A Comprehensive Evaluation of the Spatial Accuracy of Building Gaussian Splatting

Samuel McNally[1], Shabnam Jabari [1], Heather McGrath[2], Mark Masry[3]

[1] Dept. of Geodesy and Geomatics Engineering, University of New Brunswick, Canada
[2]Natural Resources Canada
[3]Modelar

Keywords: Gaussian Splatting, Point Clouds, Digital Twins, Phone LiDAR, 3D Reconstruction.

Abstract

3D building models are powerful visual tools, typically generated with well-established image-matching or LiDAR methods. However, they do not capture the view-dependent colour characteristics possible with Gaussian splatting. Despite the visual potential of Gaussian splatting, there is limited knowledge on its spatial accuracy and influencing factors, particularly for buildings. To address this gap, a two-building dataset was collected with terrestrial laser scans, images, phone LiDAR, and target points, and the visual and spatial effects of numerous factors were analyzed. These factors included the source and quality of the input camera poses and point cloud, the number of images and training iterations, and the Gaussian splat method. Gaussian splats were trained from open source and commercial image-based reconstruction methods, COLMAP and Pix4D, and phone LiDAR reconstructions. Applying Gaussian splatting to these inputs had minimal impact on the target points and the overall structure of the buildings, but the positions of Gaussians deviated from the initial point cloud, particularly before 15,000 iterations, resulting in more floaters and lower spatial accuracy. Image-based reconstruction methods outperformed phone LiDAR methods on visual and spatial metrics. Cleaning COLMAP point clouds considerably decreased Gaussian floaters, while downsampling input point clouds increased the percentage of floaters and yielded similar visual results. 2D Gaussian splatting provided geometric constraints, removing some floaters, but sacrificed visual quality. Increasing the number of images to three loops around the buildings improved visual and spatial results. Overall, the spatial accuracy of building Gaussian splatting was heavily dependent on the factors studied.

1. Introduction

3D building models enable immersive visualizations and simulations of real-world phenomena in a virtual environment. Photogrammetry and LiDAR, two well-established 3D building modelling techniques, have well-understood spatial accuracies (Wolf et al., 2014). However, the point clouds or meshes generated by these techniques often lack critical details such as reflections, shadows, and fine-grained features like vegetation. Recent developments in computer vision address these limitations and offer novel ways to create and visualize 3D models with view-angle-dependent colour characteristics, most notably Neural Radiance Fields (NeRFs) and Gaussian splats (Kerbl et al., 2023; Mildenhall et al., 2022). NeRFs represent space with a continuous function (Mildenhall et al., 2022), leading to slow rendering and challenging spatial accuracy assessment. In contrast, 3D Gaussian splats can be viewed in real time, and the discrete mean positions of Gaussians facilitate spatial accuracy assessment, all while outperforming NeRFs on visual similarity metrics (Kerbl et al., 2023).

The accuracy of NeRFs and Gaussian splats is often assessed with visual similarity metrics that do not fully capture spatial accuracy. There is a lack of understanding of the spatial accuracy of Gaussian splatting and the factors that affect its spatial accuracy, particularly for buildings. This research assessed factors affecting the spatial accuracy of building Gaussian splatting, including the source of the initial point cloud and camera poses, downsampling and cleaning the initial point cloud, the number of images and iterations, and the type of Gaussian splatting. The spatial accuracy of two image-based reconstruction methods and one phone LiDAR-based method, and their resulting Gaussian splats, were assessed against a terrestrial laser scanner point cloud and target points. Visual accuracy was assessed by comparing renderings of the Gaussian splats against test images.

2. Related Work

2.1 Image-Based 3D Reconstructions

Photogrammetry and computer vision are popular methods for creating 3D models from overlapping images. Firstly, matching points are found between the images, with information on the image location or sequence capable of aiding the matching process. Next, a bundle adjustment adjusts the orientation and position of the matched images simultaneously to determine updated values for camera parameters and points (Wolf et al., 2014). COLMAP, based on Surface-from-Motion (SfM), is a prevalent software for computer vision reconstruction research, as it is open-source and has a variety of matching options (Schönberger and Frahm, 2016), yet COLMAP outputs lack scaled coordinates. Pix4D is a common commercial photogrammetry 3D reconstruction software package (Knapitsch et al., 2017; Mora-Félix et al., 2024) with robust error analysis and automatic removal of most floaters: points that are “floating” in areas without objects. Pix4D reconstructions can also lack scale when image locations are unknown or not well known. COLMAP and Pix4D score highly on reconstruction quality (Knapitsch et al., 2017), and can create images that remove modelled radial and tangential lens distortions.

2.2 Terrestrial and Phone LiDAR

Terrestrial laser scanning (TLS) techniques can achieve sub-centimeter level accuracies (Elhashash et al., 2022; Wolf et al., 2014), thus acting as ground truth for spatial accuracy assessments of other techniques. However, they are expensive, require trained personnel, are time-consuming to collect and process, and suffer from occlusions due to the limited number of setups. In contrast, phone LiDAR allows users to move around scenes, leading to faster collection times and fewer occlusions in front of buildings. Phone LiDAR can provide a dense, scaled point cloud even in areas with few image features. However,

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-M-1-2026-31-2026 | © Author(s) 2026. CC BY 4.0 License.

range limitations and drift accumulation affect its capabilities. Phone LiDAR is understood to have a range of approximately 5 meters, making it insufficient for capturing details above the first floor of buildings. Yet for some applications, such as flood visualizations, the ground floor is the most pertinent. One benefit of the limited range is that phone LiDAR building point clouds tend to have fewer floaters than TLS and image-based point clouds due to returns from inside buildings and incorrect image matches. As the distance scanned for phone LiDAR increases, drift in phone position and orientation accumulates due to errors in the Inertial Measurement Unit (IMU) measurements and Simultaneous Localization and Mapping (SLAM) algorithms. One study found 30 centimeters of error per 60 meters of scanning, but this drift error was reduced by half with the use of a 3-axis gimbal (Tamimi, 2022). Gimbals are an affordable solution that smooths the phone's motion as users walk around scenes, decreasing the position and orientation changes of the phone, aiding IMU and SLAM algorithms. Gimbals also reduce image motion blur and are used in some computer vision datasets (Knapitsch et al., 2017).

