DOI: 10.1111/cgf.70317

EUROGRAPHICS 2026 / B. Masia and J. Thies
(Guest Editors)

COMPUTER GRAPHICS *forum*
Volume 45 (2026), Number 2

RotGS: Rotation-Guided 3D Gaussian Splatting for Turntable Sequences without Structure-from-Motion

Kyumin Kim[1], Dohae Lee[3], Hanul Baek[2], and In-Kwon Lee[2,3]

[1]Department of Intelligence Convergence, Yonsei University, Korea
[2]Department of Computer Science, Yonsei University, Korea
[3]MeTown Inc.

(a) Turntable setup (b) Comparision with previous SfM-free methods (c) 3DGS with SfM

Figure 1: *(a) Turntable setup. A fixed camera and a rotating turntable are used to capture an image sequence of the object rotating around its central axis. (b) Compared to existing SfM-free methods (Nope-NeRF, CF-3DGS, HT-3DGS, and COGS), our method produces higher-quality reconstructions in less time. (c) Background removal is essential in a turntable setup, but it leads to insufficient features, causing SfM failures. Optimizing 3DGS with camera poses from failed SfM results in lower-quality reconstructions.*

Abstract

The field of 3D reconstruction from multi-view images has advanced rapidly thanks to 3D Gaussian Splatting (3DGS), which enables efficient and photorealistic scene representation. However, optimizing 3DGS requires high-quality images from various viewpoints with accurate camera poses. The repeated collection of such data demands significant human effort, which poses a major constraint in practical applications. To address this issue, automated capturing systems that uses a turntable and fixed camera are widely employed. In a turntable setup, the background remains stationary while the object rotates. Therefore, preprocessing to remove the backgrond is essential, but the preprocessing reduces the number of reliable feature matches, which destabilizes Structure-from-Motion (SfM). This results in inaccurate camera poses, which degrades the quality of 3DGS reconstruction. We propose a novel method to optimize 3DGS in a turntable setup without SfM by leveraging the prior knowledge that objects rotate around a central axis. Unlike previous SfM-free methods that estimate camera poses for each frame, our approach reduces the complexity of optimization by representing rotations with a single global rotation axis. The estimated rotation is directly applied to the 3D Gaussians, producing motion defined as rotation flow. This rotation flow is then aligned with optical flow to provide strong geometric supervision. Through uncertainty-to-detail flow scheduling, our approach remains stable during the initial training stage when the geometry of the Gaussian set is still inaccurate. On the NeRF-Synthetic dataset and on real-world datasets captured with a turntable, our method outperforms existing SfM-free approaches in both reconstruction quality and training speed, and even demonstrates performance comparable to 3DGS optimized with precise camera poses.

CCS Concepts

• ***Computing methodologies*** → ***Reconstruction; Image-based rendering;*** *3D imaging;*

(a) Original Image (b) SfM success (c) w/o background (d) SfM failure

Figure 2: ***The effect of background removal on SfM.*** *(a) Image set with background captured by moving around the object. (b) Using the original images with the background enables SfM to leverage features from both the object and the background, thereby providing a sufficient number of features for reliable SfM. (c) Images with the background removed. (d) With the background removed images, SfM can only use features from the object itself, which reduces the number of available features and leads to SfM failure.*

1. Introduction

Reconstruction of 3D objects from images has long been a core research topic in the computer graphics community. In particular, Neural Radiance Fields (NeRF) [MST*21] have enabled photorealistic novel view synthesis, yet their slow training and inference speeds limit their practical use. Recently, 3D Gaussian Splatting (3DGS) [KKLD23] has addressed these limitations, achieving fast training, real-time rendering, and high-quality reconstruction simultaneously. As a result, 3DGS has been actively applied in areas such as entertainment and virtual reality [KL24; QXLH25]. However, similar to most existing approaches, capturing images with a handheld camera while moving around an object has several limitations, making it unsuitable for practical applications. Specifically, in commercial applications, many objects must be scanned repeatedly, and manually moving the camera to capture each object requires significant time and effort [FLY*25]. Furthermore, this approach frequently results in human errors such as out-of-focus and motion blur during capturing, which directly leads to degraded reconstruction quality [LZL*22]. Moreover, it requires a large physical space for scanning objects, making it unsuitable for indoor scanning scenarios [KJR*25]. To address these issues, the practical application employs an automated capture pipeline combining a fixed camera with a turntable [TML14; KAP*17], as shown in Figure 1 (a). This approach enables object capture in a minimal space, reduces potential human error during the process, and allows for stable and efficient capture.

In turntable setups, the background remains fixed while objects rotate, making background removal essential for camera pose estimation. Since 3DGS requires accurate camera poses for optimization, these poses are typically obtained through a preprocessing step using Structure-from-Motion (SfM) [SF16]. However, background removal can lead to an insufficient number of feature points for SfM (Figure 2 (c)). As a result, in a turntable setup, SfM often fails to estimate accurate camera poses (Figure 2 (d)), frequently degrading the reconstruction quality of 3DGS. This issue is further worsened for objects that are symmetric, textureless, or highly reflective [FLK*24]. While attaching physical markers is a common approach, it is often impractical and limits generalization. Markers often create unwanted reflections on glossy surfaces and force the collection of separate image sets—one with markers for pose estimation and another without them for 3DGS reconstruction. In addition, the computational cost of SfM is significant, which negatively impacts the overall efficiency of the reconstruction pipeline [JY25]. To address this, recent studies [FLK*24; JY25; JFV*24] have proposed approaches that jointly optimize camera poses and scene representation without SfM. However, these methods assume small camera motion, which may cause pose estimation failures under large camera motions. In addition, estimating camera poses for each frame individually results in prolonged training times. High-performance SfM methods, such as [PBPS24] have also been proposed. However, in turntable setups with a limited number of feature points, these methods still struggle to provide stable results. Several studies have utilized prior knowledge that objects rotate around a central axis to reconstruct 3D objects or estimate rotation. These approaches, however, either require specialized hardware such as structured lighting [KH12a], or achieve significantly lower reconstruction quality compared to the latest 3D reconstruction methods.

In this paper, we propose a novel method of optimizing 3DGS that does not rely on SfM, leveraging the prior knowledge that the object rotates around its central axis on a turntable. Our method jointly estimates the rotation axis and angles while optimizing the Gaussian attributes. To achieve this, we rotate the Gaussian set according to the estimated rotation and render it from a fixed camera position. The motion of the Gaussians induced during the rotation process is defined as rotation flow, and by aligning it with the optical flow [XZC*22], it provides strong geometric supervision to the model. To leverage rotation flow effectively during the early stages of training, when the Gaussian geometry is inaccurate, we propose uncertainty-to-detail flow scheduling. This approach initially considers only the flow direction, gradually incorporating the flow magnitude as the geometry becomes more accurate. Additionally, we represent the rotation of all frames using a single, shared global rotation axis, thereby reducing the complexity of the optimization process.

We evaluate the effectiveness of our proposed approach on the NeRF-Synthetic dataset [MST*21] and a turntable-based real dataset. Compared with existing SfM-free methods, our approach achieves superior reconstruction quality while also reducing training time. Furthermore, even without using SfM, our method attains comparable qualitative and quantitative performance to 3DGS that rely on SfM. We further validate the effectiveness of the proposed method through an ablation study. Our main contributions are as follows:

- We propose a novel SfM-free framework for 3DGS that leverages the turntable prior to jointly optimize Gaussian attributes, rotation axis, and rotation angles, achieving superior reconstruction quality and faster convergence.
- We propose a rotation flow derived from Gaussian motion and a training strategy that schedules the rotation flow according to Gaussian geometry, thereby providing strong geometric supervision and improved optimization stability.

2. Related Work

2.1. Turntable-based 3D Reconstruction

A turntable rotates an object around a single axis. This allows images to be easily acquired from various viewpoints and enables efficient 3D reconstruction in repetitive capture scenarios. In addition, the single-axis rotation imposes a geometric constraint, which provides structural information that can be leveraged during the reconstruction process. The early work by [FCZ98] shows that, by leveraging the geometric constraint of single-axis rotation, 3D structures can be reconstructed from image sequences without information about the camera or turntable. Subsequently, [FC04] proposes a method that combines feature points with conical projection geometry to estimate 3D coordinates from image sequences, while [ZLSH08] leverages feature points and silhouette information to reconstruct 3D rim curves, demonstrating the potential for 3D reconstruction in turntable setups. Structured-light-based approaches, such as [KH12b; YS15], estimate the rotation axis of the turntable and subsequently align the corresponding point clouds, achieving a high level of reconstruction accuracy. Previous studies required calibration objects, such as planar or cubic checkerboards, to estimate the rotation axis. In contrast, [PLS*14] proposes estimating the rotation axis without any additional calibration tools, significantly improving user convenience. Recent studies, such as [FLY*25; LZL*22; SWLY25; KAP*17], enable end-to-end 3D reconstruction for various applications, including automated capture and personal 3D printing, while addressing challenges encountered in real-world turntable setups, such as illumination changes, motion blur, and camera calibration. These studies show that turntable-based systems can achieve efficient and accurate 3D reconstruction with minimal prior information.

