The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-M-2-2023
29th CIPA Symposium "Documenting, Understanding, Preserving Cultural Heritage:
Humanities and Digital Technologies for Shaping the Future", 25–30 June 2023, Florence, Italy

NeRF FOR HERITAGE 3D RECONSTRUCTION

G. Mazzacca [1,2], A. Karami [1], S. Rigon [1], E.M. Farella [1], P. Trybala [1], F. Remondino [1]

[1] 3D Optical Metrology (3DOM) unit, Bruno Kessler Foundation (FBK), Trento, Italy
Web: https://3dom.fbk.eu - (gmazzacca, akarami, srigon, elifarella, ptrybala, remondino)@fbk.eu
[2] Dept. Mathematics, Computer Science and Physics, University of Udine, Italy

KEY WORDS: Neural Radiance Field, Heritage, 3D, Photogrammetry, AI

ABSTRACT:
Conventional or learning-based 3D reconstruction methods from images have clearly shown their potential for 3D heritage documentation. Nevertheless, Neural Radiance Field (NeRF) approaches are recently revolutionising the way a scene can be rendered or reconstructed in 3D from a set of oriented images. Therefore the paper wants to review some of the last NeRF methods applied to various cultural heritage datasets collected with smartphone videos, touristic approaches or reflex cameras. Firstly several NeRF methods are evaluated. It turned out that Instant-NGP and Nerfacto methods achieved the best outcomes, outperforming all other methods significantly. Successively qualitative and quantitative analyses are performed on various datasets, revealing the good performances of NeRF methods, in particular for areas with uniform texture or shining surfaces, as well as for small datasets of lost artefacts. This is for sure opening new frontiers for 3D documentation, visualization and communication purposes of digital heritage.

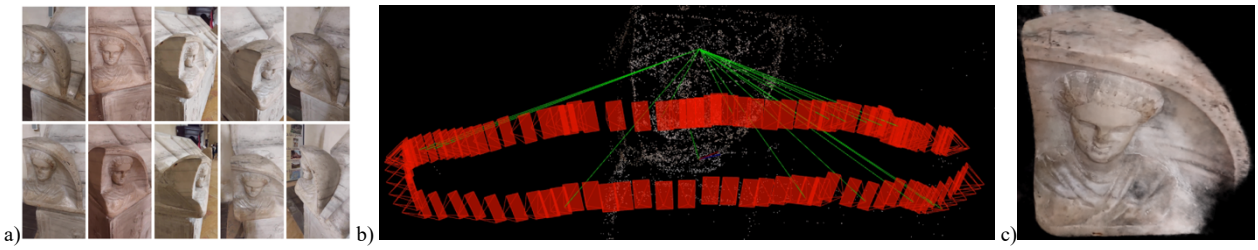

Figure 1. The NeRF method is able to optimize a continuous 5D neural radiance field representation of a scene starting from a set of oriented images. Some of the used images (a), recovered camera poses and sparse point cloud (b), and rendered 3D view from the NeRF representation (c).

1. INTRODUCTION

The 3D reconstruction and digital documentation of cultural heritage artefacts and scenes is an important task to valorize, study and safeguard, at least digitally, our patrimony. The improvements and efficiency of mass digitisation campaigns of cultural heritage have been driven mainly by the growing need for their preservation as well as by indubitable opportunities offered by digital 3D technologies, artificial intelligence (AI) methods and extended reality (XR) solutions for conservation, communication and virtual access purposes (Kniaz et al., 2019; Teruggi et al., 2021; Verhoeven et al., 2022). Nowadays, active and passive sensors, through static or mobile scanning and photogrammetric methods, provide reliable, fast and accurate 3D results (Di Stefano et al., 2021), often enriched with semantic information for further understanding and communication purposes (Grilli and Remondino, 2019; Mazzacca et al., 2022).
The photogrammetric pipeline starts from the acquisition phase, which is essential for retrieving high-quality images. Then, most of the processing steps are presently performed with automated structure from motion (SfM) approaches and multi-view stereo (MVS) algorithms (Zhou et al., 2020; Wang et al., 2021a,b).
A recent innovative approach for 3D scene reconstruction is offered by Neural Radiance Fields (NeRF - Figure 1). NeRF synthesizes novel views of complex scenes, starting from a set of oriented input images and optimizing an underlying continuous volumetric scene function (Mildenhall et al., 2020; Mueller et al., 2022). A neural radiance field is a simple fully connected network (weights of a few MB) trained to reproduce input views of a single scene using a rendering loss. The network directly maps from spatial location and viewing direction (5D input) to colour and opacity (4D output).
The aim of the paper is to shine light on emerging NeRF approaches for heritage 3D reconstruction in order to effectively use and optimize neural radiance fields to render novel photorealistic views of heritage scenes for 3D documentation, visualization and communication purposes.

