3D representation of Architectural Heritage: a comparative analysis of NeRF, Gaussian Splatting, and SfM-MVS reconstructions using low-cost sensors

Paolo Clini [1], Romina Nespeca [1], Renato Angeloni [1], Laura Coppetta [1]

[1] Dept. of Civil and Building Engineering and Architecture, Università Politecnica della Marche, Ancona, Italy – (p.clini, r.nespeca, r.angeloni)@univpm.it, l.coppetta@pm.univpm.it

Keywords: Architectural Heritage, 3D reconstruction, NeRF, Gaussian Splatting, SfM-MVS, Low-cost sensors.

Abstract

This study evaluates the performance of three image-based 3D reconstruction methods—Structure-from-Motion with Multi-View Stereo (SfM-MVS), Neural Radiance Fields (NeRF), and Gaussian Splatting (GS)—for documenting Architectural Heritage (AH). Data acquisition was conducted using low-cost sensors: a drone to capture the stone portal of the external façade and a 360° camera to document the interior spaces. NeRF and GS significantly outperformed SfM-MVS in processing time, with NeRF excelling in reconstruction completeness. However, GS faced challenges with point number control, and NeRF reconstructions exhibited artifacts and noise, particularly on flat surfaces. Accuracy assessments, using TLS point clouds as a benchmark, revealed that SfM-MVS provided the highest geometric precision for the external façade, despite minor gaps in reconstruction. In contrast, NeRF and GS fell short of the accuracy required for precise geometric documentation, with NeRF exhibiting prominent artifacts in flat or poorly detailed regions. Interior reconstructions were further limited by the higher Ground Sampling Distance (GSD) caused by the technical constraints of the 360° camera and the increased capture distance for elevated areas. In conclusion, we can affirm that while NeRF and GS demonstrate strong potential for visualization due to their rendering quality and efficiency, SfM-MVS remains the most reliable method for achieving accurate geometric documentation of AH.

1. Introduction

3D representation of AH is becoming crucial in fields like conservation, urban planning, and cultural experience. Traditional methods such as Structure-from-Motion (SfM) combined with Multi-view Stereo (MvS) have been used to create detailed 3D models from multiple images (Remondino and El-Hakim, 2006). However, newer, more efficient methods have emerged through neural rendering and machine learning advancements. Two significant approaches, Neural Radiance Fields (NeRF) and Gaussian Splatting, are now prominent alternatives, offering superior rendering quality and the ability to capture complex scene details. NeRF stand out for their ability to synthesize new views of complex scenes by optimizing a continuous volumetric function based on a set of images of the scene itself. The neural network is trained on these images with the goal of minimizing the prediction error, enabling accurate and photorealistic generation of new synthetic views of the scene. The network uses a continuous set of 5D coordinates as input, which describe spatial positions (x, y, z) and viewing directions (θ, φ), and outputs, for each point in space, the volumetric density (σ) and emitted radiance (RGB), modulated according to the viewing direction. This approach enables the generation of an image associated with a particular view by tracing a viewing ray for each pixel of the image to be synthesized. The neural network provides information about the visible scene point by point along the ray. Based on the volumetric density value, it can determine whether the ray passes through empty regions or intersects an object. When the density becomes significant, it identifies a point in the scene that contributes to the final image, determining the pixel color. By repeating this operation for each viewing direction corresponding to each pixel, the desired image can be generated. This approach also manages reflections on metallic surfaces or water, as the color depends on the viewing direction. Although the field is only a few years old, significant progress has been made in NeRF techniques. Some of the most notable advancements include Mip-NeRF (Barron et al., 2021) and Mip-NeRF 360 (Barron et al., 2022), which address challenges like anti-aliasing and handling scenes with 360-degree camera rotation. Another important development is Instant-NGP (Müller et al., 2022), which uses multi-resolution hash grids to significantly optimize rendering speed and reduce memory requirements. Gaussian Splatting aims specifically to overcome the limitations of other approaches in synthesizing new images, achieving reduced training times (Kerbl et al., 2023). The representation of a 3D scene with Gaussian Splatting resembles a point cloud, but instead of points, it uses Gaussian splats—three-dimensional Gaussian distributions that describe density and radiance in specific areas of the scene. Unlike the NeRF technique, where the computation required for empty regions is equivalent to that for regions containing objects, Gaussian splats can be concentrated in areas that require greater detail. This optimization enables rendering hundreds of images per second, allowing real-time interaction with the reconstructed scene without compromising quality. Another advantage is that the geometry of objects represented by Gaussian splats allows for scene editing operations, such as removing irrelevant portions. In (Kerbl et al., 2023), 3D anisotropic Gaussians with variable extents along different directions are introduced, offering greater flexibility compared to models based on regular grids and better adaptability for representing complex geometries. The 3D Gaussian-based scene reconstruction process begins with a point cloud, consisting of tie points previously generated through Structure from Motion (SfM). From this cloud, 3D Gaussians are generated and used to render new views, including viewpoints corresponding to the images used for reconstructing the input point cloud. The rendered output of these views is then compared with the original image, enabling iterative optimization of the Gaussians' characteristics—such as position, shape, and density—to progressively improve rendered image quality, making it as close as possible to the original.

