OC-GS: GAUSSIAN SPLATTING FOR IRREGULAR TURNTABLE CAPTURE

Anonymous authors
Paper under double-blind review

ABSTRACT

Uneven rotation and dropped frames make equal-angle assumptions unreliable for turntable reconstruction. We present OC-GS, an object-centric Gaussian splatting that refines each image's angle while maintaining a shared camera, rotation axis, and pivot. This *orbit-consistent refinement* jointly optimizes image-derived geometry and angles to reconstruct objects from sparse, irregular captures. On rendered objects with 12, 8, and 6 irregularly spaced views, OC-GS achieves mean foreground PSNR scores of 21.26, 19.36, and 15.83dB, respectively, exceeding all four evaluated pose-free Gaussian splatting baselines in each condition. Under a shared trainer, refining image-estimated angles improves mean foreground PSNR by 7.88dB over keeping those estimates fixed. An ablation study shows that both image-derived angle initialization and the shared motion model contribute to the improvement. On real captures, OC-GS's refinement increases mean foreground PSNR by 0.70dB. Results show that refining uncertain angles within a shared motion model improves reconstruction from sparse, irregular turntable captures.

1 INTRODUCTION

A turntable scanning setup is a common method that lets a fixed camera capture an object from multiple directions. It is used in plant phenotyping (Yang et al., 2024), cultural heritage digitization (Marshall et al., 2019), biological specimen digitization (Ströbel et al., 2018), and industrial surface inspection (Kaichi et al., 2018). Scanning time is critical, as some systems scan hundreds of objects a day. Reconstructing objects from fewer images can reduce capture time, but it complicates the reconstruction. Uneven rotation and dropped frames make inter-image angles unequal. We study this setting, where image order is known, but rotation angles are not.

Recent methods reconstruct Gaussian scenes without supplied camera poses. AnySplat (Jiang et al., 2025), InstantSplat (Fan et al., 2024), and SPFSplat (Huang & Mikolajczyk, 2025) use learned geometry to reconstruct from images. RotGS (Kim et al., 2026) incorporates a shared rotation axis and optical-flow supervision for turntable capture. These methods provide useful geometry and motion estimates, but irregular angular spacing remains a challenge.

A common assumption assigns angle $360^\circ j/N$ to view j, treating frame index as a measure of rotation. RotGS uses this schedule with an interpolated endpoint correction (Kim et al., 2026), which cannot represent arbitrary gaps between frames. Image-derived angles can capture unequal spacing, but keeping inaccurate estimates fixed preserves their errors. Independently optimizing every camera allows corrections that may violate the turntable motion. We instead refine individual angles within one shared rotation model.

We propose OC-GS, object-centric Gaussian splatting with *orbit-consistent refinement* (Fig. 1). Image predictions initialize the geometry and angles from a pre-trained initializer model such as VGGT (Wang et al., 2025a). OC-GS jointly optimizes the Gaussian scene, one camera, one rotation axis, one pivot, and an angle correction for each view. The shared rig maintains a consistent rotation trajectory, while the corrections adjust unequal angular gaps. Reconstruction therefore uses the captured images without requiring measured angles or constant rotation speed.

Our contributions are:

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

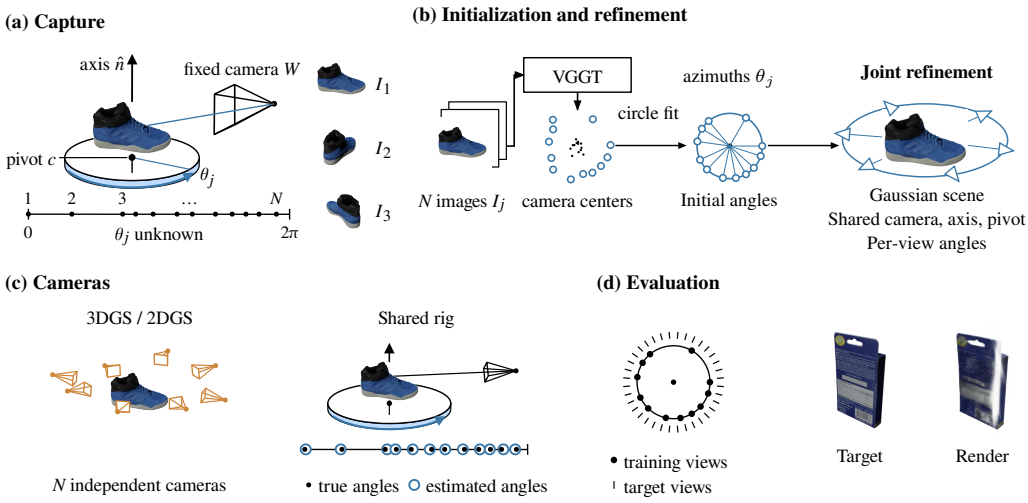

Figure 1: **OC-GS overview.** (a) A fixed camera captures an object rotating about one axis with unequal angular gaps. (b) VGGT predicts initial geometry and cameras. A circle fit to the camera centers gives starting angles. OC-GS jointly refines the Gaussian scene, shared camera, axis, pivot, and per-view angles. (c) The shared turntable rig (blue) constrains the effective cameras, unlike independent cameras (orange). (d) Reconstruction is evaluated by comparing rendered and target images.

