The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-1/W3-2023
2nd GEOBENCH Workshop on Evaluation and BENCHmarking of Sensors, Systems and GEOspatial Data
in Photogrammetry and Remote Sensing, 23–24 October 2023, Krakow, Poland

INVESTIGATION ON THE USE OF NeRF FOR HERITAGE 3D DENSE RECONSTRUCTION FOR INTERIOR SPACES

A. Murtiyoso [1]* J. Markiewicz [2], A. K. Karwel [2], P. Kot [3]

[1] Forest Resources Management, Institute of Terrestrial Ecosystems, Department of Environmental Systems Science, ETH Zurich, Switzerland – arnadi.murtiyoso@usys.ethz.ch
[2] Faculty of Geodesy and Cartography, Warsaw University of Technology, Warsaw, Poland – (jakub.markiewicz, artur.karwel)@pw.edu.pl
[3] Built Environment and Sustainable Technologies (BEST) Research Institute, Liverpool John Moores University, Liverpool, United Kingdom - p.kot@ljmu.ac.uk

Commission II

KEY WORDS: Benchmark, Heritage, Multi-View Stereo, NeRF, Photogrammetry, Point Clouds, Synthetic Data.

ABSTRACT

The concept of Neural Radiance Fields (NeRF) emerged in recent years as a method to create novel synthetic 3D viewpoints from a set of trained images. While it has several overlaps with conventional photogrammetry and especially multi-view stereo (MVS), its main point of interest is the capability to rapidly recreate objects in 3D. In this paper, we investigate the quality of point clouds generated by state-of-the-art NeRF in the context of interior spaces and compare them to four conventional MVS algorithms, of which two are commercial (Agisoft Metashape and Pix4D) and the other two open source (Patch-Match and Semi-Global Matching). Three synthetic datasets of interior scenes were created from laser scanning data with different characteristics and architectural elements. Results show that NeRF point clouds could achieve satisfactory results geometrically speaking, with an average standard deviation of 1.7 cm in interior cases where the scene dimension is roughly 25-50 m^3 in volume. However, the level of noise on the point cloud, which was considered as out of tolerance, ranges between 17-42%, meaning that the level of detail and finesse is most likely insufficient for sophisticated heritage documentation purposes, even though from a visualisation point of view the results were better. However, NeRF did show the capability to reconstruct texture less and reflective surfaces where MVS failed.

1. INTRODUCTION

Nowadays, cultural heritage is fundamental to modern societies as it preserves tangible and intangible evidence of the past. Digital technologies provide enhanced means to digitise, safeguard, and present cultural heritage assets, expanding their accessibility to a broader audience (Bocheska et al., 2023). For that purpose, nowadays, two common groups of non-invasive methods are used, namely image-based (also known as passive methods, i.e., close-range photogrammetry or multi-view stereo approaches) and range-based (i.e., Terrestrial Laser Scanning) (cf. Abbate et al., 2019; Arif and Essa, 2017; Cipriani et al., 2019, Doroszuke et al, 2022, Giżyńska et al., 2022, Murtiyoso et al., 2017, Tobiasz et al. 2022).

The rise of novel technologies based on artificial intelligence has been significant in the last decade. In 2020, the Neural Radiance Fields (NeRF) concept was introduced by Mildenhall et al. (2020). While initially developed as a solution for generating novel viewpoints in a 3D space, other researchers have quickly developed it to extract 3D models (Condorelli et al., 2021; Martin-Brualla et al., 2021). Therefore, NeRF presents a novel method of 3D reconstruction; whilst similar to the traditional photogrammetry and multi-view stereo (MVS) dense reconstruction workflow, it follows a different approach. Instead of performing a pixel-by-pixel reconstruction, NeRF predicts the level of transparency of every discrete element within a particular ray, the eponymous neural radiance field (Müller et al., 2022).

Several researchers have started experimenting with NeRF to digitally reconstruct heritage objects with the same approach as traditional photogrammetry but achieved different results depending on the case studies (Balloni et al., 2023; Croce et al., 2023; Mazzacca et al., 2023; Murtiyoso & Grussenmeyer, 2023; Vandenabeele et al., 2023). In these studies, NeRF demonstrated to work well in multiple cases, including aerial images, small objects, and outdoor environments, even though the geometric quality of the resulting point cloud is still significantly below those generated by conventional MVS algorithms. This raised the question of how well NeRF would work in the specific case of interior mapping. Indeed, conventional photogrammetry and MVS often encounter challenges in indoor environments, for example, the existence of reflective surfaces on windows and glasses. In this regard, alternative terrestrial laser scanning (TLS) is often used since it is faster and generates precise results. However, the TLS encounters the same problems facing reflective or metallic material. Furthermore, using TLS is often considered expensive within a heritage documentation context, and photogrammetry is often chosen precisely due to its lower cost. Results from Croce et al. (2023) already indicate that metallic surfaces may be reconstructed better using NeRF, but further investigations on its specific use for indoor spaces are still required.