2. RELATED WORKS

The recovery of 3D information from images is a long-lasting problem, solved for many years with conventional geometric-based approaches (Strecha et al., 2006; Goesele et al., 2007; Remondino et al., 2008, 2014; Hirschmuller, 2008; Barnes et al., 2009; Furukawa and Ponce, 2010; Jancosek and Pajdla, 2011; Bleyer et al., 2011; Rothermel et al., 2012; Schoenberger et al., 2016). Recently, learning-based 3D reconstruction methods based on point-, voxel-, mesh- or implicit (and differentiable) representations, have shown impressive results (Choy et al., 2016; Riegler et al., 2017; Chen and Zhang, 2019; Groueix et al., 2019; Wang et al., 2019; Yu and Gao, 2020), even from single images (Richter and Roth, 2018; Kniaz et al., 2019; Bath et al., 2023). Learning-based algorithms (CNN, GAN, etc.) try to infer a depth map from the set of input images, in a stereo or multi-view manner, with supervised or unsupervised approaches. Contrary to conventional methods based on handcrafted features (e.g., photometric consistency) in their cost functions, they try to reformulate the problem by also leveraging on semantic cues of the scene and learning more complex feature representations. Most methods require supervision and ground truth models,

which is often hard to obtain for real-world heritage contexts or are based on synthetic data. Therefore differentiable volumetric rendering (DVR) for implicit representations gained popularity as they can train reconstruction models from 2D images and learn implicit 3D shapes and textures (Liu et al., 2019; Niemeyer et al., 2020). Implicit representations represent shape and texture continuously and do not suffer, like voxel- and mesh-based representations, from discretization or low resolution.

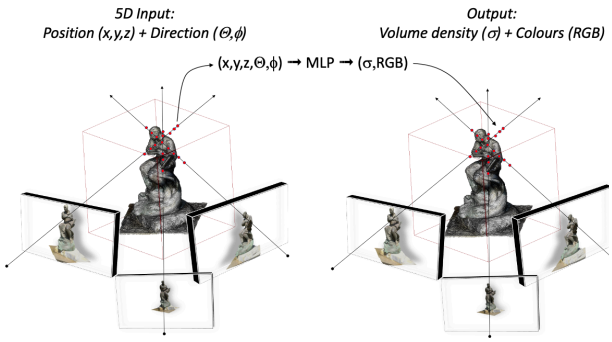

Figure 2. The basic of NeRF scene representation (built upon Mildenhall et al., 2020).

One of the last recent trends is based on neural scene representation (NeRF), which has gained popularity due to its expressiveness, speed of computation and, generally, low-memory need. Starting from the significant advance in the use of the attention mechanism (Vaswani et al., 2017), Mildenhall et al. (2020) introduced a method able to represent a scene using a deep fully-connected neural network without any convolutional layers (often referred to as a multilayer perceptron - MLP). The input for the neural network is a single continuous 5D coordinate set, i.e. spatial locations (x,y,z) and viewing directions (θ,ϕ), whereas the output is the volume density (σ) and view-dependent emitted radiance (RGB) in each direction and at each location (Figure 2). Starting from the recovered camera poses, the method is able to synthesize novel views by querying 5D coordinates along camera rays, and it uses classic volume rendering techniques to project the output colours and densities into an image. Further improvements to increase the performance of NeRF methods tackled the reduction of training time (Mueller et al., 2022), dynamic view synthesis (Pumarola 2021), limiting the number of required input images (Yu et al., 2021; Niemeyer, et al., 2022; Zhu et al., 2022), artefacts reduction (Barron et al., 2021), integration of depth supervision with sparse point clouds (Deng et al., 2022), knowledge incorporation such as Manhattan world priors (Guo et al., 2022) or monocular geometric cues (Yu et al., 2022a), upscaling to Street View (Rematas et al., 2022), large-scale (Yuanbo et al., 2022; Zhang et al., 2022;) and satellite (Roger et al., 2022) images, etc.

3. WHICH NERF?

Firstly we wanted to identify which NeRF methods were outperforming the others. Therefore we utilized SDFStudio[1] unified framework developed by Yu et al. (2022b), Nerfstudio[2] developed by Tancik et al. (2023) and NVlab[3] (Table 1). They are all consolidated neural implicit surface reconstruction approaches, enabling the development and visualization of NeRF scenes with controls and an easy workflow. In particular, the following methods were tested: Instant-NGP (Mueller et al., 2022), Nerfacto (Tancik et al., 2023), MonoSDF (Yu et al., 2022a), Tensorf (Chen et al., 2022), VolSDF (Yariv et al., 2021), Neus (Wang et al., 2021d), Unisurf (Oechsle et al., 2021), MipNeRF (Barron et al., 2022) and some variants like Neus-Facto, Mono-Neus and Mono-Unisurf.

Framework	Methods
SDFstudio	*Neus-Facto, MonoSDF, VolSDF, Mono-Neus, Neus, Mono-Unisurf, Unisurf, Mip-NeRF*
NeRFStudio	*Nerfacto, Tensorf*
NVlabs	*Instant-NGP*

Table 1. Summary of the tested frameworks and methods.

The performances of these methods were evaluated, among others [Karami et al., 2023], on the Ignatius dataset (Knapitsch et al., 2017), which contains 265 sequential images (extracted from a video at 1920x1080 px resolution). The comparison results with respect to ground truth data are reported in Figure 3. They show that Instant-NGP and Nerfacto methods achieved the best outcomes, with an error of approximately 1 cm and 1.5 cm, respectively, outperforming all other methods.

Instant-NGP uses multi-resolution hash encoding to reconstruct implicit surfaces. It is a practical and efficient learning-based approach that automatically identifies relevant details and is built upon the Tiny-CUDA-nn framework, which is a self-contained framework designed specifically for training and Lquerying neural networks. By leveraging these advanced techniques, Instant-NGP can achieve high-quality results while maintaining a fast computation time.

Nerfacto combines the latest components of NeRF techniques to balance speed and quality while remaining flexible for future modifications. It is influenced by MipNeRF-360 (Barron et al., 2022) but with optimizations and ideas from research papers such as NeRF-- (Wang et al., 2021b), NeRF-W (Martin-Brualla et al., 2021), Ref-NeRF (Verbin et al., 2022) and Instant-NGP (Müller et al., 2022).

