

INITIAL ASSESSMENT ON THE USE OF STATE-OF-THE-ART NeRF NEURAL NETWORK 3D RECONSTRUCTION FOR HERITAGE DOCUMENTATION

A. Murtiyoso^{1*} and P. Grussenmeyer²

¹ Forest Resources Management, Institute of Terrestrial Ecosystems, Department of Environmental Systems Science, ETH Zurich, Switzerland – arnadi.murtiyoso@usys.ethz.ch

² Université de Strasbourg, CNRS, INSA Strasbourg, ICube Laboratory UMR 7357, Photogrammetry and Geomatics Group, 67000, France – pierre.grussenmeyer@insa-strasbourg.fr

Commission II, WG II/6

KEY WORDS: NeRF, neural network, AI, 3D reconstruction, cultural heritage, photogrammetry.

ABSTRACT

In recent decades, photogrammetry has re-emerged as a viable solution for heritage documentation. Developments in various computer vision methods have helped photogrammetry to compete against the laser scanning technology, eventually becoming complementary solutions for the purpose of heritage recording. In the last few years, artificial intelligence (AI) has progressively entered various domains including 3D reconstruction. The Neural Radiance Fields (NeRF) method renders a 3D scene from a series of overlapping images, similar to photogrammetry. However, instead of relying on geometrical relations between the image and world spaces, it uses neural networks to recreate the so-called radiance fields. The result is a significantly faster method of recreating 3D scenes. While not designed to generate 3D models, simple computer graphics methods can be used to convert these recreated radiance fields into the familiar point cloud. In this paper, we implemented the Nerfacto architecture to recreate two instances of heritage objects and then compared them to traditional photogrammetric multi-view stereo (MVS). While the initial hypothesis posits that NeRF is not yet capable to reach the level of accuracy and density achieved by MVS as can be observed in the results, NeRF nevertheless shows a great potential due to its fractionally faster processing speed.

1. INTRODUCTION

Documentation of tangible cultural heritage is an important point for its preservation and conservation, enabling its archiving and study in posterity. Documentation techniques have also evolved continuously: while early documentation uses sketches, drawings, and eventually photographs, the use of 3D reconstruction methods have gained traction in the last few decades. Developments in digital technology in both its hardware and software aspects have contributed significantly to this fact. The active lidar and passive photogrammetry techniques may be considered the two main contenders (or in many cases, complements) in the 3D heritage documentation domain (Matrone et al., 2020).

Photogrammetry has been used for a lot longer than lidar but remained in its shadows due to a more complicated process and the need for skilled operators. It was not until the development of multi-view stereo (MVS) and dense matching, coupled with important improvements in computing capabilities that it managed to become a strong competitor to the relatively more established lidar technology. While nowadays the photogrammetric workflow has more or less been established, several questions still remain (Murtiyoso et al., 2022). For example, the generation of dense point cloud for reflective surfaces remain difficult to be addressed using classical dense matching approaches. Recent solutions tend to gear towards the use of artificial intelligence (AI) and neural networks, for example to aid the reconstruction of textureless objects (Stathopoulou et al., 2021), to recover lost heritage from historical archives (Condorelli et al., 2020) or to improve 3D models via upsampling (Ren et al., 2022).

An interesting recent development is the use of Neural Radiance Fields (NeRF) to render a 3D scene of objects (Mildenhall et al., 2020). NeRF attempts to recreate radiance fields in the form of neural networks; it is fundamentally different from pre-existing notions of 3D models familiar to photogrammetry users. NeRF, in and on itself, represents neither a 3D mesh nor voxels but rather a density function describing the transparency or opaqueness of a certain object seen from a specific point of view (namely the images). As such, NeRF was not originally developed for 3D reconstruction but rather 3D rendering, i.e. the creation of novel points of view from existing base dataset generated by a trained neural network. However, several computer graphics techniques may be utilised to convert NeRF into either a point cloud or a 3D mesh, for example by using the classical marching cubes algorithm (Lorensen & Cline, 1987).

