FEW SHOT PHOTOGRAMETRY: A COMPARISON BETWEEN NERF AND MVS-SFM FOR THE DOCUMENTATION OF CULTURAL HERITAGE

E. Balloni[1], L. Gorgoglione[2], M. Paolanti[3], A. Mancini[1], R. Pierdicca[2]*

[1]Università Politecnica delle Marche, Dipartimento di Ingegneria dell'Informazione (DII), VRAI Lab, 60131 Ancona, Italy.
(e.balloni@pm.univpm.it, a.mancini@staff.univpm.it)
[2]Universitá Politecnica delle Marche, Dipartimento di Ingegneria Civile, Edile e dell'Architettura, 60131 Ancona, Italy.
(r.pierdicca, l.gorgoglione)@staff.univpm.it
[3]University of Macerata, Department of Political Sciences, Communication and International Relations, 62100 Macerata, Italy.
(marina.paolanti@unimc.it)

KEY WORDS: NeRF, Deep Learning, Photogrammetry, Documentation, Cultural Heritage, Artificial Intelligence, 3D Reconstruction.

ABSTRACT:

3D documentation methods for Digital Cultural Heritage (DCH) domain is a field that becomes increasingly interdisciplinary, breaking down boundaries that have long separated experts from different domains. In the past, there has been an ambiguous claim for ownership of skills, methodologies, and expertise in the heritage sciences. This study aims to contribute to the dialogue between these different disciplines by presenting a novel approach for 3D documentation of an ancient statue. The method combines TLS acquisition and MVS pipeline using images from a DJI Mavic 2 drone. Additionally, the study compares the accuracy and final product of the Deep Points (DP) and Neural Radiance Fields (NeRF) methods, using the TLS acquisition as validation ground truth. Firstly, a TLS acquisition was performed on an ancient statue using a Faro Focus 2 scanner. Next, a multi-view stereo (MVS) pipeline was adopted using 2D images captured by a Mini-2 DJI Mavic 2 drone from a distance of approximately 1 meter around the statue. Finally, the same images were used to train and run the NeRF network after being reduced by 90%. The main contribution of this paper is to improve our understanding of this method and compare the accuracy and final product of two different approaches - direct projection (DP) and NeRF - by exploiting a TLS acquisition as the validation ground truth. Results show that the NeRF approach outperforms DP in terms of accuracy and produces a more realistic final product. This paper has important implications for the field of CH preservation, as it offers a new and effective method for generating 3D models of ancient statues. This technology can help to document and preserve important cultural artifacts for future generations, while also providing new insights into the history and culture of different civilizations. Overall, the results of this study demonstrate the potential of combining TLS and NeRF for generating accurate and realistic 3D models of ancient statues.

1. INTRODUCTION

3D documentation methods for Digital Cultural Heritage (DCH) domain are overrunning across different disciplines. The well-known boundaries that for years prevented a righteous dialogue between experts from different domains are going to be culled off. In particular, the heritage sciences witnessed, among the years, to an ambiguous claim for ownership of skills, methodologies, expertise (Remondino and Rizzi, 2010). Artificial Intelligence (AI) helped new generations of researchers to investigate, regardless the compartments, whether a technology can be applied and where it can be useful to solve, with impressive results, dated issues. Embracing the philosophy of the so-called Digital Humanities (DH), the main contribution that AI-driven approaches can give is that of preserving the high-quality standards of data collection, processing and validation, reducing human intervention and thus reducing timing and costly operations (Rowley, 2021). It happened for instance for the semantic segmentation of unstructured point clouds, where Deep Neural Networks (DNNs) are now able to discriminate between architectural elements. And more, DNNs are now helping archaeologists in automatising the vectorization of 2D orthophotos with segmentation tasks. Finally, generative approaches have been able to partially solve the occlusion problem, beside