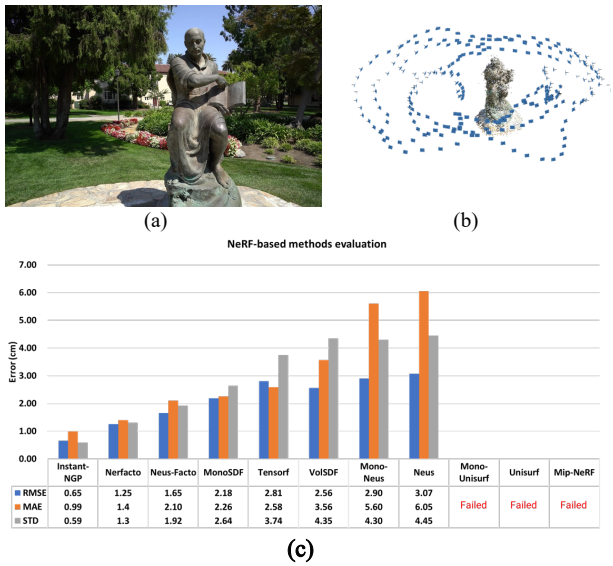

Figure 3. Sample images from the Ignatius dataset (a), recovered camera network (b) and comparison of NeRF-based methods using various criteria: RMSE, STD and MAE [cm].

[1] https://autonomousvision.github.io/sdfstudio/
[2] https://docs.nerf.studio/en/latest/
[3] https://github.com/NVlabs/instant-ngp

4. EXPERIMENTS

Following the outcomes presented in Section 3, Instant-NGP (from NVlabs) and Nerfacto (from Nerfstudio) are used to perform various experiments on some heritage datasets featuring different characteristics: availability of ground truth (GT) data (Section 4.1), presence of textureless/uniform (Section 4.2) or reflective (Section 4.3) surfaces and touristic repository of lost heritage (Section 4.4). For each dataset, the required camera poses are derived using COLMAP[4] or Agisoft Metashape[5] and ad-hoc converters[6,7] to import the camera parameters into the NeRF. After the training and rendering, a point cloud is generated and exported for analysis and visualization (Figure 4). All experiments were performed on an Alienware Aurora R12 with an 11th Gen Intel® Core™ i7-11700KF 3.60 GHz processor, 32GB of RAM and an NVIDIA GeForce RTX 3080 (10GB of VRAM).

4.1 Quantitative analysis

Geometric evaluation of NeRF-based 3D results with respect to reference ground truth (GT) or conventional Multi-View Stereo (MVS) pipelines are hereafter reported. The evaluation was performed by calculating the signed distances between the NeRF meshes and the reference one.

The first dataset consists of a smartphone video sequence (images at 960x540 px) acquired around a Mausoleum in Trento (Italy). The monument has a diameter of ca 25m and a height of ca 15m (without the basement). The acquisitions were performed below the main basement, at ca 10m distance from the object, producing occlusions. Around 200 frames were extracted to create 3D results with MVS (Colmap) and NeRF (Instant-NGP). The geometric comparison with the available Terrestrial Laser Scanner (TLS) revealed a standard deviation of ca 7.4 cm for the photogrammetric approach and ca 15 cm for the NeRF one (Figure 4, Table 2).

The second dataset consists of a smartphone video (3840x2160 px) of the remains of two arches of a structure situated in the archaeological site of Pafos (Cyprus). Approximately 180 frames, centred around a corner of the structure, and taken at a distance of roughly 10m while maintaining a parallel camera alignment to the archaeological remains, were extracted to apply a NeRF (Nerfacto) and MVS 3D reconstruction. The reference 3D data (GT) are provided by a photogrammetric dense point cloud derived from a set of images acquired with a Nikon D3X (6048x4032 px). The geometric comparisons (Figure 5, Table 2) indicate a similar standard deviation (less than 5 cm), although the Nerfacto output presents significantly more noise and details loss along the object's surfaces.

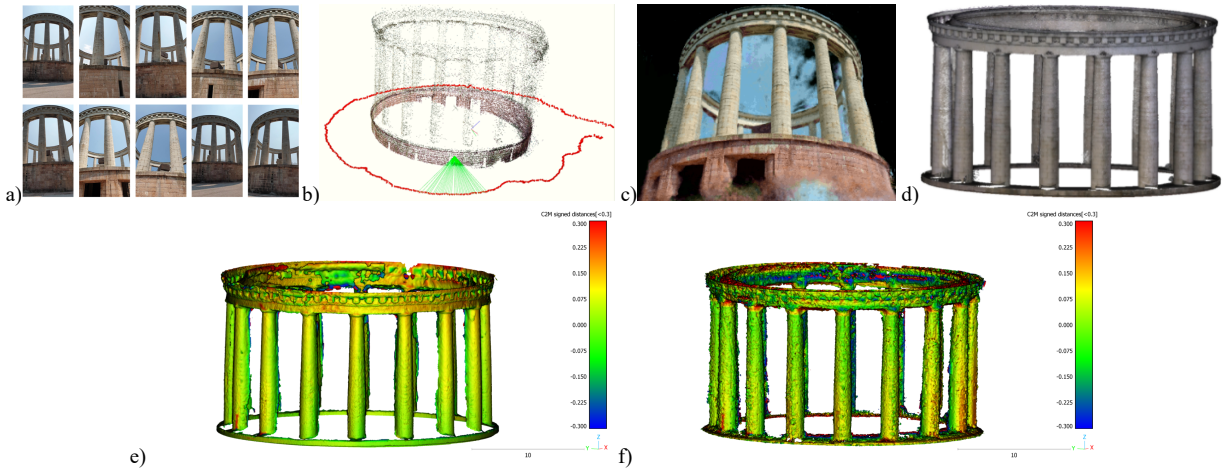

Figure 4. The smartphone-based Mausoleum (Doss Trento) dataset: some of the used images (a), derived camera network (b), InstantNGP NeRF 3D result (c) and TLS ground truth data (d). Geometric comparison of the photogrammetric (e) and NeRF (f) 3D results with respect to the TLS GT data (scalar field unit in meters).

