real scenes with a unified perspective, and spliced the two modal images diagonally. This visualization method helps to observe the correspondence between the two modalities in the spatial structure at the same time. It can be seen that whether it is geometric edges (such as car body contours), or high-frequency details (such as keyboards, bottle caps, etc.), the two modalities show highly consistent rendering results. This further demonstrates the cross-modal expression ability of BiNAR in the modeling process and its good ability to maintain fine-grained geo-

Figure 2. Pixel-level aligned rendering of the whole scenes. The edges of the cross-modal splicing in ten typical scenes remain highly coherent and free of ghosting, which intuitively verifies the sub-pixel geometric consistency achieved by BiNAR in both RGB and IR modes.

Table 1. Comparison of rendering quality on RGBT-Scenes dataset. "TG" stands for ThermalGaussian. The best results are marked in bold, and the symbol "×" indicates that the method cannot reconstruct this scene.

Metric	Method	IR Scenes										
		Dimsum	DailyStuff	Ebike	Road Block	Truck	Rotary Kiln	Building	Iron Ingot	Parterre	Land Scape	avg.
PSNR↑	mip-splatting	25.73	19.23	21.39	26.05	25.35	27.12	24.28	29.35	23.28	20.00	24.18
	Thermal3D-GS	26.38	17.95	×	25.49	26.41	×	25.54	29.74	24.42	x	25.13
	TG(MFTG)	26.14	18.94	22.30	26.57	26.39	26.70	26.09	29.84	23.96	20.11	24.70
	TG(MSMG)	**26.44**	20.79	**23.74**	26.11	25.79	26.77	26.92	30.00	22.68	21.05	25.03
	TG(OMMG)	26.34	21.97	23.52	**26.99**	25.48	26.29	26.60	29.85	26.40	22.25	25.57
	ours	26.40	**22.13**	23.57	26.92	**26.13**	**27.93**	**27.34**	**30.42**	**27.01**	**24.25**	**26.21**
SSIM↑	mip-splatting	0.881	0.790	0.811	0.911	0.848	0.925	0.850	0.890	0.854	0.799	0.856
	Thermal3D-GS	0.892	0.784	×	0.904	0.877	×	0.878	0.896	0.878	x	0.873
	TG(MFTG)	0.886	0.797	0.841	0.909	**0.879**	0.925	0.887	0.897	0.863	0.802	0.869
	TG(MSMG)	0.892	0.827	**0.877**	0.918	0.873	0.928	0.900	0.903	0.864	0.838	0.882
	TG(OMMG)	0.889	0.835	0.873	**0.920**	0.864	0.922	0.890	0.897	0.897	0.852	0.884
	ours	**0.893**	**0.845**	0.870	**0.920**	**0.879**	**0.935**	**0.905**	**0.908**	**0.908**	**0.880**	**0.894**
LPIPS↓	mip-splatting	0.127	0.267	0.306	0.212	0.157	0.125	0.247	0.095	0.233	0.329	0.210
	Thermal3D-GS	0.123	0.285	×	0.211	0.133	×	0.215	0.089	0.205	x	0.180
	TG(MFTG)	0.128	0.263	0.239	0.210	0.136	0.127	0.210	0.089	0.222	0.329	0.195
	TG(MSMG)	0.128	0.206	0.198	0.221	0.145	0.148	0.200	0.086	0.231	0.307	0.187
	TG(OMMG)	0.125	0.205	**0.196**	0.200	0.218	0.126	0.184	0.088	0.177	0.257	0.178
	ours	**0.121**	**0.196**	0.201	**0.182**	**0.128**	**0.113**	**0.172**	**0.079**	**0.169**	**0.226**	**0.159**