providing unprecedented opportunities to create data-sets to be shared with the communities (Pierdicca and Paolanti, 2022). By the way, the path towards a full exploitation of AI in the field of DCH is still tortuous and uncertain, requiring huge efforts and, as said before, a sane interdisciplinary collaboration among heterogeneous backgrounds. This is the philosophy underling this research work, where the NeRF (Neural Radiance Field) Networks (Martin-Brualla et al., 2021) are exploited in the CH domain, and compared with Geomatic approaches like MVS-SfM and Terrestrial Laser Scanner. An important aspect of NeRFs is its training and inference speed: the original NeRF took about 1-2 days to train a single scene. Many improvements have been made in this regard, mainly by improving the sampling strategy by reducing MLPs' parameters, resulting in smaller MLPs sizes and, thus, faster training, at the cost of a higher memory consumption. One of the first re-implementation of the original NeRF leaning toward speed was JaxNeRF (Deng et al., 2020), which used Google Jax to create a slightly faster and more suited for distributed computing NeRF implementation. Many works followed, such as Neural Sparse Voxel Fields (NSVF) (Liu et al., 2021), which developed a voxel-based NeRF that models the scene as a set of radiance fields bounded by voxels. This approach was faster than the original implementation but was very memory intensive. A speed up of about ten times from the original NeRF was made

---

* r.pierdicca@staff.univpm.it

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

by (Lindell et al., 2021), which approximated the volume rendering step, allowing it to use much fewer samples, resulting in a much faster network, although with a slight decrease in quality. Another work, Deterministic Integration for Volume Rendering (DIVeR) (Wu et al., 2022), approached the task from a different angle, by reversing the order of volume sampling and MLP evaluation to obtain results that, in terms of quality, outperform methods such as PlenOctrees (Yu et al., 2021), KiloNeRF (Reiser et al., 2021) and FastNeRF (Garbin et al., 2021), while maintaining a comparable speed. Finally, the most recent improvement concerning speed was made in Instant-Neural Graphics Primitives (Instant-NGP) (Müller et al., 2022), in which authors use a multiresolution hash encoding trained to reduce the MLP size; this results in a much faster training and rendering, achieving, in a matter of seconds, the same results of the previous NeRF models. The flexibility and speed of Instant-NGP is the fundamental reason as why it was chosen to perform the experiments, as such a speed improvement can be groundbreaking in the DCH field.

The motivation behind our experiment propitiates from the following research question: can AI replace standards protocols of data acquisition and processing, without reducing the quality? Or better, to what extent? It is well-known in fact that the Digitization of cultural goods (let's for a moment forget the scale of representation, despite fundamental) is nowadays entrusted on TLS - accurate, but time consuming and extremely expensive- and on Multi View Stereo Matching (namely Digital Photogrammetry or DP) – less accurate, but extremely productive and foremost low-cost. The main reason that endeavours archaeologists, historians and experts in general in using such technologies is ancestral: never forget the past. The 3D point cloud is the more accurate replica of a cultural good, the sole method discovered to recreate, virtually, an artefact, making it, de-facto, immortal. Its consequent 3D model is the instrument to analyse it, to mould it, and to instantiate further processing like sectioning, representing and understanding. Finally, cutting edge visualization tool like Virtual Reality make it sharable for the whole mankind. All in all, we are at the stage of striving the discovery of new methods to, as said before, reduce the human intervention. A NeRF is a fully-connected neural network that can generate novel views of complex 3D scenes, based on a partial set of 2D images. It is trained to use a rendering loss to reproduce input views of a scene. It works by taking input images representing a scene and interpolating between them to render one complete scene. NeRF is a highly effective way to generate images for synthetic data. In the literature, the benefit of this new technique in the CH panorama is still partially unexplored, apart from some recent works (Condorelli et al., 2021). First, a TLS acquisition of an ancient statue -with a Faro Focus 2- was performed. Then, the MVS pipeline was adopted, using 2D images from a Mini-2 DJI Mavic 2 shooting pictures all around the statue from about 1 m distance. Finally, the same pictures were used, reduced of 90% to train and run the NeRF network.

Considering the existing literature, the main contributions of this paper is to improve the knowledge, arguing over such method, and by comparing DP and NeRF in terms of accuracy and final product as a whole, exploiting a TLS acquisition as the validation ground truth.

