The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-2/W12-2026
11th Intl. Workshop 3D-ARCH "3D Virtual Reconstruction and Visualization of Complex Architectures", 10–12 February 2026, Ancona, Italy

Metric Assessment of 3D Gaussian Splatting for UAV-Based Urban Heritage Reconstruction

Widiatmoko A. Fadilah [1], Arnadi Murtiyoso [1], Tania Landes [1], Pierre Grussenmeyer [1]

[1] Université de Strasbourg, CNRS, INSA Strasbourg, ICube Laboratory UMR 7357,
Photogrammetry and Geomatics Group, 67000, Strasbourg, France.
(widiatmoko-azis.fadilah, arnadi.murtiyoso, tania.landes, pierre.grussenmeyer)@insa-strasbourg.fr

Keywords: Photogrammetry, MVS, Radiance Field, Gaussian Splatting, Urban Heritage.

Abstract

Urban heritage increasingly faces threats from natural disasters and anthropogenic activities, demanding preservation techniques that balance affordability, speed, and accuracy. Real-time rendering with 3D Gaussian Splatting (3DGS) offers immense potential for rapid heritage documentation; however optimal training configurations especially for aerial nadiral images remain limited. This study compares three 3DGS implementations: native 3DGS, Postshot Splat3, and MCMC variants against conventional multi-view stereo using an 80-meter altitude drone aerial imagery. Results reveal some interesting key findings where moderate downsampling of 1/8 resolution can substantially outperform full resolution training. Postshot Splat3 achieved optimal photometric quality while training 21x faster than native implementations. However, geometric analysis showed that rendering quality and geometric accuracy are not correlated. Furthermore, the result suggested that MCMC configurations failed to compete with native 3DGS and Postshot Splat3. These findings suggest that 3DGS-based heritage documentation requires a rather conservative resolution selection and optimized implementation for efficiency to allow cost effective preservation while maintaining dimensional fidelity.

1. Introduction

Urban heritage represents invaluable assets by portraying human achievements over centuries. However, they face constant threats from natural and human activities. Systematic conservation effort is therefore necessary to help the preservation purposes. Digital documentation has emerged as a viable strategy, driven by rapid advances in sensor technology and data processing software, which enable sophisticated approaches to heritage recording.

Among recent developments, radiance field methods such as 3D Gaussian Splatting (3DGS) present an interesting solution. Introduced by Kerbl et al. (2023), 3DGS is a scene representation method using 3D Gaussians that preserves the desirable properties of continuous volumetric fields for scene optimization while avoiding unnecessary computation in the empty space. Through opacity and position optimization, the Gaussians can be created, destroyed, or displaced to flexibly represent complex geometry. Thus, it enables real-time rendering capabilities that are essential for 3D mapping applications.

Recent applications of 3DGS show mixed results. Murtiyoso and Macher (2025) demonstrated promising orthophoto generation capability compared to conventional Multi View Stereo (MVS), though noting challenges with large, continuous surfaces. Jiang and Shao (2023) found 3DGS enhanced Unmanned Aerial Vehicles (UAV)-based reconstruction quality with reduced processing time. Conversely, Haitz et al. (2024), found that traditional photogrammetry outperformed modern methods (NeRF, 3DGS) for nadir imagery of large-scale urban scenes across various metrics.

To serve as a baseline process and dataset for our future research, this study investigates several 3DGS training strategies for reconstructing urban heritage sites. Meanwhile, urban heritage is selected as they present challenges in distinguishing historical structures from surrounding standard urban environments. We evaluate 3DGS performance on UAV datasets and assess both their quantitative and qualitative results through comparative analysis with MVS as a reference.

2. Related Works

Traditional photogrammetric reconstruction methods will be reviewed, followed by modern radiance field approaches that enable photorealistic novel view synthesis. The core concepts and variants of 3DGS employed in this study are discussed, emphasizing their distinctions from conventional techniques.

2.1 Traditional Photogrammetric Reconstruction

Photogrammetric reconstruction through Structure from Motion (SfM) and Multi-View Stereo (MVS) has been the predominant approach for heritage documentation over the past decades (Capolupo, 2021). SfM is a photogrammetric technique that simultaneously recovers camera interior and exterior orientations from overlapping imagery through automated feature matching and bundle adjustment optimization, determining 3D camera positions and sparse point clouds with minimal user input, providing the geometric foundation for MVS (Granshaw, 2020). MVS constitutes the dense reconstruction phase that generates detailed 3D point clouds from calibrated multi-view imagery by establishing pixel-wise correspondences across images to triangulate dense surface representations (Furukawa and Ponce, 2010; Granshaw, 2020). The resulting explicit geometric representations such as point clouds or triangulated meshes from MVS allows metric measurements and geometric analysis.

2.2 Radiance Field Methods for Novel View Synthesis

Recent advances in neural scene representation have introduced radiance fields as an alternative paradigm for 3D reconstruction and rendering. A radiance field represents scene content as a continuous volumetric function mapping 3D spatial coordinates and viewing directions to color and density values, enabling photorealistic novel view synthesis through volume rendering (Mildenhall et al., 2020; Torralba et al., 2024). There are several computing architecture of radiance fields with Neural Radiance Fields (NeRFs) by Mildenhall et al., (2020) and 3DGS by Kerbl et al., (2023) being the most prominent. For this study, we focus on the latter architecture, which has better computational efficiency compared to NeRF and is suitable for interactive heritage visualization due to its faster rendering capability. We evaluate the native implementation by Kerbl et al. (2023)

https://doi.org/10.5194/isprs-archives-XLVIII-2-W12-2026-143-2026 | © Author(s) 2026. CC BY 4.0 License.

alongside recent developed training strategies in challenging scenarios typical of heritage documentation.