Since its first introduction in 2020 NeRF has gained a lot of attention and new implementations improved its speed and efficiency in the span of two years. Mip-NeRF was a follow-up to the original NeRF paper, increasing its processing speed and quality (Barron et al., 2021). Instant-NGP (Müller et al., 2022) was able to recreate a 3D scene in seconds. An interesting first attempt to use the concept for heritage documentation was presented in Sun et al. (2022) based on earlier work on NeRF-W (Martin-Brualla et al., 2021) where it was applied to unsorted internet images. Finally, Condorelli et al. (2021) presented a preliminary attempt in comparing NeRF to traditional MVS point cloud in the context of heritage documentation.

In this paper, we attempt to assess the use of state-of-the-art NeRF technology for heritage documentation purposes. Three main aspects will be discussed: (1) geometric quality of the

* Corresponding author

point cloud, (2) completeness of the point cloud, and (3) the density of the point cloud. In addition, the processing time vis-à-vis classical MVS-based dense matching will also be briefly discussed. In the final section of the paper, thoughts will be presented on the identified limitations and potential for the technology to help document tangible cultural heritage in 3D.

2. METHODOLOGY

Where NeRF differs from traditional photogrammetry is in its way of recreating (or more accurately, predicting) the 3D scene and by extension the 3D model. Whereas MVS in its most basic setup performs area-based image matching (Remondino et al., 2014), NeRF determines, based on the input training data, a density function of the 3D object or the so-called radiance fields. The radiance field consists not only of density, but may also include other information including RGB values or even semantic attributes (Mildenhall et al., 2020). It is also view-dependent; this means that different viewpoint may give different information, whether density, RGB or others.

The NeRF workflow from a technical point of view is similar to that of MVS photogrammetry: overlapping images were taken of an object. In photogrammetry the poses of these images were then computed in a robust geometric operation (the bundle adjustment) and dense image matching was thereafter applied to generate a point cloud and subsequently a mesh. NeRF instead feeds these images into its neural network. However, NeRF in an on itself does not perform image orientation; even though it requires the input images to be pre-oriented. In this regard, NeRF may be considered as an alternative to traditional dense matching step in the photogrammetric workflow. Figure 1 shows the overall workflow of both NeRF and MVS as it is implemented in this study.

Due to these requirements, the workflow implemented in this paper starts with the acquisition of multiple overlapping photos of an object, similar to classical close-range photogrammetry. The images were then oriented using Agisoft Metashape and then fed into Nerfstudio (<https://docs.nerf.studio/>, accessed 11 April 2023). Nerfstudio is a simplified platform and interface to several implementations of NeRF; in this paper the Nerfacto (Tancik et al., 2022) variant was used for the experiments. After training and recreation of the 3D scene by Nerfstudio, a marching cubes algorithm was implemented to convert the radiance fields into a 3D point cloud. Conversely, the oriented images were thereafter used to also generate a dense point cloud using Metashape using the "High" preset. A comparison

between the two point clouds was then performed using the software CloudCompare (<https://www.danielgm.net/cc/> accessed 11 April 2023).

In this study, two heritage objects were recorded and compared. These two datasets include: dataset A comprising 24 images of the Reformers' Wall in the city of Geneva (Switzerland), and dataset B made of 85 images of the main entrance to the Bern Minster (Bern, Switzerland) with a Gothic tympanum depicting the Last Judgement. Dataset A is a simpler case reminiscent of the case of classical stereopairs set along one axis with overall also simpler object details, while dataset B presented a more complex object due to the various sculptures and ornate upper part of the tympanum. Both datasets were processed in Agisoft Metashape and then their point clouds compared to the ones generated by Nerfacto, with future test also planned for implementation in Instant-NGP.

Theoretical ground sampling distances (GSD) of 1.3 mm for dataset A and between 1 and 3.5 mm for dataset B were calculated to manage the expected result for the geometric quality. A Nikon Z50 mirrorless camera with a 24 mm lens was used to take 24 images for dataset A, while 85 images were taken using the same camera with a 28 mm lens setting.

