The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-M-9-2025
30th CIPA Symposium "Heritage Conservation from Bits:
From Digital Documentation to Data-driven Heritage Conservation", 25–29 August 2025, Seoul, Republic of Korea

Gaussian Splatting for Facade Orthophoto Generation – Comparison with MVS and TLS

Arnadi Murtiyoso[1], Hélène Macher[1]

[1]Université de Strasbourg, INSA Strasbourg, CNRS, ICube Laboratory UMR 7357, Photogrammetry and Geomatics Group,
67000 Strasbourg, France – (arnadi.murtiyoso, helene.macher)@insa-strasbourg.fr

Keywords: Orthophoto, Facade, Photogrammetry, Gaussian Splatting, MVS, TLS.

Abstract

The notion of radiance fields is a very recent topic which has gained much traction in the past couple of years. Among some of the well known techniques, 3D Gaussian splatting (3DGS) seems to generate promising results in terms of 3D reconstruction. In the domain of heritage documentation, orthophotos are a standard product usually generated from photogrammetry by projecting image pixels into a 3D surface created from depth maps or point clouds. In this paper we investigate the possibility to substitute the conventional method for 3D surface generation, either terrestrial laser scanning (TLS) or multi-view stereo (MVS), with 3DGS. The paper will attempt to answer two questions: (1) whether this novel method is a viable solution for orthophoto generation within the heritage documentation context, and (2) at what point is 3DGS worth the time and effort to reach an acceptable orthophoto when compared to conventional methods. To do this, we perform multiple analysis on both the geometric quality of point clouds and the orthophotos themselves, using conventional TLS data as our reference. Our results indicate that 3DGS is not only promising, but indeed a viable alternative to conventional MVS with a potential gain of processing time between 12% up to 33% to reach a comparable quality. In the scenarios tested, both of these gains came with an average point cloud standard deviation of 2 cm.

1. Introduction

Digital archiving of cultural heritage is becoming ever more important today, due to risks from both natural and anthropological sources. Metric documentation of cultural heritage provides reference points for preservation and in some cases, reconstruction of built heritage (Mouaddib et al., 2023). Although the documentation process is traditionally done using 2D maps, drawings, and photographs, modern techniques nowadays tend to use 3D digital technologies such as laser scanning and photogrammetry. A particular subset of the heritage documentation process is the creation of 2D orthophotos of building facades. This product of 3D documentation remains a very useful tool not only for visual inspections, but also for quick metric measurements (Fino et al., 2023; Choi et al., 2018).

Orthophotos are created by projecting image pixels with the help of a reference surface (Barazzetti et al., 2014). This eliminates image displacements and creates an orthogonally projected image, as opposed to the central projection of images created by standard cameras. In order to create an accurate orthophoto, it is therefore necessary to obtain a good reference surface and a set of oriented images. For close range applications, as far the reference surface is concerned a 3D mesh generated by point clouds is usually used for this purpose. In a standard photogrammetric workflow, this mesh is generated by Multi-View Stereo (MVS) which is a by-product of the process. However it may also come from a lidar or laser scanner provided that the point cloud is properly registered to the oriented images.

A common bottleneck in the orthophoto creation process is the processing time, especially when using the standard photogrammetric workflow. The combined time of 3D mesh generation and orthophoto generation often presents the majority of processing time. Novel neural radiance fields-based methods claim to tackle this problem (Murtiyoso and Grussenmeyer, 2023), first with the Neural Radiance Fields (NeRF) and later on using the more explicit 3D Gaussian splatting (3DGS) methods which is the subject of this research. While previous other studies have proven this point, it still remains to be seen if radiance fields methods such as 3DGS can provide a similar quality, both visual and geometric, to conventional methods. This is particularly important for heritage documentation applications where orthophotos are often used as an analysis tool.

In the case of facades, the 3DGS method is actually not an ideal solution as far as 3D model reconstruction is concerned, partly due to its 3D rendering approach of using adaptive densification. However, for façade orthophoto generation, an overly dense point cloud is often unnecessary. Thus, we argue that 3DGS may still be used to this end while reducing the processing time. In this experiment, we wish to answer several questions, including (1) whether the **3DGS method can generate a facade orthophoto of a comparable quality** to conventional MVS methods for heritage documentation requirements, and (2) to which point it is possible to extract **a good enough result from 3DGS for this purpose** in order to determine if the gain in processing time is worthwhile. To do this, we also compared the orthophoto to the one generated using a terrestrial laser scanner (TLS).