2.2. Structure from Motion

Structure-from-Motion (SfM) [SF16] is a representative method for reconstructing 3D structures and camera poses from image sequences. Incremental SfM methods [Wu13; SSS06; AFS*11; PNF*08], such as COLMAP [SF16], estimate camera poses and 3D points by gradually adding images, providing high accuracy and robustness. However, they are prone to drift due to cumulative errors during image registration and often require considerable processing time. To address these issues, global SfM approaches [WS14; CT15; MMM13] have been proposed, which estimate all camera poses simultaneously, offering high efficiency. However, they are sensitive to initialization and outliers. Recently, high-performance methods such as GLOMAP [PBPS24] have emerged, achieving both the accuracy of incremental SfM and the efficiency of global SfM. Additionally, SfM pipelines leveraging learning-based local features, including L2-Net [TFW17] and SuperPoint [DMR18] [FKW*19], have also been actively studied. Furthermore, methods that leverage both monocular depth and surface normals enable more robust camera pose estimation, even in low-texture regions or under wide-baseline conditions [PSSP25]. However, due to the lack of reliable correspondences in a turntable setup, even state-of-the-art SfM methods may result in unstable camera pose estimation or the reconstruction of incorrect geometry. In this work, to eliminate the impact of SfM preprocessing on the overall optimization pipeline, we propose an approach that reconstructs objects on a turntable without relying on SfM.

2.3. Joint Reconstruction and Camera Pose Estimation

Several studies have proposed jointly optimizing 3D representations and camera poses [YFB*21; WWX*21; LMTL21; CRSL22; BWL*23]. NeRFmm [WWX*21] assigns learnable parameters to each frame, enabling the simultaneous optimization of scene representation and camera poses. In addition, Nope-NeRF [BWL*23] introduces depth priors to address the limitation that relying solely on RGB loss fails to ensure stable optimization under large camera motions. However, as with other NeRF-based methods, these approaches suffer from excessively long training and inference times. Since both the scene representation and camera poses must be optimized, reconstructing a single scene can take several days. To overcome these limitations, recent works have proposed approaches based on 3D Gaussian Splatting (3DGS) [KKLD23; FLK*24; JY25; JFV*24; SPL24; SCF*25; GWK*24]. For instance, CF-3DGS [FLK*24] proposes a method that assumes small inter-frame camera motion and estimates relative camera poses between consecutive frames. In contrast, HT-3DGS [JY25] employs video frame interpolation [KJL*22] to reduce camera motion in cases of large inter-frame movement. In addition, COGS [JFV*24] leverages feature matching across images to provide geometric cues, enabling stable optimization even under sparse viewpoints. Nevertheless, these approaches still require the individual estimation of camera poses for every frame. Consequently, despite the fast training speed of 3DGS, optimization time remains lengthy, and camera pose estimation failures can occur, leading to a significant degradation in reconstruction quality. More recently, dense Simultaneous Localization and Mapping based methods [YQX*24; YLGO23; MSL*25] and pose-free feed-forward approaches [JMX*25; YLX*24; SZLP24; XGS25; HJS*24; KYP*25; YSL*25] have been explored to further accelerate inference speeds. Although these approaches offer rapid reconstruction, their output quality often remains insufficient for high-fidelity real-world object scanning. In our work, instead of estimating per-frame camera poses, we represent object rotation by optimizing a global rotation axis shared across all frames. This approach reduces the complexity of the optimization problem, enabling both faster training and higher reconstruction quality compared to existing methods.

3. Method

We propose RotGS, which leverages the prior that an object rotates around a single axis on a turntable, in which Structure-from-Motion [SF16] often fails. We first provide a brief overview of 3D Gaussian Splatting [KKLD23] in Section 3.1. In Section 3.2, we introduce the transformation that rotates the Gaussian set around a central axis and render multi-view images from a fixed camera. Section 3.3 explains how rotation flow from Gaussian motion is compared with optical flow to provide strong geometric signals. In Section 3.4, we present uncertainty-to-detail flow scheduling, which enables the use of rotation flow even during the initial training stage when the Gaussian geometry is still inaccurate. Finally,

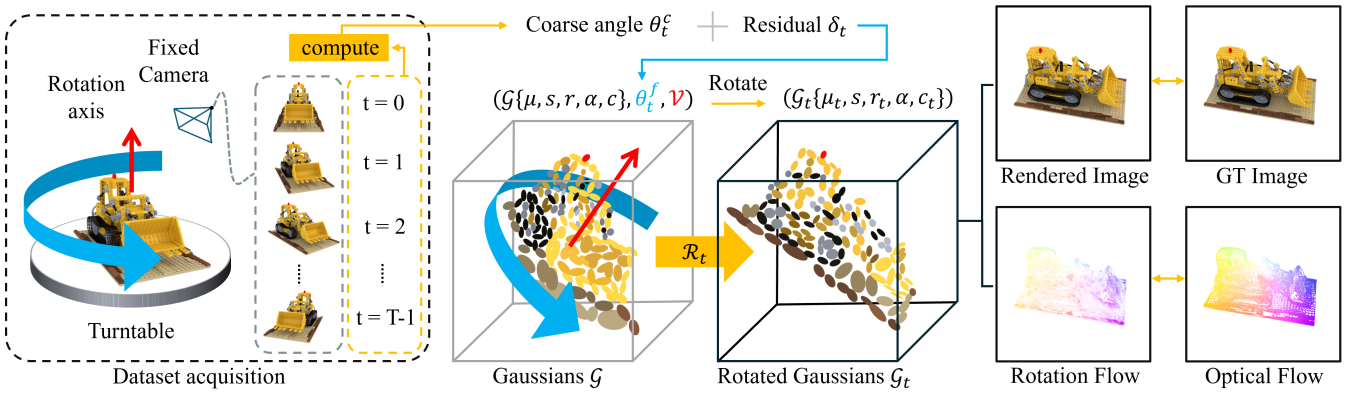


Figure 3: ***Overview of RotGS.*** *We acquire a turntable sequence by capturing an object on a turntable with a fixed camera. For each image, calculate the coarse angle θ_t^c at timestep t, then add the per-frame residual δ_t (the small rotation angle error at each timestep) to obtain the fine angle θ_t^f. Then, the Gaussian $\mathcal{G}$ is rotated using the fine angle and the global axis $\mathcal{V}$ to get the rotated Gaussian $\mathcal{G}_t$. The rendered image and the rotation flow induced during the rotation process are used to compute the loss, which is then backpropagated to jointly optimize the global axis, the residual, and Gaussian attributes.*

Section 3.5 introduces the extension of RotGS to multi-camera systems.

3.1. Preliminary: 3D Gaussian Splatting

3D Gaussian Splatting (3DGS) [KKLD23] represents a scene as a set of explicit 3D Gaussians, in contrast to implicit representations such as NeRF [MST*21]. A single Gaussian $G(x)$ is defined as:

$$G(x) = \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right). \tag{1}$$

In 3DGS, each Gaussian G is parameterized by its centroid position $\mu \in \mathbb{R}^3$, rotation $r \in \mathbb{R}^4$, scale $s \in \mathbb{R}^3$, opacity $\alpha \in [0,1]$, and color represented by Spherical Harmonics (SH) coefficients $\mathbf{c} \in \mathbb{R}^{3(l+1)^2}$ ($G = \{\mu, s, r, \alpha, c\}$), where l denotes the degree of the SH representation. Each Gaussian is initialized based on a sparse point cloud obtained from Structure-from-Motion (SfM) [SF16], and it may subsequently be subdivided or duplicated during the optimization process to achieve a more detailed scene representation. These individual Gaussians form a Gaussian set $\mathcal{G} = \{G_0, G_1, \ldots, G_N\}$, where N denotes the total number of Gaussians. The covariance matrix of a Gaussian projected onto the camera coordinate is computed using the view transformation W and the jacobian J as follows:

$$\Sigma^{2D} = JW\Sigma W^\top J^\top. \tag{2}$$

During the rendering process, all Gaussians are sorted by depth relative to the camera, and the final pixel color is determined through alpha blending:

$$C = \sum_{i=1}^{N} c_i \alpha_i \prod_{j=1}^{i-1}(1-\alpha_j), \tag{3}$$

where opacity α is used as the weight for the specific Gaussian contributing to the final pixel. This 3DGS rendering process is significantly faster than NeRF's volume rendering while providing high visual quality, enabling real-time rendering. We leverage the advantages of 3DGS to perform efficient 3D reconstruction in a turntable setup by applying direct rotational transformations to Gaussians.