Experiments were performed on native 3DGS introduced by Kerbl et al. (2023), Markov Chain Monte Carlo (MCMC) – optimized 3DGS developed by Kheradmand et al. (2025), and Splat3 implemented in the commercial software Jawset PostShot. Fundamentally, 3DGS takes SfM points as inputs to allow Gaussians initializations. Each Gaussians has geometric parameters and appearance parameters. Geometric parameters consist of position and covariance matrix, defining location, shape, size, and orientation. Appearance parameters include spherical harmonics (SH) coefficients encoding view-dependent color and opacity (alpha value) controlling transparency of each splat (Kerbl et al., 2023). Each strategy has its own distinctive architecture that differentiates between one and another which is described as follows:

Native – This method was introduced by Kerbl et al. (2023). During the training process, native 3DGS employs adaptive density control (ADC) and differentiable tile rasterizer (DTR) to reconstruct Gaussians to achieve optimal outcome using stochastic gradient descent in minimizing its loss function. ADC allows the generation of a dense set of 3D Gaussians by pruning Gaussians with opacity less than a set threshold, cloning Gaussians in under-reconstructed areas, as well as splitting Gaussians in over-reconstructed areas. To support faster rendering and sorting, this algorithm implements DTR that allows efficient backpropagation over an arbitrary number of blended Gaussians with low memory consumptions. Therefore, the computation of loss function through raster differentiation could be done faster.

MCMC – In the original paper by Kheradmand et al. (2025), the authors consider that the pruning process through resetting Gaussians opacity to small values to remove floaters is an intricate mechanism because it requires careful hyperparameter tuning and may fail in some cases that was mentioned in the paper. It is also stated that the need for good initial input point clouds and the limitation of controlling the computation memory and budget through estimating the number of 3D Gaussians that will be generated becomes a problem especially when applying the native algorithm to the real-world scenes. Thus, the main difference with the original strategy proposed by MCMC is to replace the ADC mechanism which is more heuristic and introduce a more probabilistic approach based on the Markov chain Monte Carlo sampling method to place and optimize Gaussians using Stochastic Gradient Langevin Dynamics (SGLD).

Through SGLD, a stochastic noise can be injected only into the position of Gaussians to help it escape local minima and sample on locations that truly help reconstruct the scene. This therefore encourages Gaussians generation in an area that has higher probability (high opacity). This introduction of noise would also help in improving the result of 3DGS generation on an SfM input which suffers from missing geometry. It is achievable through relocating low opacity Gaussians to high opacity regions, unlike ADC mechanisms which delete it. At the same time, this mechanism allows generation based on the maximum number of Gaussians set in the beginning of training, encouraging a fewer number of Gaussians and allowing control over the computing resource that will be utilized.

This study uses the MCMC training strategy implemented on both LichtFeld Studio (LFS) (LichtFeld Studio, 2025) and Jawset PostShot.

Splat3 – Splat3 is developed by Jawset and can be accessed through its commercial application PostShot (PS). It is one of the radiance field profiles that can be used to reconstruct Gaussians with fine details in both foreground and background (https://www.jawset.com, accessed on: 5 January 2026). Unfortunately, because it is a commercial application, the exact implementation of the method is not open.

3. Data and Methodology

3.1 Data Acquisition

The data used in this study were obtained from an aerial drone acquisition at the Kasepuhan Palace site, Cirebon, Indonesia. The site presents several documentation challenges and constraints related to weather, specific architectural features, and dense vegetation surrounding the objects, as previously mentioned in Murtiyoso et al. (2018). The data consisted of 368 nadiral images captured at 80 m flight height.

3.2 Data Processing

The processing workflow comprised two parallel branches following initial image orientation: 3D Gaussian generation using 3DGS and dense point cloud reconstruction using conventional MVS. Camera orientation was performed in Agisoft Metashape using high-quality settings, with the oriented cameras and sparse point cloud subsequently exported to the COLMAP format as required for 3DGS processing. The orientation result can be seen in Figure 1. The computation resulted in 1.1 pix. RMS reprojection error with a total of around 1.5 million tie points. The Metashape workflow was continued independently to generate a dense point cloud serving as the reference baseline for comparative assessment.

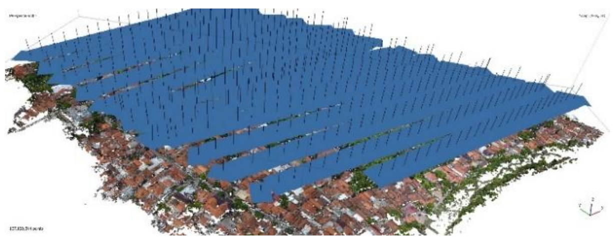


Figure 1. Oriented input images.

3DGS training and processing were conducted on a workstation with the following specifications: Intel Xeon Gold 5220R (2.20 GHz), 64 GB RAM, and NVIDIA Quadro RTX 5000 (16 GB VRAM, Compute Capability 7.5). Software requirements varied by implementation: native 3DGS requires Compute Capability ≥7.0 and 24 GB VRAM; LFS requires ≥7.5 and 8 GB VRAM; PS requires ≥7.5 with unspecified VRAM requirements.

To ensure comparability, defaults parameters were used across all selected training strategies. Given hardware constraints, images were downsampled to 1/8 and 1/4 resolution for native 3DGS and LFS. Additionally, 1/2 resolution (LFS) and full resolution (PS) were tested to leverage their optimized algorithms.