2. State of the Art

The concept of radiance fields is a novel method to represent a 3D scene that differs from conventional point clouds. While point clouds store geometric data in the form of static 3D cartesian coordinates with the possibility to attach semantic attributes to them, radiance fields also consider the fact that a point may be observed differently depending on the viewpoint. The nature of each point in a radiance field representation may be described in two ways. It can be explicit, as in the case of 3DGS (Kerbl et al., 2023), where points are represented as the eponymous 3D Gaussians. Alternatively, it can be implicit, as seen in NeRF (Mildenhall et al., 2020), where information is encoded as a continuous density function. In both cases, the focus is on 3D rendering via a learning-based process, i.e. using

the original images to train a loss function to correctly render the scenes.

NeRF provides a more implicit representation, where it tries to go through a ray while performing density field sampling along the ray. 3DGS on the other hand, provides a more straightforward solution by representing the radiance field as a 3D Gaussian, an anisotropically scaled kernel with parameters like colour, opacity, and size. During the rendering process, i.e. the display of the 3D scene on the 2D screen, NeRF goes through a ray marching process to give the results. 3DGS simply "splats" all the contributing Gaussians to 2D, hence the naming of the method. When converted into conventional point clouds, NeRF performs a density thresholding and uses the marching cubes algorithm to extract 3D points from radiance fields having a certain density value (Murtiyoso and Grussenmeyer, 2023). 3DGS on the other hand simply extracts the Gaussians themselves directly. Due to these points, 3DGS presents the advantage of very fast rendering, especially compared to the training time of NeRF or the traditional MVS process.

Furthermore, 3DGS also employs adaptive densification and pruning of its initial points. This tackles the problem of both under and over representation of objects. For this reason, 3DGS is able to reconstruct small details where conventional, pixel-size based Multi-View Stereo (MVS) may fail. However, since 3DGS also typically uses sparse Structure from Motion (SfM) points as initialization, it does not usually produce dense points. In terms of rendering, this is usually compensated by the fact that the Gaussian representation is anisotropic; however, this presents challenges when reconstructing large, continuous surfaces.

Radiance fields in general has been used for heritage documentation in several studies (Croce et al., 2023; Clini et al., 2024). Both Basso et al. (2024) and Balloni et al. (2024) also wrote extensively on the topic of this emerging technology and performing useful comparisons between different types of radiance fields. 3DGS for orthophoto generation itself is not a new topic, as shown in Chen et al. (2024) following similar justification as presented in this paper but oriented towards aerial images. Previtali et al. (2024) similarly performed a comparison, geared towards the use of low-cost sensors and non-Lambertian surfaces. This paper's novelty will concern the value of the 3DGS method for real and production-oriented heritage documentation projects.

3. Data and Research Methodology

The data used in this study were obtained from an acquisition campaign conducted as part of the archaeological study of the tower known as *"Tour des Français"*, part of the old defensive work of Strasbourg old town (*"Ponts Couverts"*). The tower's construction was completed in the first half of the 13th century. The objective of the campaign was to create archival documentation, including orthophotos and sections, for the archaeological study of the tower. The data were collected using a Z+F IMAGER 5016 TLS for both the interior and exterior of the tower, complemented by an exterior photo campaign taken using the Canon EOS R5 for better rendering of the orthophotos. Additionally, the data were georeferenced via control points. Facade orthophotos were generated by applying photos onto a 3D mesh constructed from the laser scanning data. This was then used as our reference data when comparing them to photogrammetry, both using MVS and 3DGS.

Three types of analysis were performed. First, the quality of the point cloud from 3DGS was put to test. 3DGS in and of itself does not necessarily generate point clouds per se; indeed, it generates 3D Gaussians which are anisotropic and volumetric in nature. In this paper, we retrieve only the position attribute of each Gaussians and thus created a 3D point cloud. A total of eight 3DGS point clouds, referred to as stages, were obtained. Each stage was obtained from a different point of the reconstruction process, starting with 3,000 iteration steps and increasing by an increment of 3,000 steps up to 24,000 steps. These 3D point clouds were then meshed to create a reference surface for the orthophoto generation. To test the quality of this 3DGS-converted point clouds, the TLS point cloud was used as a reference. The reference point cloud was first meshed to enable the computation of signed distances for the compared point cloud. This process was performed using the 3D software CloudCompare's cloud-to-mesh (C2M) function. From this analysis, three metrics were derived: the average error to show the accuracy of each generated point, the standard deviation to give an idea on the level of noise present, and a completeness percentage to quantify the presence of holes or unreconstucted parts.