- **Image-derived initialization for irregular capture:** predicted geometry to initialize the scene and fit a circle to predicted camera centers to estimate non-uniform angles (Sec. 2.2).
- **Per-view corrections to angular spacing:** we learn a correction for each view and subtract the mean correction to adjust relative spacing without introducing a redundant common rotation (Sec. 2.2).
- **Orbit-consistent joint refinement of shape and motion:** we optimize the angles, Gaussian scene, camera, axis, and pivot together under the turntable constraint. Component ablations show improvements over fixed angle estimates (Sec. 3.2).

2 OC-GS: ORBIT-CONSISTENT REFINEMENT

OC-GS refines image-derived geometry and angles under a shared turntable model. Each image has its own rotation angle, while all views share one camera, rotation axis, and pivot.

2.1 TURNTABLE RIG

We take N ordered RGB images I_j, foreground masks, and camera intrinsics K of an object completing a turn about a fixed axis. The angle of each image is unknown. Our goal is to recover a canonical Gaussian object $\mathcal{G}$ and cameras for novel-view rendering. We do not evaluate segmentation or intrinsic calibration.

Let W be the physical camera's world-to-camera transform, $\hat{n}$ the unit rotation-axis direction, and c a pivot on that axis, expressed in world coordinates. For image I_j, the object rotates by θ_j. Using 4×4 homogeneous transforms acting on column vectors, the object motion A_j and effective object-to-camera transform W_j are

$$A_j = T(c)R(\hat{n}, \theta_j)T(-c), \qquad W_j = W A_j. \tag{1}$$

where $T(v)$ denotes translation by v, and $R(\hat{n}, \theta_j)$ denotes rotation about the axis through the origin in direction $\hat{n}$. The product first moves the pivot to the origin, then rotates the object and translates it back. W maps the result into camera coordinates. Gaussian centers and local frames share the same rigid motion (App. A). This model groups all effective cameras into one trajectory while allowing unequal angular gaps.

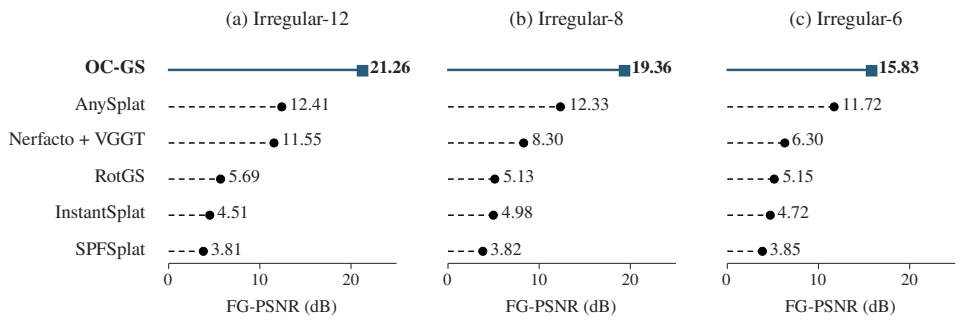

Figure 2: **Reconstruction quality under irregular capture.** Mean target FG-PSNR with 12, 8, and 6 views (higher is better). Objects must have valid camera alignment in all five component configurations and all three seeds. Each baseline's alignment failures are also excluded. Methods retain their own training settings.

2.2 Angle sources and spacing errors

Initial angles. For experiments on synthetic objects, we use VGGT (Wang et al., 2025a) to initialize both geometry and cameras, keeping its predictions fixed across variants to isolate the effect of angle refinement. We estimate angles by fitting a circle to the predicted camera centers, using the first image to define the reference angle (App. A). For real captures, we use AnySplat-derived angles (Jiang et al., 2025). OC-GS can use any initializers, but we use the best-performing ones (Sec. 3.3).

Angle refinement. OC-GS learns an angle correction for each image to improve the estimated spacing. Adding the same correction to every angle would only rotate the object, so we subtract the mean correction to adjust relative angles. We jointly optimize the corrections with the Gaussian scene, camera, axis, and pivot, while keeping the intrinsics fixed. Training uses an L_1 photometric loss with opacity and scale regularization. Initialization and optimization use only the input images (App. A).

2.3 Comparison variants.

We compare five settings in the same 2DGS (Huang et al., 2024) to distinguish the effects of angle refinement and shared motion. The **independent-camera variant** optimizes each camera separately. Within the shared rig, the **uniform-angle variant** assumes equal spacing, the **frozen-angle variant** keeps image-derived angles fixed, and **OC-GS** refines them. The **true-angle reference** uses exact angles but still estimates the axis and pivot.

We also use various spacings and image counts to validate the method. **Regular-12** and **Irregular-12** each contain twelve views, with equal or unequal angular gaps, respectively. **Irregular-8** and **Irregular-6** use fewer views with unequal gaps. Each configuration runs for 30k steps with three random seeds (App. A).

Evaluation evaluates reconstruction quality on 36 target views. We measure foreground pixel accuracy with FG-PSNR, perceptual similarity with LPIPS (Zhang et al., 2018), and silhouette overlap with IoU. Relative rotation error (RRE) compares predicted and reference rotations between camera pairs. For each object and capture condition, we average within runs and then across three different seeds. Metric definitions are given in App. B.