Focusing on CH documentation, (Balloni et al., 2023) demonstrates the potential of using these methods for an accurate and detailed 3D reconstruction. It also highlights the importance of carefully choosing the most appropriate approach based on the specific object and research question. (Murtiyoso and Grussenmeyer, 2023) explores the use of NeRF for cultural heritage documentation, demonstrating faster processing times and high point fidelity, with completeness scores of 88.7% and 92.2% for complex and simple datasets, respectively. Despite its advantages, NeRF still requires image orientation through overlapping images. Therefore, this does not reduce the field effort and time required to perform data acquisition. Moreover, it underlines that, in terms of geometric precision, NeRF is still a long way from fulfilling the requirements of high level of detail documentation. The amount of noise generated in the resulting point cloud is still too important. In (Basso et al., 2024) NeRF and GS reconstruction results of the exterior of an AH are compared, assessing that NeRF excels in exporting point clouds and meshes and effectively addresses gaps from missing survey data. In contrast, GS offers superior texture quality, brightness management, and realistic rendering of materials, light, and shadows, along with a better depth effect, though detail reconstruction remains imperfect in both models. (Croce et al., 2024) compares NeRF and photogrammetry for CH documentation. Results show that NeRF preserves completeness and material description better than photogrammetry when input data and image resolution are limited. Therefore, NeRF is recommended for low-resolution or limited-image datasets, especially in large-scale aerial surveys needed for emergency response when rapid data acquisition is essential.

Within this framework, our paper compares NeRF, Gaussian Splatting, and the traditional SfM-MvS approach for 3D architectural reconstruction using low-cost sensors, focusing on the strengths and limitations of each method regarding accuracy, computational requirements, and real-world applicability. By assessing these techniques on architectural datasets, we aim to determine the best approach to represent large-scale structures and intricate details in diverse environments, e.g. interiors and exteriors.