The seconds analysis was performed directly on the orthophotos. To this end, orthophotos were created using the 3DGS point clouds from different stages of reconstruction. Only five 3DGS point clouds were used to create the orthophotos. These five were chosen by using the completeness parameter; only those with a completeness level above 60% were used. The analysis then overlays the orthophotos with the one generated by TLS point clouds as a reference. A standard image class prediction confusion matrix and metrics analysis were then performed to quantify the results. An orthophoto generated by the conventional photogrammetric method was included in this analysis as a comparison, generated by the software Agisoft Metashape.

Finally, a third analysis was performed by choosing the best result from 3DGS and comparing it to an orthophoto generated directly by the TLS. This orthophoto does not use the close range images as the other ones, but rather the images taken directly by the TLS. It has the big advantage of providing an orthophoto very quickly after data acquisition. The focus of this analysis is therefore on the utility of such rapidly generated orthophoto for the purposes of heritage documentation. In this analysis, the orthophoto generated by MVS was again included as a comparison. All the processing in this paper was performed using the same NVIDIA RTX 3000 Ada Generation Laptop GPU.

4. Results and Discussions

4.1 Quality of 3DGS reconstruction

Results show that 3DGS can generate 3D renderings very quickly. Processing time may vary depending on where the training process is stopped, but a more or less complete representation of the facade was achieved after 15k steps, around the 12 minutes mark (see Figure 1) with a completeness rate of more than 80%. By the 18k-steps mark (around 18 minutes) the completeness rate begins to plateau at around 90%.

In terms of geometric precision, the resulting point clouds converted from Gaussians were compared to a 3D mesh surface generated by TLS data to quantify signed distances between the point clouds and a reference surface. Similarly to the observation on Figure 1, both the average error and standard deviation