3. RESULTS AND DISCUSSIONS

Results showed that Nerfacto generated 3D renders in seconds, with training process totalling longer but achieving a stable result after a few minutes. Point cloud conversion was near instantaneous, but as can be seen in Table 1, noise is still prominent in both datasets as evidenced by high values of standard deviation when compared to the reference Metashape point cloud.

Three quality parameters will be considered for each individual dataset. First, to check the geometric quality the value of mean error (\bar{x}) and standard deviation (σ) from an M3C2 (Lague et al., 2013) comparison performed in CloudCompare were used to assess Nerfacto's results, in terms of systematic error and presence of noise respectively. As both point clouds originate from the same network of oriented images, the value of \bar{x} is expected to be close to zero. This hypothesis proved to be more or less correct as seen in Table 1, where dataset A registered a value of 2 mm whereas dataset B yielded a value of 1 mm. These values are well within the expected order of precision and tolerated threshold.

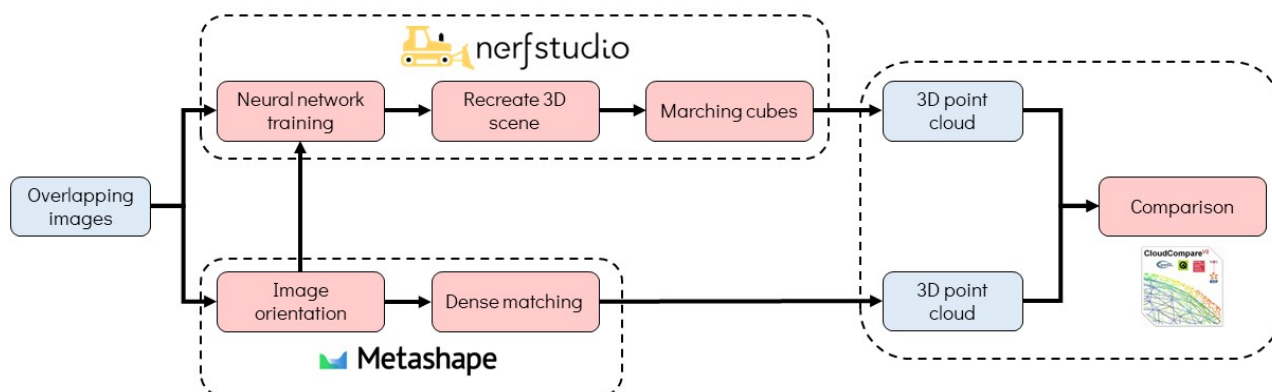


Figure 1. Flowchart of the experiment design used in this study. Nerfstudio (<https://docs.nerf.studio/>, accessed 11 April 2023) is a simplified interface supporting several implementations of NeRF, of which Nerfacto (Tancik et al., 2022) is used in this case.