3.2. Joint Rotation and Gaussian Optimization

Our method overview is illustrated in Figure 3. In this work, we use unposed images obtained from a single turntable rotation (Figure 3, left), where backgrounds are removed using Rembg [Gat]. We estimate the rotation axis shared across all frames and the rotation angle of each frame, and use this to rotate the Gaussian set to render multi-view images. In this study, which does not use an SfM-based approach, the camera poses are fixed to the identity matrix, $\mathbf{I}$, for all frames, and the Gaussian set is initialized from a random point cloud. The rotation axis, $\mathcal{V}$, is represented by position $\mathbf{a} \in \mathbb{R}^3$, and direction, $\mathbf{v} \in \mathbb{R}^2$, initialized as the scene center and the camera's up vector. Since the turntable rotates at constant speed, the coarse rotation angle, θ_t^c, for each frame can be calculated using the frame's timestep t ($\theta_t^c = 2\pi \cdot \frac{t}{T}, t = 0, 1, \ldots, T-1$). Here, T denotes the total number of images. However, in real-world capture scenarios, obtaining an image set in which the object has rotated exactly once is challenging. Therefore, when rotating a Gaussian set using only coarse rotation angles, small residuals occur, resulting in the rendering of images with slightly different views. To correct this, the fine rotation angle, θ_t^f, is estimated using the residual, δ_t, from each timestep ($\theta_t^f = \theta_t^c + \delta_t$). δ_t is obtained via linear interpolation over time of a single learnable parameter Ω ($\delta_t = \Omega \cdot \frac{t}{T}$).

Gaussian Rotation. During the optimization process, an explicit rotation transformation, $\mathcal{R}$, is applied to the 3D Gaussian set, $\mathcal{G}$, to obtain the Gaussian set, $\mathcal{G}_t$, corresponding to each timestep (Figure 3 middle). The attributes of a Gaussian affected by the rotation transformation include its position, μ, rotation, r, and SH coefficients, c. Since these attributes are explicitly defined, they can be directly rotated. Specifically, the global quaternion, $\mathbf{q}$, is computed using $\mathcal{V}$ and θ_t^f estimated by the model, and the following operations are applied to μ and r to obtain the postion, μ_t, and rotation,

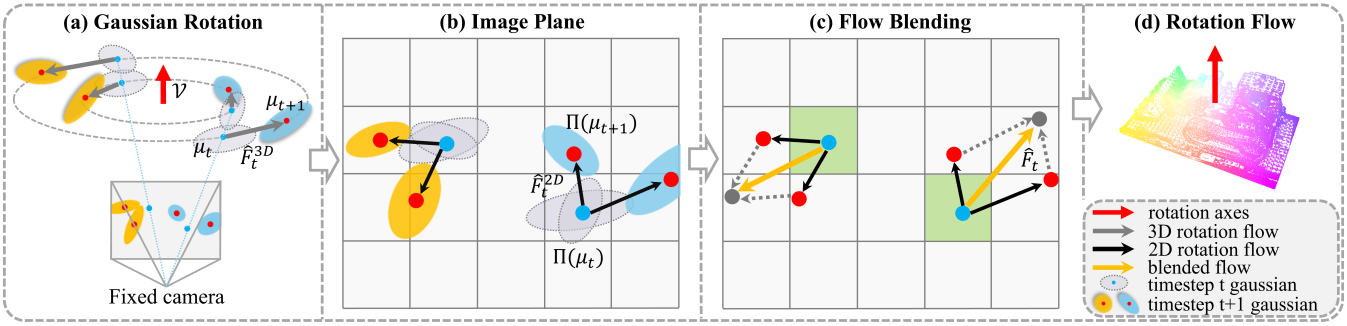


Figure 4: ***Rotation flow computation.*** *Gaussians are rotated according to the estimated rotation between consecutive timesteps, and the resulting 3D rotation flow is computed from the positional changes. The flow is then projected onto the image plane to obtain a 2D rotation flow. Per-pixel blended flow is generated by alpha blending the 2D flows of individual Gaussians, and a background mask is applied to remove irrelevant regions. The final rotation flow is compared with the optical flow to provide geometric supervision.*

r_t, corresponding to timestep, t:

$$\boldsymbol{\mu}_t = \mathcal{R}_\mu(\mathcal{V}, \theta_t^f, \boldsymbol{\mu}) = \mathbf{q} \cdot (\boldsymbol{\mu} - \mathbf{a}) \cdot \mathbf{q}^{-1} + \mathbf{a}, \tag{4}$$

$$\mathbf{r}_t = \mathcal{R}_\mathbf{r}(\mathcal{V}, \theta_t^f, \mathbf{r}) = \mathbf{q} \cdot \mathbf{r} \cdot \mathbf{q}^{-1}. \tag{5}$$

For rotating positions, considering cases where the object's rotation axis is not centered on the scene, the Gaussian set is first translated to the origin of world space relative to the estimated rotation axis center, $\mathbf{a}$, then rotated, and finally translated back to its original position. Similar to position and rotation, the SH coefficient, c, is direction-sensitive and must be rotated along with the Gaussian rotation. Accordingly, the lth level SH coefficient, c_t^l, at timestep, t, is given by:

$$\mathbf{c}_t^l = \mathcal{R}_\mathbf{c}(\mathcal{V}, \theta_t^f, \mathbf{c}^l) = \left\{ \mathcal{D}^l(\mathbf{q}) \cdot \mathbf{c}^l \;\middle|\; l = 1, 2, \ldots \right\}. \tag{6}$$

Here, $\mathcal{D}^l$ denotes the Wigner-D matrix [Wig12], which represents the rotation transformation of the lth spherical harmonic basis. These transformations are implemented in a differentiable manner, allowing integration into the end-to-end optimization process. Finally, the obtained set of parameters, $\mathcal{G}_t = \{\boldsymbol{\mu}_t, \mathbf{s}, \mathbf{r}_t, \boldsymbol{\alpha}, \mathbf{c}_t\}$, is rendered at the fixed camera pose, $\mathbf{I}$, and compared with the ground-truth (GT) image to jointly optimize the Gaussian set, $\mathcal{G}$, rotation axis, $\mathcal{V}$, and rotation angle residual, Ω (Figure 3 right).

3.3. Rotation Flow

3DGS is optimized solely with an RGB loss. However, because the RGB loss provides gradients only from pixel-level photometric similarity, it does not explicitly capture the geometric motion induced by rotation. As a consequence, the model may converge to an incorrect rotation axis during the initial stages of training, or the Gaussians may overfit while the rotation axis itself remains unoptimized. In our approach, inspired by previous work [GXC*24; ZLC*24], we propose a rotation flow derived from the motion of a rotating 3D Gaussian. By comparing this rotation flow with optical flow [XZC*22], we provide direct geometric supervision for rotation. Previous approaches based on optical flow do not model Gaussian motion as a rigid body transformation. As a result, they must