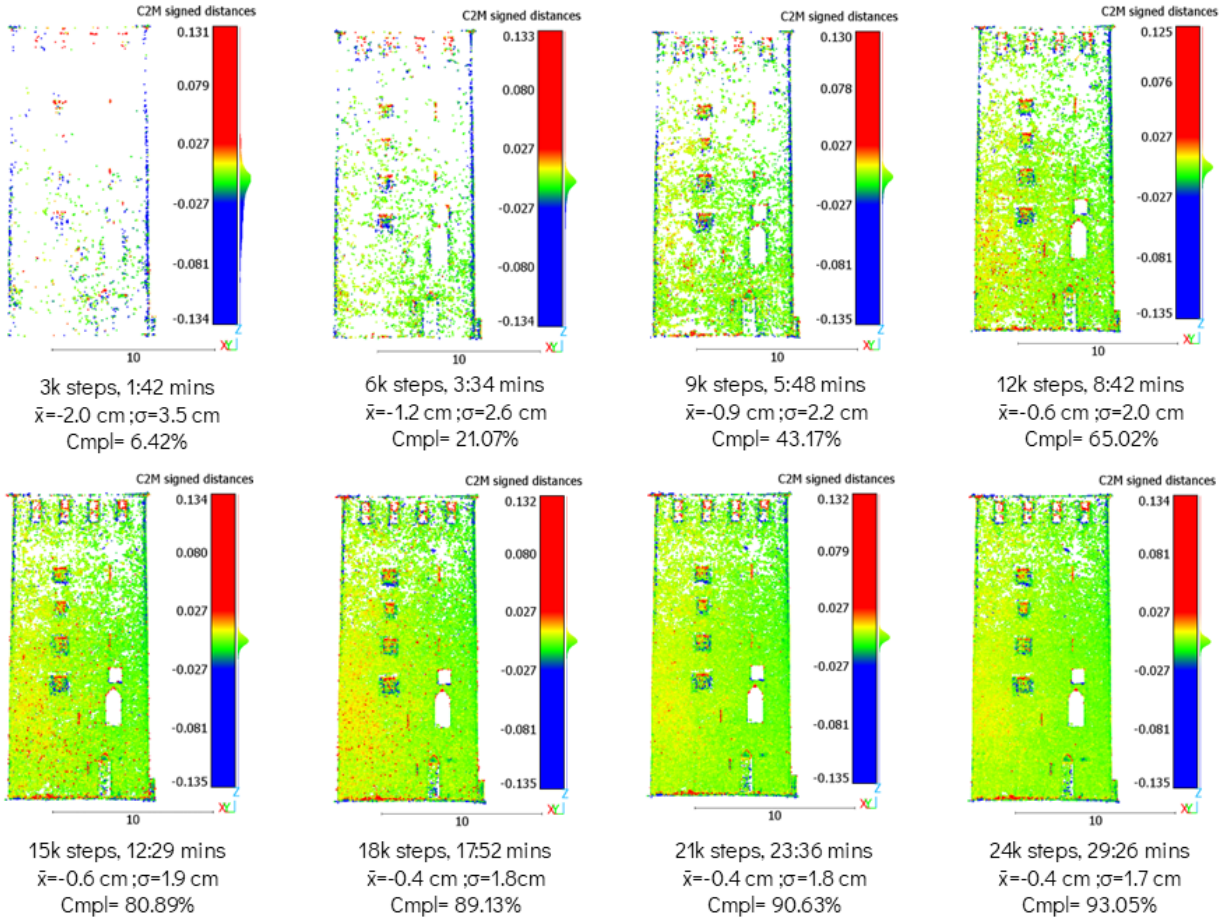
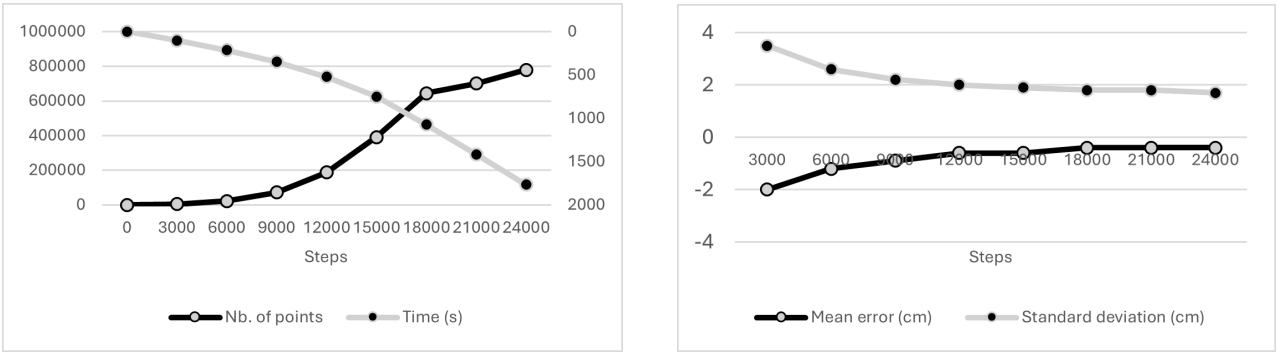


Figure 1. A visualization of the cloud-to-mesh analysis for eight stages of the 3DGS training process. $\bar{x}$ denotes mean error, σ denotes standard deviation, and `Cmpl` denotes completeness relative to the reference TLS point cloud.



(a) (b)

Figure 2. (a) Graph showing the processing time for 3DGS rendering on the right Y axis and the number of points on the left Y axis; (b) graph showing the mean error and standard deviation of the point clouds relative to the reference TLS data.

starts to stabilise around the 18k-steps mark; specifically at -0.4 cm and 1.8 cm, respectively (Figure 2).

To answer the question of which cut-off time to take into account and therefore attempt to answer the second research question of the paper, it is possible to take into account both visual and quantitative clues. Visually, the point cloud from 3DGS is more or less complete at the 15k stage, although starting from the 12k cut-off the completeness level is already above 60%. Visual inspection is nevertheless important due to the nature of the object, in which the incomplete part of the object may include important parts for the orthophoto generation further down the workflow.

The average error and standard deviation of the 12k and 15k stages remain slightly higher than the remaining three (18k, 21k, and 24k steps) but are nevertheless still within a sufficient threshold of 2 cm. Considering these observations, it was then decided to take the 3DGS point clouds from the 12k stage up to the 24k stage to be further processed into orthophotos.