In this paper, we propose a systematic benchmarking of NeRF results in the context of the indoor mapping of heritage sites. Three synthetic datasets were generated to this end from existing TLS data. The choice to use synthetic data was because the TLS generate good point clouds of the object and may thus be used as a reference for the experiments presented in this paper. The following sections shall discuss the methodology before showing the results and discussing the outcomes. The paper will end with a conclusion and recommendations for future work.

* Corresponding author

https://doi.org/10.5194/isprs-archives-XLVIII-1-W3-2023-115-2023

2. METHODOLOGY

2.1 Creation of synthetic dataset

To assess the quality of the generation of dense point clouds using an approach based on MVS and NeRF algorithms, it was decided to prepare synthetic data using the software to generate "Virtual Images" based on the Synthetic Images Simulator (Markiewicz et al., 2023).

Historic interiors at the Royal Palace in Warsaw and the Museum of King Jan III's Palace in Wilanów were selected as test sites. The following factors drove the choice of these test sites:

- **Benchmark 1** - historic 17th-century basements at the Royal Castle in Warsaw (Fig. 1a) without decorative structures consisting of bricks held together by mortar. They exhibit an asymmetrical form featuring arched ceilings, reaching a maximum height of around 3.2 metres and a minimum of approximately 2.1 metres.
- **Benchmark 2** - Museum of King Jan III's Palace in Wilanów - "The Chamber with a Parrot" (Fig. 1b) is characterised by its minimal adornments and the absence of bas-reliefs, facets, or textiles on the walls. Instead, patterns were painted on the walls in this test site to simulate spatial effects.
- **Benchmark 3** - Museum of King Jan III's Palace in Wilanów - "The Queen's Bedroom" (Fig. 1c) was characterised by intricate geometric patterns, including lavish ornaments, bas-reliefs, and facets. Additionally, the room featured mirrors adorned with golden frames, decorative fireplaces, and various fabrics adorning the walls, among other decorative elements.

Point clouds acquired from two phase scanners, Z+F 5006f (Benchmark 1 and Benchmark 2) and Z+F 5003 (Benchmark 3), were used as reference data to generate virtual images.

The images were distributed on a section of the sphere with a radius of 1.1 m placed in the central part of the point cloud. The distance between the images in a row of 19.5 cm and the distance between rows of 19.5 cm were assumed. This made it possible to generate 108 images for each benchmark with a resolution of 2192 x 1316 px, a focal length of 939.71 px and a camera lens angle of 70 degrees. All images were free of geometric distortion. An illustration of the three benchmark test sites is presented in Figure 1.

2.2 Creation and comparison of point clouds

In this paper, a specific emphasis was placed on evaluating the dense point cloud resulting from NeRF. For references, the point clouds were generated by four algorithms, namely Agisoft Metashape, Pix4D, and openMVS2, using both the Patch-Match (PM) and Semi-Global Matching (SGM) methods. The point cloud from NeRF was generated using Nerfstudio (https://docs.nerf.studio/, last accessed 13 July 2023), a visual interface containing various open implementations of the method. Specifically, the Nerfacto (Tancik et al., 2022) method was used in this paper. The parameters used to generate dense point clouds are listed in Table 1.

It is worth mentioning that NeRF, in general, including the Nerfacto implementation used in this paper, does not compute image orientation. NeRF naturally assumes that the exterior orientation parameters of the images were known beforehand. In this sense, NeRF is more or less analogous to MVS within the greater photogrammetric workflow, in which the main purpose is to create a dense point cloud. Therefore, the images used during the NeRF phase were first oriented using Agisoft Metashape; the exterior orientation parameters were thereafter fed into Nerfacto. This also implies that the absolute dimensions of the Nerfacto point cloud will naturally follow the absolute orientation setup in the Metashape project, therefore eliminating the need for absolute scaling and point cloud registration.