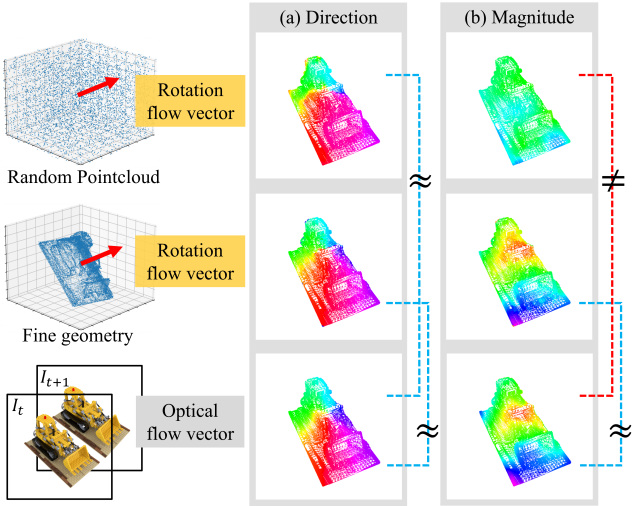


Figure 5: ***Motivation for UDFS.*** *Applying rotation with an accurate axis (red arrow) shows that while the rotation flow magnitude depends heavily on geometry, the flow direction remains reliable even for random point clouds.*

account for both positional changes and variations in scale and rotation attributes. Moreover, optical flow is influenced by both object motion and camera motion, which requires additional computation to disentangle the two. In contrast, our approach assumes a fixed camera and a rotating object, so the optical flow between time t and $t+1$ can be fully explained by the rigid transformation induced by the object's rotation. Therefore, the rotation flow can be calculated solely based on changes in the Gaussian position, thereby significantly reducing computational complexity. Specifically, at timestep t, when the Gaussian position μ is rotated by the rotation axis $\mathcal{V}$ and rotation angle θ_t^f estimated by the model (Eq. 4), the 3D flow occurs (Figure 4 (a)):

$$\hat{\mathbf{F}}_t^{3D} = \mu_{t+1} - \mu_t. \tag{7}$$

Projecting this 3D flow into the pixel space of a fixed camera allows us to obtain the corresponding 2D flow for each Gaussian (Figure 4 (b)):

$$\hat{\mathbf{F}}_t^{2D} = \Pi(\mu_{t+1}) - \Pi(\mu_t), \tag{8}$$

where Π is the transformation that projects a 3D point into the camera's pixel space. Similarly to using alpha blending to calculate the final pixel color in 3DGS [KKLD23] (Eq. 3), the 2D flow is weighted averaged based on each Gaussian's contribution to the pixel to compute the blended flow (Figure 4 (c)):

$$\hat{\mathbf{F}}_t = \left(\sum_{i=1}^{N} \hat{\mathbf{F}}_{i,t}^{2D} \alpha_i \prod_{j=1}^{i-1} (1-\alpha_j) \right) \odot \mathbf{M}_t, \tag{9}$$

Then, to ensure only the flow caused by the object's motion is used for supervision, we apply a background mask [Gat] $\mathbf{M}_t$ to get the final rotation flow (Figure 4 (d)). We define the flow loss by comparing the rotation flow $\hat{\mathbf{F}}_t$ with the optical flow $\mathbf{F}_t$ obtained from a pre-trained model [XZC*22].

3.4. Uncertainty-to-Detail Flow Scheduling

In this work, the Gaussian is initialized with a random point cloud, causing the geometric structure of Gaussian points to differ significantly from the target object during the initial training stages. This inaccurate geometry can induce erroneous supervision signals in the rotation flow-based loss. Therefore, we propose uncertainty-to-detail flow scheduling (UDFS) to enable effective flow supervision even with inaccurate geometry during the early training phase. Rotation flow (Section 3.3) is computed by rotating the position of each Gaussian based on the estimated rotation axis and angle. This provides an important clue reflecting how the model estimates rotation, even if the Gaussian's geometric structure differs from the actual one (Figure 5). In particular, the direction of the rotation flow vector can be utilized as a crucial signal to supervise the accuracy of the rotation axis estimated by the model when the Gaussian set has an inaccurate geometric structure (Figure 5 (a)). On the other hand, the magnitude of the flow vector is sensitive to the accuracy of the Gaussian geometry, and using it directly may provide incorrect supervision signals to the model (Figure 5 (b)). Accordingly, during the early stages of training, a cosine similarity loss considering only the flow direction is used:

$$\mathcal{L}_{\text{flow-dir}} = 1 - \frac{\hat{\mathbf{F}}t \cdot \mathbf{F}t}{|\hat{\mathbf{F}}t| \cdot |\mathbf{F}t|}. \tag{10}$$

As training progresses and the geometry of the Gaussian becomes more precise, we transition to an L1 loss that includes the magnitude of the flow:

$$\mathcal{L}_{\text{flow-all}} = \|\hat{\mathbf{F}}_t - \mathbf{F}_t\|_1. \tag{11}$$

The final flow loss $\mathcal{L}_{\text{flow}}$ is defined as follows, applying an exponential weight $\lambda(k)$ according to the training iteration k:

$$\mathcal{L}_{\text{flow}} = \lambda(k) \cdot \mathcal{L}_{\text{flow-dir}} + (1-\lambda(k)) \cdot \mathcal{L}_{\text{flow-all}}, \tag{12}$$

$$\lambda(k) = \exp\left(-\frac{k}{\tau}\right). \tag{13}$$

Here, τ is a hyperparameter controlling the scheduling. At the beginning of training, when the Gaussian set's geometric structure is inaccurate, flow loss relies primarily on the flow direction. As the geometry gradually becomes more accurate, the flow magnitude is incorporated as well, allowing the model to learn more detailed information. Finally, the total loss is defined as the weighted sum of the RGB loss and the flow loss:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{rgb}} \cdot \mathcal{L}_{\text{rgb}} + \lambda_{\text{flow}} \cdot \mathcal{L}_{\text{flow}}. \tag{14}$$

3.5. Expanding to Multi-Camera System

To reconstruct the complete 3D shape of an object, including its top and bottom surfaces, a multi-camera setup with N cameras fixed at various heights is required. In this system, each camera $n \in \{0, 1, \ldots, N-1\}$ captures a distinct image sequence from its fixed position. To maintain our rotation-based framework, we set all camera poses to the Identity matrix $\mathbf{I}$. This implies that the physical viewpoint differences caused by varying camera heights are instead modeled by the unique orientation of the rotation axis $\mathcal{V}_n$ within each camera's local coordinate system. To handle these viewpoint differences, we introduce an axis alignment stage before the Gaussian rotation described in Section 3.2. Axis alignment is the process of transforming the Gaussian set $\mathcal{G}$ to correspond to the n-th camera's viewpoint by aligning its rotation axis $\mathcal{V}_n$ with a reference axis. (the rotation axis of the 0-th camera). Subsequent steps are the same as for a single camera system: rotate the Gaussian set around the estimated rotation axis to render the multi-view image. At timestep t for the nth camera, the fine angle $\theta_{n,t}^f$ is computed by adding the per-frame residual to the coarse angle. The per-frame residual is obtained by interpolating the residual of the currently rendering camera Ω_n over time. The coarse angle shares the same value across all cameras. By comparing the rendered image with the GT image and aligning the rotation flow from Gaussian rotation with the optical flow, we jointly optimize the per-camera rotation axes $\mathcal{V}_n$, residual Ω_n, as well as the Gaussian attributes $\mathcal{G}$ shared across all cameras. The transformation during the axis alignment stage is implemented the same way as the Gaussian rotation explained in Section 3.2, enabling easy extension of RotGS to the multi-camera system.

4. Experiments

4.1. Implementation Details

In our implementation, most hyperparameter settings follow 3D Gaussian Splatting (3DGS) [KKLD23]. The camera is fixed at $(0,0,d)$ facing the world origin $(0,0,0)$, where $d=5$ is an arbitrary number. The initial point cloud is randomly sampled within a cube centered at the origin, with an edge length is d. Considering training speed and efficiency, only the $k=5$ Gaussians closest to the camera are used to compute the rotation flow for each pixel. Additionally, when calculating the flow loss, the full-resolution flow map is downsampled by a factor of 4. Specifically, we obtain a single rotation flow via bilinear interpolation on a 4×4 region of the original flow map and compute the flow loss by comparing it to the equally downsampled optical flow [XZC*22]. The values of k and the resolution of the flow map are adjustable factors that control the trade-off between computational efficiency and accuracy. All experiments were performed on an NVIDIA RTX 3090 GPU.