4.2 Comparison of generated orthophotos

As previously mentioned, five stages of the 3DGS reconstruction were chosen to be used to create orthophotos. The main parameter for choosing was the completeness percentage (more than 60 %). These point clouds were meshed and orthophotos were created using Metashape with a pixel resolution of 2 mm. In addition, another orthophoto was generated by the standard photogrammetric process of Metashape, using the "High" quality for the mesh creation to be used in comparison to the other 3DGS-based orthophotos. For the reference orthophoto, the TLS point cloud was meshed and an identical orthophoto of 2 mm-resolution was also generated using Metashape. In order to analyse the results, each of the 3DGS orthophotos as well as that from MVS was overlaid to the one from TLS data, and both the confusion matrices and metrics were computed from this procedure. As metrics, the standard parameters of accuracy, precision, recall, and F1 score were computed (see Figure 3).

In Figure 3, green pixels indicate false positives which may be interpreted as noise in the orthophoto. Conversely, violet pixels denote false negatives which may be interpreted as holes in the image. True positive pixels represent the correct projection of a pixel (indicated by the class `Obj`), while true negatives are the white pixels of the orthophoto background (indicated by the class `Bckgr`).

The 12k stage orthophoto in particular showed a significant presence of false negatives, a fact that stems from the incomplete nature of the input point cloud. This can also be seen in its recall value which yielded a relatively low value of 0.79. Nevertheless, the quality of existing pixels are good as indicated by the high precision value of 0.98. As may be expected from our preliminary observation in Figure 2, the amount of true positives increased as the 3DGS iteration steps increase.

The question that needed to be answered was therefore: at what point of the 3DGS reconstruction can its results be considered comparable to that from conventional MVS, here represented by Metashape. To this end, Figure 4 places the metrics of Metashape as a benchmark superposed with the same values from the five stages of 3DGS orthophotos. As regards to the precision, a comparable result was achieved quite early from the 15k stage already. Recall achieved a comparable value to Metashape at the 18k stage, while both accuracy and F1-score, which may represent a compromise value between precision and recall, achieved similar results to Metashape at the 21k stage, although values from the 18k stage were already very close to the one from MVS. Overall, 3DGS was able to achieve the results of MVS, at least in terms of orthophoto quality, after 18k steps.

4.3 Comparison to direct TLS-generated orthophoto

An orthophoto was also generated using only TLS data, by projecting the colorized point cloud onto a plane. The orthophoto creation process in the Z+F Laser-Control software involved several key steps. First, only two scans were selected because of their better image exposure, as the remaining scans were either over-exposed or under-exposed. Next, three reference points were defined to establish the working plane. A bounding box was then created around this plane, extending 50 cm on either side, to frame the area of interest. For the orthophoto settings, a pixel size of 2 mm was chosen to allow direct comparison with the other orthophotos. A gap filling option was enabled to address missing data.

Figure 5 presents a comparison between the four generation methods namely TLS, TLS + Metashape, Metashape (High) and 3DGS. A zoomed-in view of a tower window is included to highlight the differences more clearly. It is worth noting that the TLS orthophoto allows for the identification of stone contours but appears more blurred compared to the others. Additionally, due to the angle of incidence between the scanner and the top of the object, some points were not captured, resulting in white pixels in those areas.