3 Results

We evaluate reconstruction quality, the contribution of angle refinement, and computational cost. We compare OC-GS with existing methods, then examine its components in ablation experiments on synthetic objects. We also test refinement on real captures and introduce known angle errors to determine when it helps. Finally, we study the reasons for limited baseline performance and compare runtime and memory use.

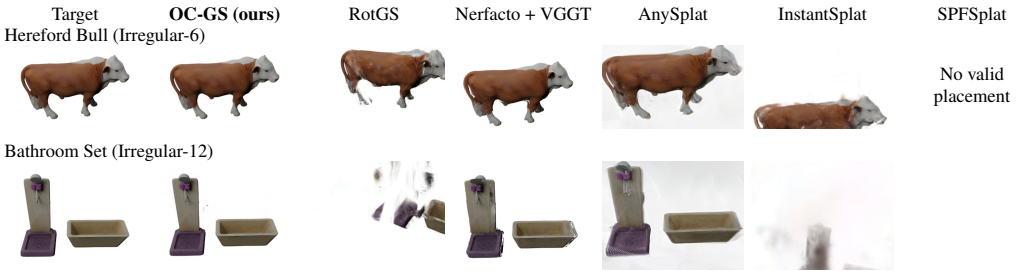

Figure 3: **Selected reconstructions under irregular capture.** Hereford Bull with six views and Bathroom Set with twelve, shown at the first target view with the same crop across methods. SPFSplat has no valid placement for Hereford Bull.

3.1 Comparison with existing reconstruction methods

We compare OC-GS with five baseline methods and two calibrated references on 100 objects from Google Scanned Objects (Downs et al., 2022). The four capture conditions use 12 regularly spaced views or 12, 8, or 6 irregularly spaced views. Within each condition, all methods are evaluated on the same 36 target views (App. C).

The baselines are AnySplat (Jiang et al., 2025), SPFSplat (Huang & Mikolajczyk, 2025), InstantSplat (Fan et al., 2024), RotGS (Kim et al., 2026), and Nerfacto (Tancik et al., 2023). We use the released versions of AnySplat, SPFSplat, InstantSplat, and RotGS. Nerfacto also runs for 30k steps, starting from cameras predicted by VGGT (Wang et al., 2025a) for the input views. Moreover, to evaluate two neural scene representation models such as Nerfacto and Splatfacto, we provide the true training cameras to run their published pipelines.

OC-GS leads under irregular capture. With 12, 8, and 6 irregularly spaced views, OC-GS achieved mean foreground PSNR scores of 21.26, 19.36, and 15.83dB, respectively, outperforming all evaluated pose-free GS methods at each view count (Fig. 2). When both methods yield valid results, OC-GS exceeded RotGS by at least 10.25dB across these conditions and AnySplat by 4.11dB with six views (Tab. A3). Our method, OC-GS, achieves the highest FG-PSNR across all irregular image settings compared with other baselines, and it scores, on average, 54.5% better than AnySplat, which has the second-best performance. Reconstruction quality decreased as fewer views were available, but OC-GS remained ahead of these baselines.

The qualitative examples show similar performance gaps (Fig. 3). OC-GS preserves the bull's leg positions and silhouette, while several baselines produce distorted or displaced reconstructions. For the Bathroom Set, OC-GS reconstructs the main parts of the scene, while baseline results show missing parts or scattered visuals.

Comparison with calibrated references. Comparisons with the two neural scene representations show the same trend. OC-GS outperforms Splatfacto with true cameras at eight and six irregular views. With twelve views, Splatfacto performs better under both regular and irregular capture. Across cases with valid camera alignment for both methods, OC-GS achieves 1.43dB higher mean FG-PSNR and 0.103 lower mean object-crop LPIPS (Tab. A4).

3.2 Which components account for the improvement?

Comparisons with external methods combine differences in initialization, camera models, and optimization. To isolate the effect of angle refinement, we compare variants using the same input images, VGGT predictions, trainer, and training budget. The key comparison holds the estimated angles fixed in one variant and optimizes them jointly with the scene in the other.

The shared rig alone is not enough. With estimated angles held fixed, the shared rig achieves 11.99dB mean FG-PSNR, below the 12.42dB of independent cameras (Tab. A1). Their relative rotation errors are also similar: 12.46° and 12.11°, respectively. This shows that the shared motion constraint alone does not compensate for inaccurate angles. Under regular capture, uniform angles

Table 1: **Effect of the shared rig and angle refinement.** FG-PSNR with the same trainer and VGGT predictions. Uniform angles assume equal spacing. Frozen angles retain image-derived estimates, while OC-GS refines them. The reference receives true angles. Means use the same valid objects per condition across all five variants and three seeds. Bold and underline indicate the best and second-best pose-free variants.

