Enhancing Initialization for 3D Gaussian Splatting

Sanghyun Hahn[1], Hyounjin Kim[1*]
[1]Department of Mechanical and Aerospace Engineering, SNU
{steve0221, hjinkim}@snu.ac.kr

Abstract

Novel view synthesis is an important task in computer vision and graphics. While Neural Radiance Fields (NeRF) has been an effective approach for this task, 3D Gaussian Splatting (3DGS) achieved real-time rendering and differentiable representations. However, 3DGS is highly dependent to the input point cloud, with only limited quantitative research done on how the geometry of the input point cloud affects the performance of 3DGS. To address this gap, we propose ***Geometric Error****, which indicates the difference of geometry between the point cloud and the ground truth. Additionally, we propose* ***Combined Initialization****, which is inspired from the observation that random initialization shows good results where feature matching is challenging. This method integrates the strengths of Structure from Motion (SfM) and Random initialization, by adding random initial points to the sparse region of the SfM generated point cloud. Finally, we propose* ***Confidence-Aware Opacity Initialization****, which assigns high initial opacity to points with low reconstruction error. Our approach outperforms SfM initialization for indoor to semi-indoor scenes in novel view synthesis, while also generating more consistent quality images across different viewpoints.*

1. Introduction

3D scene reconstruction is a crucial task in computer vision and graphics, since it plays a critical role in various ranges such as augmented/virtual reality, autonomous driving, and robotics. Neural Radiance Fields (NERF) [14] is a powerful method for this task, utilizing volumetric rendering with radiance fields. Inspired by NeRF, numerous follow-up researches were made [4, 21, 22] enhancing the novel view synthesis results (NVS) through different approaches.

However, NeRF based methods have limitations regarding their computational cost and rendering speed. To address these issues, 3D Gaussian Splatting (3DGS) [9] was introduced, offering a boosted rendering speed by representing each scene with a sum of 3D Gaussians. This method supports real-time rendering and is also differentiable, allowing gradient based approaches for optimization.

Despite its advantages, 3DGS is highly dependent to its initial point cloud, typically generated by Structure from Motion (SfM) [19]. The performance of 3DGS drop significantly in areas where SfM struggles to match features, resulting in sparse or inaccurate point cloud outputs.

In this work, we introduce **Geometric Error**, which indicates the geometric difference of a point cloud to its ground truth, and investigates the effect of Geometric Error on the performance of 3DGS. Also, we discover that random initialization can outperform initialization by Structure from Motion (SfM) [19] in regions where feature matching is challenging. Inspired from this observation, we propose **Combined Initialization**, which rectifies the SfM generated point cloud by removing outliers and adding initial points to sparse regions. Furthermore, we introduce **Confidence-Aware Opacity Initialization**, which addresses the inefficiency of initializing all opacities to a single heuristic value. Our method improves the performance of 3DGS for indoor to semi-indoor scenes, while generating more consistent quality images across different viewpoints.

2. Related works

2.1. Point Cloud Distance Metrics

Point cloud distance metrics are widely used in the domains such as computer vision, robotics, and 3D reconstruction, where quantifying the difference of two point sets is crucial. Given two point clouds $P = \{p_i\}_{i=1}^{N}$ and $Q = \{q_i\}_{i=1}^{M} \subset \mathbb{R}^3$, there are multiple metrics for calculating the distance between them.

The Chamfer Distance [2] is the average distance between pairs of the nearest neighbors:

$$d_{\text{Chamfer}}(P,Q) = \frac{1}{N}\sum_{p\in P}\|p - NN(p,Q)\| + \frac{1}{M}\sum_{q\in Q}\|q - NN(q,P)\|$$

where $NN(x,P) = \text{argmin}_{x'\in P}||x - x'||$ is the nearest neighbor function.