Metric	Method	RGB Scenes										
		Dimsum	DailyStuff	Ebike	Road Block	Truck	Rotary Kiln	Building	Iron Ingot	Parterre	Land Scape	avg.
PSNR↑	mip-splatting	23.20	19.84	25.02	27.79	22.60	20.17	19.95	23.69	22.57	20.21	22.50
	TG(MFTG)	23.79	20.89	26.82	27.72	23.71	20.96	21.56	23.81	24.07	19.83	23.32
	TG(MSMG)	**24.72**	22.10	**27.65**	27.93	23.38	22.63	23.84	23.95	25.38	20.67	24.23
	TG(OMMG)	24.16	21.48	27.22	28.97	23.64	23.35	23.80	24.33	25.61	21.80	24.44
	ours	24.54	**22.16**	27.22	**29.02**	**24.08**	**23.77**	**24.41**	**25.28**	**26.22**	**24.24**	**25.09**
SSIM↑	mip-splatting	0.835	0.719	0.879	0.904	0.808	0.762	0.747	0.856	0.826	0.688	0.80
	TG(MFTG)	0.845	0.764	0.899	0.906	0.825	0.768	0.793	0.873	0.857	0.692	0.822
	TG(MSMG)	0.861	**0.811**	**0.920**	0.901	0.836	0.811	0.849	0.873	0.848	0.690	0.840
	TG(OMMG)	0.856	0.792	0.915	0.916	0.838	0.822	0.847	0.882	0.861	0.738	0.847
	ours	**0.866**	0.810	0.916	**0.929**	**0.853**	**0.831**	**0.854**	**0.894**	**0.876**	**0.788**	**0.862**
LPIPS↓	mip-splatting	0.201	0.331	0.198	0.208	0.229	0.229	0.300	0.207	0.226	0.289	0.242
	TG(MFTG)	0.196	0.285	0.174	0.207	0.224	0.225	0.233	0.189	0.190	0.288	0.221
	TG(MSMG)	0.204	0.255	0.174	0.278	0.239	0.205	**0.171**	0.218	0.249	0.349	0.234
	TG(OMMG)	0.196	0.259	0.164	0.219	0.218	0.183	0.172	0.186	0.193	0.268	0.206
	ours	**0.188**	**0.248**	**0.157**	**0.182**	**0.208**	**0.177**	0.174	**0.173**	**0.176**	**0.226**	**0.191**

metric alignment.

4.3. Pixel-aligned Rendering

As shown in the Fig. 2, we rendered the RGB and IR channels in ten real scenes with a unified perspective, and spliced the two modal images diagonally. This visualization method helps to observe the correspondence between the two modalities in the spatial structure at the same time. It can be seen that whether it is geometric edges (such as car body contours), or high-frequency details (such as keyboards, bottle caps, etc.), the two modalities show highly consistent rendering results. This further demonstrates the cross-modal expression ability of BiNAR in the modeling process and its good ability to maintain fine-grained geometric alignment.

4.4. Ablation Studies

In order to fully verify the advantages of BiNAR, we conducted three types of ablation experiments: modality ablation experiments, optimization strategy ablation experiments and calibration ablation experiments. Tab. 2, Tab. 3 and Tab. 4 show the ablation experiment results for all

Table 2. Modality ablation for all scenes. The best results are in bold.

Metric	Method	IR Scenes									
		Desktop	UAV	Kettles	K-bike	Car	Bicycle	Computer	Aircon	Apples	Bottles
PSNR↑	w/o IR	37.82	39.24	38.94	32.17	32.39	37.28	41.29	38.83	40.35	40.17
	w/o RGB	40.78	40.28	37.75	34.08	31.50	38.75	43.73	41.94	44.89	44.80
	BiNAR	**41.72**	**41.38**	**40.97**	**37.26**	**35.21**	**38.88**	**45.80**	**42.30**	**45.13**	**45.04**
SSIM↑	w/o IR	0.971	0.986	0.985	0.952	0.956	0.976	0.993	0.987	0.989	0.987
	w/o RGB	0.977	0.986	0.984	0.974	0.956	0.982	0.992	0.992	0.993	**0.993**
	BiNAR	**0.978**	**0.989**	**0.987**	**0.976**	**0.968**	**0.983**	**0.995**	**0.993**	**0.994**	**0.993**
LPIPS↓	w/o IR	0.248	0.166	0.177	0.245	0.268	0.194	0.052	0.081	0.103	0.117
	w/o RGB	0.235	0.156	0.269	0.150	0.258	**0.166**	0.053	**0.055**	0.092	**0.092**
	BiNAR	**0.230**	**0.153**	**0.171**	**0.134**	**0.220**	0.182	**0.038**	**0.055**	**0.090**	**0.092**
Metric	Method	RGB Scenes									
		Desktop	UAV	Kettles	E-bike	Car	Bicycle	Computer	Aircon	Apples	Bottles
PSNR↑	w/o IR	32.12	36.36	39.85	29.39	29.14	29.12	36.68	39.26	40.02	36.32
	w/o RGB	23.08	23.23	24.66	22.39	21.60	22.09	24.36	29.33	30.60	28.61
	BiNAR	**32.15**	**36.62**	**40.00**	**29.56**	**29.35**	**29.63**	**37.28**	**39.64**	**40.40**	**36.53**
SSIM↑	w/o IR	**0.966**	0.986	**0.991**	0.919	0.933	0.947	0.984	0.986	0.987	0.977
	w/o RGB	0.786	0.891	0.903	0.545	0.727	0.597	0.795	0.898	0.919	0.883
	BiNAR	**0.966**	**0.987**	**0.991**	**0.924**	**0.934**	**0.949**	**0.985**	**0.987**	**0.988**	**0.978**
LPIPS↓	w/o IR	**0.058**	0.085	0.054	0.098	0.106	0.065	0.032	0.047	0.066	0.074
	w/o RGB	0.359	0.342	0.176	0.502	0.414	0.457	0.354	0.265	0.262	0.287
	BiNAR	**0.058**	**0.077**	**0.051**	**0.092**	**0.104**	**0.062**	**0.031**	**0.046**	**0.065**	**0.073**