2.3 Gaussian Splatting

3D Gaussian splatting is a novel visualization method that may play a pivotal role in the future of 3D modelling. In 3D Gaussian splatting, a series of mathematical objects (Gaussians) follow 3D Gaussian functions with opacity decreasing from the center to model a scene. Each Gaussian has a mean position, orientation, shape, size, colour coefficients, and opacity. Gaussians are conventionally initialized with a sparse point cloud from COLMAP (Kerbl et al., 2023). Next, Gaussians are iteratively updated with stochastic gradient descent by comparing renderings with input images to improve scene representation. During training, parameters that define Gaussians are updated, including positions, and an adaptive densification strategy adds and removes Gaussians.

Despite the evident potential of Gaussian splatting, it still has spatial limitations and high memory and processing requirements. Gaussian splatting models surfaces with a series of continuous Gaussians, often making surfaces appear jagged (Huang et al., 2024). Outdoor scenes pose challenges for Gaussian splats due to sky pixels, illumination differences, and moving objects, resulting in floaters and artifacts. Typically, Gaussian splatting attempts to model every pixel in the input images, even though some objects may be far away from the camera locations, such as sky pixels.

2.4 Visual Accuracy Assessment

The quality of Gaussian splats is typically evaluated solely with visual similarity metrics, comparing renderings of Gaussian splats against images. The three most common visual similarity metrics for evaluating Gaussian splats are Peak Signal to Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS) (Chen and Wang, 2025; Huang et al., 2024; Kerbl et al., 2023; Mildenhall et al., 2022; Zhou et al., 2024). PSNR takes the logarithm of the maximum intensity in the image squared divided by the mean squared error (Setiadi, 2021). To calculate SSIM, differences in the mean value (luminance), the standard deviations (contrast), and the correlation coefficient (structure) are averaged for each band, then multiplied together (Horé and Ziou, 2013). LPIPS attempts to capture the similarity between features of two images and is considered to be closer to how humans perceive image similarity by comparing extracted features from a pre-trained convolutional neural network (Zhang et al., 2018).

2.5 Point Cloud Spatial Accuracy

To evaluate the spatial accuracy of point clouds, they are typically registered to a reference point cloud with the Iterative Closest Point Algorithm (ICP) or manual point picking, then assessed with cloud-to-cloud (C2C) or target-based comparisons (Askar and Sternberg, 2023; Luetzenburg et al., 2021; Wolf et al., 2014). Mean C2C distances average the minimum distances between every point in the test point cloud and the reference cloud (Billi et al., 2025). Root Mean Square (RMS) squares these minimum distances between clouds. Chamfer distance performs a two-way comparison of squared minimum distances between point clouds averaging both directions together. The percentage of points within distance thresholds quantifies the number of points within tolerances (Billi et al., 2025), and points outside of these thresholds are often floaters. Target-based comparisons frequently use Root-Mean-Squared-Error (RMSE), the square root of the average square of the differences between the target points in the test point cloud and the known positions of the targets (Billi et al., 2025).

The spatial accuracy of Gaussian splatting is still under investigation. During training, the positions of Gaussians are updated with other attributes, leading to the Gaussian positions deviating from the initial point cloud as the number of iterations increases. Another potential factor affecting the spatial accuracy of Gaussian splatting is the source and quality of the initial point cloud and camera poses. Adding more images could alter the image-based reconstruction and provide more view angles of the Gaussians during training. The density of the initial point cloud affects the number of initial Gaussians and could impact results. The impact of these factors were analyzed in this research. Although the mean positions of Gaussians do not fully represent the accuracy of the splat, they are commonly used in the literature (Xiong et al., 2024; Yang et al., 2025).

2D Gaussian Splatting (GS) is a well-referenced method that improves surface representation by flattening 3D Gaussians into 2D disks and incorporating a regularization term and normal consistency terms to encourage Gaussians to align with Gaussians in the same neighbourhood (Huang et al., 2024). The authors of 2D GS reported a notable improvement in spatial accuracy for chamfer distance on the DTU dataset (Huang et al., 2024). However, the DTU dataset only contains miniature building models and precise camera positions from a robotic arm (Jensen et al., 2014). Yet the full-scale buildings and errors in camera positions found in real-world datasets may alter these results.

2.6 Gaussian Splatting Datasets

Although there is extensive Gaussian splatting research on indoor scenes, small objects, and street view datasets (Huang et al., 2024; Zhou et al., 2024), there is a lack of building datasets in the literature with TLS, phone LiDAR, and images to assess the spatial accuracy of the Gaussian splats. Most datasets also do not include target points, despite the importance of target points for spatial accuracy assessment. The image overlap is not always consistent within or between datasets, making comparisons more difficult. Phone LiDAR apps can capture an image and LiDAR frame when users move or rotate the phone a specified amount, leading to evenly spaced images. Phone LiDAR apps show users in real-time the collected point cloud or mesh, ensuring the capture of the entire area of interest.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-M-1-2026-31-2026 | © Author(s) 2026. CC BY 4.0 License.