Furthermore, to evaluate the relationship between training duration, reconstruction quality, and Gaussian count, models were saved every 1,000 iterations up to 15,000 (15K). This endpoint was selected based on preliminary testing showing that on our dataset, native 3DGS reached its maximum Gaussian count (default cap) by 15K iterations. Beyond this point, training primarily refined existing Gaussians (orientation and covariance) with minimal visual improvement (Figure 2). Additionally, the maximum number of Gaussians were also set based on this output. Thus, 20M Gaussians were set as the maximum cap value during the training of PS Splat3, PS MCMC, and LFS MCMC.

Moreover, several optimizations available in native 3DGS were tested during the training such as depth regularization (DR), exposure compensation (EC), and Adam optimizer. DR and EC consecutively are used to address untextured regions and to handle lighting variations across images (Kerbl et al., 2024). Meanwhile, Adam optimizer was applied on all native 3DGS configurations for improved convergence efficiency (Mallick et al., 2024). These configurations were tested to obtain the best possible result from native 3DGS and investigate their effects on computational performance and final output. The following Table 1 summarizes all training configurations.

Method	Resolution(s)	DR	EC	Notes
Native	1/8, 1/4	Y/N	Y/N	Default parameters, Adam optimizer, ablations study on DR/EC
PS Splat3	1/8, 1/4, 1/1	-	-	Default parameters
PS MCMC	1/1,	-	-	Default parameters
LFS MCMC	1/8, 1/2	-	-	Default parameters

Table 1. Summary of training configurations used in this study. Note that all configurations are stopped at 15K iterations; DR and EC are available only on native 3DGS.

3.3 Analysis Method

The performance of 3DGS training strategies is compared against MVS through four complementary assessments:

Computational Efficiency – Training time is measured only from the generation of 3DGS model, excluding shared SfM preprocessing. For Metashape, dense point cloud generation time is recorded for comparison.

Rendering Quality – Novel view synthesis quality is assessed between 3DGS variants using held-out test images with three metrics (Ding et al., 2021): **PSNR** (Peak Signal-to-Noise Ratio) measures pixel-wise accuracy, **SSIM** (Structural Similarity Index) evaluates structural similarity, and **LPIPS** (Learned Perceptual Image Patch Similarity) quantifies perceptual quality. Higher PSNR/SSIM and lower LPIPS indicate better quality. Metashape is excluded as MVS does not support novel view rendering.

Qualitative Assessment – Visual reconstruction quality is evaluated through structured observation of detail preservation, completeness, and artifact presence, with emphasis on architecturally complex regions and vegetation-occluded areas.

Geometric Accuracy – Point cloud accuracy is evaluated using M3C2 distance analysis (James et al., 2017) in CloudCompare, with Metashape MVS output serving as reference. Mean distance (bias) and standard deviation (precision) are reported. Since native 3DGS and LFS outputs align directly with SfM coordinates, no transformation or registration is necessary. Meanwhile PS coordinates are corrected using logged transformation parameters.

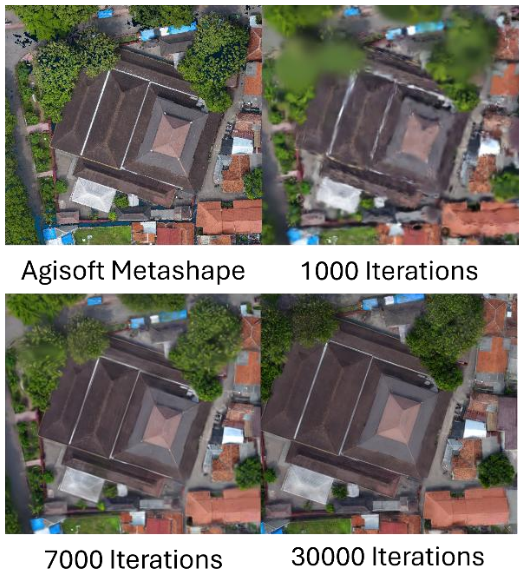


Figure 2. Visual comparison output between MVS's point clouds and 3DGS's rendered Gaussians on several iteration outputs, demonstrating progressive refinement towards photorealistic quality.

4. Results and Discussions

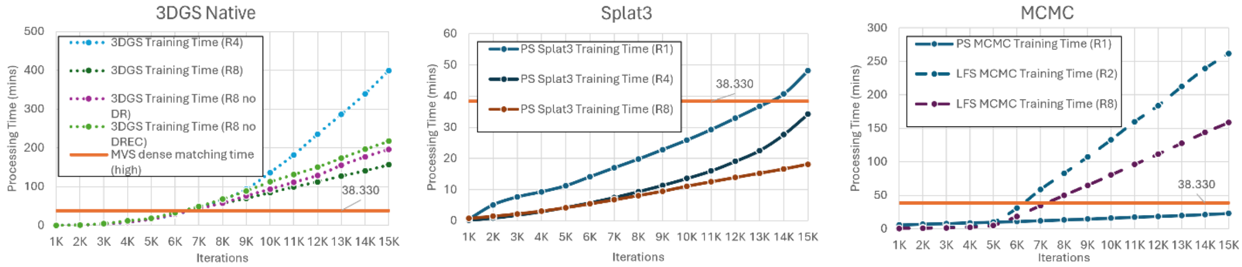
In this section experiment results will be presented through assessment comparison between three 3DGS training strategies mentioned in the previous section. The results will be organized according to the categories established in Section 3.3.