2. Methodology

2.1 Case study

The Church of Santa Maria della Piazza was selected as a case study for this article due to its unique architectural characteristics, which make it an exemplary representation of architectural heritage (Figure 1). Located near the port area, the Church of *Santa Maria della Piazza* is regarded as Ancona's earliest and most significant place of worship. The city's first cathedral, dedicated to Saint Stephen and built outside the ancient city walls, has been identified in the remains of a Paleo-Christian Basilica uncovered during archaeological excavations beneath the current structure. Originally modest in form and decoration, the church now features a tripartite façade crafted by stonecutter Master Filippo in 1210, as noted in the inscription on the entrance portal's lunette. The lunette is adorned with four tiers of blind arches and includes repurposed elements such as the praying Virgin, the bust of the Archangel Gabriel, a peacock, and polychrome chancel slabs. Additional decorative features include a pair of lions, two-tailed sirens, doves, griffins, and a series of majolica basins. The large, splayed entrance portal is framed by a richly decorated frieze with scrolls and vine leaves, alternating with an elaborate array of human and animal figures drawn from medieval bestiaries. In the keystone, Christ the Judge is depicted. Inside, the church is divided into three naves by octagonal pillars and features a roof structure with trusses. The presbytery ends with an apse containing a wooden crucifix from the 16th century.
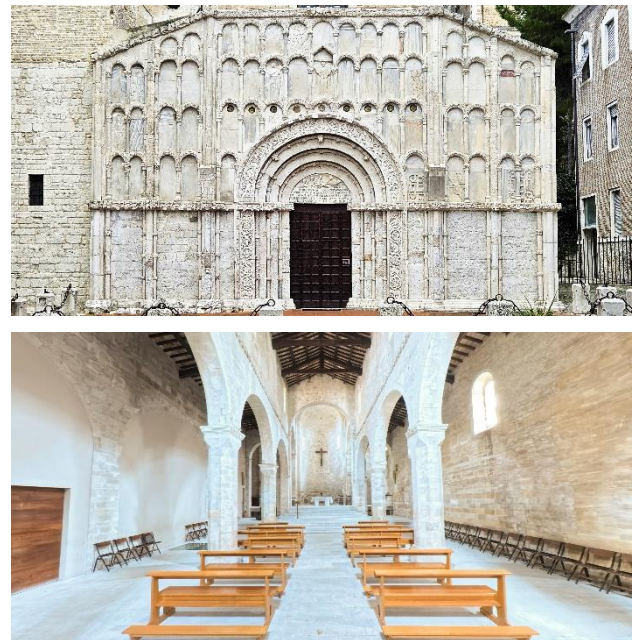

Figure 1. Ancona, Church of Santa Maria della Piazza. View of the exterior and interior.

2.2 Data collection and processing

Two low-cost tools were selected for data collection: a drone to document the exterior stone portal and a 360° camera to capture the interior space. To assess the accuracy of the reconstructions generated from the data acquired by these sensors, a Leica RTC360 laser scanner was also employed, conducting a scan of the façade along with 12 additional scans to reconstruct the church's interior. 76 images were captured from approximately 4 m with a DJI Mavic Mini UAV, equipped with a 1/2.3" CMOS sensor offering a 12 MP resolution, resulting in a GSD of 1,2 mm. For the interior, an Insta360 x4 360° camera was used to capture 48 360° images (11904x5952 px) according to the scheme in Figure 2, resulting in a GSD of 4,5 mm. For image acquisition by drone, the GSD was set according to the Historic England Geospatial Survey Specifications for Cultural Heritage. These guidelines recommend a GSD below 2mm for a 1:50 representation scale, which is appropriate for the documentation of an AH. By contrast, for panoramic images, the GSD was 4.5mm—a higher value, determined by the distance from the building's covering elements, i.e. trusses and vaults, due to the ground-based image capture.