	FG-PSNR (dB) ↑			
Configuration	Regular-12	Irregular-12	Irregular-8	Irregular-6
No shared rig	13.07	<u>12.48</u>	<u>12.47</u>	<u>11.61</u>
Rig + uniform angles	**26.29**	6.54	6.65	6.28
Rig + frozen estimated angles	12.04	12.19	12.13	11.61
OC-GS	<u>22.75</u>	**21.26**	**19.36**	**15.83**
Rig + true angles (reference)	26.16	25.22	22.24	18.19
Gain from refining the estimated angles				
Mean (dB)	+10.71	+9.08	+7.23	+4.22

achieve 26.29dB, compared with 12.04dB for frozen image-derived angles in the same rig (Tab. 1). Accurate angular spacing is therefore important even when all views share one motion model.

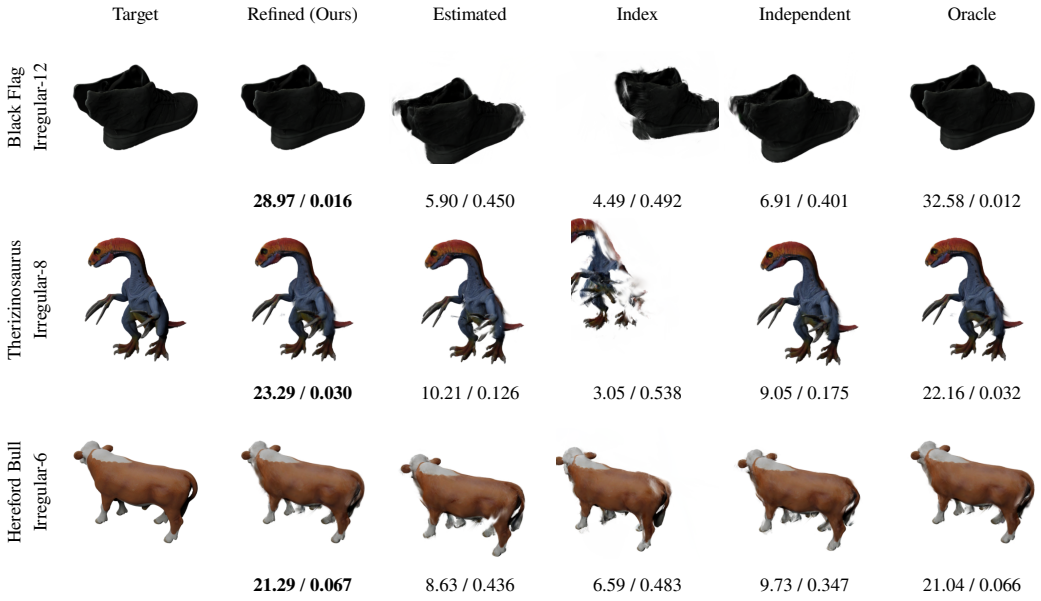

Figure 4: **Selected examples of improvement with OC-GS.** Black Flag (12 views), Therizinosaurus (8), and Hereford Bull (6) were selected to illustrate reconstruction gains. Each row uses the same target view and crop across methods. Scores report FG-PSNR (dB) / object-crop LPIPS. Bold indicates the best valid pose-free result.

Joint angle refinement improves reconstruction. Jointly refining the estimated angles raises mean FG-PSNR to 19.88dB, an average improvement of 7.88dB over keeping them fixed. Refinement improves the mean in every capture condition (Tab. 1), and the median improvement is 7.00dB. Mean object-crop LPIPS decreases from 0.281 to 0.142, while silhouette IoU increases from 0.718 to 0.860 (Tab. A1). Relative rotation error also decreases, and angular-spacing RMSE falls from $10.10°$ to $8.53°$. The qualitative examples show clearer silhouettes and more complete object parts after refinement (Fig. 4).

Initialization and the shared rig both matter. We use two additional ablations to test the roles of image-derived initial angles and the shared rig (Tab. 2). When initialized with uniform angles, refinement improves FG-PSNR by 1.96dB over keeping those angles fixed, but remains 5.80dB below OC-GS. This result supports the use of image-derived angles as the starting point for refinement.

For independent cameras, increasing both camera learning rates tenfold improves FG-PSNR by only 0.23dB, leaving a 6.92dB gap to OC-GS. At the tested settings, faster camera optimization does not

recover the benefit of refining image-derived angles within the shared rig. For these ablations, we first average over valid conditions for each object, then average equally across objects.

Table 2: **Refinement without image-derived angles or without the rig.** Uniform angles + refinement removes image-derived angular initialization. Bold marks the best mean per column using FG-PSNR (dB).

Configuration	Regular-12 ↑	Irregular-12 ↑	Irregular-8 ↑	Irregular-6 ↑
Uniform angles + refinement	**23.39**	12.78	10.65	7.57
Independent cameras (higher rate)	13.30	12.75	12.70	11.83
OC-GS	22.75	**21.26**	**19.36**	**15.83**

The gain persists across evaluation and loss choices. The average refinement gain remains between 6.9 and 8.1dB when we change how alignment failures are handled, apply stricter camera-alignment criteria, or exclude target views close to the training views (App. B and App. E). In a separate experiment with three L_1/SSIM loss weightings, refinement improves mean FG-PSNR in all four capture conditions under every weighting. Average gains range from 4.88 to 7.68dB (App. E).