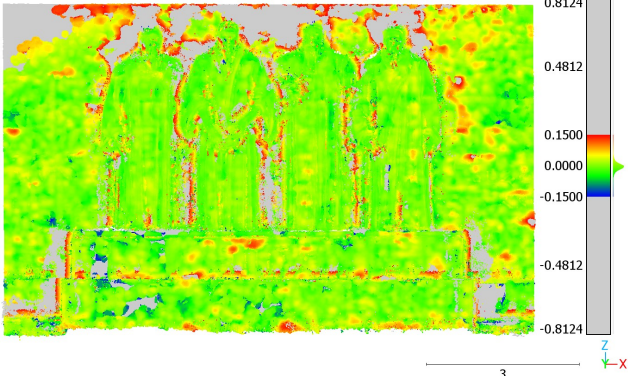
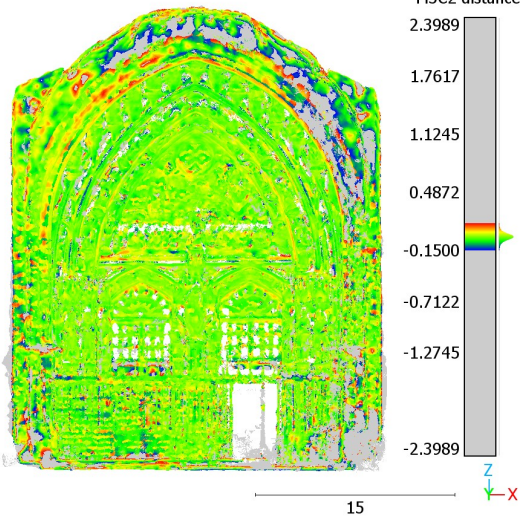
	Dataset A (24 images)	Dataset B (85 images)
Metashape	 <p style="text-align: center;"><i>8,914,633 points</i></p>	 <p style="text-align: center;"><i>11,569,500 points</i></p>
Nerfacto	 <p style="text-align: center;"><i>1,202,095 points</i></p>	 <p style="text-align: center;"><i>7,174,407 points</i></p>
M3C2 analysis	 <p style="text-align: center;">$\bar{x} = 0.2 \text{ cm}; \sigma = \pm 3.5 \text{ cm}; \text{completeness } 92.2\%$</p>	 <p style="text-align: center;">$\bar{x} = 0.1 \text{ cm}; \sigma = \pm 5.0 \text{ cm}; \text{completeness } 88.7\%$</p>

Table 1. Comparison of Metashape and Nerfacto on the two datasets used in this study. The third row shows results from the MC3D analysis performed on CloudCompare. The completeness value represents the percentage of inlier points from the Nerfacto point cloud as related to the reference Metashape point cloud.


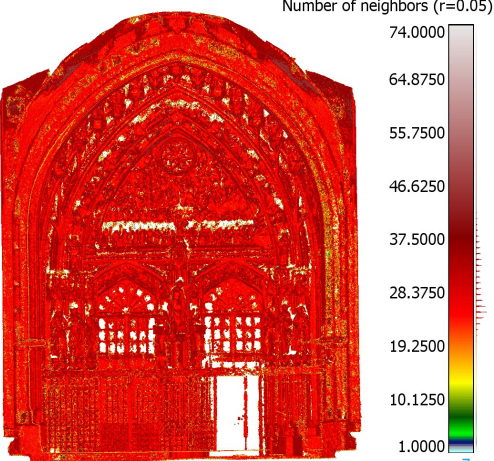
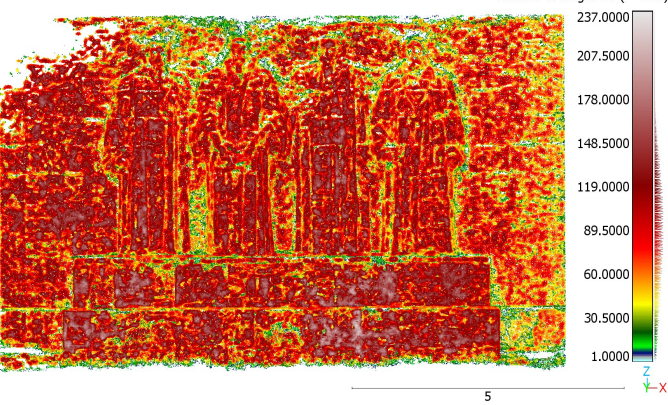
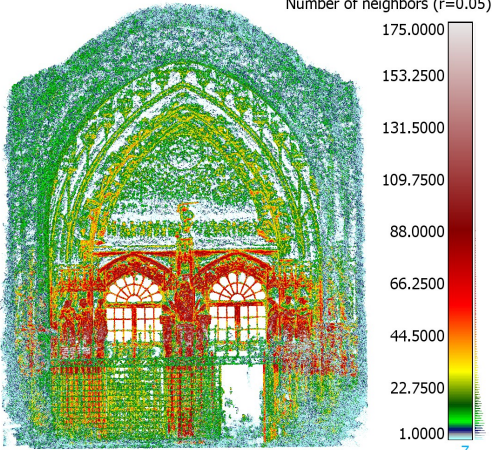
	Dataset A	Dataset B
Metashape	 <p>Average number of neighbours: 821.93 points</p>	 <p>Average number of neighbours: 29.40 points</p>
Nerfacto	 <p>Average number of neighbours: 86.75 points</p>	 <p>Average number of neighbours: 34.00 points</p>