Both image datasets were separately imported into Agisoft Metashape and processed using the software's standard 3D reconstruction pipeline. By selecting the highest available settings, a point cloud consisting of 115,8 million points was generated for the stone portal and a point cloud consisting of 93,6 million points for the interior. Then, both image datasets were also processed using both NeRF and Gaussian Splatting techniques, specifically with Nerfacto and 3D Gaussian Splatting methods, respectively. The workflow for processing was as follows: initially, the images were processed with COLMAP's SfM algorithms (Schönberger and Frahm, 2016) to extract camera parameters and poses. This step is essential for both Nerfacto and 3D Gaussian Splatting reconstructions. The key difference between the two in this operation lies in the dataset preparation, which must adhere to specific compatibility requirements. Once prepared, training can commence, during which a real-time viewer enables monitoring of all processing stages.
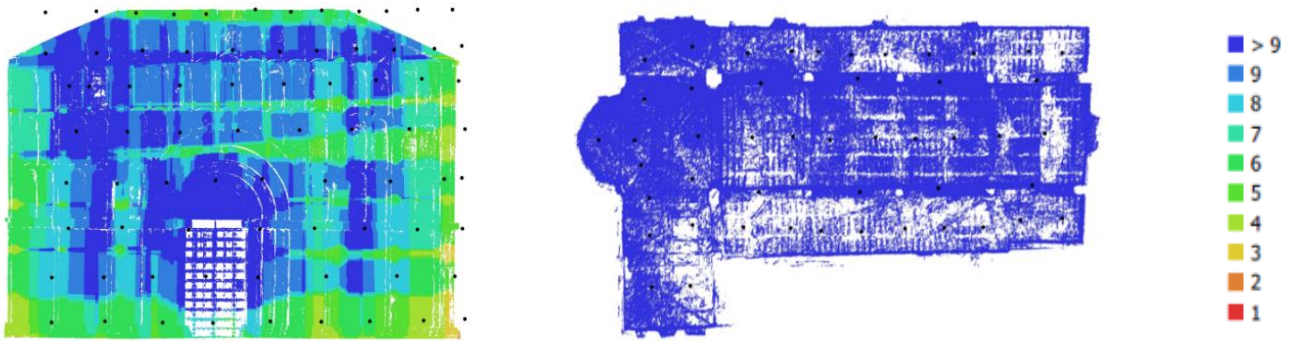
Figure 2. Ancona, Church of Santa Maria della Piazza. Left: overlay of images related to the external façade. Right: overlay of 360° images of the interior.
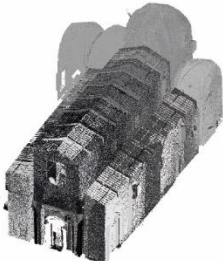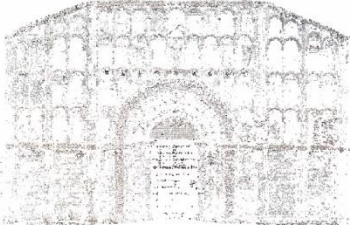
	Exterior	Interior
TLS	11.400.000 points	200.000.000 points
SfM-MVS	115.900.000 points	93.600.000 points
NeRF	85.700.000 points	93.200.000 points
Gaussian Splatting	40.000 points	165.000 points

Table 1. Point Clouds of the exterior and interior of the Church.

After training and the reconstruction of the 3D scene by both Nerfacto and 3D Gaussian Splatting, the resulting radiance fields and splats were converted into a 3D point cloud. Nerfstudio provides the option to adjust the number of points in the generated point cloud, which was set to align with the point cloud derived from photogrammetry. On the other hand, Gaussian Splatting 3D does not offer this customization, resulting in an automatically generated point cloud with 55k points for the stone portal and 166k points for the interior (Table 1).

3. Results and discussion

The results were analysed through two approaches: qualitative and quantitative investigations. A preliminary step involved conducting a visual inspection of the obtained reconstructions, examining the point clouds of both the interior and exterior in their entirety, as well as through horizontal and vertical slices. Regarding the reconstruction of the external façade, an overall view indicates that the NeRF-generated model is the most complete. However, it is also characterized by a high level of noise, with particularly noticeable artifacts in the flat areas of the small arches (Figure 3).