The paper is organized as follows: Section 2 gives details on the state of art on AI techniques applied to DCH domain; Section 3 presents the methods chosen and also explains the rules used for the decision making of this approach. Section 4 presents the results, the performance comparison of the algorithms used and some discussions. Section 5 is devoted to the conclusions and our future works in this direction.

2. STATE-OF-THE-ART

This section briefly reviews some relevant background works concerning AI techniques for the analysis and processing of digital representation of CH. As stated in the introduction, AI algorithms can support and speed up a variety of activities linked to architecture, building or civil engineering. Currently, AI and more in detail its subsets ML and DL are used and applied on areas close to the architectural scale and to CH (Wysocki et al., 2023) even if their use is still limited, since most of CH literature shows a tendency to rely on statistical toolboxes, which are commonly applied as a "black-box" on small CH datasets that are not generally publicly available (Fiorucci et al., 2020). It is also possible to notice that in the last years several initiatives have been made for promoting CH and AI techniques are applied to improve the visiting experience.

Examples of this promotion are the application of AI methods and NLP to support CH institutions and organizations. In (Machidon et al., 2020), an approach based on natural processing language (NLP) has been proposed to retrieve CH resources from Europeana[1]. In particular, the authors designed a solution to improve the accessibility and search accuracy of DCH resources from Europeana through a system that integrates AI, NLP, web services, and APIs. Dou et al. have proposed a knowledge graph for Intangible Cultural Heritage (ICH) which aimed to extract information from ICH text data using NLP techniques to support their organization, management and protection (Dou et al., 2018).

AI have also used to analyze large amounts of historical data and identify patterns that can help preserve and better understand CH. Image processing techniques have been used in the DCH domain for several purposes. For example, (Felicetti et al., 2021), developed Mo.Se. (Mosaic Segmentation), an algorithm that exploits DL and image segmentation techniques to overcome the labour and intensive procedure of extracting information and labelled tesserae from ancient mosaics. The proposed methodology combined U-Net 3 Network with the Watershed algorithm. The approach was tested in the pavement of St. Stephen's Church in Umm ar-Rasas, a Jordan archaeological site, located 30 km southeast of the city of Madaba (Jordan). Hurtut et al. proposed a method for the analysis of the pictorial content of line drawings by the use of the geometrical information of stroke contours. The authors showed that the developed method could be successfully applied for the indexing of line drawings in a retrieval framework (Hurtut et al., 2011).

Focusing more on architectural heritage, it is well established that the safeguarding and maintenance of historic architectural heritage has become crucial in preserving and protecting them from warfare, environmental impacts, calamities, and man-made debacles. The likelihood of these dangers is magnified by the perpetual progression of chemical alteration enacted upon all monuments. Surveying technologies like photogrammetry and laser scanning are currently employed for data collection, establishing them as conventional techniques for three-dimensional documentation of heritage properties (Grilli and Remondino,

---

[1] https://www.europeana.eu/

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

2020). Shalunts et al. designed an approach based on clustering and learning of local features to classify the architectural style of facade windows (Bebis et al., 2011). In this context, the association of semantic information to the point clouds leads to a description of CH, expediting the phase of data interpretation and management. According to (Pierdicca et al., 2020), DL algorithms had great potential to this regard. DL techniques are suitably adopted for directly handling the raw data of point clouds without an intermediate processing that allows a more regular representation. An example is the work of Malinverni et al. (Malinverni et al., 2019) that exploited PointNet++ (Qi et al., 2017) to semantically segment 3D point clouds of CH dataset. A newly dataset was specifically collected to deal with CH data and manually labelled by domain experts: ArCH dataset (Matrone et al., 2020). The specific goal of this paper was to demonstrate the effectiveness of a DL framework specialized in point clouds semantic segmentation to tackle with CH-related point clouds. Inspired by the great results obtained in (Wang et al., 2019), which introduced a module called EdgeConv, that constructs a local neighborhood graph and applies convolution-like operations and developed a new DL model named DGCNN (Dynamic Graph Convolutional Neural Network), dynamically updates the graph, changing the set of k-nearest neighbors of a point from layer to layer of the network, the same authors made an extension of the previous work and exploited the novelties offered by the DGCNN. Thus in (Pierdicca et al., 2020) they proposed a modified version of DGCNN by adding relevant features such as normal and HSV encoded color. This improved version aimed at facilitating the management of DCH assets that have complex geometries, extremely variable and defined with a high level of detail.