3. Methodology

Since there is a lack of building datasets for Gaussian splatting with images, phone LiDAR, terrestrial laser scans, and targets, a two-building dataset was created with those components with a focus on residential size buildings. Gaussian splats were trained from the images, camera poses, and point clouds generated by Pix4D, COLMAP, and phone LiDAR. The spatial accuracy of the reconstructions and Gaussian splats were evaluated by comparing the point clouds against terrestrial laser scans and target points. Visual similarity was assessed by comparing renderings of the Gaussian splats against test images. An overview of the methodology is shown in Figure 1. This research took a quantitative experimental approach. The independent variables were spatial accuracy and visual similarity metrics. Based on initial tests and the literature, the selected factors to analyze were the source of the initialization point cloud and camera poses (phone LiDAR, COLMAP, and Pix4D), cleaning and downsampling the initialization point cloud, the number of training iterations and images, and the Gaussian splatting method (3DGS and 2DGS).

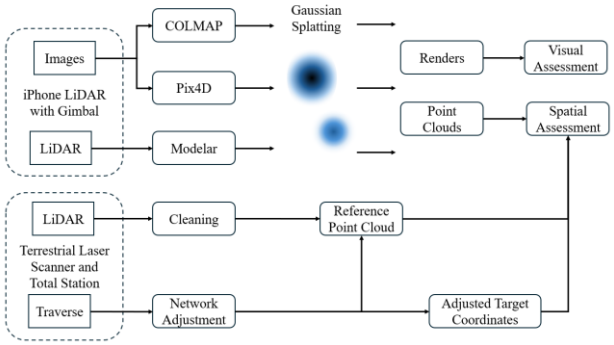

Figure 1: Methodology Overview

3.1 Data Collection

Two buildings, building 1 and building 2, with perimeters of 20m and 40m respectively, were selected that had few occlusions and clear lines of sight around the buildings. Firstly, intervisible spikes were set for control points in a loop around each building, ensuring the entire building exteriors were visible from these points, resulting in four control points offset from building corners. Next, three to four black and white checkerboard targets were placed on each wall of the two buildings at varying heights, ensuring a minimum of six targets were visible from each setup and each target was visible from multiple setups for redundancy. Subsequently, a Trimble SX12 total station and terrestrial laser scanner was set up on the control points to perform a closed-loop traverse, measure the targets, and scan the building.

For phone LiDAR, an iPhone 16 Pro with a DJI Osmo Mobile 6 three-axis gimbal was utilized. By comparing overlap points in the phone LiDAR point cloud, scanning with the gimbal decreased coordinate differences of overlap points by half, supporting the findings of Tamimi (2022), and it also reduced motion blur in the images. To train Gaussian splats directly from phone LiDAR outputs, the outputs must include images, camera poses, and a point cloud. We used the free app Modelar, which provides all the necessary outputs and numerous scan settings. To ensure sufficient overlap for LiDAR frames and images, a LiDAR frame was captured when the phone was rotated 8 degrees or moved 10 centimeters, and an image was captured every three LiDAR frames (see Figures 2 and 3). The buildings were scanned all the way around, with an attempt to capture points as high on the facades as possible, as shown in Figure 2. Each building was scanned with phone LiDAR four times, with the run with the least sun exposure selected for processing.

Figure 2: Phone LiDAR Acquisition: Point Cloud and Image Locations for Building 2

Figure 3: Sample Image Sequence for Building 2

3.2 Ground Truth Processing

The total station observations (see Figure 4) were adjusted in Trimble Business Center to produce final coordinates for the target points and the manually cleaned terrestrial laser scanner point cloud. The network adjustment resulted in semi-major axes of the 95% confidence coordinate ellipses for the control points and targets all under 3.3mm for both buildings.

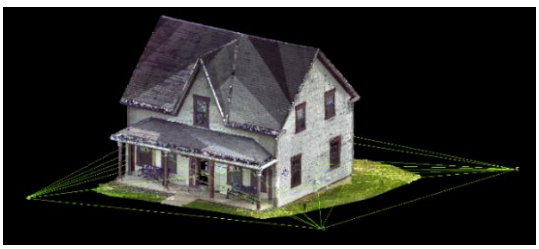

Figure 4: Total Station Measurements and TLS Point Cloud for Building 2

3.3 Image-Based 3D Reconstructions

Images, COLMAP format camera poses, and point clouds were exported from the Modelar app. The overlap areas at the end of the building loop in the phone LiDAR point cloud and overlap images were removed. In total, for the selected run, there were 93 images for building 1 and 201 images for building 2. Pix4D Mapper and COLMAP were selected for this project to align with existing research and to compare open-source and commercial software packages for 3D reconstruction. To enable sequential image matching, timestamp information was added to the Modelar image metadata based on the image names. Pix4D and COLMAP 3D reconstructions resulted in point clouds, camera poses, and undistorted images in an arbitrary coordinate system (see Figure 5). Background points in the Pix4D, Modelar, and COLMAP point clouds outside of the extent of the reference point cloud were removed. The Pix4D, COLMAP, and phone LiDAR point clouds were cleaned manually. The target point

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-M-1-2026-31-2026 | © Author(s) 2026. CC BY 4.0 License.

coordinates were not added to these reconstructions to keep them separate for spatial accuracy assessment.

Figure 5: COLMAP (left) and Pix4D (right) Reconstructions for Building 2

Pix4D provides text files with a rotation matrix R and a translation matrix T for each image. These camera poses were converted to world-to-camera COLMAP format text files for Gaussian splatting. The COLMAP translation vector t, which transforms world positions to camera positions was computed with Equation 1 (Schönberger, 2025a):

$$t = -RT \quad (1)$$

The rotation matrix R was converted to a Hamilton convention quaternion with the function utilized in COLMAP (Schönberger, 2025b). The camera locations and orientations were verified visually against the point cloud.