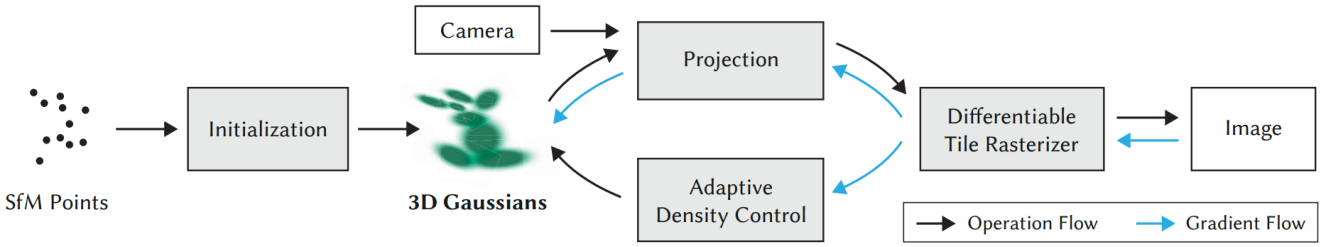

Figure 1. An overview of the 3D Gaussian Splatting pipeline. The Gaussians are initialized from Structure from Motion, then optimized through reducing projection error and Adaptive Density Control.

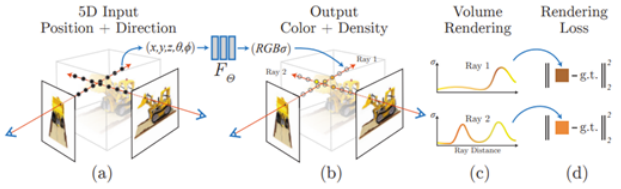

Figure 2. An overview of Neural Radiance Fields (NeRF). NeRF utilizes MLP and volumetric rendering for affective Novel View Synthesis.

The Hausdorff distance [5] is the maximum distance between any pair of the nearest neighbors:

$$d_{\text{Chamfer}}(P,Q) = \frac{1}{2}\max_{p\in P}\|p - NN(p,Q)\| + \frac{1}{2}\max_{q\in Q}\|q - NN(q,P)\|$$

The Earth Mover's distance, or Wasserstein distance [15], is the average distance between pairs according to an optimal correspondence $\pi \in \mathbb{R}^{n\times m}$:

$$d_{\text{EM}}(P,Q) = \min_{\pi\in(P,Q)} \sum_{p_i\in P}\sum_{q_i\in Q} \pi_{i,j}\|p_i - q_i\|$$

where $\pi_{i,j}$ is a number between 0 and 1 that denotes the correspondence between p_i and q_i. Each of these metrics exhibit different strengths and weaknesses, making them suitable under different conditions.

2.2. Neural Radiance Fields (NeRF)

Neural Radiance Fields [14] is a popular NVS method that exploits Multi-Layer Perceptrons (MLP) and volumetric rendering for generating novel view images. The NeRF network takes the 3D coordinate $\mathbf{x} = (x, y, z)$ and viewpoint direction $\mathbf{d} = (\theta, \phi)$ as its input and returns the color c and opacity σ as its output. NeRF renders the color of the desired viewpoint by blending the colors with opacity as a coefficient along the ray, as displayed in Fig. 2

The transmittance at depth t along the selected ray is represented as:

$$T(t) = \exp\left(-\int_{t_n}^{t}\sigma(r(s))\,ds\right) \tag{1}$$

which is the cumulative opacity from the near bound t_n to the depth t. Expanding this approach, the expected color can be written as follows:

$$C(r) = \int_{t_n}^{t_f} T(t)\sigma(r(t))c(r(t),\mathbf{d})\,dt \tag{2}$$

Various follow-up research exists which are based on the idea of neural radiance fields volumetric rendering, expanding the NeRF framework. The metrics mostly aim to achieve faster training time, or a higher quality NVS result by different ray-marching methods, encodings, or MLPs. [1, 4, 21, 22]

2.3. 3D Gaussian Splatting

3D Gaussian Splatting (3DGS) [10] represents a 3D scene with a large number of 3D Gaussians. The parameters of 3DGS are the Gaussian covariance matrix Σ, Gaussian center $\mathbf{x} = (x, y, z)$, opacity α, and color $\mathbf{c}$. The covariance matrix is represented as $\Sigma = RSS^TR^T$, where Σ, S, and R are the 3D covariance matrix, scaling matrix, and rotation matrix, respectively.

The covariance matrix and center defince each Gaussian as:

$$G(\mathbf{x}) = \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

where $\boldsymbol{\mu}$ is the center of the Gaussian.