These observations are corroborated by horizontal slices and the vertical slice performed at the center of the façade. Notably, these analyses reveal that while the SfM-MVS point cloud is less complete than the one generated by NeRF, it is the closest to the reference model obtained through TLS. At the upper portion of the wooden door, the vertical slice reveals an evident artifact in the NeRF reconstruction. In the upper small arches, NeRF accurately reconstructs the statue of the praying Virgin but produces a significantly less accurate geometry for the flat arch located above it (Figure 4a). Due to the reduced number of points, a visual comparison between GS and the other reconstructions proves challenging. However, it is evident that the GS reconstruction is concentrated in areas with more complex geometry, whereas the flat sections exhibit an extremely sparse point distribution. Similar conclusions can be drawn from the visual inspection of the results obtained for the reconstruction of the interior environment. Even when using 360° images, NeRF stands out as the most efficient solution for a complete reconstruction of the investigated scene. However, as particularly evident from the slices, it also exhibits higher levels of noise and greater deviations from the TLS data compared to the results achieved with SfM-MVS (Figure 4b). The analysis then focused on a specific structural element—a pillar—further confirming the superior completeness of the NeRF reconstruction, while highlighting persistent difficulties in accurately reconstructing flat surfaces (Figure 4c).

For quantitative comparison, the point clouds from SfM-MVS, NeRF, and 3D Gaussian Splatting were compared in CloudCompare using the TLS point clouds as a reference. Both for the exterior and interior, the following parameters were considered for evaluating each reconstruction. First, to assess geometric accuracy, the mean error ($\bar{x}$) and standard deviation (σ) values from an M3C2 comparison (Lague et al., 2013) were used to evaluate image-based reconstructions in terms of systematic error and noise presence, respectively.
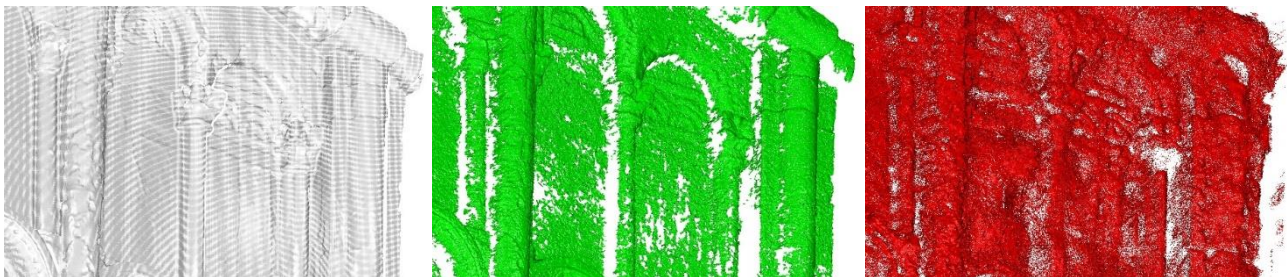

Figure 3. Visual comparison of a detail of the external façade. From left to right: TLS, SfM-MVS, and NeRF point clouds.

Figure 4. From left to right, vertical slices depicting: the external façade, highlighting the praying Virgin statue (a); the transept, focusing on the vault above the altar (b); and a pillar (c). The TLS point cloud is shown in black, SfM-MVS in green, and NeRF in red.

Second, a density analysis was carried out by calculating the number of neighbors within a 5 cm radius sphere for each point in the Metashape, Nerfacto, and 3D Gaussian Splatting point clouds. The results are presented in Table 2 and Table 3.

Focusing on the exterior, the geometric accuracy was evaluated comparing the three image-based reconstruction with a single TLS scan carried out at 10 m from the façade, setting the point density 3.1@10m. Preliminarily, reconstructed elements outside the façade were deleted. None of the three methods yielded a mean error higher than 2 mm, with Gaussian Splatting showing the highest. It is however more interesting to observe the value of the standard deviation, as for the mean error, SfM-MVS reconstruction generated the lowest value (1.5 cm), GS the medium (3.8 cm) and NeRF the highest (4.5 cm). This is more than 20 times higher than the GSD (2 mm). The density analysis of the external façade highlighted a very homogeneous distribution for the SfM-MVS.