3.4 Gaussian Splatting

Gaussian splats were trained with an NVIDIA RTX 4500 GPU. The popular method of keeping 1 in 8 images separate for testing was adopted for all tests to determine model performance on novel views (Barron et al., 2022). Due to memory limitations, the Pix4D and phone LiDAR point clouds were downsampled to 1 centimeter, except the tests described in section 4.5, and images were downsampled to a width of 1600 pixels for all tests. Since COLMAP is a common reconstruction method in Gaussian splatting literature, the Gaussian splat method, the number of images, and the number of iterations were tested on the COLMAP reconstructions. The original implementation of Gaussian splatting, developed by Kerbl et al. (2023), acted as a baseline. 2D GS was tested against this baseline to compare a method with geometric constraints (Huang et al., 2024). The comparisons for the number of images were only performed on the smaller building (building 1) due to memory limitations. While testing the number of images, one to four runs of LiDAR images were used, resulting in 109, 235, 353, and 477 images. All other tests were performed with one run. Gaussian splats were trained for 30,000 iterations, since it is the default number of training iterations, and the spatial and visual metrics levelled off before this point, indicating that additional training was unnecessary. To assess Gaussian splat spatial accuracy during training, increasing training intervals were chosen for the cleaned COLMAP reconstruction of building 2: 500, 1000, 1500, 2000, 3000, 5000, 7000, 10000, 15000, 20000, 25000, and 30000.

3.5 Visual and Spatial Accuracy Assessment

After changing each independent variable, the visual and spatial accuracy of the inputs and the Gaussian splats were assessed. To align with computer vision literature, Gaussian splat visual quality was evaluated with PSNR, LPIPS, and SSIM (Xiong et al., 2024; Yang et al., 2025). Gaussian splat mesh generation from depth maps alters the underlying geometry of the Gaussians by rejecting Gaussians behind objects. Thus, mean positions were chosen as a higher fidelity way to capture the spatial accuracy of the Gaussians. The input point clouds and Gaussian splat mean positions were roughly aligned manually and finely aligned with ICP to the TLS point cloud. The scale was held fixed during alignment for the phone LiDAR point clouds and resulting Gaussian splats, but was allowed to change for the image-based reconstructions, since they lacked scale.

For an overall comparison to the reference TLS cloud, mean and RMS C2C distances were computed. Chamfer distance was not selected as it is a two-way comparison that requires the reference and test point clouds to cover the same area, when the reference cloud may cover more of the building, such as the roof, than the test point clouds. To analyze the number of floaters, the percentages of points within 0.05-, 0.1-, and 0.5-meter thresholds were computed. To attain a spatial accuracy metric independent of floaters that captures the overall structure of the building, target RMSE was selected. Zeroth-order spherical harmonic coefficients were a sufficient approximation for colour to pick out target points in the Gaussian splats. The picked target points were compared with their adjusted coordinates. There is some inherent error to point picking, based on which point is chosen and whether a point is in the center of the target, but these differences were found to be only a few millimeters for image-based reconstructions.

4. Results and Discussion

The 3D reconstructions were tested for their spatial accuracy, and the Gaussian splats were tested for visual and spatial accuracy, as shown in Table 1. The best result for each building is highlighted in bold.

Independent Variables			3D Reconstruction Spatial Accuracy (m)						Gaussian Splat Spatial Accuracy (m)						Gaussian Splat Visual Metrics		
Building	Method	Clean	Target	C2C Metrics		% In Threshold			Target	C2C Metrics		% In Threshold			Pixel		Feature
			RMSE	Mean	RMS	0.5	0.1	0.05	RMSE	Mean	RMS	0.5	0.1	0.05	SSIM ↑	PSNR ↑	LPIPS ↓
1	Modelar	×	0.099	0.072	0.076	**100**	89.5	17.0	N/A	0.064	0.121	99.6	82.4	50.2	0.614	19.486	0.385
		✓	0.099	0.072	0.076	**100**	89.5	17.1	N/A	**0.062**	**0.087**	**99.7**	82.5	50.3	0.613	19.156	0.387
	Pix4D	×	0.092	0.028	0.041	**100**	96.4	83.6	N/A	0.417	1.486	91.3	82.4	65.6	0.685	17.913	0.359
		✓	0.086	0.027	0.036	**100**	98.5	83.7	N/A	0.445	1.532	92.3	84.2	67.1	0.686	18.257	0.360
	COLMAP	×	**0.020**	0.796	3.451	87.3	81.9	77.8	**0.018**	0.679	3.145	93.5	89.8	81.2	**0.766**	**22.283**	**0.294**
		✓	0.021	**0.016**	**0.027**	**100**	**98.8**	**94.3**	0.019	0.136	0.867	97.8	**93.9**	**84.7**	0.761	22.198	0.303
2	Modelar	×	0.172	0.112	0.131	**100**	43.7	17.9	N/A	0.133	0.248	98.9	49.2	26.2	0.500	17.337	0.399
		✓	0.173	0.108	0.125	**100**	46.5	18.5	N/A	0.126	0.175	99.1	49.7	26.6	0.501	17.385	0.398
	Pix4D	×	0.011	**0.012**	**0.022**	**100**	**99.1**	**97.3**	0.015	0.050	0.249	99.2	93.3	80.9	0.766	20.289	0.266
		✓	**0.010**	**0.012**	**0.022**	**100**	**99.1**	**97.3**	0.013	**0.036**	**0.115**	**99.7**	**93.8**	**81.4**	0.779	20.922	**0.252**
	COLMAP	×	0.016	0.021	0.124	99.6	98.3	96.0	0.009	0.104	1.149	98.9	91.5	80.7	**0.780**	**23.023**	0.260
		✓	0.012	**0.012**	0.025	**100**	99.0	96.8	**0.008**	0.055	0.374	99.2	91.6	80.8	0.776	22.666	0.263

Table 1: 3D Reconstruction and Gaussian Splat Visual and Spatial Results

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-M-1-2026-31-2026 | © Author(s) 2026. CC BY 4.0 License.

4.1 3D Reconstruction Spatial Accuracies

The spatial accuracy of the 3D reconstruction point clouds varied considerably depending on the type of reconstruction, the building, and point cloud cleaning. Image-based reconstructions outperformed phone LiDAR for both buildings in terms of points within 0.1 and 0.05m thresholds and target RMSE. With a doubling in the building perimeter, the target RMSE and C2C metrics for phone LiDAR increased by a factor between 1.5 and 2 times, indicating drift accumulation.