4.1 Computational Efficiency

Figure 2 shows the visual comparison between the reference dense point cloud and the native 3DGS rendered output at several iterations. The visualization demonstrates how 3DGS algorithms gradually generate photorealistic scene representations. This preliminary assessment indicates that native 3DGS achieves visually comparable quality to Metashape's MVS by approximately 10K iterations, with diminishing visual improvements thereafter. Meanwhile, computational efficiency between several 3DGS training configurations can be reflected in Figure 3.

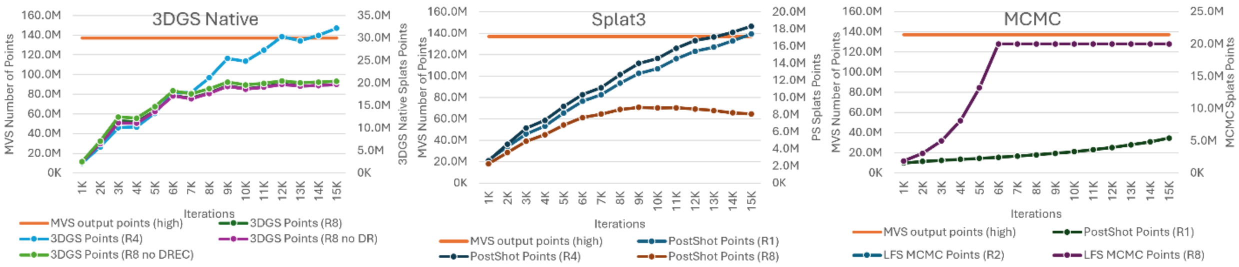
Figure 3 compares training times required to reach 15K iterations across all tested configurations, normalized against Metashape's MVS processing time (38 minutes). The figure presents several training configurations with variety of resolutions: 1/1 resolution (R1), 1/2 resolution (R2), 1/4 resolution (R4), and 1/8 resolution (R8). At R8, native 3DGS, PS Splat3, and LFS MCMC required 5.0, 0.5, and 4.1 times the MVS processing time, respectively. At higher resolutions (R4), native 3DGS and LFS MCMC exhibited significantly longer training times at 10.4 and 6.8 times the MVS baseline. In contrast, PS's optimized implementations demonstrated competitive performance: Splat3 at full resolution (R1) required only 1.3 times the MVS time, while MCMC at R1 completed training 1.7 times faster than MVS. Overall, native 3DGS and LFS MCMC underperformed compared to Metashape's MVS in terms of computational efficiency, while PS implementations achieved comparable or superior processing times despite operating at higher resolutions.

https://doi.org/10.5194/isprs-archives-XLVIII-2-W12-2026-143-2026 | © Author(s) 2026. CC BY 4.0 License.



(a) (b) (c)

Figure 3. Processing time comparisons across training iterations for different 3DGS implementations compared to Metashape's MVS dense point cloud as reference. (a) Native 3DGS configurations with DR and EC ablation study. (b) PS Splat3 at different resolution settings. (c) MCMC implementations from different software, LFS and PS. The horizontal line (orange line), in every chart shows the MVS processing time at 38.3 mins.



(a) (b) (c)

Figure 4. Number of Gaussian counts comparisons across training iterations for different 3DGS implementations compared to Metashape's MVS dense point cloud as reference. (a) Native 3DGS configurations with DR and EC ablation study. (b) PS Splat3 at different resolution settings. (c) MCMC implementations from different software, LFS and PS. The horizontal line (orange line), in every chart shows the MVS number of points at 140M points.

Figure 4 presents Gaussians count evolution across training iterations in three panels, with each panel showing the Metashape's MVS point count at >140M points as reference. Native 3DGS configurations at the same resolution exhibit similar growth patterns, plateauing around 20M Gaussians after 10K iterations with minor variations between configurations (Figure 4a). Notably, the configuration without DR and EC produced the highest gaussian count at this plateau, suggesting these optimizations improve scene representation efficiency by reducing redundancy while maintaining reconstruction quality. This is also confirmed in Table 2 which shows a small SSIM difference between configurations that apply these optimizations.

Among R8 configurations, PS Splat3 generated the fewest Gaussians (~8M), despite a maximum cap of 20M, while following a similar plateau behaviour after 10K iterations (Figure 4b). In contrast, LFS MCMC demonstrated aggressive growth, reaching the 20M cap by 6K iterations and remaining constant thereafter (Figure 4c). An interesting finding emerges when comparing MCMC implementations from different software, where PS MCMC at full resolution produced only ~5M Gaussians, significantly fewer than LFS MCMC at lower resolutions. This difference indicates implementation specific optimizations or parameter tuning strategies between the two MCMC variants, despite both claiming to use the same underlying algorithm.

At higher resolutions, LFS MCMC maintained similar Gaussians counts and growth patterns regardless of resolution, presented with different lines which are stacked on top of another (Figure 4c). Meanwhile, native 3DGS and PS Splat3 produced proportionally more Gaussians at higher resolution. A key distinction appears between native 3DGS and PS Splat3, because native 3DGS does not have maximum gaussian cap in its default implementation, it reached approximately 32M Gaussians, whereas PS Splat3 remained constrained below the 20M cap. An unexpected finding which requires further investigation is that PS Splat3 at R4 generated more Gaussians (~20M) than at full resolution (~18M), contradicting the typical expectation that higher resolution demand more primitives for detail preservations. This may be attributed to PS's adaptive densification strategy responding differently to resolution-dependent gradient patterns.