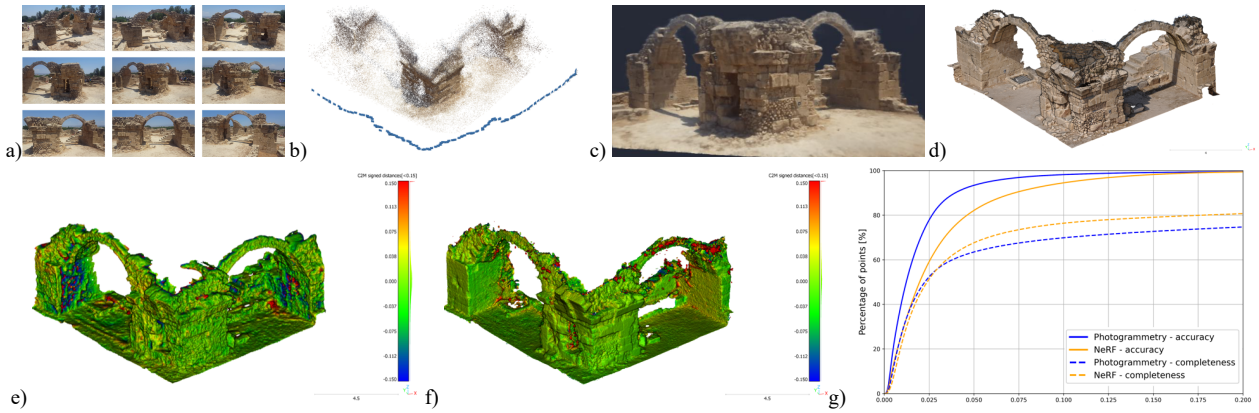

Figure 5. Some frames extracted from the video of the monument in Pafos (a), recovered camera network (b), 3D view of the Nerfacto (c) and photogrammetric (d) reconstruction from the extracted frames. Geometric comparison of NeRF (e) and MVS (f) with respect to GT (scalar field unit in meters). Accuracy and completeness analysis (g).

[4] https://colmap.github.io/
[5] https://www.agisoft.com/
[6] colmap2nerf.py: https://github.com/NVlabs/instant-ngp/tree/master/scripts
[7] agi2nerf: https://github.com/EnricoAhlers/agi2nerf

Figure 6. Piazza Duomo dataset with the recovered camera network (a), the 3D view of Nerfacto result (b) and the completeness evaluation (photogrammetry in blue, NeRF in orange) on the building façades (c). An orthographic view of the photogrammetric (d) and NeRF (e) point clouds, highlighting the capabilities of NeRF to better handle textureless areas.

	Photogrammetric model			***NeRF model***		
	Mean	*Std*	*Time*	*Mean*	*Std*	*Time*
Doss	3.4	7.4	ca 63 min	5.7	15	ca 30 sec
Pafos	0.6	4.8	ca 26 min	-0.1	4.7	ca 3 min

Table 2. Quantitative analyses [cm] for the Doss (Figure 4) and Pafos (Figure 5) datasets and processing time (3000 epochs for the NeRF approaches.

The Pafos dataset was also used to calculate accuracy and completeness (often named as precision and recall, respectively) following the approaches of Knapitsch et al. (2017) and Nocerino et al. (2020). The two metrics were computed with respect to the photogrammetric (Nikon) 3D model. Figure 5g shows how the video-based photogrammetric reconstruction is more accurate whereas the NeRF 3D model has a higher completeness.

4.2 Textureless surfaces

Conventional SfM and MVS methods normally meet problems while performing 3D reconstruction of surfaces with uniform colours or textureless areas.
A dataset of 20 high-resolution images (6048x4032 px) taken with a Nikon D3X was acquired on some buildings in the Trento's Duomo square (Italy). The images were captured at ground level, at varying distances from the building facades which have evenly painted plasters. The MVS processing was done in Metashape whereas Nerfacto was used for the NeRF 3D reconstruction (Figure 6). The 3D result generated with NeRF seems to be more complete, with higher density and more consistent point distribution in the challenging areas.
As no real ground truth data were available, the completeness is computed with an approach for a planar-like surface built upon Knapitsch et al. (2017). First, both point clouds are cropped to the common area of interest. A reference plane is determined by fitting a plane to a downsampled photogrammetric point cloud using a least-squares approach. Both point clouds are then projected onto this plane and their 3D coordinates are reduced to 2D in a new coordinate frame defined by the plane and the projections of the original Y and Z axes on it. In this new reference frame, a ground truth polygon of the complete façade is defined by constructing a concave hull of all evaluated point clouds. To evaluate the completeness, for each point in the evaluated point clouds, a buffer is calculated at a series of distance thresholds τ. The resulting polygons are merged into single geometries for each τ and cropped to a common extent within the ground truth polygon. The completeness function $C(\tau)$ is then defined as the ratio between the area of the polygon obtained for a certain τ and the total area of the reference façade polygon. The results show NeRF outperforming photogrammetry at a 1cm distance threshold by 10pp of the completeness metric (Figure 6c).