For rendering the 3D scene as novel view images, the Gaussians are projected onto the desired 2D camera plane as an affine approximation by $\Sigma' = JW\Sigma W^TJ^T$. Here, J is the Jacobian matrix of the Camera-to-Image coordinate transform, and W is the World-to-Camera coordinate transform matrix.

The color of each Gaussian is viewpoint-dependent and represented using Spherical Harmonics (SH), which takes the viewpoint direction $\mathbf{d} = (\theta, \phi)$ as the input. Through

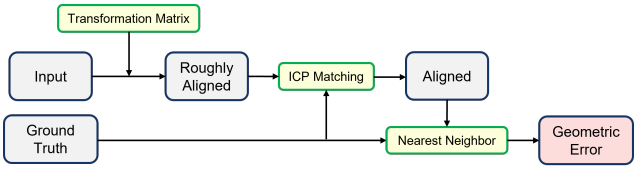

Figure 3. Pipeline of the geometric error calculation algorithm.

Method	PSNR↑	SSIM↑	LPIPS↓	Init. Points
COLMAP	28.21	0.891	0.185	74625
Random_0.01	27.63	0.881	0.203	746
Random_0.1	27.24	0.876	0.210	7462
Random_1	27.15	0.867	0.221	74625
Random_10	27.19	0.867	0.223	746250

Table 1. Comparison of SSIM, PSNR, LPIPS, and Initial Points for COLMAP and Random configurations. The colored cells highlight the best and second best results.

the training process, the degree of SH increases gradually for detailed colors. Finally, the 2D projected Gaussians are rendered onto the image plane by alpha blending, determining the color of each pixel:

$$C(\mathbf{x}) = \sum_{i \in N} c_i \alpha_i G^{2D}(\mathbf{x}) \prod_{j=1}^{i-1} (1 - \alpha_j G^{2D}(\mathbf{x}))$$

Through the optimization process, the Gaussians are adjusted using Adaptive Density Control. Redundant Gaussians are deleted, and essential Gaussians are cloned/split for handling poorly reconstructed regions and floating Gaussians that cause blurry reconstructions. The pipeline overview of 3DGS is visualized in Fig. 1.

3. Method

3.1. Geometric Error

COLMAP [17], a widely used Structure from Motion [19] library, is the most common method for initializing point clouds for 3DGS. However, since the performance of COLMAP is highly dependent to the matched features, scenes involving repeated patterns or sparse views often produce noisy and sparse point clouds. Although It is well known that the accuracy of the initial point cloud affects the performance of 3DGS greatly, no quantitative research has been made to explain this relation. Therefore, we evaluated how the geometric error of the initial point cloud - compared to ground truth point cloud - affects the novel view synthesis results.

The geometric error calculation of a point cloud given its ground truth is as follows. First, the point cloud is aligned to the ground truth through the Iterative Closest Point (ICP) [3] algorithm. Then, we exploit a single-sided

Figure 4. Comparison of Novel View Synthesis results with Random Initialization (Left) and COLMAP Initialization (Right). Random Initialization shows better reconstruction for the chairs, which is a repeated pattern where feature matching is challenging.

version of the Chamfer Distance, finding the nearest neighbor for each point on the aligned input point cloud from the ground truth. This process can be formally defined as:

$$E_{geo}(T \circ P, P_{GT}) = \frac{1}{N} \sum_{p^t \in T \circ P} \|p^t - NN(p^t, P_{GT})\|$$

where $T \in SE(3)$ denotes the transformation matrix obtained by the ICP matching algorithm, P is the reconstructed point cloud, and P_{GT} represents the Ground Truth point cloud.

For outdoor scenes, the geometric error is calculated by cropping points inside the ground truth bounding box from the reconstructed point cloud. This is necessary since the ground truth does not provide geometric information for the background. The overall pipeline of the Geometric Error calculation algorithm is visualized in Fig. 3.

3.2. Combined Initialization

Initializing 3DGS with randomly generated point clouds generally uncommon. This only happens when providing an initial point cloud through preprocessing is nearly impossible due to sparse or noisy input images. Tab. 1 shows the random initialization results for the *Church* scene from the Tanks and Temples [11] dataset, where COLMAP Initialization outperforms Random Initialization overall. However, through this experiment, we discovered that random initialization may outperform COLMAP for certain renderings.