However, this may be attributed to an automatic application of a decimation filter by the software at the conclusion of the point cloud processing. SfM-MVS unreconstructed areas are mostly limited to specific sections of the wooden panels on the entrance door, characterized by dark, uniformly colored wood, where no key points were identified during the photogrammetric process. The NeRF reconstruction displays a density distribution that closely reflects the distribution of images used for the 3D reconstruction, with higher densities observed in areas captured by a greater number of images. Notably, no unreconstructed regions are present. By contrast, the GS reconstruction reveals several unreconstructed regions in areas with flat geometry. The point density increases proportionally with the complexity of the geometry. Consequently, the point cloud representation aligns with the distribution of Gaussian splats, which are smaller and more concentrated in regions of greater geometric detail.

Focusing on the quantitative analysis of the interior, a point cloud derived from the alignment of 12 TLS scans was used as a reference. These scans were conducted with a resolution of 3.1mm@10m, achieving an RMS alignment error of less than 1mm. In the comparison phase, the highest mean error was observed in the SfM-MVS reconstruction (7mm), which was lower for GS (6mm) and NeRF (4mm). However, when examining the standard deviation, GS recorded the highest value (25.8cm), while NeRF and SfM-MVS showed lower values of 18.2cm and 19.6cm, respectively. In terms of density, NeRF and SfM-MVS provided similar average values, but SfM-MVS demonstrated greater uniformity. Specific evaluations were conducted on certain portions of the interior space. For instance, in the longitudinal right section of the building, SfM-MVS reconstruction exhibited a lower mean error compared to NeRF, contrary to the results observed in the transverse section facing the apse, located near the altar. In this section, it was noted that the SfM-MVS reconstruction displayed the highest errors concentrated on the left side, where the reconstruction was also less dense. Finally, an analysis of the ceiling revealed missing portions in all reconstructions, with NeRF once again proving to be the most complete, while its error values were comparable to those obtained through SfM-MVS.
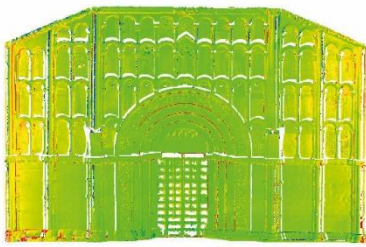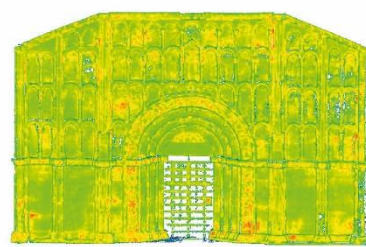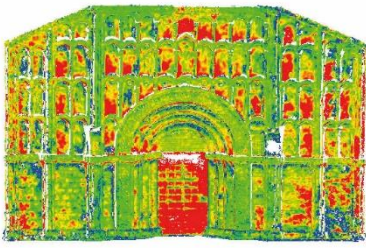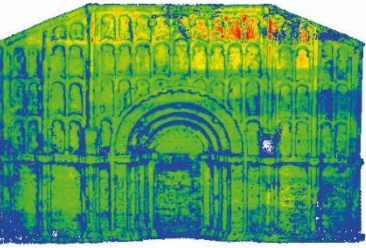