Comparison and analysis were made using the software CloudCompare (https://www.danielgm.net/cc/, last accessed 7 September 2023) to assess the geometric quality (via the M3C2 tool) relative to the four MVS point clouds. In this analysis, the NeRF point cloud was compared against each of the four MVS-generated point clouds using the statistical average and standard deviation values of the signed distances created by the M3C2 algorithm. Due to the use of the same exterior orientation parameters, registration was unnecessary. The average M3C2 distance error is nearly zero and virtually negligible in all the tested cases. Using the standard deviation values, however, we attempted to quantify the presence of noise.

Furthermore, an outlier analysis was performed by looking for the nearest neighbour correspondence of each NeRF points on the reference point clouds. Points without correspondence were considered as outliers and represented as percentage points. This parameter enabled us to assess the rate at which NeRF could create a faithful representation of each benchmark relative to each of the four photogrammetric references, which, in this case, were assumed to be of a higher quality. A point cloud density analysis was also conducted to compare the results from the different sources. This analysis computed the number of neighbouring points within a sphere of 1 cm^3 radius for each point in the NeRF point clouds. The density value is a simple but valuable parameter to assess the distribution of points in the point cloud.

Software	Quality Parameter (Image resolution)	Filtering	Number views	Fusion Mode	Iters	Matching window size
Agisoft Metashape	Ultra High (full resolution)	Depth filtering (Mild)	x	x	x	x
Pix4D	Full image size	x	x	x	x	9x9
OpenMVS-PM	1	0 (Disable)	0 - all neighbour views available	0	10	x
OpenMVS-SGM	1	0 (Disable)	0 - all neighbour views available	-1 -2	10	x

Table 1 The parameters for point cloud generation used for the MVS point clouds to be used as references.