Fig. 4 illustrates the novel view synthesis result for a frame containing repeated chairs. In renderings where the repeated patterns cover the majority of the image, Random Initialization outperformed COLMAP initialization. This is because COLMAP struggles to match features for the chairs, leading to empty regions in the initial pointcloud. In contrast, Random Initialization provides initial points near

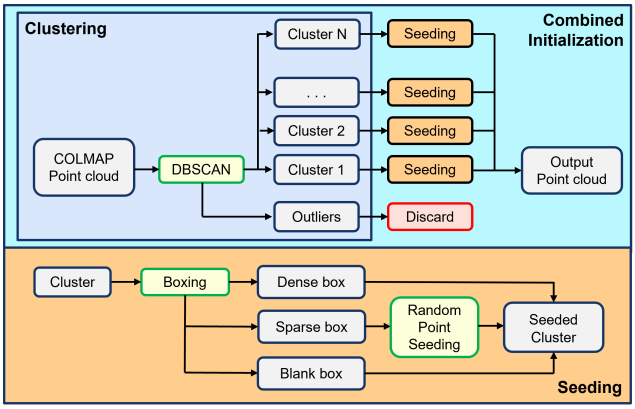

Figure 5. Pipeline of the Combined Initial point cloud generation process.

the actual chair positions, enabling better reconstruction of the chairs through Adaptive Density Control.

Inspired from this observation, we suggest a **Combined Initialization** metric that integrates COLMAP and Random Initialization. The method starts with generating a point cloud through SfM. DBSCAN [6] is applied to the SfM generated point cloud, clustering them into multiple groups. Points that belong to clusters with fewer than 0.01% of the total points are classified as noise and removed. This preprocessing step, referred to as **Clustering**, significantly reduces the geometric error, decreasing it to 15% of the original value for the *Church* scene.

After clustering, Principal Component Analysis (PCA) is applied to each cluster of the remaining point cloud, determining the three principal axes. Then, the points are projected onto these axes to calculated its outer bounding box of the cluster. The bounding box is split into 1000 boxes of the same size by dividing each edge into 10 equal segments. For each sub-box, the number of points inside the box is calculated. If a sub-box contains fewer points than 1/10000 of the total cluster point count—equivalent to 1/10 of the average point count—random points are generated within the sub-box. The number of generated points is set to 1/5 of the average cluster point count. This process, referred to as **Seeding**, generates initial points in sparsely reconstructed regions, providing ADC with a closer starting point for the cloning step. The overall pipeline of the Combined Initialization process is displayed in Fig. 5.

3.3. Confidence-Aware Opacity Initialization

The opacity of Gaussians (α) is a crucial factor that significantly influences the performance of 3DGS. In the vanilla 3DGS algorithm, the opacity of all Gaussians are initalized to 0.1. Through the optimization process, most opacities converge to values below 0.1, while a small subset approaches values near 1.0. However, because opacity values

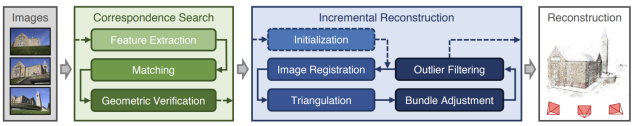

Figure 6. An overview of the Structure-from-Motion pipeline. The pipeline consists of four major steps: Feature Matching, Triangulation, Perspective-n-Point (PnP), and Bundle Adjustment (BA).

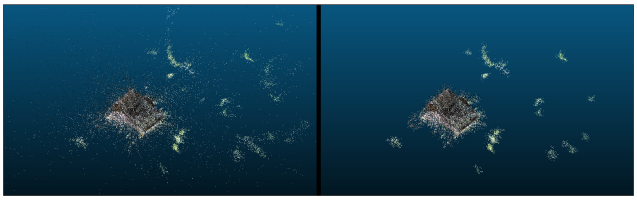

Figure 7. Comparison of the COLMAP generated point cloud (Left) and the Combined initial point cloud, rectified through the Clustering and Seeding process (Right), generated from the *Meetingroom* scene.

change only incrementally at each training iteration, initializing all values to 0.1 is inefficient. In order to resolve this issue, we suggest **Confidence-Aware Opacity Initialization** integrating the confidence of the initial points into the opacity of Gaussians.