Refinement narrows the gap to true angles. Refinement reduces, but does not eliminate, the gap to the true-angle reference. With twelve irregular views, uniform angles produce FG-PSNR 18.68dB below the reference. Image-derived angles close 5.65dB of this gap, and joint refinement adds another 9.08dB. With six views, refinement still provides a median improvement of 4.23dB over frozen estimates, although mean quality remains below the true-angle reference.

3.3 Real-capture evidence and initialization

We next validate whether angle refinement also improves reconstruction on the RotGS released dataset, which consists of real captures. We compare the initial angle estimates and optimized angles during the reconstruction using 24, 17, 13, and 7 training views. Both settings use the same initial angles estimated by AnySplat. We show that optimizing these angles improves mean FG-PSNR by 0.70dB and lowers mean LPIPS at every view count (App. D).

Initializers on synthetic and real captures. We further study the effectiveness of different initializers with OC-GS on synthetic and real captured datasets. VGGT shows the best performance, which outperforms the second-best, VGGT-Ω, by 0.86 dB (Tab. 3). On the RotGS released dataset, AnySplat outperforms the second-best π^3 by 0.36dB (Tab. 4).

Table 3: **OC-GS with different initializers.** We show the performance of different initializers and report them with the mean FG-PSNR (dB ↑). The best is bolded, which is from VGGT.

OC-GS initializer	Regular-12	Irregular-12	Irregular-8	Irregular-6
VGGT	**22.75**	**21.26**	**19.36**	**15.83**
AnySplat	19.88	20.28	18.30	15.10
π^3	19.64	20.45	18.61	15.19
VGGT-Ω	21.19	20.13	18.87	15.57
Depth Anything 3	20.24	19.77	18.19	15.28
CUT3R	17.26	18.45	15.71	14.70
MapAnything	7.81	8.39	6.70	5.82
DUSt3R	13.35	15.42	14.36	11.10
MASt3R-SfM	17.89	14.43	8.46	7.07
Index angles (VGGT geometry/rig)	15.55	11.33	10.13	9.28

3.4 When does angular refinement help?

Because we use pretrained models for the angle, they can be inaccurate due to the input image or the initialization method. To ablate those effects, we study the effect of angle error directly by adding noise of known magnitude to the ground-truth angles. We then compare keeping the noisy angles unchanged with optimizing them during reconstruction. Both settings use the same twelve regularly spaced input views, VGGT-initialized geometry, and an estimated rotation axis.

Table 4: **Pose initialization and camera models on real captures.** Mean FG-PSNR (dB), evaluated using each reconstruction's cameras. The named methods provide initial poses. The OC-GS rig keeps its azimuths fixed, while independent cameras are optimized from the predictions. Bold and underline indicate the best and second-best predictors within each block at the displayed precision.

Configuration / training views	24	17	13	7
OC-GS shared rig, initialized azimuths frozen				
AnySplat azimuths	**22.62**	**21.53**	**19.50**	**15.73**
π^3 azimuths	<u>21.82</u>	<u>20.91</u>	**19.50**	<u>15.72</u>
Index angles (no network)	21.20	17.12	16.40	13.56
Per-view cameras, optimized from				
π^3	<u>20.45</u>	<u>18.70</u>	**16.75**	**14.03**
AnySplat	**20.67**	**18.91**	<u>15.36</u>	<u>13.82</u>
Depth Anything 3	18.52	15.49	14.30	13.53
VGGT	17.52	16.08	13.97	11.56
VGGT-Ω	17.51	15.89	14.87	10.38
CUT3R	16.73	14.93	12.66	9.33
MASt3R-SfM	17.23	13.96	10.71	9.85
MapAnything	16.38	13.40	12.17	9.15
DUSt3R	14.99	13.45	11.05	7.33
Index angles (no network)	20.79	17.08	15.83	12.12

We remove the average angular offset introduced by the noise, leaving only errors in the relative angles between views. We measure these errors using their root-mean-square (RMS) value and test levels from 0° (exact angles) to 10° (Tab. 5).

Table 5: **Angle-error experiment.** RMS spacing error is the root-mean-square angular noise in degrees after removing a common rotation. `Frozen` keeps the supplied angles fixed, while `Refined` optimizes them. Δ reports the refinement gain.

RMS spacing error	Frozen ↑	Refined ↑	Δ (dB)
0° (exact angles)	26.04	22.88	-3.16
0.5°	18.44	23.19	$+4.75$
1°	15.89	22.92	$+7.03$
2°	13.50	22.60	$+9.10$
3°	12.36	22.71	$+10.35$
10°	9.08	20.33	$+11.25$

Angle refinement improves reconstruction at all tested error levels from 0.5° to 10° RMS (Tab. 5). Compared with keeping the initial angles unchanged, the FG-PSNR gain improves from 4.75dB at 0.5° to 11.25dB at 10°. At the largest error, refinement raises FG-PSNR from 9.08dB to 20.33dB. These results support refining imperfect angle estimates in our reconstruction setting. When the initial angles are exact, keeping them unchanged improves performance by 3.16dB (App. E).

Also, with exact initial angles, both variants fit the training images similarly well, but refinement produces worse results on unseen views. This is consistent with overfitting and shows why training-image quality alone is insufficient to assess reconstruction (App. E).