5. Conclusions

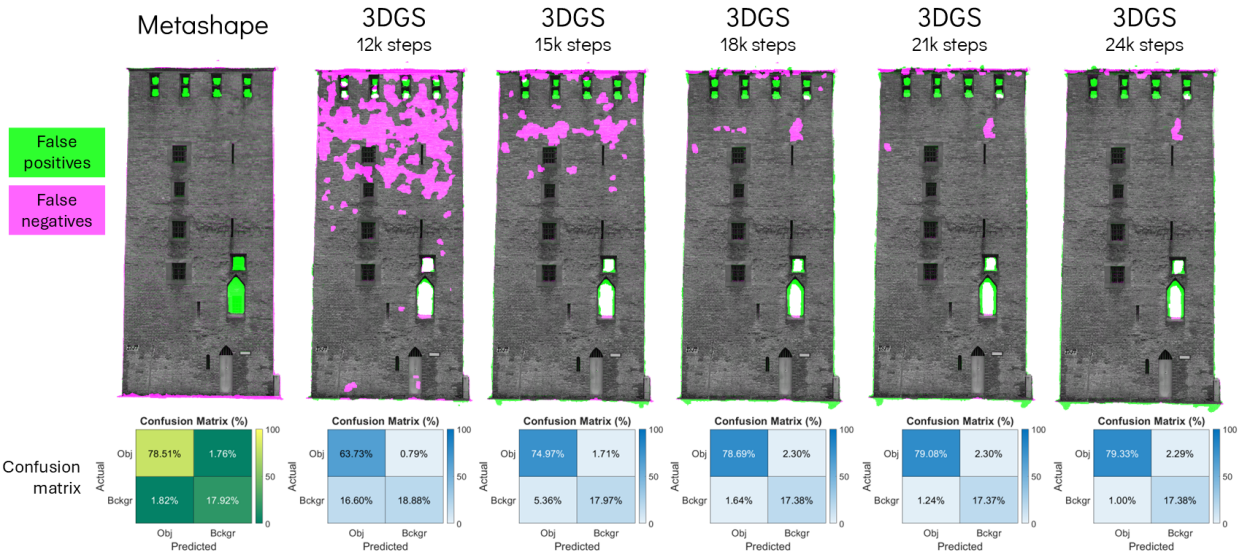
In this paper, we presented a comprehensive comparison of the use of 3D Gaussian splatting for facade orthophoto generation. Results show that as regards to the point clouds, a stable result in terms of average error and standard deviation was achieved after 15k steps. When translated into orthophotos, the 18k and 21k stages indicate the best compromise between processing time and quality, with regards to the benchmarked orthophoto generated by conventional MVS via Metashape.

It is also worth noting that Metashape, using identical computer specifications, generated depth maps (a precursor step to point cloud generation) in 26:47 minutes using the "High" setting. This setting means that the depth map was created by subsampling the input images by a factor of 2. In this regard, the results of 3DGS after 18k steps provided a gain of 33.30% in processing time, with a point cloud completeness rate of 89.13% and an orthophoto F1-score of 0.96 compared to Metashape's 0.98. If we go to the results from 21k steps, the gain in processing time is reduced to 11.88% but the point cloud completeness improves slightly to 90.63% and an orthophoto F1-score of 0.98. Furthermore, as can be seen in Figure 6, the processing time fitted a second degree polynomical curve almost perfectly when plotted against the number of iteration steps (R^2=0.993). In this sense it is possible to predict the processing time for further 3DGS stages. Additional results show that direct orthophoto generation from the TLS software actually possesses enough quality for a rapid examination of the object. However, the quality of the pixels is naturally not sufficient for proper texturing in the case of high precision orthophoto requirements.

These observations show that 3D Gaussian splatting is indeed a viable solution for facade orthophoto generation. 3DGS has shown to be **able to generate a facade orthophoto of a comparable quality** to traditional MVS. It also achieved a **good enough result with a gain of time up to 33.30%** when compared once again to MVS. There remains a caveat, nevertheless, that the experiments presented in this paper was only implemented on one case and one facade. While the present conclusion drawn gives already indications to the feasibility of 3D Gaussian splatting, future work should be conducted to complement this conclusion with other datasets.

Acknowledgments

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).



	Metashape	3DGS 12k steps	3DGS 15k steps	3DGS 18k steps	3DGS 21k steps	3DGS 24k steps
Accuracy	**0.9642**	0.8261	0.9293	0.9607	0.9645	0.9671
Precision	**0.9781**	0.9878	0.9777	0.9716	0.9717	0.9719
Recall	**0.9774**	0.7934	0.9333	0.9796	0.9845	0.9876
F1 Score	**0.9777**	0.8800	0.9550	0.9756	0.9781	0.9797

Figure 3. Orthophotos generated by point clouds from MVS (Metashape) and 3DGS compared against TLS point cloud, with their respective confusion matrices and metrics. For the confusion matrix, the `Obj` class denotes the existing object i.e. the facade in question, while the `Bckgr` class denotes the image’s white background.