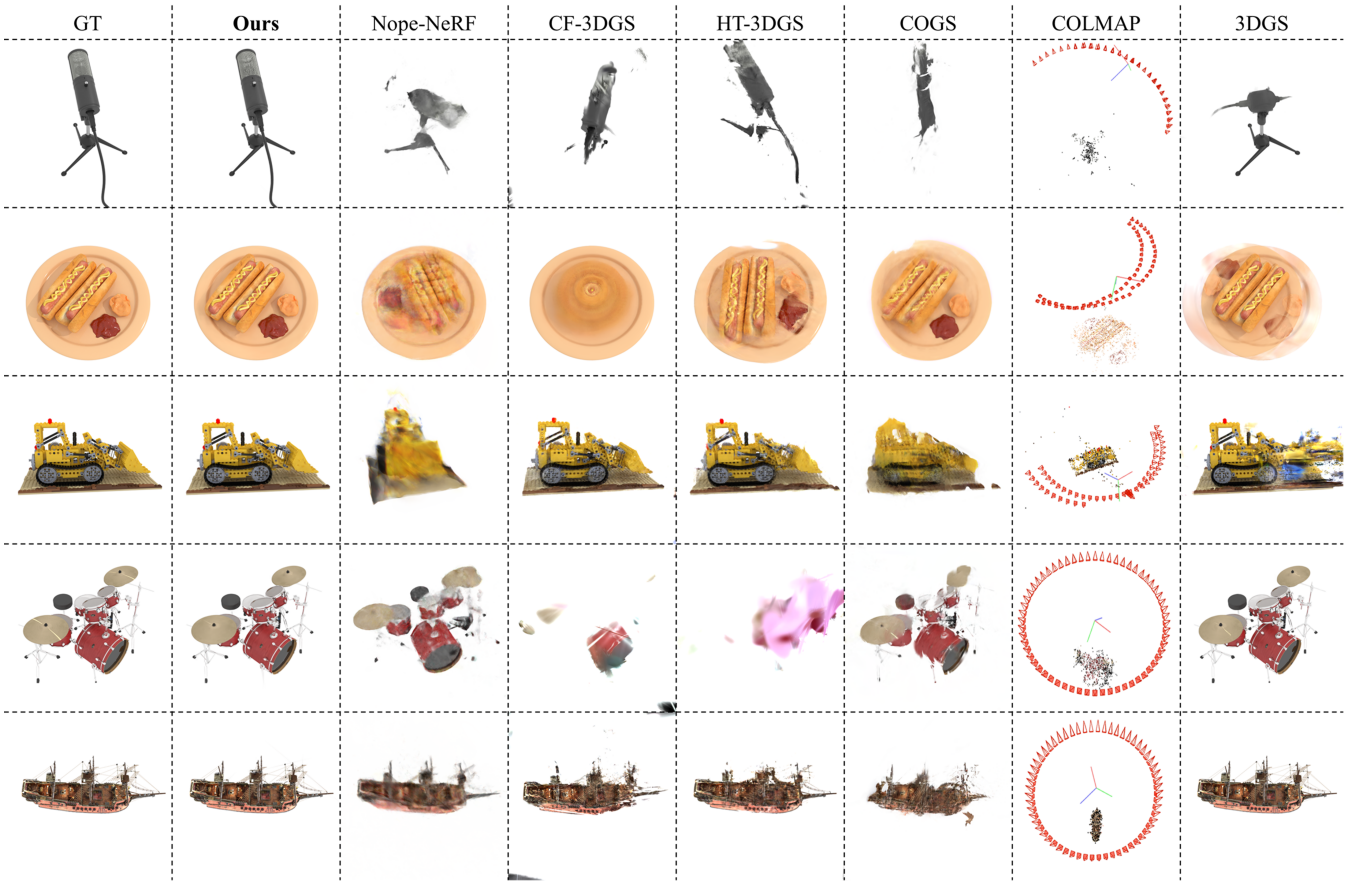


Figure 6: ***Qualitative Comparison with Optimization-based*** *Methods. Comparison of Unposed Methods (Nope-NeRF, CF-3DGS, HT-3DGS, COGS), 3DGS with SfM, and our method. Column 7 visualizes camera poses and point clouds from COLMAP. Rows 1–3 show cases where SfM failed, and rows 4–5 show cases where SfM succeeded. Unposed Methods fail to estimate accurate camera poses, resulting in artifacts. 3DGS performs well when SfM succeeds but with degraded quality when SfM fails. Our method achieves high-quality reconstruction in all cases by accurately estimating the turntable rotation.*

4.2. Datasets

Synthetic Datasets. To construct synthetic datasets that mimic real-world turntable setup while enabling quantitative analysis, we utilized the 3D models from the NeRF Synthetic Dataset [MST*21]. Using Blender, we rotated the 3D model at a constant speed around a single axis and rendered images at regular intervals from a fixed camera position. To reproduce the residual between coarse and fine rotation angles, Gaussian noise with a mean of 0 and standard deviation of 5 is added to the fine rotation angle. For example, if the fine rotation angle is 360°, the coarse rotation angle is randomly set to 355°, requiring the model to estimate a total residual of 5°. For the rotation axis, three different rotation axes (25°, 45°, 75°) were set for each 3D model to evaluate the model's generalization performance under various rotation conditions.

Real Datasets. Real-world datasets were acquired using a rotating turntable and tripod. Twenty-four different objects were placed on the ComXim turntable V1.0. The turntable was rotated at a constant speed and image sequences were captured from a fixed position using an iPhone 15 Pro mounted on a tripod. When the user determined that the object had completed one full rotation, they stopped the turntable. During this process, a residual occurred between the coarse and the fine rotation angle. We measured the residual manually and found it to be between 5° and 10°. The camera was positioned at the location where each object was best visible, resulting in each object having a different rotation condition.

4.3. Comparison with Unposed Methods

Per-scene Optimization. We compared existing optimization-based methods [BWL*23; FLK*24; JY25; JFV*24] and our proposed method based on qualitative and quantitative metrics. The comparison includes standard image quality metrics such as PSNR [HZ10], SSIM [WBSS04], LPIPS [ZIE*18], and total training time. For unposed methods, since the camera pose of the test view is unknown, we followed the procedure proposed in [WWX*21; BWL*23]. In detail, we initialized the test view camera pose with the closest train view, then minimized photometric error while freezing the trained NeRF [MST*21] or 3DGS [KKLD23] model to estimate the test view camera pose. In our method, the fine angle

Table 1: ***Quantitative comparison with unposed methods.***

	PSNR↑	SSIM↑	LPIPS↓	Times↓
Ours	**34.02**	**0.975**	**0.033**	15m
Nope-NeRF	18.52	0.828	0.241	11h
CF-3DGS	18.32	0.830	0.187	26m
HT-3DGS	19.60	0.847	0.165	34m
COGS	20.19	0.853	0.152	6h
AnySplat	19.27	0.830	0.364	**1m**

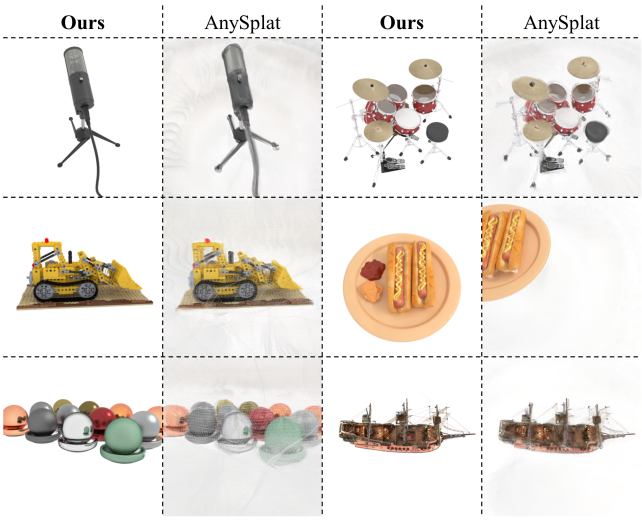


Figure 7: ***Qualitative Comparison with Feed-forward Methods.*** *Comparison of Unposed feed-forward method (AnySplat) and ours.*

is calculated by interpolating the trained residual to the corresponding timestep and adding it to the coarse angle to obtain the rotation for the test view. Subsequently, the Gaussian set is rotated using the trained axis and fine angle to render the test view. As shown in Table 1, our method consistently outperforms existing methods across all metrics. Particularly, in the qualitative comparison results shown in Figure 6, existing methods produce artifacts due to inaccuracies in camera pose estimation, whereas the proposed method achieves precise and realistic 3D reconstruction by accurately estimating the object's rotation axis and angle. Furthermore, our method offers advantages in terms of training efficiency. Previous approaches require estimating camera parameters for each frame, resulting in a large number of optimization targets and thus being disadvantageous in terms of training stability and speed. In contrast, our approach sets the rotation axis and rotation angle residuals as global parameters shared across all frames, thereby reducing the complexity of the overall optimization process and showing improvements in both training stability and time.

Feed-forward method. We compare our approach with AnySplat [JMX*25], an advanced pose-free feed-forward method. As shown in Tab. 1, while AnySplat achieves faster inference through its one-step architecture, our method provides significantly higher reconstruction quality. This performance gap stems from the fundamental difference between generalized prediction and per-scene optimization. Feed-forward models attempt to regress 3DGS parameters in a single pass, which often results in "averaged" outputs that lack high-frequency details. This leads to the oversmoothed textures and blurred edges shown in Fig. 7. In contrast, our iterative optimization refines Gaussian attributes to fit the specific geometry and appearance of each object. Although more time-consuming, our approach captures fine-grained textures and complex materials that a single forward pass struggles to represent. Furthermore, by incorporating a rotation-guided constraint, we provide a robust geometric prior, ensuring superior multi-view consistency and sharper details in turntable sequences.