4.3 Reflective surfaces

Conventional SfM and MVS methods face problems if reflective and shining or transparent surfaces have to be digitized. A dataset of about 60 high-resolution images (6048x4032 px) was captured with a NIKON D750 camera at the MAG museum in Riva del Garda (Italy). The object (Figure 7a) is a small bronze statue featuring reflective surfaces and a transparent basement. The images were acquired by rotating the object and capturing images from both parallel and oblique points of view (Figure 7b).

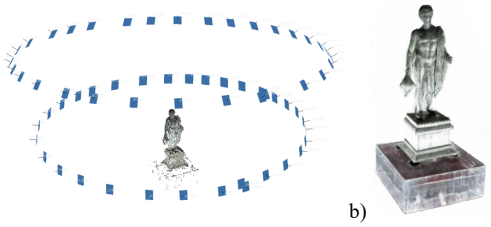

Figure 7. Camera network (a) and closeup of the 3D object (b).

Figure 8 shows the 3D results of the conventional MVS in Metashape, Nerfacto in NerfStudio and Instant-NGP in NVlab. The bronze surface was nicely reconstructed by MVS and Nerfacto methods but not by Instant-NGP. It is however noticeable how the Nerfacto point cloud contains a significant amount of noise at the border of the object. Probably with shorter baselines (i.e. more images), both NeRF methods would have

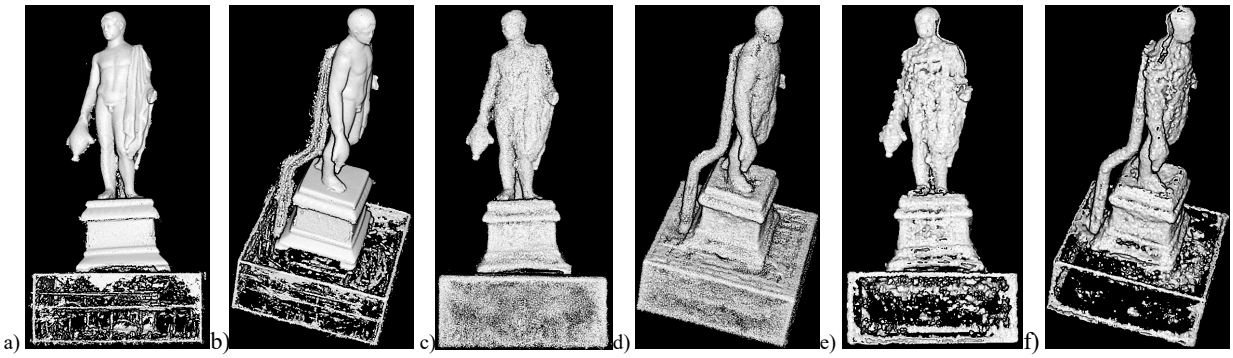

Figure 8. Photogrammetric MVS (a-b), Nerfacto (c-d) and Instant-NGP (e-f) point clouds of the bronze statue with a transparent basement and support.

performed better. Nerfacto outperformed in reconstructing the transparent pedestal and back-support. This highlights the potential of (some) NeRF methods in reconstructing transparent surfaces in a variety of contexts. The computational time for MVS was ca 8 min, ca 3 min for the Nerfacto and ca 40 sec for Instant-NGP (3000 epochs).

4.4 Unconstrained touristic images (Photo-tourism)

Photogrammetry has been often used to reconstruct lost heritage objects or monuments by using tourist or archival photos (Gruen et al., 2004). The potential of NeRF methods was tested on a set of ca 30 unordered touristic images taken from the online repository REKREI[8] (Vincent et al., 2015, 2016) focused on the Temple of Baalshamin in Palmyra, a monument destroyed in 2015 (Figure 9a). The dataset contains images of varying resolutions and distances, most focusing on the temple's frontal part. For the processing, Colmap was applied as conventional MVS approach. On the other hand, the Photo-tourism implementation in NerfStudio, probably similar to NeRF-W (Martin-Brualla et al., 2021), was chosen as it was developed to handle unconstrained image collections "in the wild" and different camera models. The processing for Colmap took ca 30 min whereas the NeRF approach needed ca 2 min. Due to the limited number of images from the sides and back, both approaches failed to reconstruct those parts. As shown in Figure 9d, the Colmap dense point cloud shows lower density and completeness in a few areas compared to the NeRF results (Figure 9e), such as the columns' bases and the inner part of the facade. This is possibly due to the large baselines or inconsistencies among the images, caused by differences in acquisition conditions as well the front columns casting ever-changing shadows on the inner façade. However, the NeRF dense point cloud is noisier compared to the dense point cloud derived in Colmap.