Table 2. Density analysis as performed in CloudCompare. In this analysis, the number of neighbours were counted for each point constituting the point clouds within a sphere with a set radius of 5 cm.

Secondly, a measure of completeness was used. This value was obtained by overlaying the two point clouds. By assuming that the Metashape point cloud is by default more complete, each point in the Nerfacto point cloud is projected to it. Missing points are then considered as incomplete points and a percentage point representing the completeness rate was computed based on this information. Thirdly and finally, a density analysis was performed by computing the number of neighbours for each point in the Metashape and Nerfacto point clouds within a sphere with a radius of 5 cm. The first and second parameters are shown in Table 1, while the third parameter is described in Table 2.

For dataset A, Nerfacto generated around 1.2 million points with notable upper parts lacking in point density. This may be due to the fact that the images were taken from a lower point of view and oriented upwards; thus, the upper parts of the wall present the farthest distance from the camera. It is therefore interesting to note that object-to-sensor distance may play a role in determining the quality of the point cloud. It managed nevertheless to attain a completeness score of 92.2%. In the third row of Table 1, missing and incomplete parts of the object are coloured grey. In terms of geometric accuracy, it yielded a

mean error of 2 mm which is consistent with the initial hypothesis. It is however more interesting to observe the value of the standard deviation, of which a value of 3.5 cm was registered. This is more than 25 times higher than the theoretical GSD. Furthermore, considering a 2.7σ tolerance as is usual in surveying applications, it is about 10 times higher than the tolerated threshold. This is also evident visually, as can be seen in Table 1.

A similar observation can be seen in dataset B, where Nerfacto managed to generate 7.1 million points. The mean error is again consistent with the initial assumption, but the standard deviation reached 5 cm, which is around 15 times larger than the GSD. Within dataset B, missing parts can also be seen where they are expected to be, namely image blind spots and higher parts where the object to sensor distance is much higher. Indeed, it registered a completeness value of 88.7% which is nevertheless quite good considering the dimensions of the object (height of around 25 meters). An interesting visual observation in Table 1 may show, however, a discrepancy as regards to the density level of the point cloud in several areas.

Thirdly and finally, in order to ascertain the quality of the point cloud density, Table 2 shows a density analysis as performed in the software CloudCompare. In this analysis, the number of neighbouring points were counted for each point in the point cloud within a spherical radius of 5 cm. In both dataset A and B, Metashape showed a very homogeneous density throughout the resulting point clouds. While this may be due to point cloud post processing algorithms inside Metashape itself, it still presents a clean result throughout as far as point cloud density is concerned. With Nerfacto, a more heterogeneous density can be observed visually from Table 2.

In dataset A, Nerfacto registered a very low average of only 86.75 neighbours while Metashape gave an almost ten times denser value. The higher density area is concentrated on the left, central, and lower parts of the object while the upper parts registered lower densities. This finding is consistent with the previous completeness analysis.

For dataset B, Nerfacto yielded an average number of neighbours of 34 points. Interestingly, Metashape gave a slightly lower value of 29.40 points. Visual inspection indicate that Nerfacto generated very dense point clouds in several specific parts of the object, namely the central pillar. A cropped image showing more details of this central pillar can be seen in Figure 2. Note the higher density of the central part of the dataset, which may be explained by the fact that most of the images in the dataset is centred on this part, in line with keeping a convergent geometry commonly practised in close range photogrammetry. This induces more radiance fields on the said area, thereby generating a much denser part of the point cloud.