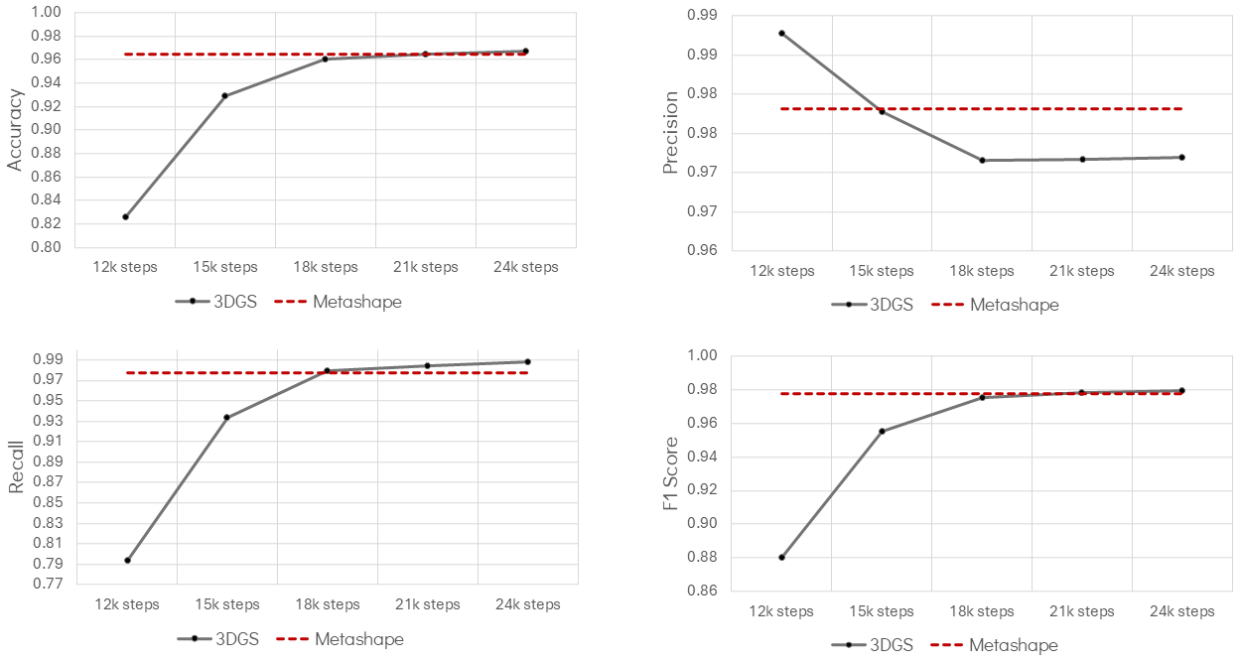


Figure 4. Metrics derived from the confusion matrices with values from Metashape used as threshold.

References

Balloni, E., Ceka, D., Pierdicca, R., Paolanti, M., Mancini, A., Zingaretti, P., 2024. Comparative assessment of Neural Rendering methods for the 3D reconstruction of complex heritage sites in the inner areas of the Marche region-Italy. *Digital Applications in Archaeology and Cultural Heritage*, 35, e00371.

Barazzetti, L., Brumana, R., Oreni, D., Previtali, M., Roncoroni, F., 2014. True-orthophoto generation from UAV images: Implementation of a combined photogrammetric and computer vision approach. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, II-5, 57-63.

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. *The International Archives of the Photo-*