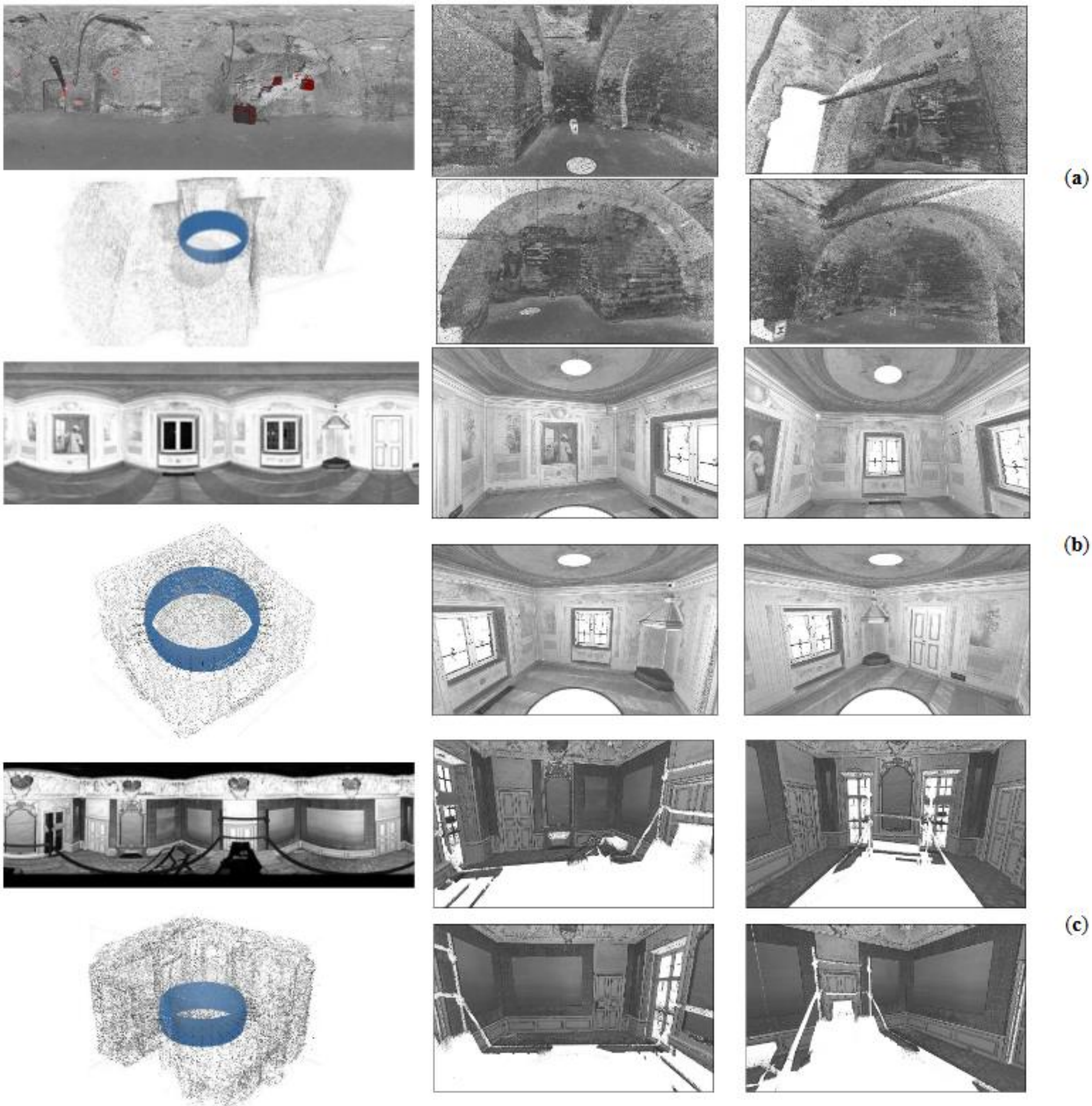


Figure 1 The virtual images used for dense point reconstruction: (a) Benchmark 1 - tin-roofed palace at the Royal Castle in Warsaw, (b) Benchmark 2 and (c) Benchmark 3 – rooms in the museum of King Jan III's Palace in Wilanów (Markiewicz et al., 2023).

Finally, a short analysis of the capability of NeRF to reconstruct reflective surfaces was performed on a part of Benchmark 2. This analysis assumed a flat surface as the ideal reference and compared all five point clouds (four MVS and one NeRF). A "completeness" value was also computed for each method by generating orthophotos of the windows and counting the number of non-white/empty pixels. In all analyses, a tolerance of 5 cm was imposed.

3. RESULTS AND DISCUSSIONS

3.1 Generated point clouds

The point clouds generated by the four MVS-based algorithms and NeRF are shown in Figure 2. Based on visual inspection, all MVS solutions allowed for generating high-quality point clouds. A notable observation is that Metashape was able to generate more points on texture less surfaces compared to the other solutions, with OpenMVS-SGM in second place. Pix4D generated the less dense point cloud, even with an ultra high setting.

The point cloud generated by Nerfacto can be seen to possess an elevated noise level and outliers. Even on flat surfaces with texture, Nerfacto generated noisy point clouds, if not slightly more complete than the other MVS results except for Metashape. Regarding computation time, Nerfacto required time to train its neural network using the 108 images for each benchmark. However, the training time is still faster than conventional dense matching by at least half when using the exact GPU specifications. It is also worth noting that Nerfacto generated radiance fields; the point cloud was converted using the marching cube method.