In Figure 2, several parts of the pillar gave a very high density value in Nerfacto's point cloud, averaging on 60 points. In some points, the number of neighbours even reach 204. The points with the highest density seem to be concentrated on object borders, while flat surfaces generally have reduced density. This is in sharp contrast to the same part as reconstructed by MVS in Metashape, where a very homogeneous density can be observed all throughout the central pillar. However, it is also worth noting that the results from Metashape is most likely subsampled in a post-processing step after the dense matching to obtain this homogeneous density.

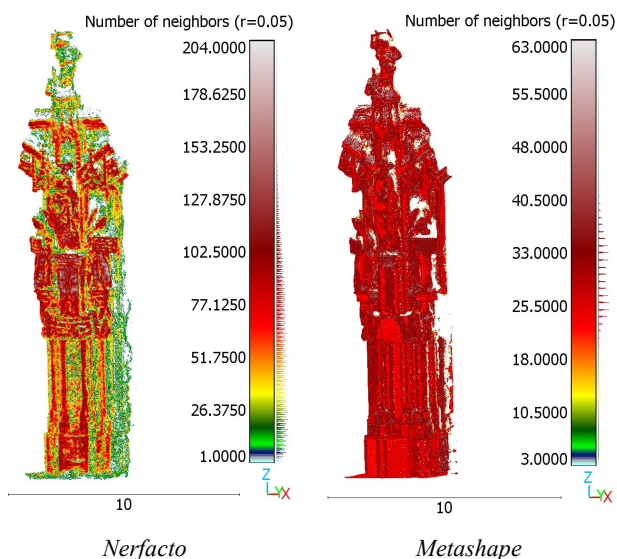


Figure 2. A more detailed view of the central pillar point cloud as generated by Nerfacto and Metashape.

4. CONCLUSIONS

This paper presented a preliminary test on the use of the novel NeRF method specifically for use in cultural heritage objects. The method performed well in terms of processing time, which is a fraction of the time required by traditional MVS. It is also accurate as far as point fidelity is concerned, as evidenced by the values of M3C2 mean error. It also fared well in the completeness test, despite encountering complex cases such as the ornate dataset B, scoring 88.7%. The simpler dataset A gave a slightly albeit even higher score of 92.2%.

Several limitations can however be identified within the context of heritage documentation from the authors' point of view. The current state of NeRF still requires image orientation, which in turn requires multiple overlapping images. Therefore, this does not reduce the field effort and time required to perform data acquisition. The orientation process also relies on two solutions: either (1) a traditional SfM photogrammetry process or (2) using the help of solid state lidar (SSL), for example as is available on modern iPad© and iPhone© devices. However, the field of NeRF is evolving in an exponential rate, and it is very probable that newer innovations may quickly remedy these limitations in the near future.

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. This is reminiscent of the results obtained by the Apple© SSL technology (Losé et al., 2022; Murtyoso et al., 2021), another novel technology which has recently seen applications also in heritage documentation. The density of the point cloud is also too heterogeneous for a robust documentation purpose, although slight modification to the data acquisition procedure (e.g., increasing the number of images on other areas of interest) may rectify this problem. As NeRF becomes more widespread and its workflow more streamlined in the future, these problems are expected to be more manageable. Indeed, a similar process can be observed during the early days of dense matching (Murtyoso et al., 2016).

Future investigation will involve tests on other variants of NeRF and comparison with other established methods for 3D heritage documentation, such as laser scanning. It is also interesting to note that NeRF emerged virtually in parallel to developments in miniaturised solid state lidar showing a very promising future for low-cost 3D heritage documentation. The nature of NeRF itself then raises the question on how the word "heritage documentation" would be defined in the future, specifically when used to refer to metric documentation.