More recently, an ambitious task of applying AI techniques in DCH domain has been tackled: 3D reconstruction starting from images. These aspect involves several challenges such as the identification of the starting images from the vast archive material and the ability of photogrammetric algorithms to work with numerically reduced and low-quality images. Some approaches, such as (Vicini et al., 2022), include the usage of Signed Distance Functions to represent 3D objects. More recently, the Neural Radiance Field approach emerged; Neural Radiance Fields have proven to be capable of achieving photo-realistic results in complex scenes and, thus, gained a lot of research interest (Mildenhall et al., 2021). In the span of 2 years, more than 200 preprints on arXiv have been registered (Gao et al., 2022), with the aim of improving the original architecture in different aspects. Some significant works include Mip-NeRF (Barron et al., 2021), which used cone tracing instead of the ray tracing of the standard NeRF; Ref-NeRF (Verbin et al., 2021), which was built starting from Mip-NeRF to model reflective surfaces more accurately.

This approach has already been adopted in (Condorelli et al., 2021). The authors assesed the experiments on two different dataset specifically collected by them: flower and tower dataset. In particular, the last one includes images which were acquired by the same authors during a survey on the place. They compared the results obtained by the application of NeRF (Mildenhall et al., 2021) with the traditional photogrammetry pipelines based on Structure-from-Motion (SfM) and Multi-View-Stereo (MVS) open-source algorithms (Schonberger and Frahm, 2016a).

With respect to the above mentioned state-of-the-art works, our approach consists in creating a novel pipeline that leverages the NeRF architecture and, in particular, Instant-NGP for the cre-

ation of 3D scene representations and subsequent mesh generation.

3. MATERIALS AND METHODS

The comparison between NeRF and MVS-SfM reconstruction methods was performed by using a specific case study to conduct our tests. A statue was chosen for the experiments, namely the Alberico Gentili statue, located in San Ginesio (Macerata, Italy). From this monument, 2 datasets were acquired, the first one composed of images taken from UAV, the second composed of spherical panoramas taken from TLS. As the dataset used for the NeRF pipeline (UAV dataset) was not intended to be used with NeRF initially, this work can be also useful to prove how a dataset not originally acquired to be used with a NeRF approach can be suitable anyway. A detailed description of the created datasets follows.

3.1 Dataset

In the presented case study, data acquisition was carried out using two different techniques: photogrammetry and laser scanning. For photogrammetry, a Mini-2 DJI Mavic 2 drone was used as the UAV system. The images were then processed using SfM-MVS techniques, which estimated the 3D structure from 2D image sequences. For laser scanning, a terrestrial laser scanning system (TLS), specifically the Faro Focus 2, was used to acquire four spherical panoramas. Each panorama had a high resolution of 10240x5120 pixels, enabling the generation of highly accurate and detailed point clouds. The use of both techniques allowed for a more comprehensive analysis of the site and an assessment of the accuracy and reliability of the data obtained using each method, resulting in a more complete and reliable dataset for research. Table 1 lists the main data collected and the methodology used, including data from the NeRf methodology, as the starting dataset was the same. Table 1 shows the different datasets and acquisition methods. In particular, the UAV dataset that was used in the experiments is comprised of 224 images of the Alberico Gentili Statue, taken with a Mini-2 DJI Mavic 2, with a size of 4000×3000 pixels in JPG format. The TLS dataset is comprised of 4 spherical panoramas, taken with a Faro Focus 2, with an image size of 10420x5120. In Figure 1 some examples extracted from the UAV dataset are shown (after the relative orientation, thus providing the sparse point cloud).