4.2 Rendering Metrics and Qualitative Assessment

As we can see in Figure 4, compared to the MVS dense point cloud (~140M), all 3DGS variants demonstrated substantial memory efficiency. The highest Gaussians count observed is ~32M for native 3DGS at R4, representing a 4.4x reduction in scene primitives. Meanwhile, most configurations operated with considerably fewer Gaussians, ranging from 5M to 20M. This compactness, combined with 3DGS's real-time rendering capability, poses a significant advantage over traditional MVS representation for applications requiring interactive visualization or real-time processing.

However, Gaussians count alone does not determine reconstruction quality. Fewer Gaussians may indicate either efficient scene representation or insufficient capture of details. More Gaussians could represent either better detail preservation or redundant primitives. The wide variation in Gaussians counts of this study raises a question about the relationship between primitive count and reconstruction fidelity. To evaluate whether these configurations achieve favorable quality-efficiency trade-offs, quantitative rendering metrics (PSNR, SSIM, and LPIPS) are presented in Table 2.

Rendering Metrics – From Table 2, we can see that native 3DGS configurations at R8 demonstrate strong performance, with the standard configuration where DR and EC are enabled achieve high score of SSIM (~0.81) and PSNR (~25.99 dB). The ablation study demonstrates that DR and EC enhance scene representation efficiency. The configurations employing both optimizations achieve comparable performance to those using EC alone, while configurations lacking these optimizations exhibit slightly degraded performance. As we mentioned previously, the marginal SSIM difference of 0.011 between optimized and non-optimized configurations, combined with the observed reduction in Gaussians count (Figure 4), demonstrates that DR and EC enable more efficient scene representation, where they achieve comparable visual fidelity with approximately 5-10% fewer Gaussians.

Training Strategy	SSIM ↑	PSNR ↑	LPIPS ↓
3DGS R8 no DR EC	0.80	24.96	0.18
3DGS R8 no DR	0.81	26.11	0.17
3DGS R8	0.81	25.99	0.17
3DGS R4	0.70	24.06	0.29
PS Splat3	0.48	N/A	N/A
PS Splat3 R4	0.80	N/A	N/A
PS Splat3 R8	0.93	N/A	N/A
PS MCMC	0.40	N/A	N/A
LFS MCMC R2	0.49	18.21	0.67
LFS MCMC R8	0.36	17.54	0.54

Table 2. Rendering metric comparisons between several training strategies. Higher PSNR/SSIM and lower LPIPS are better.

Among all tested configurations, PS Splat3 at R8 achieved the highest SSIM score, indicating superior photometric reconstruction quality despite having only ~8M Gaussians. This result suggests that PS's proprietary densification strategy effectively balances detail preservation with memory usage. However, the absence of PSNR and LPIPS metrics for PS limits comprehensive quality assessment. This is because PS's software does not expose this metrics in its standard output, providing only SSIM values for model evaluation. Thus, preventing direct multi-metric validation against other implementations, despite not diminishing the validity of SSIM-based comparisons.

A critical finding emerges across both native 3DGS and PS Splat3 implementation, whereas higher resolutions consistently produced worse rendering quality than moderate downsampling levels. Native 3DGS at R4 produces substantially worse rendering metrics than the R8 configuration, despite generating 60% more Gaussians. Therefore, showing a 14% decrease in SSIM, 1.93 dB drop in PSNR, and 70% increase in perceptual error. Similarly, PS Splat3 exhibits a non-monotonic relationship between resolution and quality, as R8 achieved the highest SSIM and R4 produced moderate quality (SSIM ≈ 0.80). Meanwhile, full resolution resulted in the poorest performance, showing 84% quality degradation from the optimal R8 configuration.

This result contradicts the intuitive expectation where higher-resolution inputs should enable better reconstruction quality. However, recent literature shows similar observations in 3DGS training, particularly for aerial and large-scale scenes. Multiple studies reporting that moderate downsampling, i.e., R4 to R8 outperforms full resolution in both quantitative metrics and training stability. Some discussions mentioned that adjusting learning rates might produce better results, which suggests a suboptimal performance at higher resolutions when using default hyperparameters. A study by (Zhao et al., 2025), specifically stated that, without tuning optimization, hyperparameters can lead to unstable and inefficient training. The default learning rates of 3DGS were calibrated in the 1-1.6K pixel range, which explain why a steep performance drop at full resolution of our dataset (3840 pixel).

Moreover, Table 2 clearly shows that all MCMC implementations are substantially inferior in terms of rendering quality compared to native 3DGS and PS Splat3. With SSIM values ranging from 0.36 to 0.49 which are way too low than the better performing configurations that range from 0.80 to 0.93. LFS MCMC at R8 recorded the poorest metrics overall. PS MCMC at full resolution and LFS MCMC at R2, reached SSIM score which remained far below acceptable reconstruction quality. This shows that the poor performance persists regardless of Gaussian count, with both 5M and 20M Gaussians exhibiting similar photometric failures.

These results indicate systematic failure rather than software specific issues, as each MCMC implementation produced similarly poor outcomes. The failure correlates with the "explosion artifacts" and far-field floaters observed in qualitative assessment which can be seen in Figure 6. MCMC's unbounded stochastic exploration mechanism generates Gaussians far from the actual scene geometry. As mentioned in (Yao et al., 2026), large distances and sparse viewing angles can cause the model to easily produce floaters and have overgrowth issues. Thus, MCMC's probabilistic sampling exacerbates this problem rather than mitigates. The problem is also added with the fact that in aerial nadir scene lacks sufficient geometric constraints to prevent divergence into physically implausible regions. Furthermore, MCMC's relocation mechanism moving Gaussians to far-field regions which not contributing to meaningful scene reconstruction, consistent with the explosion artifacts observations.