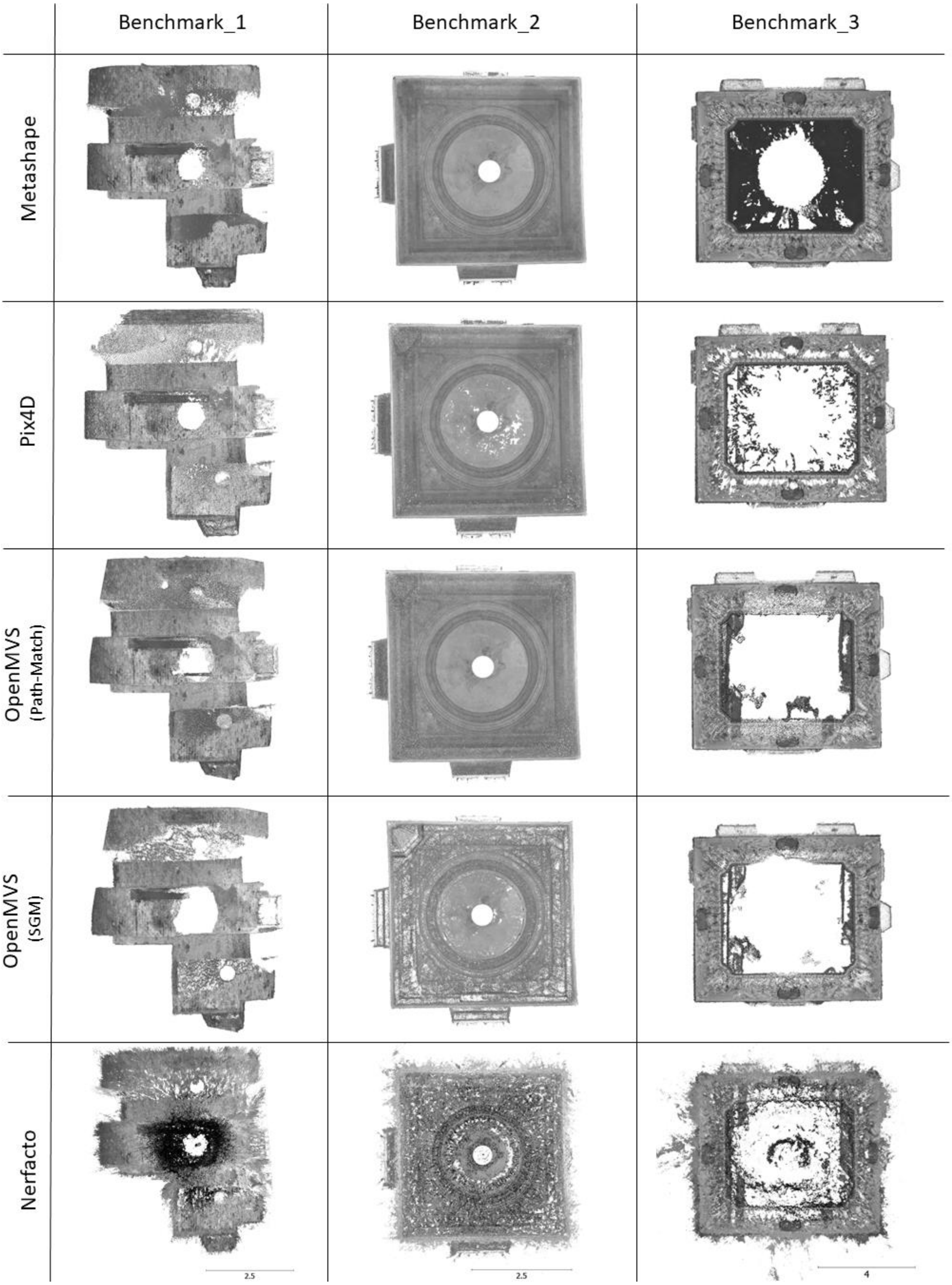
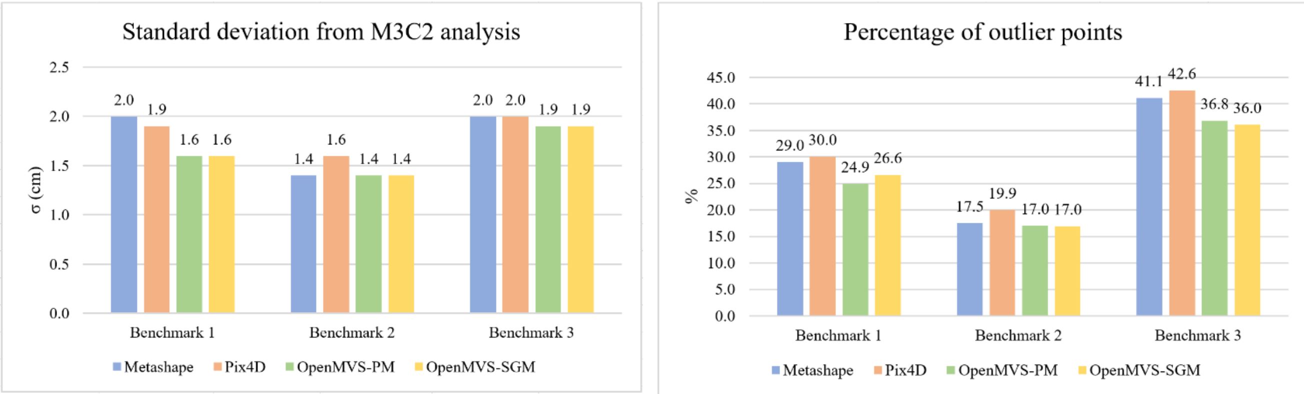
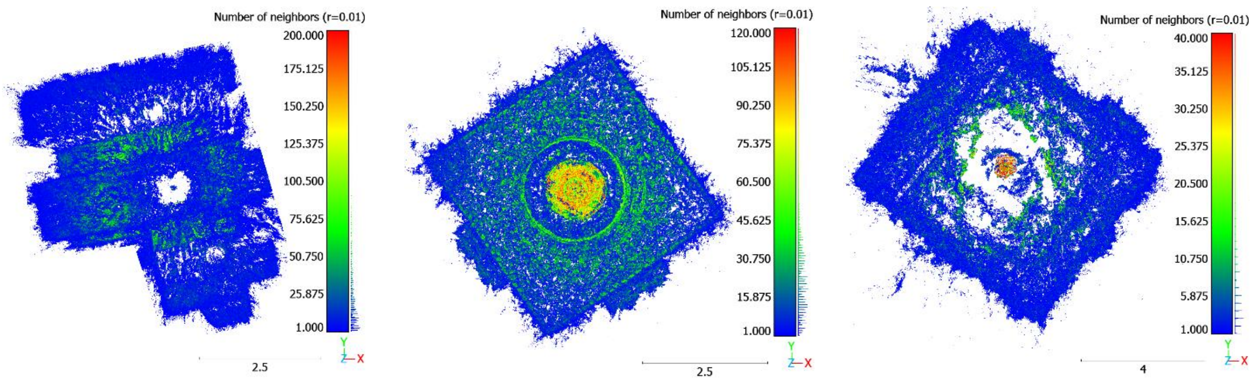