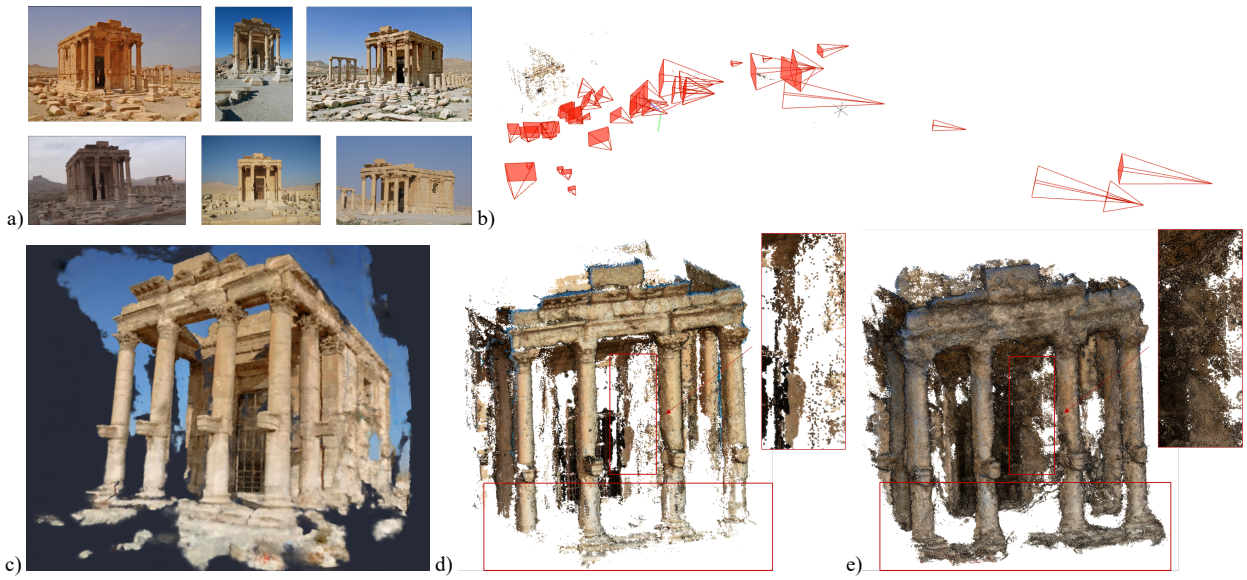

Figure 9. Some of the REKREI images utilised for the 3D reconstruction of the Palmyra temple (a) and the recovered camera network (b). NeRF-W 3D view (c) and visual comparison of the photogrammetric (b) and NeRF (c) 3D results.

[8] https://rekrei.org/

5. CONCLUSIONS AND FUTURE WORKS

The work presented an investigation of NeRF methods for heritage 3D reconstruction. Qualitative and quantitative results reported the capabilities of neural radiance fields to derive quite accurate 3D models from a set of images. Textureless, transparent and reflective surfaces were also considered as well as low- and high-resolution images, acquired with smartphones or reflex cameras.

Instant-NGP and Nerfacto were primarily utilised as they show the best performances on a typical historical monument. Additionally, the NeRF-W method was employed to process an unstructured collection of touristic images representing a heritage site that has been destroyed.

The quantitative analyses indicate a comparable level of accuracy to the dense point cloud generated through conventional MVS methods, with Colmap having a slightly better accuracy although requiring more processing time. Moreover, NeRF methods appear to perform better in scenarios where conventional MVS techniques usually struggle. Even if more tests are surely needed, their performances on textureless surfaces and transparent objects seem very promising. Surely, time-wise, the NeRF approach is generally faster than a MVS approach.

This article serves as an initial evaluation of NeRF capabilities in producing cultural heritage 3D contents. In the next phase of our research, we will narrow our focus to specific tasks to obtain a more comprehensive understanding of the behaviour and true potential of various NeRF methods in the cultural heritage domain. In particular, we will:

- Investigate the impact of image quality and quantity on the accuracy and completeness of NeRF-based 3D reconstructions of cultural heritage objects;
- Perform an extended assessment of NeRF capabilities to accurately reconstruct reflective and transparent surfaces;
- Evaluate reliable approach to remove background which is not part of the area/object we want to digitally reconstruct;
- Explore NeRF's potential in accurately reconstructing cultural heritage objects from tourist datasets with unconstrained acquisition conditions, focusing in particular on of lost monuments;
- Finalize the NeRFBK dataset (https://github.com/3DOM-FBK/NeRFBK) for benchmarking NeRF methods in various contexts and scenarios (heritage, industry, urban, etc.).

ACKNOWLEDGMENTS

Authors are thankful to Matthew Vincent for supporting the data collection within the REKREI database.

REFERENCES

Barnes, C., Shechtman, E., Finkelstein, A., and Goldman, D. B. Patchmatch: A randomized correspondence algorithm for structural image editing. *ACM Trans. Graph*., 28(3):24, 2009.

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. Proc. *ICCV*, pp. 5855-5864.

Barron, J. T., Mildenhall, B., Verbin, D., Srinivasan, P. P., Hedman, P., 2022. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. Proc. *CVPR*, pp. 5470-5479.

Bath S.F., Birkl, R., Wofk, D., Wonka, P., Müller, M., 2023. ZoeDepth: Zero-shot Transfer by Combining Relative and Metric Depth. *arXiv*:2302.12288.

Bleyer, M., Rhemann, C., Rother, C., 2011. PatchMatch Stereo-Stereo Matching with Slanted Support Windows. Proc. *BMVC*.

Chen, Z., Zhang, H., 2019. Learning implicit fields for generative shape modeling. Proc. *CVPR.*

Chen, A., Xu, Z., Geiger, A., Yu, J., Su, H., 2022. TensoRF: Tensorial Radiance Fields. Proc. *ECCV*.

Choy, C.B., Xu, D., Gwak, J., Chen, K., Savarese, S., 2016. 3D-R2N2: A unified approach for single and multi-view 3D object reconstruction. Proc. *ECCV*.