The intuition underlying this approach is that if a point is accurately reconstructed, it should contribute strongly to the rendering process and therefore have high opacity. We initialize opacities based on the reconstruction error from the 3D reconstruction process, assigning higher opacity values (closer to 1.0) to points with lower reconstruction errors.

COLMAP follows the SfM pipeline outlined in [18], as shown in Fig. 6. The pipeline consists of four major steps: Feature Matching, Triangulation, Perspective-*n*-Point (PnP), and Bundle Adjustment (BA). Feature Matching establishes correspondences between images, Triangulation estimates depth for each point, and PnP computes camera poses. The final reconstruction error is obtained from the BA process, which jointly optimizes the 3D point coordinates and camera poses, given the initial estimates from the preceding steps.

The reprojection error of the i-th 3D point with respect to the j-th camera is defined as the Euclidean distance between the 2D ground-truth image observation $\mathbf{x}_{ij}$ and the projection of the 3D reconstructed point $\hat{\mathbf{x}}_{ij}$:

$$e_{ij} = \|\mathbf{x}_{ij} - \hat{\mathbf{x}}_{ij}\|_2 \tag{3}$$

The objective of the optimization process is to minimize the total reprojection error over all cameras and points:

$$E = \sum_{i,j} e_{ij} \tag{4}$$

Initialization Method	PSNR↑	SSIM↑	LPIPS↓	Init. Points	Output Points	Inital Geo. Error	Output Geo. Error
COLMAP	28.21	0.891	0.185	74625	1,566,408	0.4700	0.1240
GT_random_sampled	28.28	0.895	0.180	74625	1,663,144	0.0000	0.0510
GT_noisy_001	28.08	0.894	0.182	74625	1,641,607	0.0120	0.0515
GT_noisy_005	28.01	0.890	0.188	74625	1,559,763	0.0519	0.0520
GT_noisy_01	27.96	0.889	0.192	74625	1,489,835	0.0734	0.0534

Table 2. Geometric error and novel view synthesis results for different initial point clouds using the Tanks and Temples *Church* dataset. The entries labeled `GT_xxx` correspond to randomly sampled points from the ground truth point cloud, with a Gaussian noise of mean 0 and standard deviation `xxx` added to it. The colored cells highlight the best and second best results.

Scene	Init. Points	Clustered	Seeded
Church	74625	-3346 (-4.5%)	+861 (+1.2%)
Meetingroom	207591	-15713 (-7.1%)	+7443 (+3.7%)
Barn	93512	-8466 (-9.1%)	+2279 (+2.4%)

Table 3. Difference of the number of points through the Clustering and Seeding process. Clustered denotes the number of outliers removed by Clustering, and Seeded denotes the number of random points generated by Seeding.

Method	Exec. Time (s)	# of Points	PSNR	SSIM	LPIPS
COLMAP	61.4	12897	0.9395	28.21	0.1455
hloc	34.5	8654	0.9552	30.56	0.1313

Table 4. Comparison of the geometric reconstruction process, COLMAP and hloc (Superpoint + SuperGlue). hloc outperforms COLMAP both in computational efficiency and Novel View Synthesis quality.

After optimization using the Levenberg–Marquardt algorithm [12, 13], the final per-point reprojection error e_i is obtained by averaging over all cameras observing the i-th point:

$$e_i = \frac{1}{N_i} \sum_{j=1}^{N_i} e_{ij} \tag{5}$$

Since this reprojection error is unbounded, it must be transformed into a confidence value suitable for opacity initialization. First, the values e_i are min-max normalized to the range $[0, 1]$. Then, an activation function is applied to invert the values such that lower errors correspond to higher confidence. We test two different activation functions, **inverse** and **sigmoid**:

$$f_{inv}(e) = \frac{1}{a + bx}, \quad f_{sigm} = \frac{1}{c + e^{dx}} \tag{6}$$

where a, b, c,and d are hyperparameters controlling the shape and steepness of the function.

The resulting confidence scores are then normalized to the range $[0.05, 0.9]$ and used to initialize the opacity values. The lower bound of 0.05 corresponds to the threshold below which Gaussians are removed in the Adaptive Density Control procedure, while 0.9 is chosen as a practical upper limit since no Gaussians reach full opacity (i.e., 1.0).