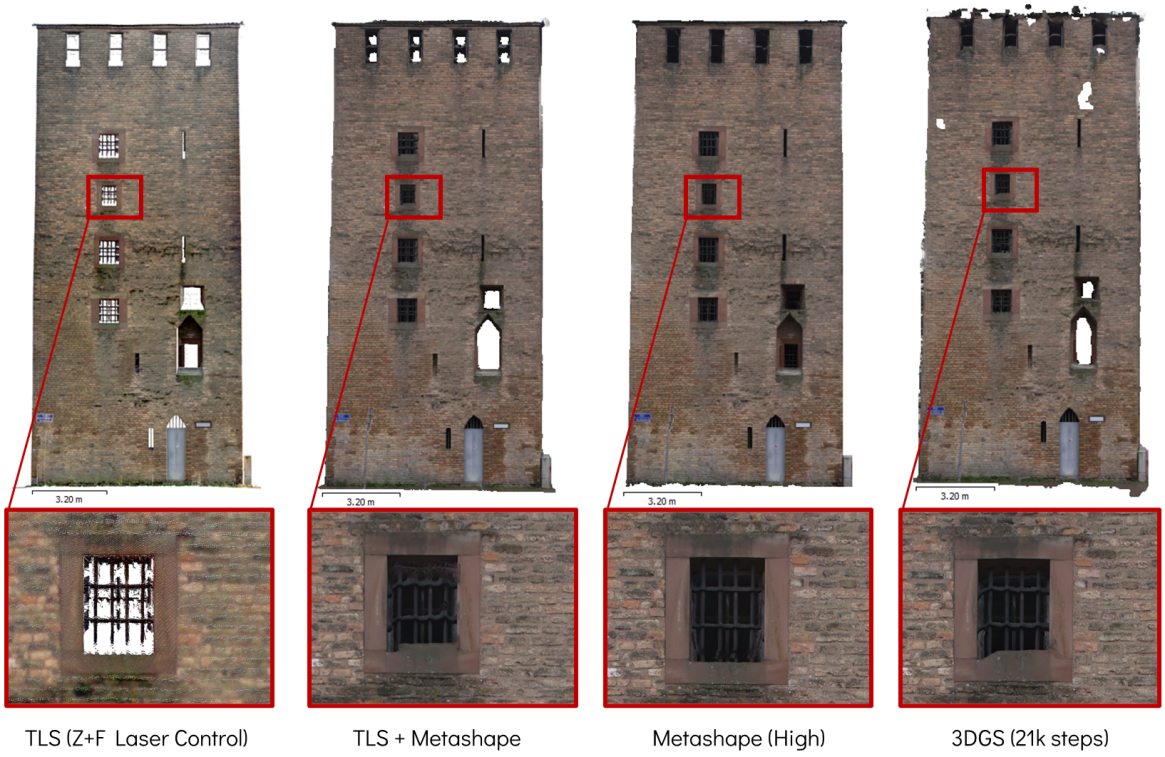


Figure 5. Visual comparison of the four orthophoto generation methods.

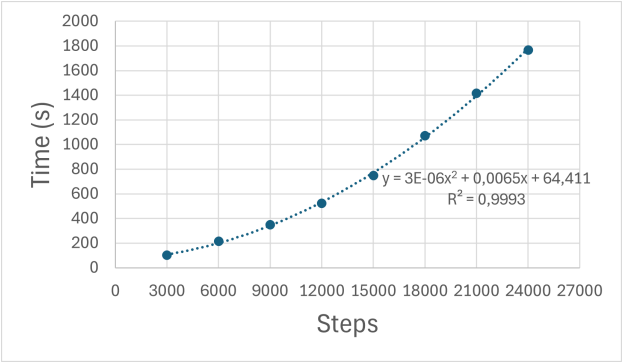


Figure 6. Processing time plotted against the number of steps and fitted into a second degree polynomial curve (R^2=0.993).

grammetry, Remote Sensing and Spatial Information Sciences, 57–64.

Chen, S., Yan, Q., Qu, Y., Gao, W., Yang, J., Deng, F., 2024. Ortho-NeRF: generating a true digital orthophoto map using the neural radiance field from unmanned aerial vehicle images. *Geo-spatial Information Science*, 1-20.

Choi, J., Yeum, C. M., Dyke, S. J., Jahanshahi, M. R., 2018. Computer-Aided Approach for Rapid Post-Event Visual Evaluation of a Building Façade. *Sensors*, 18, 3017.

Clini, P., Nespeca, R., Angeloni, R., Coppetta, L., 2024. 3D representation of Architectural Heritage: a comparative analysis of NeRF, Gaussian Splatting, and SfM-MVS reconstructions using low-cost sensors. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-2/W8-2024, 93-99.

Croce, V., Caroti, G., Luca, L. D., Piemonte, A., 2023. Neural radiance fields (nerf): review and potential applications to digital cultural heritage. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII, 25–30.

Fino, M. D., Galantucci, R. A., Fatiguso, F., 2023. Condition Assessment of Heritage Buildings via Photogrammetry: A Scoping Review from the Perspective of Decision Makers. *Heritage*, 6, 7031-7067.

Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G., 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics*, 42.

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. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12346 LNCS, 405-421.

Mouaddib, E. M., Pamart, A., Pierrot-Deseilligny, M., Girardeau-Montaut, D., 2023. 2D/3D data fusion for the comparative analysis of the vaults of Notre-Dame de Paris before and after the fire. *Journal of Cultural Heritage*.

Murtiyoso, A., Grussenmeyer, P., 2023. Initial assessment on the use of state-of-the-art NeRF neural network 3D reconstruction for heritage documentation. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-M-2-2023, 1113-1118.

Previtali, M., Barazzetti, L., Roncoroni, F., 2024. Orthophoto generation with gaussian splatting: mitigating reflective surface artifacts in imagery from low-cost sensors. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-2/W8-2024, 371-378.