Table 2: ***Quantitative comparison of our method and 3DGS on SfM success/failure cases.*** *The criteria for SfM success and failure are described in Sec. 4.4.*

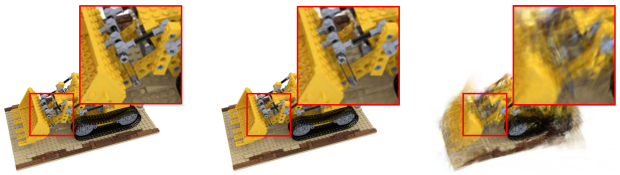
Data	Model	PSNR↑	SSIM↑	LPIPS↓
SfM failure	Ours	**33.86**	**0.977**	**0.027**
SfM failure	3DGS	23.73	0.937	0.075
SfM success	Ours	34.40	0.976	0.036
SfM success	3DGS	**35.23**	**0.979**	**0.026**

Table 3: ***Ablation study.*** *Comparison of our full method with variants excluding axis optimization, residual refinement, flow loss, or UDFS.*

	PSNR↑	SSIM↑	LPIPS↓
Ours	**34.02**	**0.975**	**0.033**
w/o axis optimization	22.94	0.877	0.142
w/o residual refinement	30.17	0.946	0.063
w/o flow loss	32.62	0.966	0.044
w/o UDFS	30.19	0.947	0.066

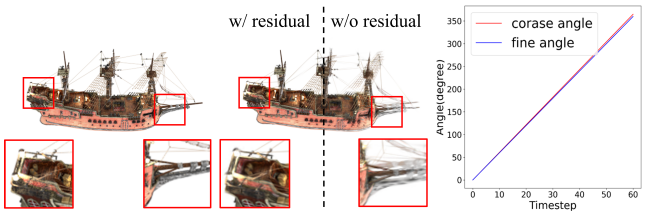
4.4. Comparison with 3D Gaussian Splatting using SfM

We compared our method with 3DGS using SfM [SF16]. The datasets were classified into successful or failed SfM cases based on two conditions: (1) the number of registered images matched the number of input images, and (2) the cameras followed a consistent circular trajectory. The second condition is evaluated based on the camera centers obtained from SfM. First, principal component analysis is applied to the camera centers. If the variance along the third principal component is less than 1% of the total variance, the camera centers are considered to lie approximately on a single plane. When planarity is confirmed, each center is projected onto this plane, and the mean of the projected 2D coordinates is taken as the circle center. If the distances from the projected points to the circle center have a coefficient of variation of 0.05 or less, the cameras are considered to follow a circular trajectory, and SfM succeeds. If either of the two conditions (complete registration of all views and circular trajectory) is not satisfied, the data is classified as a SfM failure. This enables a quantitative comparison between ours and 3DGS using SfM. In Table 2, we present a quantitative comparison by separating the datasets into cases where SfM suc-



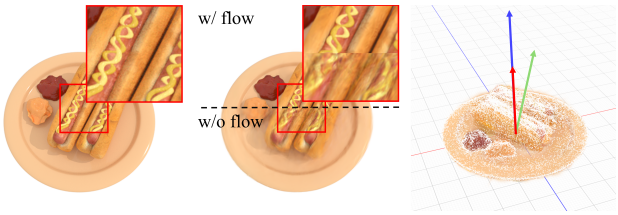
(a) Ground Truth (b) w/ axis optimizaiton (c) w/o axis optimization

Figure 8: ***Ablation study: axis optimization.*** *Gaussian overfits to the incorrect rotation axis, causing artifacts.*



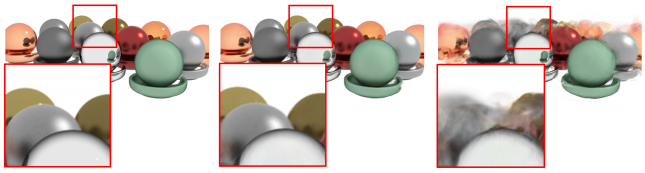
(a) Ground Truth (b) residual ablation (c) angle residual

Figure 9: ***Ablation study: residual refinement.*** *The red line in (c) represents the coarse angle, the blue line represents the fine angle. Although the difference is only about 5 degrees, the reconstruction quality significantly decreases when residual refinement is not applied ((b) right).*



(a) Ground Truth (b) flow ablation (c) axes

Figure 10: ***Ablation study: flow loss.*** *(c) Arrows indicate rotation axes: blue for the axis with flow loss, red for the GT axis, and green for the axis without flow loss. Without the flow loss, the model-estimated rotation axis may fail to reach the GT axis.*



(a) Ground Truth (b) w/ UDFS (c) w/o UDFS

Figure 11: ***Ablation study: Uncertainty to detail flow scheduling.*** *Without UDFS, incorrect supervision signals during the initial training stage result in reduced reconstruction quality.*

ceeded and failed. For a fair comparison, the SfM failure data used in the quantitative comparison was the case that cameras did not follow an exact circular trajectory, although all cameras were registered. Therefore, the number of images used in the training process for both 3DGS and ours is the same. In the case of SfM failure, 3DGS optimized with incorrect camera poses may show artifacts such as overlapping objects or multiple object appearances, or render from entirely incorrect views (Figure 6 (row 1-3)). Additionally, image registration failure leads to fewer images being used for training than the actual number of images, resulting in lower reconstruction quality. In contrast, ours can render accurate views by precisely estimating rotation, demonstrating superior qualitative and quantitative performance compared to 3DGS models optimized with incorrect camera poses. Comparisons on datasets where SfM succeeded are shown in Figure 6 (rows 4–5) and Table 2. For datasets where SfM succeeded, both 3DGS and our method demonstrated high-quality reconstruction. Quantitatively, our method differed by only approximately 0.8 dB from 3DGS trained with accurate SfM camera poses. This demonstrates that our SfM-free approach robustly estimates the rotation axis and angle, achieving reconstruction quality comparable to 3DGS trained with precise camera poses.

4.5. Ablation Study

Global Axis Optimization. To demonstrate the importance of rotation axis alignment, we conduct an experiment where the rotation axis is initialized as the camera's up vector and then fixed. If the rotation axis is not optimized, the reconstruction quality is significantly degraded (Figure 8 (c)). This is because the Gaussian set rotates in an incorrect direction due to the wrong rotation axis, but the Gaussian attributes were overfitted to make a rendered image as similar as possible to the GT image.

Residual Refinement. To evaluate the effectiveness of residual refinement, we conducted an experiment where Gaussians were rotated using only coarse rotation angles without optimizing the residual. In this case, the residual present in each frame causes the image to be rendered from slightly different views, significantly negatively impacting reconstruction. For example, as shown in Figure 9 (c), even when the fine angle and coarse angle had a difference of approximately 5°, a significant performance difference of about 4 dB was observed based on PSNR, and visually, the results showed low quality (Figure 9 (b)). The rotation angle residual is represented as a single learnable parameter, requiring only a small computational overhead. However, it significantly impacts reconstruction performance.

Rotation Flow loss. To validate the effectiveness of rotation flow loss, experiments were conducted by removing the flow loss term. Using only the RGB loss provided insufficient gradients for proper alignment of the rotation axis, resulting in cases where optimization of the rotation axis stalled midway (Figure 10 (c)). In contrast, incorporating the flow loss supplied stronger geometric supervision. This mitigated the issue of the rotation axis falling into local minima and enabled more accurate estimation of the rotation axis.

Uncertainty-to-Detail Flow Scheduling UDFS, proposed to effectively utilize flow loss even in the initial training stages when the Gaussian geometric structure is inaccurate, is important for ensuring training stability. To verify this, the model was trained from

the beginning using both the direction and magnitude of the rotation flow vector for supervision. As shown in Figure 11, removing UDFS resulted in a significant performance drop. This occurred because the vector magnitude of the rotation flow obtained from the inaccurate geometry of the Gaussian set provided incorrect supervision signals to the model, causing the rotation axis to misalign in the wrong direction.