Concerning the NeRF experiments, the UAV dataset has been used, as images were found to be suitable for NeRF scene representation; to prove the effectiveness of the chosen network, the UAV dataset has been reduced to 38 images (about 15% of the total), with the aim of achieving comparable results.

3.2 Photogrammetric pipeline

The documentation and modeling of cultural objects and sites is often carried out using two different methods: photogrammetry and laser scanning. These methods differ significantly in the way data is captured. Photogrammetry uses Structure from Motion (SfM) technology, which involves acquiring images from different positions and angles to reconstruct a three-dimensional model of the scene or object. Laser scanning, on the other hand, directly acquires data from three-dimensional point clouds using a laser, without using SfM technology. Both methods are suitable for different purposes and situations. Photogrammetry is particularly useful when covering a large area

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

Acquisition system	Sensor	Dataset		Methodology
		Image number	Image size	
UAV	Mini-2 DJI Mavic 2	228	4000x3000	SfM
TLS	Faro Focus 2	4 spherical panoramas	10240x5120	Laser scan
UAV	Mini-2 DJI Mavic 2	38	2400x1800	NeRF

Table 1. Description of dataset acquisition and methodology used
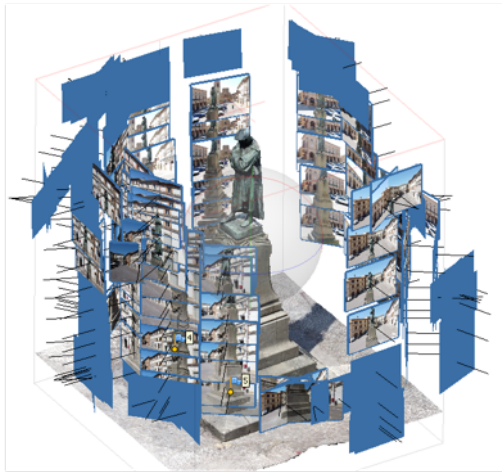
Figure 1. Dataset images and SfM image orientation

or when accessing difficult-to-reach objects since images can be taken from different angles and positions. Image-based 3D reconstruction techniques are considered cost-effective and efficient for producing high-quality 3D digital models of real-world objects in terms of hardware requirements, knowledge background, and man-hours. Typically, at least two images with common features are required, and 3D data accompanied by texture information can be derived through perspective or projective geometry formulations. These methods, mainly computer vision (Remondino and El-Hakim, 2006), are generally preferred for lost objects, monuments, or architectures with regular or complex geometric shapes, small objects with free-form shape, mapping applications, deformation analysis, and time or location constraints for data acquisition and processing. Laser scanning, on the other hand, is particularly useful when obtaining a large amount of three-dimensional data in a very precise and detailed way. In the context of cultural heritage, generating complete digital representations of the scene captured, often in the form of 3D surfaces, requires great attention to the quality of the surface models or meshes. These models must be highly detailed and sufficiently accurate, especially for metric applications. Surface generation can be seen as an integrated problem in the complete 3D reconstruction pipeline, and thus visibility information (pixel similarity and image orientation) is exploited in the meshing procedure, contributing to an optimal photo-consistent mesh.

3.3 NeRF workflow

A much faster version of the original NeRF network has been used, namely Instant Neural Graphics Primitives with a Multiresolution Hash Encoding (Instant-NGP[2]) (Müller et al., 2022); this network allows for accurate training within a short period of time, besides enriching the bulk of knowledge in the field (Mildenhall et al., 2021). In particular, the choice of Instant-NGP was made due to its training and inference speed, which is

an key factor in the CH field, given that, in a classic MVS-SfM approach, this can be a really time-consuming task. Also, NeRF allows for the generation of novel views, so a dataset with a low example number can be suitable.

An Anaconda[3] environment running Python 3.9.15 has been created, installing all the necessary packages to run the network. CUDA Toolkit[4] version 11.8 has been used, as Instant-NGP leverages the tiny-cuda-nn framework[5] for the multiresolution hash input encoding.