Figure 5. Characteristic 3DGS artifacts across training configurations.

Qualitative Assessments – Figure 5 presents characteristic 3DGS artifacts appearing in the reconstructions after 15K iterations. Visual quality is evaluated through detailed inspection of representative regions.

Native 3DGS demonstrates the most photorealistic output, particularly at higher resolutions (R4), with accurate highdefinition texture reproduction on roof tiles, vegetation canopies, and architectural details. However, the R8 configuration exhibits characteristic Gaussian texture artifacts manifesting as blob-like patterns on uniform surfaces, such as ground pavements, roof tiles, and dense vegetation. These artifacts result from insufficient Gaussian density to represent high-frequency texture details at reduced resolution. Despite artifacts, scene completeness remains high with no missing geometry.

PS Splat3 produces clean reconstructions free from Gaussian texture artifacts across all resolutions, indicating superior densification strategy. Visual quality scales predictably with resolution, R8 shows softer edges and reduced fine detail (individual roof tiles, ornamental patterns), while R4 and R1 progressively recover sharpness. The absence of blob-like artifacts suggests Postshot's algorithm better handles the texture-geometry decomposition inherent to 3DGS optimization.

MCMC implementations achieve complete scene coverage without obvious missing regions but suffer severe visual fidelity degradation. LFS MCMC R2 exhibits blurred textures on vegetation canopies and roof structures, with color bleeding between adjacent surfaces indicating poor Gaussian localization. LFS MCMC R8 presents a noisy appearance on flat color regions that lacking texture detail, which is likely due to low-resolution input limitations and MCMC's failure to adequately densify surface details on complete nadiral images.

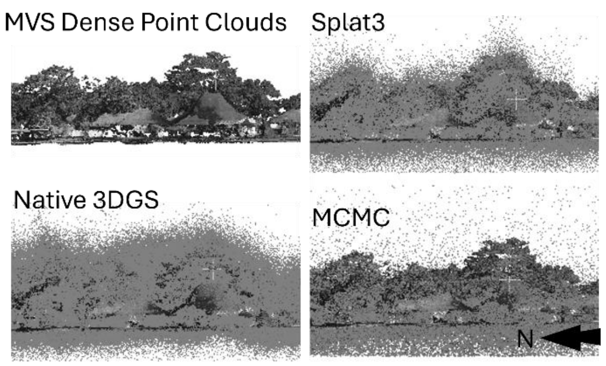


Figure 6. Point cloud comparison of 3DGS training strategies after cropping to analysis bounding box.

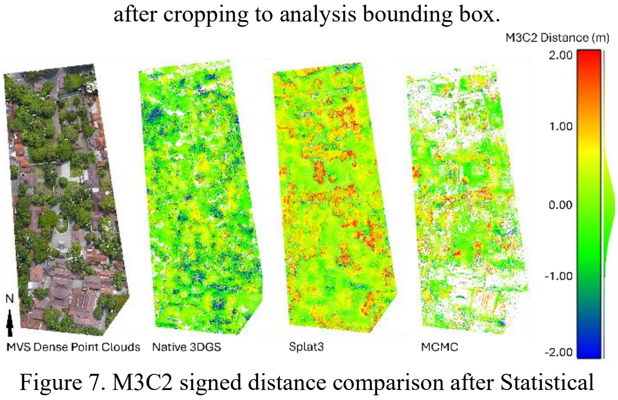


Figure 7. M3C2 signed distance comparison after Statistical Outlier Removal and normalized standard deviation filtering.

Furthermore, Figure 6 reveals point cloud spatial distribution among methods. Native 3DGS generates extensive floating artifacts detached from scene geometry, while PS Splat3 supresses these artifacts, positioning Gaussians closer to actual surfaces. This is also reflected in Table 3 where all original mean distance from PS implementations has lower mean and standard deviation, which explaining superior visual fidelity. From the figure, MCMC exhibits characteristics “explosion artifacts” even after the result is cropped to the analysis bounding box. These geometric failures directly correlate with MCMC’s poor rendering quality, as mispositioned Gaussians cannot reconstruct coherent surface appearance regardless of optimization iterations.

Moreover, Figure 7 depicts M3C2 signed distance analysis after Statistical Outlier Removal (SOR) and normalized filtering. The figure shows the result after filtering process is applied to the M3C2 output. Native 3DGS shows systematic negative bias (blue area) on steep roofs and vegetations. Splat3 shows positive bias (red area) on identical features. MCMC displays extensive filtered region (white area) indicating geometric reconstruction failure on complex structures.

While Figures 5 to 7 provide qualitative assessment of rendering quality and spatial distribution, quantitative geometric accuracy assessment requires systematic analysis. The following section will present detailed M3C2 distance statistics (Table 3) across all configurations to complement visual observation we just discussed

4.3 Geometric Accuracy

To evaluate geometric accuracy, M3C2 distance analysis was performed comparing each 3DGS reconstruction against the Metashape’s MVS dense point cloud as reference. Given that 3DGS representations inherently generate floating artifacts and outliers beyond the actual scene geometry, a three-stage cleaning pipeline was applied prior to M3C2 computation: (1) bounding box cropping to the scene extent, (2) SOR filtering to eliminate noise, and (3) normalized distance filtering based on average standard deviation to remove remaining anomalies. This preprocessing ensures that geometric comparisons reflect true surface reconstruction accuracy rather than artifact density.