Interestingly, for image-based reconstructions, the larger building, building 2, showed improved spatial accuracy on all metrics. This indicates that factors beyond building size affect the spatial accuracy of these reconstructions and could be explained by more wall features on building 2. The Pix4D reconstruction of building 1 had matching issues. For all other image-based reconstructions, the target RMSEs were all in the one to two centimeter range, indicating that the overall structure of the building was well preserved.

Cleaning the initial point cloud had a negligible impact on target RMSE, with changes of only a few millimeters present. However, cleaning the initial point cloud led to major changes in the C2C metrics and threshold percentages, indicating that cleaning can remove floaters, but may not improve the underlying building structure. The COLMAP reconstructions, in particular, had lower percentages of points within 0.5m of the TLS point cloud at 87.3% and 99.6%, indicating the presence of floaters far from the TLS scan that skewed the C2C metrics, versus 100% for Pix4D and Modelar. This is expected since Pix4D has some automatic cleaning, and Modelar has fewer floaters due to its limited range. After cleaning, Pix4D and COLMAP achieved very similar spatial accuracy results for building 2.

4.2 Impact of Gaussian Splatting

When applying Gaussian splatting, notable changes to the spatial accuracy of the constructions were observed, as shown in Table 1. These changes could be impacted by errors in the camera poses, point cloud, and splatting process. On nearly all Gaussian splats, the C2C metrics and the percentage thresholds deteriorated from the reconstruction. With Gaussian splatting, the target RMSE changed less than 8 millimeters for all tests where the targets were visible, with some positive and some negative changes. These findings indicate that Gaussian splatting preserves the underlying structure of the buildings but leads to floaters impacting C2C and threshold metrics. The visual results differed from the spatial results, indicating that the visual similarity metrics failed to fully capture spatial accuracy.

4.3 Input Reconstruction Method

Image-based reconstruction Gaussian splats outperformed phone LiDAR Gaussian splats on visual metrics. This could be explained by phone LiDAR drift, errors in the image to point cloud registration, and unmodeled lens distortion in the phone LiDAR images as Modelar samples the Apple camera feed without applying additional distortion corrections. The targets were not all visible in the phone LiDAR Gaussian splats, preventing target RMSE computation. Contrasting the visual results, phone LiDAR Gaussian splats achieved better C2C metrics and percentages within the 0.5-meter threshold for the smaller building, where less drift had accumulated. Interestingly, some phone LiDAR metrics improved with Gaussian splatting, such as the percentage of points within 0.05 meters for both buildings, the percentage of points within 0.1 meters for building 2, and mean C2C metrics for building 1. A possible explanation is that Gaussian splatting corrected some of the initial point cloud drift, but floaters caused the RMS C2C to decrease.

Depending on the building and the metric, Pix4D or COLMAP Gaussian splats performed better. On building 2, the COLMAP and Pix4D Gaussian splats had similar SSIM and LPIPS, and visually looked nearly identical, and COLMAP had a higher PSNR. The warped geometry in the Pix4D reconstruction of building 1 resulted in the Gaussian splat having lower visual similarity and threshold percentages than COLMAP overall. On all COLMAP and Pix4D inputs except for the noisy COLMAP reconstruction for building 1, applying Gaussian splatting increased C2C metrics by a factor of 3 or more and decreased the percentage of points within distance thresholds by up to 18%. Without geometric constraints, applying Gaussian splatting to image-based reconstructions of buildings appears to lead to a notable decrease in spatial accuracy from the reconstruction.

4.4 Point Cloud Cleaning

Cleaning the initial point cloud had a marginal effect on visual similarity metrics, with maximum changes of 0.013, 0.633, and 0.014 for SSIM, PSNR, and LPIPS. Spatially, cleaning had a negligible impact on target RMSE with a maximum change of 2 millimeters, but markedly improved the C2C distances and the percentage of points within thresholds, indicating that cleaning reduces the number of floaters but has little impact on the overall building structure. Cleaning the COLMAP point cloud resulted in the most pronounced improvement in C2C metrics and threshold percentages since it was the noisiest input point cloud. Although there was some difficulty differentiating between background objects and floaters, most points far from the TLS point cloud were floaters. Cleaning the input point clouds for building 1 had more of an impact, which could be explained by more sky and background pixels in the images for building 1, since Gaussian splatting attempts to model all pixels.

4.5 Point Cloud Downsampling

Gaussian splats were trained from cleaned Modelar and Pix4D reconstructions for building 1 using all points to compare with the 1cm downsampling used in other tests. Keeping the full point count yielded similar visual results, with marginally worse pixel-based metrics and marginally better feature-based metrics (see Table 2). With larger numbers of initial Gaussians, the gradients are spread over more Gaussians, leading to fewer training updates to each Gaussian. Increasing the quantity of input points increased the number and percentage of Gaussians in similar positions to the originals, improving C2C metrics and thresholds for Pix4D splats, and lowering mean C2C and thresholds for Modelar splats. With Modelar, Gaussian splatting may not have corrected as much of the phone LiDAR drift due to the presence of more input points.

Gaussian Splat Input	Gaussian Splat Spatial Accuracy (m)					Splat Visual Metrics		
	C2C Metrics		% In Threshold			SSIM	PSNR	LPIPS
	Mean	RMS	0.5	0.1	0.05			
Modelar 1cm	**0.062**	0.087	99.7%	82.5%	50.3%	0.613	**19.156**	0.387
Modelar Full	0.073	**0.085**	**99.9%**	76.3%	30.4%	0.571	19.144	0.381
Pix4D 1cm	0.445	1.532	92.3%	84.2%	67.1%	**0.686**	18.257	0.360
Pix4D Full	0.085	0.613	99.2%	**92.8%**	**71.6%**	0.653	18.082	**0.356**

Table 2: Gaussian Splat Results With and Without Downsampling for Building 1

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-M-1-2026-31-2026 | © Author(s) 2026. CC BY 4.0 License.