4. Experiment

Dataset For experiments, we used the Tanks and Temples dataset [11] along with the DTU dataset [8]. Tanks and Temples provide a more challenging scene for reconstruction, while DTU is convenient environment for experiments. For §4.1 we selected *Church*, an indoor scene from Tanks and Temples since it is given the ground truth point cloud and is a challenging scene to reconstruct. Odd number indices were used as the input, and the model was trained and tested on the same data. For the evaluation in §4.2, we used two additional scenes: *Meetingroom*, an semi-indoor scene containing windows with views of the exterior, and *Barn*, an outdoor scene. Following the evaluation process of Mip-NeRF [1], images with indices that are multiples of 8 were excluded in the training process. For the evaluation in §4.3, we selected four scenes from DTU with the same Mip-NeRF style test-train split.

Evaluation Metrics We use three types of metrics to evaluate image quality: Peak Signal-to-Noise Ratio (PSNR) [7], Structural Similarity Similarity Index Metric (SSIM) [20], and Learned Perceptual Image Patch Similarity (LPIPS) [23]. Higher PSNR and SSIM values, along with lower LPIPS values indicate better model performance.

Computational Cost All scenes were trained on a single NVIDIA RTX 3060 GPU, with its training time varying from 40 min to 60 min, varying on the scene size. Using hloc [16] to replace COLMAP [17] decreases the execution time by 50% as shown in Tab. 4 The Clustering and Seeding process costs under 1 minutes in total. Furthermore, the combined initialization of the point cloud does not introduce additional computational overhead during the 3DGS training process, since it does not increase the number of points as visualized in Tab. 3.

Scene	Church			Meetingroom			Barn			Std. Between Views		
Initialization Method	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR	SSIM	LPIPS
COLMAP [17]	19.13	0.711	0.310	25.15	0.865	0.225	23.48	0.776	0.282	4.208	0.123	0.115
Random	18.18	0.692	0.333	24.41	0.851	0.250	19.52	0.731	0.343	3.872	0.121	0.113
Clustered	18.96	0.711	0.310	25.09	0.865	0.225	23.09	0.771	0.288	4.102	0.122	0.115
Combined Initialization	19.16	0.713	0.309	25.17	0.866	0.224	23.24	0.776	0.282	4.175	0.120	0.114

Table 5. Quantitative comparison of novel view synthesis on the Tanks and Temples Dataset. The colored cells highlight the best and second best results. Standard deviation between views refer to the standard deviation of the quality metrics across the test views, with values averaged across the three datasets. A lower standard deviation indicates that the quality of the rendered images is more consistent and balanced across different views.

4.1. Geometric Error

Tab. 2 presents the geometric error calculations for the *Church* scene. The original COLMAP point cloud was used as the baseline and compared against point clouds sampled from the ground truth, with Gaussian noise of mean zero and different standard deviations added. The point cloud sampled randomly from the ground truth, which has zero initial geometric error, shows the best NVS results. As gaussian noise is added to the ground truth, the NVS performace degrades proportionally, with the increase of the output geometric error. This suggests that higher geometric error of the initial point cloud results the Adaptive Density Control process to struggle in placing new Gaussians accurately, resulting a downgrade in novel view synthesis quality.

One thing to note is that the COLMAP initialization has the highest initial and output geometric error among these metrics, but shows reasonable NVS results. This suggests that the features extracted and match by COLMAP are more impactful in the 3DGS optimization process, leaving room for future research.

4.2. Combined Initialization

Four initialization metrics were evaluated: COLMAP, Random, Clustered, Combined. The random initial point cloud was obtained by randomly sampling the same number of points to the COLMAP generated point cloud from the ground truth. Clustered denotes the point cloud generated by the Clustering process, and Combined Initialization denotes the final rectified point cloud through Clustering and Seeding.

Tab. 3 visualizes the changes in the number of points throughout the point cloud rectification process. The indoor scene, *Church*, shows the smallest change in the point count, since most part of the scene in densely reconstructed. The outdoor scene, *Barn*, shows the largest decrease in the number of points. This is due to the sparse reconstruction of outdoor objects in the large scene, which the Clustering process defines as outliers and removing them from the point cloud.

To test the randomness of Combined initialization, the values of the Church scene was calculated as the average of three different trials. Since each box is small enough, the random points can be considered as uniform, not showing a significant difference over each trial.