Table 2. M3C2 comparison and density analysis of the point clouds of the exterior of the Church.
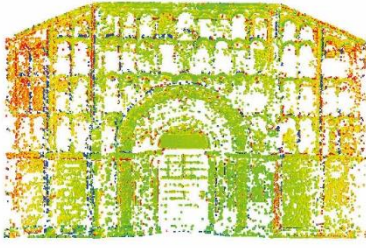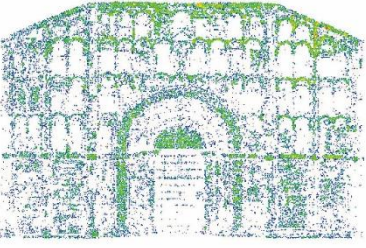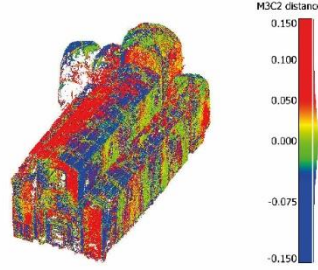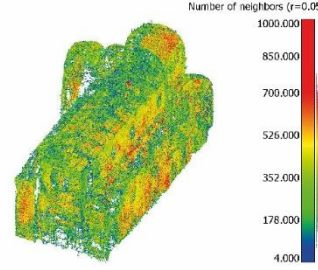
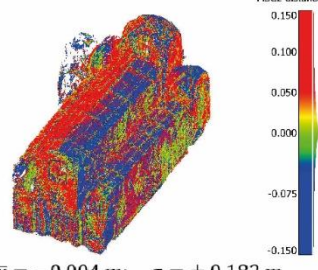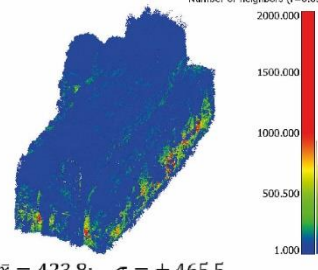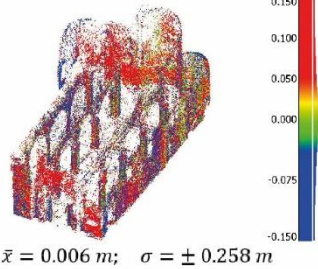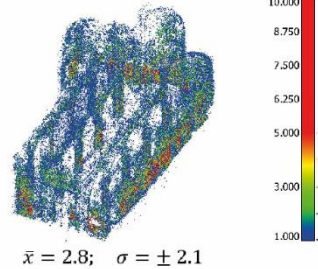| Comparison with TLS from M3C2	Density analysis (number of neighbors r = 0.05)
SfM-MVS $\bar{x} = 0.007\ m;\quad \sigma = \pm\ 0.196\ m$	$\bar{x} = 447.6;\quad \sigma = \pm\ 172.5$
NeRF $\bar{x} = -0.004\ m;\quad \sigma = \pm\ 0.182\ m$	$\bar{x} = 423.8;\quad \sigma = \pm\ 465.5$
Gaussian Splatting $\bar{x} = 0.006\ m;\quad \sigma = \pm\ 0.258\ m$	$\bar{x} = 2.8;\quad \sigma = \pm\ 2.1$

Table 3. M3C2 comparison and density analysis of the point clouds of the interior of the Church.

4. Conclusion

This paper compares the results of different image-based 3D reconstructions applied to AH. Specifically, the study explored the use of SfM-MVS, NeRF, and GS methods using images acquired with low-cost sensors featuring different capabilities: a drone for capturing the external façade and a 360° camera for the interior spaces. The selection of these tools was driven by the objective of maximizing efficiency during the data acquisition phase. The drone was used to focus on the façade elements, enabling documentation of decorative features even at height, while the 360° camera was utilized indoors to simultaneously capture walls, flooring, and roof structures in each shot.

Considering processing time, both NeRF and GS significantly outperformed traditional MVS, completing tasks in a fraction of the time. Moreover, NeRF excelled in completeness, producing reconstructions without missing parts. On the other hand, GS's limitation in controlling the point number prevents it from generating point clouds as dense as those produced by MVS or NeRF.

Focusing on density values, the results suggest distinct dependencies for each method. In SfM-MVS and GS, the density of reconstruction is tied to the presence of features, with areas such as the arches or stratigraphically detailed masonry being reconstructed more densely. In contrast, NeRF appears to rely more on the quantity of input images, with denser reconstructions corresponding to areas where image overlaps are prominent.

Moving to an evaluation of the reconstruction accuracy, for the external façade, SfM-MVS achieves acceptable error, although areas reconstructed from a smaller number of images exhibit greater deviations, as highlighted from the comparison with the TLS point cloud. In contrast, both the NeRF and GS reconstructions fail to provide representations suitable for the graphical documentation of AH. In particular, NeRF is affected by artifacts concentrated on flat surfaces and in the portions not reconstructed by SfM-MVS. Focusing on the interior, none of the methods proved capable of delivering an adequately accurate representation of the investigated architecture. This outcome is undoubtedly influenced by the higher GSD, which was affected by the characteristics of the 360° camera used and the capture distance, particularly for the upper parts of the structure.