Independent Variables		3D Reconstruction Spatial Accuracy (m)						Gaussian Splat Spatial Accuracy (m)						Gaussian Splat Visual Metrics		
Number of Images	Clean	Target	C2C Metrics		% In Threshold			Target	C2C Metrics		% In Threshold			Pixel		Feature
		RMSE	Mean	RMS	0.5	0.1	0.05	RMSE	Mean	RMS	0.5	0.1	0.05	SSIM ↑	PSNR ↑	LPIPS ↓
109	×	0.020	0.796	3.451	87.3	81.9	77.8	0.018	0.679	3.145	93.5	89.8	81.2	0.766	22.283	0.294
	✓	0.021	0.016	0.027	**100**	98.8	94.3	0.019	0.136	0.867	97.8	93.9	84.7	0.761	22.198	0.303
235	×	0.014	0.770	4.256	91.0	88.4	86.0	0.013	1.299	4.321	86.8	85.2	81.0	0.769	22.573	0.286
	✓	0.013	**0.011**	**0.018**	**100**	99.4	94.9	0.015	0.065	0.376	97.6	95.6	90.8	0.757	21.624	0.302
353	×	0.013	0.515	3.563	93.4	90.6	88.1	0.014	0.874	3.556	90.4	88.8	84.5	**0.792**	**23.509**	0.269
	✓	**0.012**	0.012	**0.018**	**100**	**99.6**	**97.5**	0.015	**0.043**	**0.208**	**98.2**	**96.1**	**91.2**	0.778	22.157	0.287
477	×	**0.012**	0.697	4.382	93.4	90.8	88.3	**0.009**	0.959	3.848	89.4	87.7	82.8	0.785	23.425	**0.254**
	✓	0.014	0.012	0.019	**100**	99.5	97.4	0.011	0.056	0.318	98.0	**96.1**	90.7	0.772	22.412	0.273

Table 3: COLMAP Reconstruction and Gaussian Splat Visual and Spatial Results with Changes in the Number of Images

4.6 Number of Images

Increasing the number of images was tested with COLMAP for building 1 with one to four loops of images around the building, with the results shown in Table 3. Increasing the number of images to 353 images improved spatial and visual results for the COLMAP reconstructions and the resulting Gaussian splats. Beyond 353 images, adding additional images led to marginally lower C2C metrics and percentages within thresholds; however, it led to a higher target RMSE. Visually, as shown in Figure 6, increasing the number of images from 109 to 235 seems to have improved the representation of building facades and removed some floaters, and increasing the number of images beyond 235 improved the representation of other features, such as grass and sky. Cleaning the initial point cloud improved spatial performance and worsened visual performance on all four tests.

Figure 6: Visual Assessment of Cleaned COLMAP Gaussian Splats for Building 1

Increasing the number of images beyond 353 images led to marginally lower visual accuracies, perhaps due to lighting changes during image capture. Increasing the number of images decreased the discrepancy between the train and test visual accuracies, since with more training images, there are training images taken closer to the test images and more viewpoints to train the model. Continually adding images is not always practical due to time and memory constraints. Increasing the number of images fourfold increased the number of input points and Gaussians, increasing the number of Gaussians from 889K to 1032K with cleaning and 945K to 1140K without cleaning.

4.7 2D GS vs 3D GS

For building 2, 2D GS was compared against 3D GS for the cleaned COLMAP reconstruction (see Table 4). With 2D GS, facades were thinner and more well-defined, as shown in Figure 7, improving the percentages of points within thresholds and C2C metrics. Interestingly, 2D GS did not lower target RMSE, indicating that the picked average depth from 3D GS was similar to or of better quality than 2D GS. Despite the improvements in spatial accuracy, 2D GS had lower visual similarity scores, sacrificing visual quality for geometric accuracy with fewer floaters, indicating a trade-off between visual and spatial accuracy when choosing the Gaussian splat method.

	Gaussian Splat Spatial Accuracy (m)						Splat Visual Metrics		
Splat Method	Target	C2C Metrics		% In Threshold			SSIM	PSNR	LPIPS
	RMSE	Mean	RMS	0.5	0.1	0.05			
3D GS	**0.008**	0.055	0.374	99.2	91.6	80.8	**0.776**	**22.666**	**0.263**
2D GS	0.014	**0.037**	**0.178**	**99.4**	**94.8**	**84.7**	0.754	22.073	0.309

Table 4: Impact of 2D GS for Building 2 on the Cleaned COLMAP Reconstruction

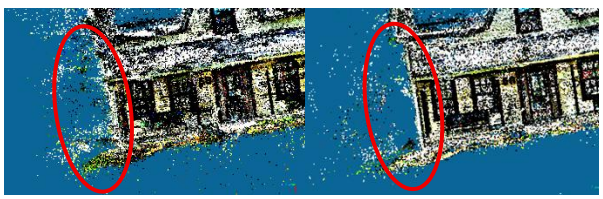

Figure 7: 3D (left) and 2D (right) GS Comparison on Building 2

4.8 Number of Training Iterations

The number of training iterations had a sizable impact on the visual and spatial accuracy of the cleaned COLMAP Gaussian splats for building 2 (see Figures 8 and 9). Increasing the number of iterations increased the visual similarity metrics until it levelled off near 25,000 iterations. The C2C metrics and floater percentages increased sharply near the start of training, with notable changes detected at only 500 iterations, before densification. Interestingly, the percentage of points above distance thresholds and the mean C2C distance peaked near 3,000 iterations despite the number of points outside these thresholds still increasing, perhaps due to further densification of building points. The RMS C2C curve had a local maximum near 3,000 iterations, but continued to increase as Gaussian splatting continued to push some Gaussians farther from the building throughout the training process. The discrepancy between RMS and mean C2C curve shapes could be explained due the higher sensitivity to outliers far from the building when squaring the distances. The densification stopped at 15,000 iterations, leading to the percentage of points inside thresholds levelling off.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-M-1-2026-31-2026 | © Author(s) 2026. CC BY 4.0 License.