Figure 8. Qualitative results of COLMAP initialized 3DGS (Left) and Combined initialized 3DGS (Right). Combined initialization shows improved representations of the back of the chairs, shown inside the red box.

The evaluation results on novel view synthesis are visualized in Tab. 5 and Fig. 8. Our method outperforms the COLMAP initialization for indoor scenes, where most points with small clusters, which are usually generated outside of the building, can be treated as outliers. Removing the outliers and adding random seeds in sparsely reconstructed regions with repeated patterns enhances the NVS results. As the scene contains more outdoor regions, the effectiveness of our method decreases, since clusters with small quantities may also be inliers since the scene is much more complicated. Therefore, our method achieves similar to better results in *Meetingroom*, and worse results in *Barn* compared to COLMAP initialization.

In addition, our Combined Initialization method showed reduced standard deviations across different viewpoints for novel view image renderings. This suggests that the point cloud rectification process is capable of generating more consistent quality images across different rendering viewpoints.

Scene	24			37			40			55			Average		
Method	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Uniform	24.64	0.928	0.110	24.80	0.895	0.173	23.82	0.877	0.207	27.84	0.879	0.283	25.47	0.884	0.221
Inverse	24.37	0.929	0.110	25.02	0.905	0.169	25.12	0.889	0.197	28.02	0.882	0.282	26.18	0.892	0.216
Sigmoid	24.96	0.930	0.109	25.07	0.904	0.168	24.41	0.887	0.197	27.77	0.876	0.283	25.75	0.889	0.216

Table 6. Qualitative comparison of novel view synthesis results across four scenes from the DTU dataset. *Uniform* refers to the baseline opacity initialization with a constant value of 0.1, while *Inverse* and *Sigmoid* denote confidence-aware opaciy initializations based on the activation functions described in Eq. (6). The best and second best values per column are highlighted in colors.

Method	PSNR↑	SSIM↑	LPIPS↓
COLMAP	19.13	0.711	0.310
Combined	19.16 (+ 0.03)	0.713 (+ 0.02)	0.309 (+ 0.01)
CAO	19.29 (+ 0.16)	0.720 (+ 0.09)	0.305 (+ 0.05)
Combined + CAO	19.30 (+ 0.17)	0.722 (+ 0.11)	0.305 (+ 0.05)

Table 7. Ablation study for Combined Initialization and Confidence-Aware Opacity Initialization (CAO) on the *Church* scene from the Tanks and Temples dataset.

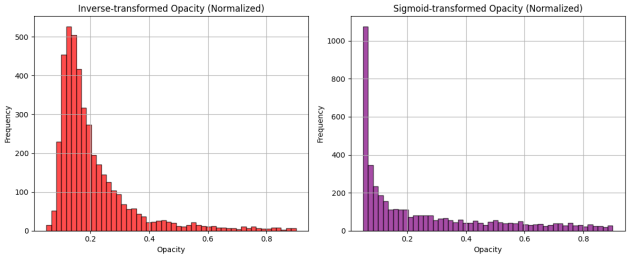

Figure 9. Visualization of the opacity distribution after applying the Inverse activation function (Left) and Sigmoid activation function (Right). The minimum and maximum values are normalized to 0.05 and 0.9, respectively.

4.3. Confidence-Aware Opacity Initialization

To evaluate the affect of Confidence-Aware Opacity Initialization, we compare three different initialization strategies: *Uniform*, where opacities are equally initialized to 0.1; and *Inverse* and *Sigmoid*, where opacities are initialized following the activation functions defined in Eq. (6). The hyperparameters were set to $a = 1, b = 10, c = 1, d = 15$.

Tab. 6 presents the evaluation results on novel view synthesis in four scenes of the DTU Dataset. While both metrics that integrate 3D reconstruction confidence into opacity initialization enhanced the evaluation result, selecting the activation function as *Inverse* outperformed *Sigmoid*.

This mainly results from the difference of the opacity distribution after applying each activation function, displayed in Fig. 9. While the Sigmoid activation function initializes the majority of the opacities near the minimum value, 0.05, the Inverse activation function allows a more equally distributed initialization. The distributed opacities offers the 3DGS algorithm a more diverse optimization path towards the optimal opacities, enhancing the quality of the results.