3.5 Sources of error in the baseline reconstructions

We run additional experiments to investigate why the baselines perform worse on unseen views. These validate camera consistency, the ability to reproduce training images, and the spatial distribution of reconstructed geometry.

RotGS: sensitivity to the rotation axis. RotGS reproduces its training images well when rendered through its estimated cameras and is also designed for the turntable setup, but quality decreases when the same scene is rendered through aligned reference cameras. In a separate regular-capture experiment on five objects, fixing the rotation axis to its true value increases mean FG-PSNR from 15.97 to 29.45dB, with improvements on every object. This result suggests that axis estimation can limit reconstruction even when the input views are regularly spaced (App. F.1).

Image-based baselines: camera errors and limited training fit. For AnySplat and InstantSplat, replacing inferred camera parameters with calibrated ones while holding the reconstructed scene fixed reduces training-image quality. This indicates inconsistency between the reconstructed scene and the true cameras. SPFSplat also struggles to fit the training images, even with calibrated intrinsics. Correcting InstantSplat's camera export improves target-view quality only slightly (App. F.2).

Nerfacto: good training fit, limited unseen views quality. Nerfacto is able to fit sparse training images well while producing unseen views of limited visual quality, even with true cameras. In the reconstructions studied here, much of the predicted content lies outside the object region. In a four-object experiment with true cameras and known scene bounds, replacing scene contraction with a bounded field increases mean target FG-PSNR from 6.80 to 17.78dB under irregular capture.

Runtime and memory use. We also evaluate the cost comparison. OC-GS achieves the highest mean FG-PSNR and uses the least peak GPU memory, 1.51 GB (Tab. 6). Fitting takes 16.2 minutes, less than RotGS and Nerfacto. InstantSplat and the feed-forward methods are faster but have lower reconstruction quality. VGGT inference is precomputed for OC-GS.

Table 6: **Reconstruction quality, runtime, and peak GPU memory.** Quality uses valid benchmark runs. Cost uses separate profiling runs. VGGT inference is excluded from runtime and memory for OC-GS and Nerfacto. Bold and underline indicate the best and second-best values at the displayed precision.

Method	Type	Time (min) ↓	Peak VRAM (GB) ↓
OC-GS	Per-scene	16.2	**1.51**
RotGS (index schedule)	Per-scene	20.8	<u>2.20</u>
Nerfacto + VGGT poses	Per-scene	23.7	3.05
InstantSplat	Per-scene	<u>2.5</u>	6.69
AnySplat	Feed-forward	**0.5**	8.31
SPFSplat	Feed-forward	**0.5**	6.06

4 LIMITATIONS

OC-GS assumes a fixed axis with minimal axis noise (App. F.1) and depends on initialization(VGGT). However, we expect the future models to improve reconstruction.

5 CONCLUSION

OC-GS reconstructs objects from sparse, irregular turntable images by jointly optimizing image-derived angles and Gaussian geometry under a shared motion model. It outperforms the evaluated pose-free GS baselines in all three irregular capture conditions. With the same trainer and initialization, angle refinement adds 9.08dB FG-PSNR with twelve irregular views and 4.22dB with just six views, compared with keeping the estimated angles fixed (Tab. 1). Moreover, OC-GS uses 31% less peak GPU memory and takes 22% less fitting time than RotGS, excluding VGGT inference (Tab. 6). Future work will adapt refinement to initialization quality and evaluate its use in plant phenotyping as a downstream task.

6 RELATED WORK

Gaussian reconstruction from sparse images. 3DGS (Kerbl et al., 2023) and 2DGS (Huang et al., 2024) optimize Gaussian scenes from images with known poses, often estimated by COLMAP (Schonberger & Frahm, 2016). InstantSplat (Fan et al., 2024) combines learned geometry with joint Gaussian and camera optimization. FreeSplatter (Xu et al., 2025), AnySplat (Jiang et al., 2025), and SPFSplat (Huang & Mikolajczyk, 2025) reconstruct Gaussian scenes without supplied poses. OC-GS uses the turntable motion to constrain reconstruction: all views share one camera, axis, and pivot, with a separate angle for each image.

Learning geometry and camera poses. DUSt3R (Wang et al., 2024) predicts dense point maps, and MASt3R (Leroy et al., 2024) adds matching features for geometric alignment. VGGT (Wang et al., 2025a) predicts cameras and scene geometry jointly. π^3 (Wang et al., 2026b) removes the

need for a designated reference view, while CUT3R (Wang et al., 2025b) maintains a persistent state across images. Depth Anything 3 (Lin et al., 2025), MapAnything (Keetha et al., 2026), and VGGT-Ω (Wang et al., 2026a) provide additional sources of geometry and camera estimates. OC-GS converts predicted cameras into initial turntable angles and refines them using the captured images.

Joint camera and scene refinement. NeRF– (Wang et al., 2021), BARF (Lin et al., 2021), and SCNeRF (Jeong et al., 2021) jointly optimize camera poses and scene radiance. CamP (Park et al., 2023) improves camera optimization through preconditioning. SPARF (Truong et al., 2023) uses correspondence and depth consistency, while PoRF (Bian et al., 2024) learns a pose residual field. NeRS (Zhang et al., 2021), SAMURAI (Boss et al., 2022), and NeROIC (Kuang et al., 2022) reconstruct objects from image collections with uncertain cameras. OC-GS constrains all camera updates to a shared rotation trajectory throughout joint scene and motion optimization.