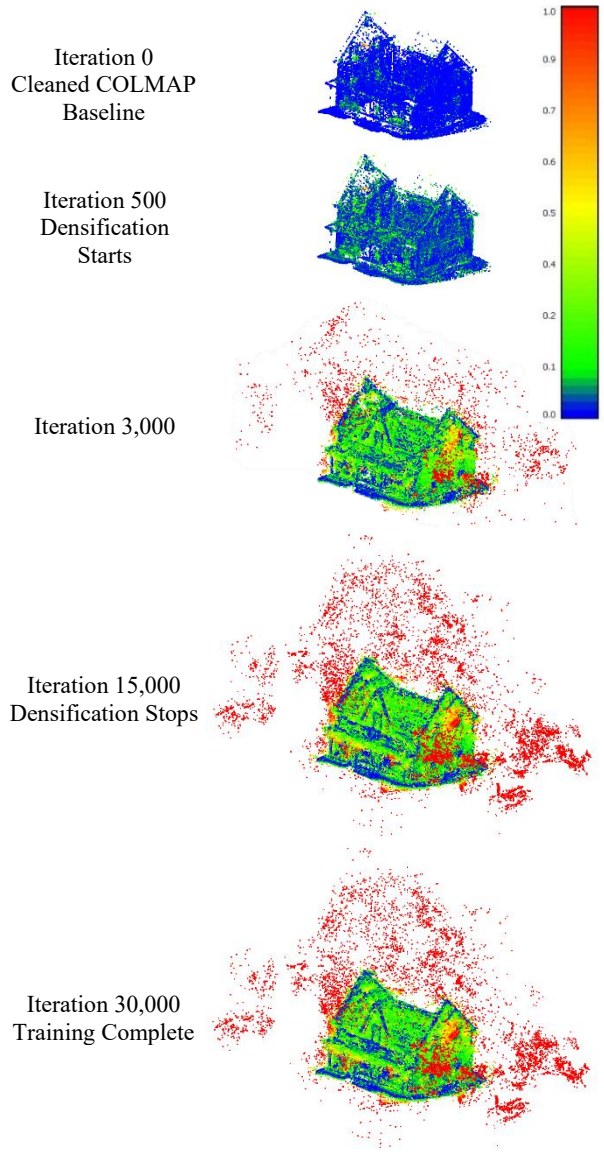

Figure 8: Deviation of Gaussian Positions from the Reference Point Cloud During Training in Meters

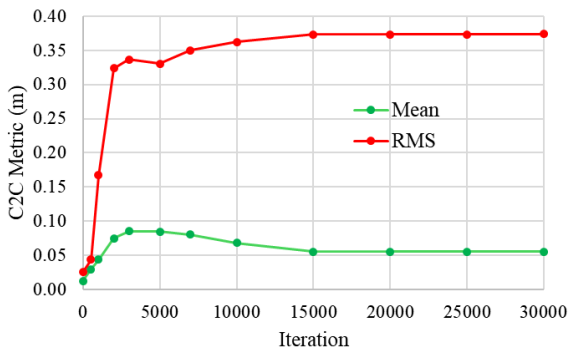

Figure 9: Impact of Iterations on Spatial Accuracy Metrics

4.9 Limitations and Avenues for Further Research

Although several factors affecting the spatial accuracy of building Gaussian splats were analyzed, additional potential factors that could affect accuracy include building facade type, image downsampling, image overlap and positions, and sampling Gaussians instead of only using the mean position. The size of the buildings studied was limited by the memory of the phone LiDAR app and the capability of image matching of COLMAP and Pix4D for buildings with repetitive features. Gaussian splatting is also limited by computer memory and processing, requiring downsampled point clouds and images, and capping the number of images used. Ideally, to produce the most realistic and spatially accurate Gaussian splats, the image quantity and the variety of image locations would increase. Removing sky pixels with image segmentation may decrease the number of floaters in sky regions and provide improved estimates for the visual quality of buildings. Newer end-to-end deep learning reconstruction methods may improve image-based reconstruction results and subsequent Gaussian splats, particularly for large buildings. The findings from this study may differ in regions with buildings with more complex geometry or taller buildings, where adding drone imagery could supplement terrestrial data capture.

Phone LiDAR drift had a sizable impact on subsequent Gaussian splats, even with the use of a gimbal. Future research could delve into decreasing the drift present in phone LiDAR by fusing it with image-based reconstruction methods. The lack of scale in the image-based reconstructions led to unscaled Gaussian splats, which would need to be addressed if measurements were desired, as there is rarely TLS ground truth available. The noise in the COLMAP reconstruction could be reduced by incorporating depth information from phone LiDAR.

5. Conclusion

The spatial accuracy of building Gaussian splatting was heavily dependent on the quality and source of the input point cloud and camera poses, the number of images and training iterations, and the Gaussian splat method. The IMU and SLAM algorithms in phone LiDAR were insufficient at correcting drift, with image-based reconstructions producing higher quality reconstructions and Gaussian splats. Despite this, phone LiDAR still offered a scaled Gaussian splat and images suitable for building Gaussian splatting. The open-source COLMAP had the best performance on visual similarity metrics. However, the default COLMAP sparse point cloud used in Gaussian splatting research is noisy, increasing the number of floaters and decreasing the spatial accuracy for building Gaussian splats. Pix4D produced a cleaner initial point cloud than COLMAP and still scored highly on visual metrics for building 2, but it struggled to match images on the building with fewer features, decreasing its visual and spatial quality for building 1. Downsampling the initial point cloud had little impact on visual metrics and led to a higher percentage of Gaussians farther from the original points. Increasing the number of images improved visual and spatial results until three loops around the building were completed. The positions of Gaussians deviated from the initial point cloud, particularly during the early training stages. 2D GS provided some geometric constraints, but it was still insufficient in terms of floaters and sacrificed visual similarity. For the best spatial building Gaussian splat, a cleaned COLMAP reconstruction with images taken at different locations with geometric constraints would be ideal. This research showed that building Gaussian splatting has yet to become a mapping tool, but it is still a tremendous visualization tool, and its spatial accuracy may improve in the future as more techniques are developed to restrict the movement of Gaussians.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-M-1-2026-31-2026 | © Author(s) 2026. CC BY 4.0 License.