In order to prepare the dataset for training, camera parameters must be extracted from the images, as both camera positions and images are needed for the network to create the scene representation. To achieve this goal, the COLMAP[6] software has been used, as it implements SfM techniques (Schönberger and Frahm, 2016b) for estimating three-dimensional structures from two-dimensional image sequences. For usage with COLMAP, dataset images have been resized to 2400 width and 1800 height, as it has proved to speed up the generation of the camera parameters while maintaining high image resolution. Both the camera positions and the image set are then fed to the network to start the training process. Then, the 5D output for each image (x, y, z spatial location coordinates and θ, ϕ viewing direction) is given as input to the network, which outputs view-dependent emitted radiance (RGB) and a volume density σ. Finally, through classic volume rendering techniques, the output is projected on an image, which enable the computation of a loss to train effectively. The result is the 3D scene representation. After the scene is trained, a mesh can be extracted by using the marching cubes algorithm (Lorensen and Cline, 1987).

4. RESULTS AND DISCUSSION

Different tests have been performed, changing the number of training steps to better assess the goodness of the network. Figure 3 shows the different results at different training steps (1.000, 2.000, 5.000, 20.000, 35.000). Table 2 shows the various training times. Due to the speed of convergence of the network, 35.000 steps proved to be more than enough to reach a good result (training with more steps resulted in very similar outputs). The network was trained on a PC with a NVIDIA RTX 2070 Super GPU with 8 GB of VRAM, a Ryzen 5 3600 CPU and 16 GB of RAM.

By increasing the number of steps, the network is capable of representing the monument with good quality and accuracy. These results show the suitability of the application of this method for a faster and accurate training or a longer and extremely accurate training. A demo video showcasing the results obtained
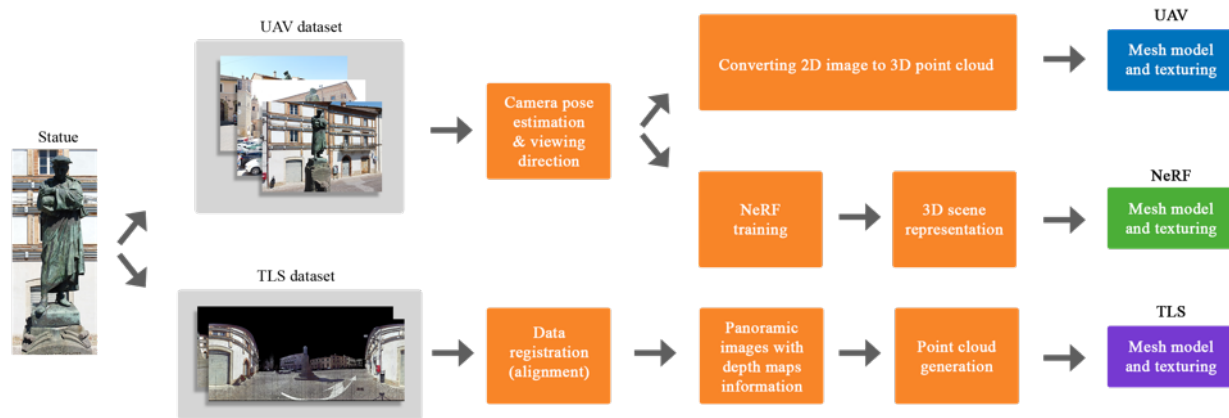
---