Table 3. Optimization strategy ablation for all scenes. The best results are in bold.

Metric	Method	IR Scenes									
		Desktop	UAV	Kettles	E-bike	Car	Bicycle	Computer	Aircon	Apples	Bottles
PSNR↑	RGB First	35.62	36.57	38.65	34.48	31.95	35.35	43.78	35.93	44.10	43.30
	IR First	25.55	24.17	23.54	22.52	23.49	22.41	31.49	22.42	27.00	28.74
	BiNAR	**41.72**	**41.38**	**40.97**	**37.26**	**35.21**	**38.88**	**45.80**	**42.30**	**45.13**	**45.04**
SSIM↑	RGB First	0.968	0.979	0.982	0.970	0.953	0.971	0.990	0.980	0.992	0.989
	IR First	0.862	0.894	0.882	0.771	0.865	0.838	0.952	0.909	0.933	0.959
	BiNAR	**0.978**	**0.989**	**0.987**	**0.976**	**0.968**	**0.983**	**0.995**	**0.993**	**0.994**	**0.993**
LPIPS↓	RGB First	0.250	0.184	0.176	0.138	0.236	0.220	0.052	0.099	0.095	0.100
	IR First	0.441	0.318	0.322	0.484	0.424	0.444	0.278	0.318	0.295	0.256
	BiNAR	**0.230**	**0.153**	**0.171**	**0.134**	**0.220**	**0.182**	**0.038**	**0.055**	**0.090**	**0.092**
Metric	Method	RGB Scenes									
		Desktop	UAV	Kettles	E-bike	Car	Bicycle	Computer	Aircon	Apples	Bottles
PSNR↑	RGB First	27.84	29.11	28.71	19.62	19.26	24.19	23.55	25.18	25.09	24.60
	IR First	30.92	34.38	35.34	26.63	26.25	28.33	34.63	38.34	39.83	36.33
	BiNAR	**32.15**	**36.62**	**40.00**	**29.56**	**29.35**	**29.63**	**37.28**	**39.64**	**40.40**	**36.53**
SSIM↑	RGB First	0.945	0.954	0.957	0.473	0.650	0.914	0.850	0.927	0.902	0.900
	IR First	0.958	0.983	0.968	0.893	0.888	0.944	0.980	0.983	0.987	0.977
	BiNAR	**0.966**	**0.987**	**0.991**	**0.924**	**0.934**	**0.949**	**0.985**	**0.987**	**0.988**	**0.978**
LPIPS↓	RGB First	0.092	0.179	0.173	0.419	0.393	0.109	0.194	0.151	0.242	0.192
	IR First	0.068	0.097	0.077	0.129	0.156	0.070	0.035	0.054	0.069	0.076
	BiNAR	**0.058**	**0.077**	**0.051**	**0.092**	**0.104**	**0.062**	**0.031**	**0.046**	**0.065**	**0.073**

Table 4. Calibration ablation for all scenes. The best results are in bold.