Figure 2 Resulting point clouds from the four reference MVS solutions (Metashape, Pix4D, OpenMVS-PatchMatch and OpenMVS-SGM) and the tested Nerfacto method.



(a) (b)

Figure 3 Graphs showing the results of the M3C2 analysis, where the four MVS point clouds were used as reference against the Nerfacto point cloud: (a) standard deviation of the signed distances and (b) percentage of points considered out of tolerance.



Average = 8.7 pts/cm³ Average = 6.4 pts/cm³ Average = 1.9 pts/cm³

(a) (b) (c)

Figure 4 Density analysis of the NeRF point clouds for (a) Benchmark 1, (b) Benchmark 2, and (c) Benchmark 3.

3.2 Geometric comparison (M3C2 analysis)

Figure 3 displays two histograms of the M3C2 analysis, one showing the standard deviation values and the other the percentage of points considered as outliers within the NeRF point cloud and concerning each MVS reference point cloud. An overall average value of 1.7 cm was achieved in terms of standard deviation. The highest values were obtained from Benchmarks 1 and 3, which may be influenced by the more complex form of Benchmark 1 and the presence of decorative and reflective surfaces in Benchmark 3. The standard deviation values may reflect the noise magnitude on the NeRF point clouds within the set tolerance; the 1.7 cm average value indicates that NeRF struggled to reconstruct well-defined elements. In this case, there is no perceptible difference between using any of the four MVS point clouds as a reference.

While the first part of Figure 3 described the presence of noise within the set tolerance of 5 cm, the second part attempted to quantify the number of points considered out of tolerance. Globally, the average outlier amounted to 28.2%, ranging from 17.0% (Benchmark 2, both OpenMVS) to 42.6% (Benchmark 3, Pix4D). In this regard, the highest percentage of outliers can be observed on Benchmark 3, with an average value of 39.1% of the points across the four references. This can also be perceived visually from Figure 2. The presence of reflective and textureless surfaces in Benchmark 3 may be the main reason for this, as MVS tended to leave uncertain surfaces as holes in the point cloud. In contrast, NeRF attempted to reconstruct it at the expense of creating more noisy points. Conversely, Benchmark 2 showed the lowest percentage of outlier points at an average of 17.8% across the four MVS references, although this value is still relatively high. Benchmark 2 has smaller dimensions than the other two, which may contribute to this result.

In general, observations show that NeRF provides a heterogenous point cloud density, with problems creating well-defined elements of a real-world object. The presence of noise is also considerable, as indicated by high outlier levels, particularly on Benchmark 3, although processing time is faster than traditional MVS. In both standard deviation and outlier-related analyses, the choice of MVS reference does not reflect an important change to the received values.

3.3 Density analysis of NeRF point clouds

The density analysis was performed solely on the NeRF point clouds, as seen in Figure 4. The average density of the point clouds ranged between 1.9 (Benchmark 3) and 8.7 points (Benchmark 1) per cm³. The very low value for Benchmark 3 again may be attributed to the fact that the space included many reflective surfaces, while Benchmark 1, which included bricks, provided more texture.