REFERENCES

- Barron, J. T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R., & Srinivasan, P. P. (2021). Mip-NeRF: A Multiscale Representation for Anti-Aliasing Neural Radiance Fields. ICCV 2021.
- Condorelli, F., Rinaudo, F., Salvatore, F., & Tagliaventi, S. (2020). A neural networks approach to detecting lost heritage in historical video. ISPRS International Journal of Geo-Information, 9(5).
- Condorelli, F., Rinaudo, F., Salvatore, F., & Tagliaventi, S. (2021). A comparison between 3D reconstruction using nerf neural networks and MVS algorithms on cultural heritage images. International Archives of the Photogrammetry, Remote

- Sensing and Spatial Information Sciences - ISPRS Archives, 43(B2-2021), 565–570.
- 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.
- Lorensen, W. E., & Cline, H. E. (1987). Marching Cubes: A High Resolution 3D Surface Construction Algorithm. *Computer Graphics*, 21(4), 163–169.
- Losé, L. T., Spreafico, A., Chiabrando, F., & Tonolo, F. G. (2022). Apple LiDAR Sensor for 3D Surveying: Tests and Results in the Cultural Heritage Domain. *Remote Sensing*, 14, 4157.
- Martin-Brualla, R., Radwan, N., Sajjadi, M. S. M., Barron, J. T., Dosovitskiy, A., & Duckworth, D. (2021). NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 7206–7215.
- Matrone, F., Lingua, A., Pierdicca, R., Malinverni, E. S., Paolanti, M., Grilli, E., Remondino, F., Murtiyoso, A., & Landes, T. (2020). A benchmark for large-scale heritage point cloud semantic segmentation. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, 1419–1426.
- 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.
- Müller, T., Evans, A., Schied, C., & Keller, A. (2022). Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics*, 41(4).
- Murtiyoso, A., Grussenmeyer, P., Koehl, M., & Freville, T. (2016). Acquisition and Processing Experiences of Close Range UAV Images for the 3D Modeling of Heritage Buildings. In M. Ioannides, E. Fink, A. Moropoulou, M. Hagedorn-Saupe, A. Fresa, G. Liestøl, V. Rajcic, & P. Grussenmeyer (Eds.), *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection: 6th International Conference, EuroMed 2016, Nicosia, Cyprus, October 31 -- November 5, 2016, Proceedings, Part I* (pp. 420–431). Springer International Publishing.
- Murtiyoso, A., Grussenmeyer, P., Landes, T., & Macher, H. (2021). First assessments into the use of commercial-grade solid state lidar for low cost heritage documentation. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 43(B2-2021), 599–604.
- Murtiyoso, A., Pellis, E., Grussenmeyer, P., Landes, T., & Masiero, A. (2022). Towards Semantic Photogrammetry: Generating Semantically Rich Point Clouds from Architectural Close-Range Photogrammetry. *Sensors*, 22(3).
- Remondino, F., Spera, M. G., Nocerino, E., Menna, F., & Nex, F. (2014). State of the art in high density image matching. *The Photogrammetric Record*, 29(146), 144–166.
- Ren, Y., Chu, T., Jiao, Y., Zhou, M., Geng, G., Li, K., & Cao, X. (2022). Multi-Scale Upsampling GAN Based Hole-Filling Framework for High-Quality 3D Cultural Heritage Artifacts. *Applied Sciences (Switzerland)*, 12(9).
- Stathopoulou, E. K., Battisti, R., Cernea, D., Remondino, F., & Georgopoulos, A. (2021). Semantically derived geometric constraints for MVS reconstruction of textureless areas. *Remote Sensing*, 13(6), 1–19.
- Sun, J., Chen, X., Wang, Q., Li, Z., Averbuch-Elor, H., Zhou, X., & Snavely, N. (2022). Neural 3D Reconstruction in the Wild. *SIGGRAPH '22 Conference Proceedings*, August 7–11, 2022, Vancouver, BC, Canada, 1–9.
- Tancik, M., Weber, E., Ng, E., Li, R., Yi, B., Wang, T., Kristoffersen, A., Austin, J., Salahi, K., Ahuja, A., McAllister, D., & Kanazawa, A. (2022). Nerfstudio: A Framework for Neural Radiance Field Development. <https://Github.Com/Nerfstudio-Project/Nerfstudio>.