Metric	Method	IR Scenes									
		Desktop	UAV	Kettles	K-bike	Car	Bicycle	Computer	Aircon	Apples	Bottles
PSNR↑	w/o Calib	38.50	40.91	40.45	34.87	34.36	38.72	45.54	41.83	43.49	43.44
	BiNAR	**41.72**	**41.38**	**40.97**	**37.26**	**35.21**	**38.88**	**45.80**	**42.30**	**45.13**	**45.04**
SSIM↑	w/o Calib	0.972	0.987	0.986	0.962	0.961	0.977	0.994	0.992	0.992	0.992
	BiNAR	**0.978**	**0.989**	**0.987**	**0.976**	**0.968**	**0.983**	**0.995**	**0.993**	**0.994**	**0.993**
LPIPS↓	w/o Calib	0.242	0.158	0.176	0.174	0.243	0.206	0.046	0.058	0.095	0.096
	BiNAR	**0.230**	**0.153**	**0.171**	**0.134**	**0.220**	**0.182**	**0.038**	**0.055**	**0.090**	**0.092**
Metric	Method	RGB Scenes									
		Desktop	UAV	Kettles	E-bike	Car	Bicycle	Computer	Aircon	Apples	Bottles
PSNR↑	w/o Calib	32.02	36.36	39.76	29.02	28.81	29.39	37.17	39.39	39.65	36.31
	BiNAR	**32.15**	**36.62**	**40.00**	**29.56**	**29.35**	**29.63**	**37.28**	**39.64**	**40.40**	**36.53**
SSIM↑	w/o Calib	0.965	0.986	0.990	0.918	0.930	0.947	0.984	0.986	0.985	0.977
	BiNAR	**0.966**	**0.987**	**0.991**	**0.924**	**0.934**	**0.949**	**0.985**	**0.987**	**0.988**	**0.978**
LPIPS↓	w/o Calib	**0.058**	0.078	**0.051**	0.099	0.108	0.064	**0.031**	0.047	0.070	0.074
	BiNAR	**0.058**	**0.077**	**0.051**	**0.092**	**0.104**	**0.062**	**0.031**	**0.046**	**0.065**	**0.073**

scenes.

In the modality ablation experiment, we specifically removed the contribution of each modality in the Gaussians pose optimization process. As shown in Tab. 2, all scenes were evaluated and the performance of the three indicators of PSNR, SSIM and LPIPS was statistically analyzed. The experimental results show that when the contribution of the IR modality to the Gaussians pose optimization is removed (w/o IR), the PSNR and SSIM of the reconstruction results are reduced, while the LPIPS indicator (the lower the better) increases, indicating that relying solely on RGB optimization fails to fully capture the deep geometric information provided by IR data. Similarly, removing RGB optimization alone (w/o RGB) will also reduce the overall performance, but the reduction is different from w/o IR, reflecting the different complementarity of the information provided by the two modalities in pose optimization. This shows that the two modalities have different focuses in the optimization of Gaussians pose. The IR modality provides the enhancement of object edge and depth information, while the RGB modality plays a key role in capturing texture and details. The effect of using both for joint optimization (BiNAR) is significantly better than the independent optimization of pose by a single modality, thus ensuring the advantages of the final reconstruction result in visual quality and geometric consistency.

As shown in Tab. 3, we compared the experimental results of sequential optimization and simultaneous optimization of the two modalities. The experimental results show that whether the RGB First or IR First strategy is adopted, it is inferior to the BiNAR overall strategy in PSNR and SSIM; the LPIPS value is also higher, indicating that the sequential optimization method is difficult to take into account the information complementarity of the two modalities at the same time. The two-modal fusion and balanced optimization strategy adopted by BiNAR has a more obvious performance advantage

As shown in Tab. 4, we remove the bi-modal camera joint calibration module, and the rendering performance dropped significantly. This is because the location of the Gaussians does not match the actual camera line of sight, which leads to conflicts in the optimization process. When the difference between the poses of the two cameras increases, this conflict will also increase.

4.5. Limitations and Future Work

Our joint calibration yields an accurate relative pose RT between RGB and IR cameras, as validated by the sub-pixel reprojection error in Sec 5.2. However, the global scale alignment in Eq. (4) currently relies on manually measured distances and selected feature pairs in COLMAP. In some scenes, this distance-based scaling factor s can introduce a small global mismatch, and we apply a minor refinement of s (or T_{colmap}). Developing a fully automatic and more robust scale estimation or joint optimization of poses and scale within the reconstruction loop is an interesting direction for future work.

References

[1] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023. 1

[2] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik,

Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 1

[3] Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4104–4113, 2016. 1

[4] Zehao Yu, Anpei Chen, Binbin Huang, Torsten Sattler, and Andreas Geiger. Mip-splatting: Alias-free 3d gaussian splatting. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 19447–19456, 2024. 2