Another interesting observation is the distribution of points in each point cloud. In both Benchmarks 1 and 2, more points were generated for the parts of the scene nearest to the cameras, reaching 120-200 points per cm³. This starkly contrasts the other parts of the dataset, a majority of which registered values as low as 1 point per cm³. In all the MVS reconstructions, density values are virtually homogeneous, ranging from 8 to 12 points per cm³.

https://doi.org/10.5194/isprs-archives-XLVIII-1-W3-2023-115-2023

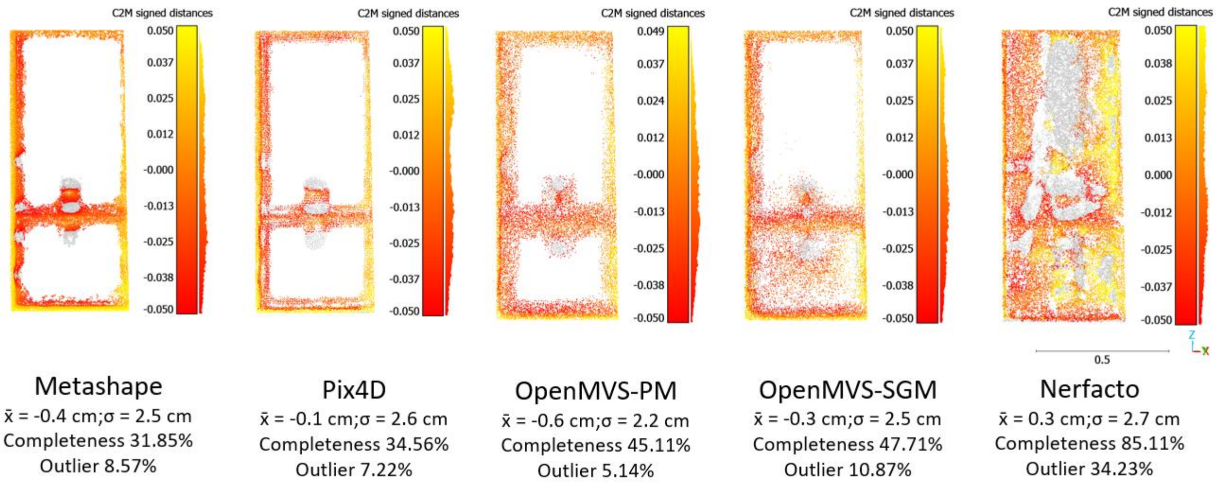


Figure 5 Comparison of a glass window in Benchmark 2 showing the capability of NeRF to reconstruct challenging surfaces, albeit with limited geometric quality.

3.4 Comparison of reflective surfaces

One advantage mentioned in other studies regarding the use of NeRF is its capability to reconstruct difficult surfaces, such as glasses or metal, which are traditionally challenging for MVS-based methods (Croce et al., 2023). In this section, a small portion of Benchmark 2 was segmented to assess this hypothesis. Specifically, a part of a glass window on Benchmark 2 was used as the sample case study.

An illustration and summary of the results of this analysis are shown in Figure 5. Visually, all MVS-based methods used in this study failed to reconstruct the window's glasses, only managing to generate points on the windowsills. However, NeRF, was able to generate points on the glass partition. To determine to what extent each solution could complete the reconstruction of this challenging architectural element, each window point cloud was converted into an orthophoto, and the number of non-white/empty pixels was counted. This led to the completeness values seen in Figure 5. These numerical values validate the visual observations that Nerfacto could reconstruct 85.11% of the window, while Metashape and Pix4D scored around 30% and OpenMVS 45%. SGM seemed to give the best result in this regard, even reconstructing the lower part of the glass windows. However, the completeness percentage does not show the geometric quality.

To quantify the geometric quality, a flat was created using RANSAC to represent the ideal flat surface of the window. All point clouds were then compared to this surface, and signed distances for each point were computed using the mesh-to-point cloud distance function in CloudCompare. All five point clouds generally gave an average error rate of 0.4 cm. However, the standard deviation of the NeRF point cloud achieved the same order of magnitude as the MVS results, showing how the NeRF results were also affected by noise.

Furthermore, using the same tolerance of 5 cm, some of the points in each point cloud were excluded and considered as outliers. All MVS point clouds except SGM generated less than 10% of outliers, with SGM yielding 10.87%. NeRF, conversely, is plagued by a large percentage of outliers; up to 34.23% of all points representing the window were considered outliers, meaning that they deviate from the ideal surface more than 5 cm.

4. CONCLUSIONS

This paper tested the feasibility of using the novel NeRF method to generate point clouds in the context of the interior mapping of heritage buildings. The results showed that NeRF is still very much dependent on conventional photogrammetry and Structure-from-Motion (SfM) for the orientation part. This means that, generally the resulting point clouds are relatively accurate. However, the quality of the dense cloud is far too noisy for proper use in heritage documentation, at least when compared to conventional MVS-based results. This was true compared to all four reference point clouds, with an average standard deviation of 2 cm and an outlier rate of 28%. The density of NeRF point clouds also wildly fluctuates depending on the object's position relative to the camera. Finally, using NeRF for reflective surfaces seemed promising if needed to be improved. Indeed, while NeRF successfully reconstructed most of the reflective surfaces, many generated points must be discarded as outliers.