Turntable geometry. Earlier work recovered shape and camera motion using silhouettes, circular-motion constraints, and conic fitting (Fitzgibbon et al., 1998; Mendonca et al., 2001; Wong & Cipolla, 2001; Jiang et al., 2003; Hernández et al., 2007). RotGS (Kim et al., 2026) combines a shared rotation axis with optical-flow alignment for Gaussian reconstruction. Its uniform index schedule and interpolated endpoint correction cannot represent arbitrary per-frame gaps. OC-GS initializes nonuniform angles from images and refines each angle around a shared axis.

Reproducibility statement

We provide detailed information in the supplementary material. Training and sampling settings are given in App. A, metrics in App. B, and baseline settings in App. C. We will release the code upon acceptance.

Ethics statement

This study uses rendered objects and real object captures and does not involve human subjects.

AI use statement

An AI writing assistant was used to revise the manuscript's grammar and wording.

References

Jia-Wang Bian, Wenjing Bian, Victor Prisacariu, and Philip Torr. Porf: Pose residual field for accurate neural surface reconstruction. In *International Conference on Learning Representations*, volume 2024, pp. 46620–46638, 2024.

Mark Boss, Andreas Engelhardt, Abhishek Kar, Yuanzhen Li, Deqing Sun, Jonathan Barron, Hendrik Lensch, and Varun Jampani. Samurai: Shape and material from unconstrained real-world arbitrary image collections. *Advances in Neural Information Processing Systems*, 35:26389–26403, 2022.

Laura Downs, Anthony Francis, Nate Koenig, Brandon Kinman, Ryan Hickman, Krista Reymann, Thomas B McHugh, and Vincent Vanhoucke. Google scanned objects: A high-quality dataset of 3d scanned household items. In *2022 international conference on robotics and automation (ICRA)*, pp. 2553–2560. Ieee, 2022.

Zhiwen Fan, Wenyan Cong, Kairun Wen, Kevin Wang, Jian Zhang, Xinghao Ding, Danfei Xu, Boris Ivanovic, Marco Pavone, Georgios Pavlakos, et al. Instantsplat: Sparse-view gaussian splatting in seconds. *arXiv preprint arXiv:2403.20309*, 2024.

Andrew W Fitzgibbon, Geoff Cross, and Andrew Zisserman. Automatic 3d model construction for turn-table sequences. In *European Workshop on 3D Structure from Multiple Images of Large-Scale Environments*, pp. 155–170. Springer, 1998.

Carlos Hernández, Francis Schmitt, and Roberto Cipolla. Silhouette coherence for camera calibration under circular motion. *IEEE transactions on pattern analysis and machine intelligence*, 29(2): 343–349, 2007.

Binbin Huang, Zehao Yu, Anpei Chen, Andreas Geiger, and Shenghua Gao. 2d gaussian splatting for geometrically accurate radiance fields. In *ACM SIGGRAPH 2024 conference papers*, pp. 1–11, 2024.

Ranran Huang and Krystian Mikolajczyk. No pose at all: Self-supervised pose-free 3d gaussian splatting from sparse views. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 27947–27957. IEEE, 2025.

Yoonwoo Jeong, Seokjun Ahn, Christopher Choy, Anima Anandkumar, Minsu Cho, and Jaesik Park. Self-calibrating neural radiance fields. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 5846–5854, 2021.

Guang Jiang, Hung-Tat Tsui, Long Quan, and Andrew Zisserman. Geometry of single axis motions using conic fitting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10): 1343–1348, 2003.

Lihan Jiang, Yucheng Mao, Linning Xu, Tao Lu, Kerui Ren, Yichen Jin, Xudong Xu, Mulin Yu, Jiangmiao Pang, Feng Zhao, et al. Anysplat: Feed-forward 3d gaussian splatting from unconstrained views. *ACM Transactions on Graphics (TOG)*, 44(6):1–16, 2025.

Tomoya Kaichi, Shohei Mori, Hideo Saito, Junichi Sugano, and Hideyuki Adachi. Multi-view surface inspection using a rotating table. *Electronic Imaging*, 30:1–5, 2018.

Nikhil Keetha, Norman Müller, Johannes Schönberger, Lorenzo Porzi, Yuchen Zhang, Tobias Fischer, Arno Knapitsch, Duncan Zauss, Ethan Weber, Nelson Antunes, et al. Mapanything: Universal feed-forward metric 3d reconstruction; map-anything. github. io. In *2026 International Conference on 3D Vision (3DV)*, pp. 499–509. IEEE, 2026.

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

Kyumin Kim, Dohae Lee, Hanul Baek, and In-Kwon Lee. Rotgs: Rotation-guided 3d gaussian splatting for turntable sequences without structure-from-motion. In *Computer Graphics Forum*, pp. e70317. Wiley Online Library, 2026.

Zhengfei Kuang, Kyle Olszewski, Menglei Chai, Zeng Huang, Panos Achlioptas, and Sergey Tulyakov. Neroic: Neural rendering of objects from online image collections. *ACM Transactions on Graphics (ToG)*, 41(4):1–12, 2022.