References

Askar, C., Sternberg, H., 2023: Use of Smartphone Lidar Technology for Low-Cost 3D Building Documentation with iPhone 13 Pro: A Comparative Analysis of Mobile Scanning Applications. *Geomatics*, 2023, 3(4), 563-579. doi.org/10.3390/geomatics3040030.

Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P., Hedman, P., 2022: Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields. Presented at the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 5460–5469. doi.org/10.1109/CVPR52688.2022.00539.

Billi, D., Caroti, G., Piemonte, A., 2025: Metric Error Assessment Regarding Geometric 3D Reconstruction of Transparent Surfaces via SfM Enhanced by 2D and 3D Gaussian Splatting. *Sensors*, 2025, 25(14), 4410. doi.org/10.3390/s25144410.

Chen, G., Wang, W., 2025. A Survey on 3D Gaussian Splatting. doi.org/10.48550/arXiv.2401.03890.

Elhashash, M., Albanwan, H., Qin, R., 2022: A Review of Mobile Mapping Systems: From Sensors to Applications. *Sensors*, 22, 4262. doi.org/10.3390/s22114262.

Horé, A., Ziou, D., 2013: Is there a relationship between peak-signal-to-noise ratio and structural similarity index measure? *IET Image Processing*, 7, 12–24. doi.org/10.1049/iet-ipr.2012.0489.

Huang, B., Yu, Z., Chen, A., Geiger, A., Gao, S., 2024: 2D Gaussian Splatting for Geometrically Accurate Radiance Fields. Presented at the SIGGRAPH '24: Special Interest Group on Computer Graphics and Interactive Techniques Conference, ACM, Denver CO USA, 1–11. doi.org/10.1145/3641519.3657428.

Jensen, R., Dahl, A., Vogiatzis, G., Tola, E., Aanaes, H., 2014: Large Scale Multi-view Stereopsis Evaluation. Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 406–413. doi.org/10.1109/CVPR.2014.59.

Kerbl, B., Kopanas, G., Leimkuehler, T., Drettakis, G., 2023: 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Trans. on Graph.*, 42, 1–14. doi.org/10.1145/3592433.

Knapitsch, A., Park, J., Zhou, Q.-Y., Koltun, V., 2017: Tanks and temples: benchmarking large-scale scene reconstruction. *ACM Trans. on Graph.*, 36, 78:1-78:13. doi.org/10.1145/3072959.3073599.

Luetzenburg, G., Kroon, A., Bjørk, A.A., 2021: Evaluation of the Apple iPhone 12 Pro LiDAR for an Application in Geosciences. *Sci. Rep., 11, 22221*. doi.org/ 10.1038/s41598-021-01763-9.

Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R., 2022: NeRF: representing scenes as neural radiance fields for view synthesis. *Commun. ACM*, 65, 99–106. doi.org/10.1145/ 3503250.

Mora-Félix, Z.D., Rangel-Peraza, J.G., Monjardín-Armenta, S.A., Sanhouse-García, A.J., 2024: Performance and precision analysis of 3D surface modeling through UAVs: validation and comparison of different photogrammetric data processing software. *Phys. Scr.*, 99, 035017.doi.org/10.1088/1402-4896/ad23ab.

Schönberger, J.L., 2025a. Output Format — COLMAP 3.13.0.dev0 | a5332f46 (2025-07-05) documentation. colmap.github.io/format.html (November 1, 2025).

Schönberger, J.L., 2025b. COLMAP. github.com/colmap/colmap (May 1, 2025).

Schönberger, J.L., Frahm, J.-M., 2016. Structure-From-Motion Revisited. Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4104–4113. doi.org/10.1109/CVPR.2016.445.

Setiadi, D.R.I.M., 2021: PSNR vs SSIM: imperceptibility quality assessment for image steganography. *Multimed. Tools Appl.*, 80, 8423–8444. doi.org/10.1007/s11042-020-10035-z.

Tamimi, R., 2022: Relative Accuracy Found Within iPhone Data Collection. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLIII-B2-2022, 303–308. doi.org/10.5194/isprs-archives-XLIII-B2-2022-303-2022.

Wolf, P.R., Dewitt, B.A., Wilkinson, B.E., 2014: Elements of Photogrammetry With Applications in GIS, Fourth. ed. McGraw-Hill Education.

Xiong, B., Ye, X., Tse, T.H.E., Han, K., Cui, S., Li, Z., 2024: SA-GS: Semantic-Aware Gaussian Splatting for Large Scene Reconstruction with Geometry Constrain. doi.org/10.48550/arXiv.2405.16923.

Yang, X., Ji, D., Li, Y., Guo, J., Guo, Y., Xie, J., 2025: SGCR: Spherical Gaussians for Efficient 3D Curve Reconstruction. doi.org/10.48550/arXiv.2505.04668.

Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O., 2018: The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. Presented at the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 586–595. doi.org/10.1109/CVPR.2018. 00068.

Zhou, X., Lin, Z., Shan, X., Wang, Y., Sun, D., Yang, M.-H., 2024. DrivingGaussian: Composite Gaussian Splatting for Surrounding Dynamic Autonomous Driving Scenes. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).