5. Limitations and Future Work

Our method also has limitations. In turntable-based object capturing, the illumination remains fixed while the object rotates, which causes variations in surface brightness across frames [FLY*25]. This violates the illumination-invariance assumption underlying standard Novel View Synthesis methods, causing the same surface point to be observed with different colors across frames. As a result, computing the RGB loss provides incorrect supervision signals to the model, potentially degrading reconstruction quality. This can be addressed by integrating our method with previous work [FLY*25], which optimizes a neural radiance representation conditioned on light rotations. Additionally, our work assumes known camera intrinsics and a turntable rotating at a constant speed. Consequently, reconstruction quality may suffer when camera intrinsics are inaccurate or when hardware defects induce variations in the turntable's rotational speed. It would be valuable to address these limitations in future work, by optimizing the camera's intrinsic parameters together with the rotation axis, angle, and Gaussian attributes, and by modeling per-frame rotation residuals as learnable parameters to compensate for differences in the turntable's rotation speed. Furthermore, our pipeline leverages off-the-shelf models for background masks [Gat] and optical flow [XZC*22]. Although these models generally provide stable results, our works' performance can be affected by the quality of their outputs. Future research could focus on enhancing robustness to potential inaccuracies in these models. These extensions are expected to establish RotGS as a standardized pipeline for turntable-based 3D reconstruction.

6. Conclusion

In this paper, we propose a novel framework for optimizing high-quality 3D Gaussian Splatting (3DGS) that does not rely on Structure-from-Motion (SfM) in a turntable setup. By leveraging the prior knowledge that an object rotates around a single axis, we show that the Gaussian attributes, rotation axis, and rotation angle can be jointly optimized. Furthermore, by aligning the rotation flow, which is induced by Gaussian motion, with the optical flow, we provide the model with strong geometric supervision. Through uncertainty-to-detail flow scheduling, this supervision can be effectively applied even in the initial stages of training, when the geometry of the Gaussian set is inaccurate. Experimental results on both synthetic and real datasets demonstrate that the proposed method outperforms existing unposed methods in terms of reconstruction quality and training efficiency, and even achieves qualitatively and quantitatively comparable performance to 3DGS optimized with precise camera poses obtained via SfM.

Acknowledgements

This work was supported by the National Research Foundation of Korea (No. RS-2024-00348094) grant funded by the Korea government (MSIT) and the Starting growth Technological R&D Program (TIPS Program, (No. RS-2024-00553521)) funded by the Ministry of SMEs and Startups(MSS, Korea) in 2025.

References

[AFS*11] AGARWAL, SAMEER, FURUKAWA, YASUTAKA, SNAVELY, NOAH, et al. "Building rome in a day". *Communications of the ACM* 54.10 (2011), 105–112 3.

[BWL*23] BIAN, WENJING, WANG, ZIRUI, LI, KEJIE, et al. "Nope-nerf: Optimising neural radiance field with no pose prior". *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023, 4160–4169 3, 7.

[CRSL22] CHNG, SHIN-FANG, RAMASINGHE, SAMEERA, SHERRAH, JAMIE, and LUCEY, SIMON. "Gaussian activated neural radiance fields for high fidelity reconstruction and pose estimation". *European Conference on Computer Vision*. Springer. 2022, 264–280 3.

[CT15] CUI, ZHAOPENG and TAN, PING. "Global structure-from-motion by similarity averaging". *Proceedings of the IEEE international conference on computer vision*. 2015, 864–872 3.

[DMR18] DETONE, DANIEL, MALISIEWICZ, TOMASZ, and RABINOVICH, ANDREW. "Superpoint: Self-supervised interest point detection and description". *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2018, 224–236 3.

[FC04] FREMONT, VINCENT and CHELLALI, RYAD. "Turntable-based 3D object reconstruction". *IEEE Conference on Cybernetics and Intelligent Systems, 2004.* Vol. 2. IEEE. 2004, 1277–1282 3.

[FCZ98] FITZGIBBON, ANDREW W, CROSS, GEOFF, and ZISSERMAN, ANDREW. "Automatic 3D model construction for turn-table sequences". *European Workshop on 3D Structure from Multiple Images of Large-Scale Environments*. Springer. 1998, 155–170 3.

[FKW*19] FAN, BIN, KONG, QINGQUN, WANG, XINCHAO, et al. "A performance evaluation of local features for image-based 3D reconstruction". *IEEE Transactions on Image Processing* 28.10 (2019), 4774–4789 3.

[FLK*24] FU, YANG, LIU, SIFEI, KULKARNI, AMEY, et al. "Colmap-free 3d gaussian splatting". *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024, 20796–20805 2, 3, 7.

[FLY*25] FAN, JIAHUI, LUAN, FUJUN, YANG, JIAN, et al. "Free Your Hands: Lightweight Turntable-Based Object Capture Pipeline". *arXiv preprint arXiv:2503.05511* (2025) 2, 3, 10.

[Gat] GATIS, DANIEL. *rembg: Rembg is a tool to remove images background*. Version 2.0.61. URL: https://github.com/danielgatis/rembg 4, 6, 10.

[GWK*24] GUO, JIAXIN, WANG, JIANGLIU, KANG, DI, et al. "Free-surgs: Sfm-free 3d gaussian splatting for surgical scene reconstruction". *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2024, 350–360 3.

[GXC*24] GAO, QUANKAI, XU, QIANGENG, CAO, ZHE, et al. "Gaussianflow: Splatting gaussian dynamics for 4d content creation". *arXiv preprint arXiv:2403.12365* (2024) 5.

[HJS*24] HONG, SUNGHWAN, JUNG, JAEWOO, SHIN, HEESEONG, et al. "Pf3plat: Pose-free feed-forward 3d gaussian splatting". *arXiv preprint arXiv:2410.22128* (2024) 3.

[HZ10] HORE, ALAIN and ZIOU, DJEMEL. "Image quality metrics: PSNR vs. SSIM". *2010 20th international conference on pattern recognition*. IEEE. 2010, 2366–2369 7.

[JFV*24] JIANG, KAIWEN, FU, YANG, VARMA T, MUKUND, et al. "A construct-optimize approach to sparse view synthesis without camera pose". *ACM SIGGRAPH 2024 Conference Papers*. 2024, 1–11 2, 3, 7.

[JMX*25] JIANG, LIHAN, MAO, YUCHENG, XU, LINNING, et al. "Anysplat: Feed-forward 3d gaussian splatting from unconstrained views". *ACM Transactions on Graphics (TOG)* 44.6 (2025), 1–16 3, 8.

[JY25] JI, BO and YAO, ANGELA. "SfM-Free 3D Gaussian splatting via hierarchical training". *Proceedings of the Computer Vision and Pattern Recognition Conference*. 2025, 21654–21663 2, 3, 7.

[KAP*17] KHAWALDEH, SAED, ALEEF, TAJWAR ABRAR, PERVAIZ, USAMA, et al. "Complete end-to-end low cost solution to a 3d scanning system with integrated turntable". *arXiv preprint arXiv:1709.02247* (2017) 2, 3.

[KH12a] KAZÓ, CSABA and HAJDER, LEVENTE. "High-quality structured-light scanning of 3D objects using turntable". *2012 IEEE 3rd International Conference on Cognitive Infocommunications (CogInfoCom)*. IEEE. 2012, 553–557 2.

[KH12b] KAZÓ, CSABA and HAJDER, LEVENTE. "High-quality structured-light scanning of 3D objects using turntable". *2012 IEEE 3rd International Conference on Cognitive Infocommunications (CogInfoCom)*. IEEE. 2012, 553–557 3.

[KJL*22] KONG, LINGTONG, JIANG, BOYUAN, LUO, DONGHAO, et al. "Ifrnet: Intermediate feature refine network for efficient frame interpolation". *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, 1969–1978 3.

[KJR*25] KU, KIBON, JUBERY, TALUKDER Z, RODRIGUEZ, ELIJAH, et al. "SC-NeRF: NeRF-based Point Cloud Reconstruction using a Stationary Camera for Agricultural Applications". *Proceedings of the Computer Vision and Pattern Recognition Conference*. 2025, 5472–5481 2.

[KKLD23] KERBL, BERNHARD, KOPANAS, GEORGIOS, LEIMKÜHLER, THOMAS, and DRETTAKIS, GEORGE. "3D Gaussian splatting for real-time radiance field rendering." *ACM Trans. Graph.* 42.4 (2023), 139–1 2–4, 6, 7.

[KL24] KIM, HYUNJEONG and LEE, IN-KWON. "Is 3dgs useful?: Comparing the effectiveness of recent reconstruction methods in vr". *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE. 2024, 71–80 2.