Finally, the merged result of Combined Initialization and Confidence-Aware Opacity Initialization (CAO) on the *Church* scene of the Tanks and Temples dataset is displayed in Tab. 7. While CAO mainly brings the performance enhancement, Combined Initialization also slightly improves the performance, and reduces the standard deviation between different views, enabling the algorithm to maintain consistent quality.

5. Conclusion

In this work, we introduced Geometric Error as an evaluation metric for point cloud, and demonstrated that reducing Geometric Error of the initial point cloud enhances the performance of 3DGS. We also proposed a point cloud rectification method that combines the strengths of COLMAP initialization and Random initialization. Evaluated on three different scenes, our method performs better in novel view synthesis tasks for indoor to semi-indoor scenes, while generating more consistent quality images across different views. Finally, we propose Confidence-Aware Opacity Initialization, which assigns high initial opacity to points with low reconstruction errors. This approach addresses the inefficiency of initializing all opacities to the same value, enhancing the performance in novel view synthesis tasks.

However, despite the improvements shown by Combined Initialization for indoor to semi-indoor scenes, our method generates degraded novel view images in outdoor scenes due to the extremely sparse features of the larger scenes. To address this limitation, we suggest that future work focusing on developing methods to classify scenes as indoor or outdoor would potentially improve the performance across a wider range of scenarios.

References

[1] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5470–5479, 2022. 2, 5

[2] M Akmal Butt and Petros Maragos. Optimum design of chamfer distance transforms. *IEEE Transactions on Image Processing*, 7(10):1477–1484, 1998. 1

[3] Dmitry Chetverikov, Dmitry Svirko, Dmitry Stepanov, and Pavel Krsek. The trimmed iterative closest point algorithm. In *2002 International Conference on Pattern Recognition*, pages 545–548. IEEE, 2002. 3

[4] Kangle Deng, Andrew Liu, Jun-Yan Zhu, and Deva Ramanan. Depth-supervised nerf: Fewer views and faster training for free. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12882–12891, 2022. 1, 2

[5] M-P Dubuisson and Anil K Jain. A modified hausdorff distance for object matching. In *Proceedings of 12th international conference on pattern recognition*, pages 566–568. IEEE, 1994. 2

[6] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, pages 226–231, 1996. 4

[7] Alain Hore and Djemel Ziou. Image quality metrics: Psnr vs. ssim. In *2010 20th international conference on pattern recognition*, pages 2366–2369. IEEE, 2010. 5

[8] Rasmus Jensen, Anders Dahl, George Vogiatzis, Engil Tola, and Henrik Aanæs. Large scale multi-view stereopsis evaluation. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 406–413. IEEE, 2014. 5

[9] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics (ToG)*, 42(4):1–14, 2023. 1

[10] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023. 2

[11] Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics (ToG)*, 36 (4):1–13, 2017. 3, 5

[12] Kenneth Levenberg. A method for the solution of certain non-linear problems in least squares. *Quarterly of applied mathematics*, 2(2):164–168, 1944. 5

[13] Donald W Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics*, 11(2):431–441, 1963. 5

[14] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Computer Vision–ECCV 2020*, pages 405–421, 2020. 1, 2

[15] Ofir Pele and Michael Werman. Fast and robust earth mover's distances. In *2009 IEEE 12th international conference on computer vision*, pages 460–467. IEEE, 2009. 2

[16] Paul-Edouard Sarlin, Cesar Cadena, Roland Siegwart, and Marcin Dymczyk. From coarse to fine: Robust hierarchical localization at large scale. In *CVPR*, 2019. 5

[17] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 3, 5, 6

[18] Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4104–4113, 2016. 4

[19] Noah Snavely, Steven M Seitz, and Richard Szeliski. Photo tourism: exploring photo collections in 3d. In *ACM siggraph 2006 papers*, pages 835–846, 2006. 1, 3

[20] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 5

[21] 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. 1, 2

[22] Qiangeng Xu, Zexiang Xu, Julien Philip, Sai Bi, Zhixin Shu, Kalyan Sunkavalli, and Ulrich Neumann. Point-nerf: Point-based neural radiance fields. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5438–5448, 2022. 1, 2

[23] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 5