Vincent Leroy, Yohann Cabon, and Jérôme Revaud. Grounding image matching in 3d with mast3r. In *European conference on computer vision*, pp. 71–91. Springer, 2024.

Chen-Hsuan Lin, Wei-Chiu Ma, Antonio Torralba, and Simon Lucey. Barf: Bundle-adjusting neural radiance fields. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 5721–5731. IEEE, 2021.

Haotong Lin, Sili Chen, Junhao Liew, Donny Y Chen, Zhenyu Li, Guang Shi, Jiashi Feng, and Bingyi Kang. Depth anything 3: Recovering the visual space from any views. *arXiv preprint arXiv:2511.10647*, 2025.

ME Marshall, AA Johnson, SJ Summerskill, Q Baird, and E Esteban. Automating photogrammetry for the 3d digitisation of small artefact collections. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42:751–757, 2019.

Paulo R. S. Mendonca, K-YK Wong, and Roberto Cipolla. Epipolar geometry from profiles under circular motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):604–616, 2001.

Keunhong Park, Philipp Henzler, Ben Mildenhall, Jonathan T Barron, and Ricardo Martin-Brualla. Camp: Camera preconditioning for neural radiance fields. *Acm Transactions on Graphics (TOG)*, 42(6):1–11, 2023.

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

Bernhard Ströbel, Sebastian Schmelzle, Nico Blüthgen, and Michael Heethoff. An automated device for the digitization and 3d modelling of insects, combining extended-depth-of-field and all-side multi-view imaging. *ZooKeys*, (759):1, 2018.

Matthew Tancik, Ethan Weber, Evonne Ng, Ruilong Li, Brent Yi, Terrance Wang, Alexander Kristoffersen, Jake Austin, Kamyar Salahi, Abhik Ahuja, et al. Nerfstudio: A modular framework for neural radiance field development. In *ACM SIGGRAPH 2023 conference proceedings*, pp. 1–12, 2023.

Prune Truong, Marie-Julie Rakotosaona, Fabian Manhardt, and Federico Tombari. Sparf: Neural radiance fields from sparse and noisy poses. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4190–4200. IEEE, 2023.

Shinji Umeyama. Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 13(4):376–380, 1991.

Jianyuan Wang, Minghao Chen, Nikita Karaev, Andrea Vedaldi, Christian Rupprecht, and David Novotny. Vggt: Visual geometry grounded transformer. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5294–5306. IEEE, 2025a.

Jianyuan Wang, Minghao Chen, Shangzhan Zhang, Nikita Karaev, Johannes Schönberger, Patrick Labatut, Piotr Bojanowski, David Novotny, Andrea Vedaldi, and Christian Rupprecht. Vggt-ω. *arXiv preprint arXiv:2605.15195*, 2026a.

Qianqian Wang, Yifei Zhang, Aleksander Holynski, Alexei A Efros, and Angjoo Kanazawa. Continuous 3d perception model with persistent state. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10510–10522. IEEE, 2025b.

Shuzhe Wang, Vincent Leroy, Yohann Cabon, Boris Chidlovskii, and Jerome Revaud. Dust3r: Geometric 3d vision made easy. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20697–20709. IEEE, 2024.

Yifan Wang, Jianjun Zhou, Haoyi Zhu, Wenzheng Chang, Yang Zhou, Zizun Li, Junyi Chen, Jiangmiao Pang, Chunhua Shen, and Tong He. \piˆ 3: Permutation-equivariant visual geometry learning. In *International Conference on Learning Representations*, volume 2026, pp. 10481–10497, 2026b.

Zirui Wang, Shangzhe Wu, Weidi Xie, Min Chen, and Victor Adrian Prisacariu. NeRF–: Neural radiance fields without known camera parameters. *arXiv preprint arXiv:2102.07064*, 2021.

K-YK Wong and Roberto Cipolla. Structure and motion from silhouettes. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 2, pp. 217–222. IEEE, 2001.

Jiale Xu, Shenghua Gao, and Ying Shan. Freesplatter: Pose-free gaussian splatting for sparse-view 3d reconstruction. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 25442–25452. IEEE, 2025.

Danni Yang, Huijun Yang, Dongfeng Liu, and Xianlin Wang. Research on automatic 3d reconstruction of plant phenotype based on multi-view images. *Computers and Electronics in Agriculture*, 220:108866, 2024. ISSN 0168-1699. doi: https://doi.org/10.1016/j.compag.2024.108866. URL https://www.sciencedirect.com/science/article/pii/S0168169924002576.

Jason Zhang, Gengshan Yang, Shubham Tulsiani, and Deva Ramanan. Ners: Neural reflectance surfaces for sparse-view 3d reconstruction in the wild. *Advances in Neural Information Processing Systems*, 34:29835–29847, 2021.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *2018 IEEE/CVF conference on computer vision and pattern recognition*, pp. 586–595. IEEE, 2018.

Tinghui Zhou, Richard Tucker, John Flynn, Graham Fyffe, and Noah Snavely. Stereo magnification: Learning view synthesis using multiplane images. *arXiv preprint arXiv:1805.09817*, 2018.