The quality of NeRF point clouds results from the bottleneck of this method, which is the conversion of radiance fields into 3D points. The use of the marching cubes method, which greatly reduces the quality of the radiance field 3D visualisation when represented as point clouds. However, active research in this topic is an ongoing trend, and better results may be expected in the near future.

Another possible path forward would be introducing semantic priors in a mixed MVS-NeRF approach. Images may be pre-emptively segmented into classes, and the appropriate dense matching method is applied according to the identified classes.

ACKNOWLEDGEMENTS

This paper was co-financed under the research grant of the Warsaw University of Technology supporting the scientific activity in the discipline of Civil Engineering and Transport

REFERENCES

Abbate, E., Sammartano, G., & Spanò, A., (2019). Prospective upon multi-source urban scale data for 3d documentation and monitoring of urban legacies. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XLII-2/W11, 11–19.

Arif, R. & Essa, K., (2017). Evolving techniques of documentation of a world heritage site in Lahore. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-2/W5, 33–40.*

Balloni, E., Gorgoglione, L., Paolanti, M., Mancini, A., & Pierdicca, R. (2023). Few shot photogrammetry : a comparison between NeRF and MVS-SfM for the documentation of cultural heritage. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-M-2-2023, 155–162.*

Bochenska, A., Kot, P., Muradov, M., Markiewicz, J., Zawieska, D., Hess, M., and Antoniou, A. (2023): Critical evaluation of cultural heritage architectural standard documentation methods across different European countries. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-M-2-2023, 251–258,

Cipriani, L., Bertacchi, S., & Bertacchi, G., (2019). An optimised workflow for the interactive experience with cultural heritage through reality-based 3D models: Cases study in archaeological and urban complexes. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-2/W11, 427–434.*

Condorelli, F., Rinaudo, F., Salvadore, 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.

Croce, V., Caroti, G., Luca, L. De, & Piemonte, A. (2023). Neural radiance fields (NeRF): review and potential applications to digital cultural heritage. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLVIII-M-2-2023, 453–460.*

Doroszuk, K., Markiewicz, J. (2022). The possibility of using close-range photogrammetry in the inventory of historic complex basements - case study. *Sensors and Machine Learning Applications* , *1*(2). https://doi.org/10.55627/smla.001.02.0014

Giżyńska, J., Komorowska, E., & Kowalczyk, M. (2022). The comparison of photogrammetric and terrestrial laser scanning methods in the documentation of small cultural heritage object – case study. *Journal of Modern Technologies for Cultural Heritage Preservation*, *1*(1). https://doi.org/10.33687/jmtchp.001.01.0013

Markiewicz, J., Kowalczyk, M., Karwel, K., Kot, P., & Markiewicz, Ł. (2023). The evaluation of structure-from-motion workflow with the TLS synthetic images simulator – the cultural heritage approach. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLVIII-M-2-2023, 1013–1020.*

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.

Mazzacca, G., Karami, A., Rigon, S., Farella, E. M., Trybala, P., & Remondino, F. (2023). NeRF for heritage 3D reconstruction. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-M-2-2023, 1051–1058.*

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. (2023). Initial assessment on the use of state-of-the-art NeRF neural network 3d reconstruction for heritage documentation. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-M-2-2023, 1113–1118.*

Murtiyoso, A., Koehl, M., Grussenmeyer, P., & Freville, T., (2017). Acquisition And Processing Protocols For Uav Images: 3D Modeling Of Historical Buildings Using Photogrammetry. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, IV-2/W2, 163–170*

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.

Tobiasz, A., Zawieska, D., & Kopiasz, J. (2022). Methodology of spatial modelling and visualisation of remains of the fortified Lusatian settlement in Biskupin based on archival data. *Journal of Modern Technologies for Cultural Heritage Preservation*, 1(1). https://doi.org/10.33687/jmtchp.001.01.0008

Vandenabeele, L., Häcki, M., & Pfister, M. (2023). Crowd-sourced surveying for building archaeology: the potential of structure from motion (SFM) and neural radiance fields (NeRF). *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLVIII-M-2-2023, 1599–1605.*