An additional consideration regarding costs pertains to the equipment used for data processing. While the images were acquired using very low-cost tools (costing less than €500), the processing of data with NeRF and GS requires high-performance computational systems, which involve significantly higher costs (exceeding €5000).

In conclusion, it can be stated that, based on images acquired with low-cost sensors, NeRF and GS are not yet capable of achieving sufficient accuracy for the geometric documentation of architectural heritage. On the other hand, it should be noted that NeRF and GS were developed as solutions for image rendering and are therefore more suitable for visualization purposes rather than measurement, as clearly demonstrated by the high level of visual detail shown in (Figure 5).

Figure 5. Above: view of the external portal generated with NeRF. Below: view of the interior generated with GS.

References

Balloni, E., Gorgoglione, L., Paolanti, M., Mancini, A., Pierdicca, R., 2023. Few shot photogrametry: a comparison between NeRF and MVS-SfM for the documentation of cultural heritage. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* XLVIII-M-2-2023, 155–162. doi.org/10.5194/isprs-archives-XLVIII-M-2-2023-155-2023.

Barron, J. T., Mildenhall, B., Verbin, D., Srinivasan, P. P., & Hedman, P., 2021. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada, 2021, pp. 5835-5844. doi: 10.1109/ICCV48922.2021.00580.

Basso, A., Condorelli, F., Giordano, A., Morena, S., Perticarini, M., 2024. Evolution of Rendering Based on Radiance Fields. The Palermo Case Study for a Comparison Between Nerf and Gaussian Splatting. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* XLVIII-2/W4-2024, 57–64. doi.org/10.5194/isprs-archives-XLVIII-2-W4-2024-57-2024.

Croce V, Billi D, Caroti G, Piemonte A, De Luca L, Véron P., 2024. Comparative Assessment of Neural Radiance Fields and Photogrammetry in Digital Heritage: Impact of Varying Image Conditions on 3D Reconstruction. *Remote Sensing*. 16(2):301. doi.org/10.3390/rs16020301.

Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G., 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Trans. Graph.* 42, 4, Article 139 (August 2023), 14 pages. doi.org/10.1145/3592433.

Lague, D., Brodu, N., Leroux, J., 2013. Accurate 3D comparison of complex topography with terrestrial laser scanner: Application to the Rangitikei canyon (N-Z). *ISPRS Journal of Photogrammetry and Remote Sensing 82, 10-26.* doi.org/10.1016/j.isprsjprs.2013.04.009.

Murtiyoso, A., Grussenmeyer, P., 2023. Initial assessment on the use of state-of-the-art nerf neural network 3d reconstruction for heritage documentation. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* XLVIII-M-2-2023, 1113–1118. doi.org/10.5194/isprs-archives-XLVIII-M-2-2023-1113-2023.

Remondino, F., El-Hakim, S., 2006. Image-based 3D Modelling: A Review. *The Photogrammetric Record 21,* 269–291. doi.org/10.1111/j.1477-9730.2006.00383.x.

Schönberger, J.L., and Frahm, J.-M., 2016. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 4104-4113. doi: 10.1109/CVPR.2016.445.

Tancik, M., Weber, E., Ng, E., Li, R., Yi, B., Wang, T., Kristoffersen, A., Austin, J., Salahi, K., Ahuja, A., Mcallister, D., Kerr, J., Kanazawa, A., 2023. Nerfstudio: A Modular Framework for Neural Radiance Field Development. *ACM SIGGRAPH 2023 Conference Proceedings (SIGGRAPH '23). Association for Computing Machinery, New York, NY, USA, Article 72, 1–12.* doi.org/10.1145/3588432.3591516.