Deng, K., Liu, A., Zhu, J.-Y., Ramanan, D., 2022. Depth-supervised NeRF: fewer views and faster training for free. Proc. *CVPR*.

Di Stefano, F., Torresani, A., Farella, E.M., Pierdicca, R., Menna, F., Remondino, F., 2021. 3D Surveying of Underground Built Heritage: Opportunities and Challenges of Mobile Technologies. *Sustainability*, Vol.13, 13289.

Furukawa, Y. and Ponce, J., 2010. Accurate, dense and robust multiview stereopsis. IEEE *Trans. PAMI*, 32(8): 1362-1376.

Goesele, M., Snavely, N., Curless, B., Hoppe, H. and Seitz, S. M., 2007. Multi-view stereo for community photo collections. Proc. *ICCV*.

Grilli, E., Remondino, F., 2019. Classification of 3D Digital Heritage. *Remote Sensing*, Vol. 11(7), 847.

Groueix, T., Fisher, M., Kim, V.G., Russell, B.C., Aubry, M., 2019. AtlasNet: A papiermache approach to learning 3D surface generation. Proc. *CVPR*.

Gruen, A., Remondino, F., Zhang, L., 2004. Photogrammetric Reconstruction of the Great Buddha of Bamiyan, Afghanistan. The Photogrammetric Record, Vol.19(107), pp. 177-199.

Guo, H., Peng, S., Lin, H., Wang, Q., Zhang, G., Bao, H., Zhou, X., 2022. Neural 3D scene reconstruction with the Manhattan-world assumption. Proc. *CVPR*.

Hirschmuller, H., 2008. Stereo processing by semiglobal matching and mutual information. IEEE *Trans. PAMI*, 30(2): 328-342.

Jancosek, M., Pajdla, T., 2011. Multi-view reconstruction preserving weakly-supported surfaces. Proc. *CVPR*.

Karami, A., Rigon, S., Yan, Z., Mazzacca, G., Remondino, F., Qin, R., 2023. A critical analysis of NeRF-based 3D reconstruction. *Remote Sensing*, in review.

Knapitsch, A., Park, J., Zhou, Q.Y. and Koltun, V., 2017. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics*, *36*(4), pp.1-13.

Kniaz, V. V., Remondino, F., Knyaz, V. A., 2019. Generative Adversarial Networks for single photo 3D reconstruction. *ISPRS*

Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W9, pp. 403-408.

Liu, S., Chen, W., Li, T., Li, H., 2019. Soft rasterizer: Differentiable rendering for unsupervised single view mesh reconstruction. Proc. *ICCV*.

Liu, L., Gu, J., Lin, K.Z., Chua, T.-S., and Theobalt, C., 2020. Neural Sparse Voxel Fields. Proc. *NIPS*.

Martin-Brualla, R., Radwan, N., Sajjadi, M. S., Barron, J. T., Dosovitskiy, A., Duckworth, D., 2021. Nerf in the wild: Neural radiance fields for unconstrained photo collections. Proc. *CVPR*, pp. 7210-7219.

Mazzacca, G., Grilli, E., Cirigliano, G.P., Remondino, F., Campana, S., 2022. Seeing among foliage with lidar and machine learning: towards a transferable archaeological pipeline. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* XLVI-2/W1-2022, 365-372.

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. Proc. *ECCV*.

Müller, T., Evans, A., Schied, C., Keller, A., 2022. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. ACM Trans. Graph., Vol. 41(4), pp. 102-117.

Niemeyer, M., Mescheder, L., Oechsle, M., Geiger, A., 2020. Differentiable Volumetric Rendering: Learning Implicit 3D Representations without 3D Supervision. Proc. *CVPR*.

Niemeyer, M., Barron J.T., Mildenhall, B., Sajjadi, M.S.M., Geiger, A., Radwan, N., 2022. RegNeRF: Regularizing Neural Radiance Fields for View Synthesis from Sparse Inputs. Proc. *CVPR*.

Nocerino, E., Stathopoulou, E. K., Rigon, S., & Remondino, F., 2020. Surface reconstruction assessment in photogrammetric applications. *Sensors*, 20(20), 5863.

Oechsle, M., Peng, S., Geiger, A., 2021. UNISURF: Unifying Neural Implicit Surfaces and Radiance Fields for Multi-View Reconstruction. Proc. *ICCV*.

Pumarola, A., Corona, E., Pons-Moll, G., Moreno-Noguer, F., 2021. D-NeRF: Neural Radiance Fields for Dynamic Scenes. Proc. *CVPR*.

Rematas, K., Liu, A., Srinivasan, P.P., Barron, J.T., Tagliasacchi, A., Funkhouser, T., Ferrari, V., 2022. Urban Radiance Fields. Proc. *CVPR*.

Remondino, F., El-Hakim, S., Gruen, A., Zhang, L., 2008. Turning images into 3D models - Development and performance analysis of image matching for detailed surface reconstruction of heritage objects. *IEEE Signal Processing Magazine*, Vol. 25(4), pp. 55-65.

Remondino, F., Spera, M.G., Nocerino, E., Menna, F., Nex, F., 2014. State of the art in high density image matching. *The Photogrammetric Record*, Vol. 29(146), pp. 144-166.

Richter, S. R. and Roth, S., 2018. Matryoshka Networks: Predicting 3D geometry via nested shape layers. Proc. *CVPR*.

Riegler, G., Ulusoy, A.O., Geiger, A., 2017. Octnet: Learning deep 3d representations at high resolutions. Proc. *CVPR*.