Table 3 presents M3C2 distance statistics across all configurations, reporting both original unfiltered clouds and cleaned/filtered results. The “Diff to MVS Point Clouds” columns show mean ($\bar{x}$) and standard deviation (σ) of signed distances, while "Diff After Normalised" presents these values normalized by the average standard deviation (2.04 m) across all methods. The retention rate indicates the percentage of points remaining after filtering, with lower retention rates suggesting higher artifact prevalence.

Native 3DGS configurations at R8 demonstrate comparable geometric accuracy regardless of DR and EC settings. The marginal difference which falls within measurement uncertainty indicates that DR and EC primarily affect photometric rendering quality rather than geometric accuracy. All three configurations retained 46-47% of points after filtering, suggesting similar artifact distribution. The systematic negative bias across all native R8 configuration indicates that 3DGS reconstructions consistently position surfaces slightly below the MVS reference. The standard deviations reflect both genuine discrepancies and residual noise after filtering.

At higher resolution, native 3DGS demonstrates a trade-off between systemic bias and variance. Where it shows large mean with low variance, despite having the highest retention rate among native configurations. This suggests that R4 produces fewer extreme outliers but more pervasive low-magnitude errors throughout the reconstruction. The rendering metric from Table 2 with this result confirms that higher resolution training without hyperparameter adjustment produce inferior reconstruction quality across all dimensions.

Diff. to MVS Point Clouds (M3C2)			Diff. After Normalised		Num. of Points / Splats			
Training Strat.	$\bar{x}$ (m) Original	σ (m) Original	$\bar{x}$ (m) Normalised	σ (m) Normalised	Original	Cleaned	Filtered	Ret.%
3DGS R8 no DREC	-0.58	2.66	-0.20	0.55	20,377,293	10,649,665	9,544,438	46.84%
3DGS R8 no DR	-0.50	2.45	-0.19	0.54	19,684,355	10,358,394	9,419,409	47.85%
3DGS R8	-0.51	2.57	-0.18	0.54	19,685,796	10,335,893	9,342,494	47.46%
3DGS R4	-0.39	1.53	-0.28	0.40	32,162,259	18,007,017	17,345,885	53.93%
PS Splat3	0.38	1.70	0.40	0.48	17,431,687	10,225,289	9,681,510	55.54%
PS Splat3 R4	0.25	1.94	0.32	0.50	18,337,960	11,165,779	10,364,523	56.52%
PS Splat3 R8	0.18	1.94	0.28	0.52	8,091,328	4,524,773	4,211,715	52.05%
PS MCMC	-0.06	1.48	-0.09	0.45	5,415,685	860,354	825,539	15.24%
LFS MCMC R2	0.04	1.13	0.00	0.38	20,000,000	13,724,982	13,405,305	67.03%
LFS MCMC R8	0.48	2.99	0.02	0.58	20,000,000	15,165,298	13,336,734	66.68%
Average Std. Dev.		2.04						

Table 3. Detailed geometric assessment result using M3C2 algorithm which compares several 3DGS training strategies with MVS output as reference.

PS Splat3 configurations on the other hand demonstrated superior geometric accuracy compared to native 3DGS across all resolutions tested. With near-zero mean bias after filtering and the lowest standard deviation among all method, PS Splat3 at R1 exhibits the best geometric accuracy. However, it has the worst rendering metric. This reveals a dissociation between rendering quality and geometric accuracy in 3DGS reconstructions. PS Splat3 at R4 presents the most balanced performance with highest retention across all methods. This rate indicates fewer severe outliers, while the positive bias suggests that PS's algorithm positions surface above rather than below the MVS reference. Meanwhile, at R8 despite having the highest rendering metric, it shows slightly lower geometric accuracy with relatively lower retention rate. Thus, these results suggest overfitting to view dependent appearance at the expense of geometric consistency.

PS MCMC presents a misleading result with near-zero filtered bias while having the lowest retention rate. Around 85% of generated points were filtered as outliers. Which likely represents only the sparse SfM initialization points that MCMC failed to adequately densify, rather than a meaningful reconstruction of a scene. Meanwhile, LFS MCMC demonstrates a poor and unbalanced performance, despite having the highest retention rate among all training configurations. LFS MCMC at R2 presents the questionable zero filtered bias and tightest standard deviation which might suggest that the algorithm only produce around a subset of the scene while failing to reconstruct complex structures (Figure 7). Meanwhile, at R8 LFS MCMC has the largest original standard deviation despite appearing geometrically accurate in mean distance. These experiments suggest that all MCMC variants present poor performance even in geometric accuracy assessment.

4.4 Experiment Limitations

Several constraints qualify the interpretation and generalizability of findings.

Geometric Reference – Dense point cloud from MVS served as the geometric reference rather than terrestrial laser scanning (TLS). MVS dense point cloud might contain systematic errors, thus reported geometric accuracies represents relative performance among implementation rather than absolute dimensional fidelity. Future validation against TLS data would strengthen accuracy claims for surveying applications.

Implementation Versions and Parameters – Native 3DGS, Postshot, and Lichtfeld Studio were evaluated using default parameters without hyperparameter optimization. The recent study by Yan et al. (2025) provides complementary benchmark analysis of advanced 3DGS variants specialized for urban scenes, exploring alternative training strategies beyond the scope of this study.

5. Conclusions

This study comprehensively evaluated three 3DGS training strategies against traditional MVS for aerial nadir photogrammetry, presenting insight into practical applications and theoretical understanding.

Through these experiments, PS Splat3 demonstrated exceptional computational efficiency, with 21 times faster and 60% fewer Gaussians than native 3DGS. Native 3DGS remained computationally intensive without optimization, while DR and EC provide small improvements. PS Splat3 at R4 emerged as the optimal balanced configuration. Meanwhile, the result revealed that MCMC's unbounded stochastic exploration is unsuitable for viewing configurations with limited angular diversity, which contradicts its claims of superior robustness to initialization quality.