[2] https://github.com/NVlabs/instant-ngp

[3] https://anaconda.org

[4] https://developer.nvidia.com/cuda-toolkit

[5] https://github.com/NVlabs/tiny-cuda-nn

[6] https://github.com/colmap/colmap

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

Figure 2. Workflow of the three methodologies: UAV, NeRF and TLS

(a) 1.000 training steps   (b) 2.000 training steps   (c) 5.000 training steps   (d) 20.000 training steps   (e) 35.000 training steps
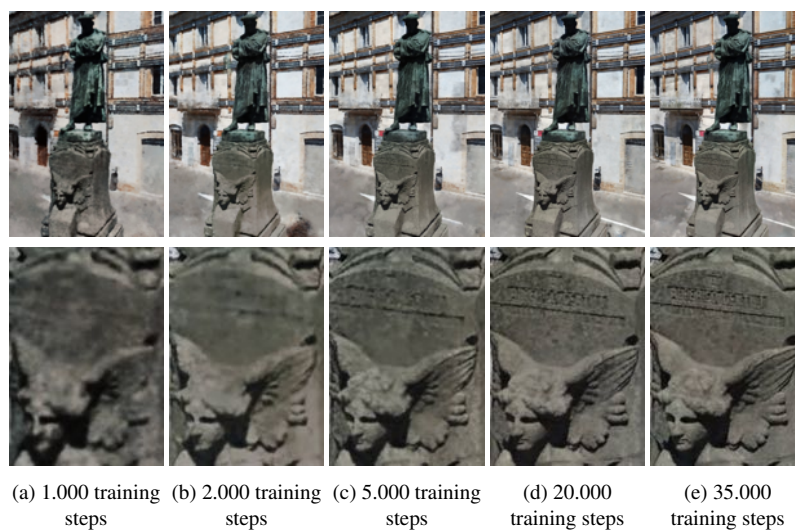
Figure 3. Qualitative results of training at different steps

Steps number	Training time
1.000	17 seconds
2.000	26 seconds
5.000	56 seconds
20.000	3 minutes 45 seconds
35.000	7 minutes 15 seconds

Table 2. Number of steps for training and their corresponding training time

from the 35.000 steps training[7] has also been created to better exhibit the effectiveness of the network used.

Starting from the 3D scene representation, a polygon mesh of the monument has been generated, by cropping the relevant portion. This can be useful as the mesh can be imported in 3D modeling tools. Figure 4 *left* shows some qualitative results. At the same time, the photogrammetric model was created, in order to analyse and understand in detail the pros and cons of each method (Figure 4 *right*).

Given the outcomes of our computation, there is the need to argue about the pros and cons of the proposed method. First of all, when dealing with cultural heritage documentation, the quality
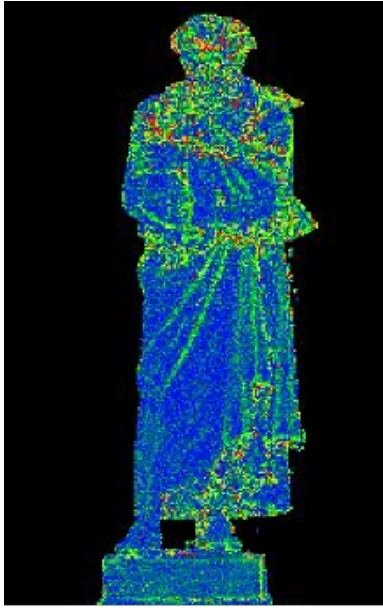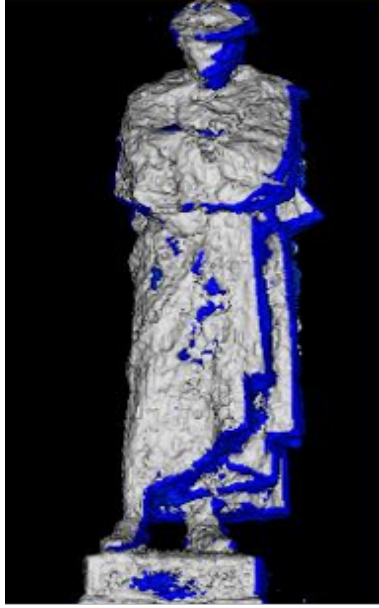