Roger, M., Facciolo, G., Ehret, T., 2022. Learning Multi-View Satellite Photogrammetry With Transient Objects and Shadow Modeling Using RPC Cameras. Proc. *CVPRW*, pp. 1310-1320.

Rothermel, M., Wenzel, K., Fritsch, D. and Haala, N., 2012. SURE: photogrammetric surface reconstruction from imagery. Proc. *LC3D Workshop*, Berlin, Germany.

Schoenberger, J., Zheng, E., Pollefeys, M., Frahm, J.-M., 2016. Pixelwise View Selection for Unstructured Multi-View Stereo. Proc. *ECCV*.

Strecha, C., Fransens, R. and Van Gool, L., 2006. Combined depth and outlier estimation in multi-view stereo. Proc. *CVPR*., Vol. 2, pp. 2394-2401.

Tancik, M., Weber, E., Ng, E., Li, R., Yi, B., Kerr, J., Kanazawa, A., 2023. Nerfstudio: A modular framework for neural radiance field development. *arXiv preprint* arXiv:2302.04264.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, L., Polosukhin, I., 2017. All you need is attention. Proc. *NIPS*.

Verbin, D., Hedman, P., Mildenhall, B., Zickler, T., Barron, J.T. and Srinivasan, P.P., 2022, June. Ref-nerf: Structured view-dependent appearance for neural radiance fields. Proc. *CVPR*, pp. 5481-5490.

Verhoeven, G., Wild, B., Schlegel, J., Wieser, M., Pfeifer, N., Wogrin, S., Eysn, L., Carloni, M., Koschiček-Krombholz, B., Molada-Tebar, A., Otepka-Schremmer, J., Ressl, C., Trognitz, M., and Watzinger, A., 2022. Project indigo – document, disseminate & analyse a graffiti-scape. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* XLVI-2/W1-2022, pp. 513-520.

Vincent, M. L., Coughenour, C., Flores Gutierrez, M., Lopez-Menchero Bendicho, V. M., Remondino, F., Frtisch, D., 2015. Crowd-sourcing the 3D digital reconstructions of lost cultural heritage. Proc. *IEEE Digital Heritage*, Vol. 1.

Vincent, M., Coughenour, C., Remondino, F., Gutierrez, M.F., Lopez-Menchero Bendicho, V.M., Fritsch, D., 2016. Rekrei: A public platform for digitally preserving lost heritage. Proc. 44th *CAA Conference*.

Teruggi, S., Grilli, E., Fassi, F., Remondino, F., 2021. 3D surveying, semantic enrichment and virtual access of large cultural heritage. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, Vol. VIII-M-1-2021, pp. 155-162.

Wang, W., Qiangeng, X., Ceylan, D., Mech, R., Neumann, U., 2019. DISN: Deep implicit surface network for high-quality single-view 3D reconstruction. Proc. *NIPS*.

Wang, F., Galliani, S., Vogel, C., Speciale, P., Pollefeys, M., 2021a. PatchmatchNet: Learned Multi-View Patchmatch Stereo. Proc. *CVPR*.

Wang, X., Wang, C., Liu, B., Zhou, X., Zhang, L., Zheng, J., Bai, X., 2021b. Multi-view stereo in the Deep Learning Era: A comprehensive review. *Display*, Vol. 70.

Wang, P., Liu, L., Liu, Y., Theobalt, C., Komura, T., Wang, W., 2021c. NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction. Proc. *NIPS*.

Wang, Z., Wu, S., Xie, W., Chen, M., & Prisacariu, V. A., 2021d. NeRF--: Neural radiance fields without known camera parameters. *arXiv preprint* arXiv:2102.07064

Yariv, L., Gu, J., Kasten, J., Lipman, J., 2021. Volume Rendering of Neural Implicit Surfaces. Proc. *NISP*.

Yu, Z., Gao, S., 2020. Fast-MVSNet: Sparse-to-Dense Multi-View Stereo with Learned Propagation and Gauss-Newton Refinement. Proc. *CVPR*.

Yu, A., Ye, V., Tancik, M., Kanazawa, A., 2021. pixelNeRF - Neural Radiance Fields from One or Few Images. Proc. *CVPR*.

Yu, Z., Peng, S., Niemeyer, M., Sattler, T., Geiger, A., 2022a. MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. Proc. *NIPS*.

Yu, Z., Chen, A., Antic, B., Peng, S. P., Bhattacharyya, A., Niemeyer, M., Tang, S., Sattler, T., Geiger, A., 2022b. SDFStudio: A Unified Framework for Surface Reconstruction.

Yuanbo, X., Linning, X., Xingang, P., Nanxuan, Z., Anyi, R., Theobalt, C., Dai, B., Lin, D., 2022. BungeeNeRF: Progressive Neural Radiance Field for Extreme Multi-scale Scene Rendering. Proc. *ECCV*.

Zhang, X., Bi, S., Sunkavalli, K., Hao, S., Zexiang, X., 2022. NeRFusion: Fusing Radiance Fields for Large-Scale Scene Reconstruction. Proc. *CVPR*.

Zhu, Z., Peng, S., Larsson, V., Xu, W., Bao, H., Cui, Z., Oswald, M.R., Pollefeys, M., 2022. NICE-SLAM: Neural Implicit Scalable Encoding for SLAM. Proc. CVPR.

Zhou, K., Meng, X., Cheng, B., 2020. Review of stereo matching algorithms based on deep learning. *Computational intelligence and neuroscience*, Vol.2020.