Furthermore, the experiment suggested that optimal implementation on aerial photogrammetry requires careful hyperparameter tuning to use high resolution settings. Otherwise, conservative downsampling would be the optimal setting for aerial 3DGS applications.

Finally, geometric analysis shows no correlation between rendering quality and measurement accuracy. This finding indicates that specialized 3DGS configurations for large-scale heritage reconstruction are necessary, either through hybrid workflows separating visualization and surveying tasks, or through novel training strategies balancing photorealism, geometric fidelity, and computational efficiency.

Acknowledgements

This paper was published within the context of the Digital Twins for Cultural Heritage (Twin4CH) Junior Professorship Chair, financed by the French National Agency for Research (*Agence nationale de la recherche* – ANR). We extend our sincere gratitude to Dr. Deni Suwardhi from Institut Teknologi Bandung and his team in acquiring and sharing the Kasepuhan Palace heritage site dataset.

References

Capolupo, A., 2021. Accuracy Assessment of Cultural Heritage Models Extracting 3D Point Cloud Geometric Features with RPAS SfM-MVS and TLS Techniques. Drones 5. https://doi.org/10.3390/drones5040145

Ding, K., Ma, K., Wang, S., Simoncelli, E.P., 2021. Comparison of Image Quality Models for Optimization of Image Processing Systems. Int J Comput Vis 129, 1258–1281. https://doi.org/10.1007/s11263-020-01419-7

Furukawa, Y., Ponce, J., 2010. Accurate, Dense, and Robust Multiview Stereopsis. IEEE Transactions on Pattern Analysis and Machine Intelligence 32, 1362–1376. https://doi.org/10.1109/TPAMI.2009.161

Granshaw, S.I., 2020. Photogrammetric terminology: fourth edition. The Photogrammetric Record 35, 143–288. https://doi.org/10.1111/phor.12314

Haitz, D., Hermann, M., Roth, A.S., Weinmann, Michael, Weinmann, Martin, 2024. The Potential of Neural Radiance Fields and 3D Gaussian Splatting for 3D Reconstruction from Aerial Imagery. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences X-2–2024, 97–104. https://doi.org/10.5194/isprs-annals-X-2-2024-97-2024

James, M.R., Robson, S., Smith, M.W., 2017. 3-D uncertainty-based topographic change detection with structure-from-motion photogrammetry: precision maps for ground control and directly georeferenced surveys. Earth Surface Processes and Landforms 42, 1769–1788. https://doi.org/10.1002/esp.4125

Jiang, C., Shao, H., 2023. Fast 3D Reconstruction of UAV Images Based on Neural Radiance Field. Applied Sciences 13, 10174. https://doi.org/10.3390/app131810174

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

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

Kerbl, B., Meuleman, A., Kopanas, G., Wimmer, M., Lanvin, A., Drettakis, G., 2024. A Hierarchical 3D Gaussian Representation for Real-Time Rendering of Very Large Datasets. https://doi.org/10.48550/arXiv.2406.12080

Kheradmand, S., Rebain, D., Sharma, G., Sun, W., Tseng, J., Isack, H., Kar, A., Tagliasacchi, A., Yi, K.M., 2025. 3D Gaussian Splatting as Markov Chain Monte Carlo. https://doi.org/10.48550/arXiv.2404.09591

LichtFeld Studio, 2025. A high-performance C++ and CUDA implementation of 3D Gaussian Splatting.

Mallick, S.S., Goel, R., Kerbl, B., Steinberger, M., Carrasco, F.V., De La Torre, F., 2024. Taming 3DGS: High-Quality Radiance Fields with Limited Resources, in: SIGGRAPH Asia 2024 Conference Papers, SA '24. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3680528.3687694

Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R., 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. https://doi.org/10.48550/arXiv.2003.08934

Murtiyoso, A., Grussenmeyer, P., Suwardhi, D., Awalludin, R., 2018. Multi-Scale and Multi-Sensor 3D Documentation of Heritage Complexes in Urban Areas. ISPRS International Journal of Geo-Information 7, 483. https://doi.org/10.3390/ijgi7120483

Murtiyoso, A., Macher, H., 2025. Gaussian Splatting for Facade Orthophoto Generation – Comparison with MVS and TLS. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLVIII-M-9–2025, 1059–1064. https://doi.org/10.5194/isprs-archives-XLVIII-M-9-2025-1059-2025

Torralba, A., Isola, P., Freeman, W.T., 2024. Foundations of Computer Vision, Adaptive Computation and Machine Learning series. MIT Press.

Yan, Z., Yin, M., Shao, Y., Remondino, F., 2025. Evaluating 3D Gaussian Splatting for Urban Scene Reconstruction. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLVIII-1-W6-2025, 251–258. https://doi.org/10.5194/isprs-archives-XLVIII-1-W6-2025-251-2025

Yao, Y., Zhang, B., Zhang, W., Gao, L., Peng, D., Li, B., Wang, Y., Wang, B., 2026. ARSGaussian: 3D Gaussian Splatting with LiDAR for aerial remote sensing novel view synthesis. ISPRS Journal of Photogrammetry and Remote Sensing 231, 288–306. https://doi.org/10.1016/j.isprsjprs.2025.10.022

Zhao, H., Weng, H., Lu, D., Li, A., Li, J., Panda, A., Xie, S., 2025. On Scaling Up 3D Gaussian Splatting Training. https://doi.org/10.48550/arXiv.2406.18533