[KYP*25] KANG, GYEONGJIN, YOO, JISANG, PARK, JIHYEON, et al. "SelfSplat: Pose-free and 3D prior-free generalizable 3D Gaussian splatting". *Proceedings of the Computer Vision and Pattern Recognition Conference*. 2025, 22012–22022 3.

[LMTL21] LIN, CHEN-HSUAN, MA, WEI-CHIU, TORRALBA, ANTONIO, and LUCEY, SIMON. "Barf: Bundle-adjusting neural radiance fields". *Proceedings of the IEEE/CVF international conference on computer vision*. 2021, 5741–5751 3.

[LZL*22] LI, ZINUO, ZHANG, ZHEN, LUO, SHENGHONG, et al. "An improved matting-SFM algorithm for 3D reconstruction of self-rotating objects". *Mathematics* 10.16 (2022), 2892 2, 3.

[MMM13] MOULON, PIERRE, MONASSE, PASCAL, and MARLET, RENAUD. "Global fusion of relative motions for robust, accurate and scalable structure from motion". *Proceedings of the IEEE international conference on computer vision*. 2013, 3248–3255 3.

[MSL*25] MEULEMAN, ANDREAS, SHAH, ISHAAN, LANVIN, ALEXANDRE, et al. "On-the-fly reconstruction for large-scale novel view synthesis from unposed images". *ACM Transactions on Graphics (TOG)* 44.4 (2025), 1–14 3.

[MST*21] MILDENHALL, BEN, SRINIVASAN, PRATUL P, TANCIK, MATTHEW, et al. "Nerf: Representing scenes as neural radiance fields for view synthesis". *Communications of the ACM* 65.1 (2021), 99–106 2, 4, 7.

[PBPS24] PAN, LINFEI, BARÁTH, DÁNIEL, POLLEFEYS, MARC, and SCHÖNBERGER, JOHANNES L. "Global structure-from-motion revisited". *European Conference on Computer Vision*. Springer. 2024, 58–77 2, 3.

[PLS*14] PANG, XUFANG, LAU, RYNSON WH, SONG, ZHAN, et al. "A tool-free calibration method for turntable-based 3D scanning systems". *IEEE computer graphics and applications* 36.1 (2014), 52–61 3.

[PNF*08] POLLEFEYS, MARC, NISTÉR, DAVID, FRAHM, J-M, et al. "Detailed real-time urban 3d reconstruction from video". *International Journal of Computer Vision* 78.2 (2008), 143–167 3.

[PSSP25] PATAKI, ZADOR, SARLIN, PAUL-EDOUARD, SCHÖNBERGER, JOHANNES L, and POLLEFEYS, MARC. "MP-SfM: Monocular Surface Priors for Robust Structure-from-Motion". *Proceedings of the Computer Vision and Pattern Recognition Conference*. 2025, 21891–21901 3.

[QXLH25] QIU, SHI, XIE, BINZHU, LIU, QIXUAN, and HENG, PHENG-ANN. "Advancing extended reality with 3d gaussian splatting: Innovations and prospects". *2025 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR)*. IEEE. 2025, 203–208 2.

[SCF*25] SHI, DONGBO, CAO, SHEN, FAN, LUBIN, et al. "Trackgs: Optimizing colmap-free 3d gaussian splatting with global track constraints". *arXiv preprint arXiv:2502.19800* (2025) 3.

[SF16] SCHONBERGER, JOHANNES L and FRAHM, JAN-MICHAEL. "Structure-from-motion revisited". *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, 4104–4113 2–4, 8.

[SPL24] SCHMIDT, CHRISTIAN, PIEKENBRINCK, JENS, and LEIBE, BASTIAN. "Look gauss, no pose: Novel view synthesis using gaussian splatting without accurate pose initialization". *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2024, 8732–8739 3.

[SSS06] SNAVELY, NOAH, SEITZ, STEVEN M, and SZELISKI, RICHARD. "Photo tourism: exploring photo collections in 3D". *ACM siggraph 2006 papers*. 2006, 835–846 3.

[SWLY25] SONG, JUNFANG, WANG, TENGJIAO, LEI, SHANZHONG, and YAN, ZHUYANG. "Improved PMVS 3D reconstruction assisted by union camera-turntable calibration". *AIP Advances* 15.7 (2025) 3.

[SZLP24] SMART, BRANDON, ZHENG, CHUANXIA, LAINA, IRO, and PRISACARIU, VICTOR ADRIAN. "Splatt3r: Zero-shot gaussian splatting from uncalibrated image pairs". *arXiv preprint arXiv:2408.13912* (2024) 3.

[TFW17] TIAN, YURUN, FAN, BIN, and WU, FUCHAO. "L2-net: Deep learning of discriminative patch descriptor in euclidean space". *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, 661–669 3.

[TML14] TAUBIN, GABRIEL, MORENO, DANIEL, and LANMAN, DOUGLAS. "3d scanning for personal 3d printing: build your own desktop 3d scanner". *ACM Siggraph 2014 Studio*. 2014, 1–66 2.

[WBSS04] WANG, ZHOU, BOVIK, ALAN C, SHEIKH, HAMID R, and SIMONCELLI, EERO P. "Image quality assessment: from error visibility to structural similarity". *IEEE transactions on image processing* 13.4 (2004), 600–612 7.

[Wig12] WIGNER, EUGENE. *Group theory: and its application to the quantum mechanics of atomic spectra*. Vol. 5. Elsevier, 2012 5.

[WS14] WILSON, KYLE and SNAVELY, NOAH. "Robust global translations with 1dsfm". *European conference on computer vision*. Springer. 2014, 61–75 3.

[Wu13] WU, CHANGCHANG. "Towards linear-time incremental structure from motion". *2013 International Conference on 3D Vision-3DV 2013*. IEEE. 2013, 127–134 3.

[WWX*21] WANG, ZIRUI, WU, SHANGZHE, XIE, WEIDI, et al. "NeRF–: Neural radiance fields without known camera parameters". (2021) 3, 7.

[XGS25] XU, JIALE, GAO, SHENGHUA, and SHAN, YING. "Freesplatter: Pose-free gaussian splatting for sparse-view 3d reconstruction". *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2025, 25442–25452 3.

[XZC*22] XU, HAOFEI, ZHANG, JING, CAI, JIANFEI, et al. "Gmflow: Learning optical flow via global matching". *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022, 8121–8130 2, 5, 6, 10.

[YFB*21] YEN-CHEN, LIN, FLORENCE, PETE, BARRON, JONATHAN T, et al. "inerf: Inverting neural radiance fields for pose estimation". *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2021, 1323–1330 3.

[YLGO23] YUGAY, VLADIMIR, LI, YUE, GEVERS, THEO, and OSWALD, MARTIN R. "Gaussian-slam: Photo-realistic dense slam with gaussian splatting". *arXiv preprint arXiv:2312.10070* (2023) 3.

[YLX*24] YE, BOTAO, LIU, SIFEI, XU, HAOFEI, et al. "No pose, no problem: Surprisingly simple 3d gaussian splats from sparse unposed images". *arXiv preprint arXiv:2410.24207* (2024) 3.

[YQX*24] YAN, CHI, QU, DELIN, XU, DAN, et al. "Gs-slam: Dense visual slam with 3d gaussian splatting". *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024, 19595–19604 3.

[YS15] YE, YUPING and SONG, ZHAN. "An accurate 3D point cloud registration approach for the turntable-based 3D scanning system". *2015 IEEE International Conference on Information and Automation*. IEEE. 2015, 982–986 3.

[YSL*25] YANG, JIANING, SAX, ALEXANDER, LIANG, KEVIN J, et al. "Fast3r: Towards 3d reconstruction of 1000+ images in one forward pass". *Proceedings of the Computer Vision and Pattern Recognition Conference*. 2025, 21924–21935 3.

[ZIE*18] ZHANG, RICHARD, ISOLA, PHILLIP, EFROS, ALEXEI A, et al. "The unreasonable effectiveness of deep features as a perceptual metric". *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, 586–595 7.

[ZLC*24] ZHU, RUIJIE, LIANG, YANZHE, CHANG, HANZHI, et al. "Motiongs: Exploring explicit motion guidance for deformable 3d gaussian splatting". *Advances in Neural Information Processing Systems* 37 (2024), 101790–101817 5.

[ZLSH08] ZHONG, HUANG, LAU, WS, SZE, WF, and HUNG, YS. "Shape recovery from turntable sequence using rim reconstruction". *Pattern recognition* 41.11 (2008), 3295–3301 3.