Figure 4. Resulting NeRF mesh

of the representation is strictly dependent on the motivation of the work. In other words, as dictated by the EU commission guidelines[8], the resolution, accuracy and detail of a digital replica needs to follow strict rules according to the output of the project. Given this aspect, it is clear that the well-established
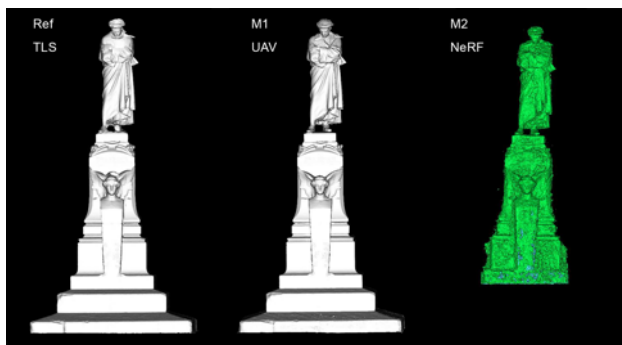
---

[7] https://www.youtube.com/watch?v=wOiX6uEur0Q

[8] https://digital-strategy.ec.europa.eu/en/library/study-quality-3d-digitisation-tangible-cultural-heritage

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

Method	Accuracy (RMS)	C2M (Cloud to Mesh - Signed distances)	Roughness
TLS- UAV	0.0310413	<0.15	0.003923
TLS- NeRF	0.0197409	<1.25	0.005446

Table 3. Evaluation metrics for quantitative analyses. Values are in mm.

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

Figure 5. Shaded surface models of evaluated datasets. From left to right: Reference - TLS , M1 - UAV, M2 - NeRF.

methods based on TLS and SfM are still outperforming the AI-Based and Generative ones. This is clearly demonstrated in Table 3, considering the roughness value achieved. Surface reconstruction methods in photogrammetric applications were evaluated using the work of (Nocerino et al., 2020) as a reference. The results of the approaches considered were evaluated using various metrics, including accuracy and roughness. In contrast, calculation time was not considered a key factor in this investigation (Table 3). Besides, performing C2M comparison, the accuracy values seems to be comparable between SfM and NeRF method. This is misleading, since the relative oreintation between the two method is performed in the same way, thus conducting to similar results; however, the main issue emerges with the creation of the dense cloud, since the NeRf method does not produce the dense cloud that is a by-product fo the algorithms. Further investigations are needed in order to extract a comparable point cloud from the NeRF method. However, the insiders knows that during the acquisition campaigns there are several problems (expeically in the architectural and archaeological fields) that hamper the surveyors to collect complete dataset. Missing images, occlusions, weather conditions and so on are often "enemies" of good quality models; in this case, NeRF method can be a winning solution to achieve a result that, despite not accurate and noisy, can be used for further processing. Finally, it is fair to say that the Photogrammetric pipeline is very time consuming for both acquisition and processing time, whilst the TLS is stil la very costy instrument; with the advent of NeRF method, in case the survey campaign has the above mentioned limitations, can be a valuable alternative.

5. CONCLUSION AND FUTURE WORKS

In conclusion, this study presents a novel approach for 3D documentation of an ancient statue, which combines TLS acquisition and MVS pipeline using images from a Mini-2 DJI Mavic 2 drone. The results of this study demonstrate the potential of using these methods for accurate and detailed 3D documentation of CH objects. Additionally, the comparison between DP and NeRF methods highlights the importance of carefully choosing the appropriate approach based on the specific object and research question. Finally, the interdisciplinary nature of this study underscores the need for continued collaboration and communication between experts from different domains to advance the field of 3D documentation for DCH. This research contributes to ongoing efforts to improve the accuracy and efficiency of 3D documentation methods, which has important implications for CH preservation and research.

Some further works could involve the usage of different algorithms for mesh extraction in the NeRF pipeline, such as Deep Marching Tetrahedra (Shen et al., 2021), that could improve mesh generation. Other additions for NeRFs could see the usage of RGB-D images as input, as providing depth information in input can result in reduced artifacts and other minor issues currently present in the generated meshes.

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.

Bebis, G., Boyle, R., Parvin, B., Koracin, D., Wang, S., Kyungnam, K., Benes, B., Moreland, K., Borst, C., DiVerdi, S. et al., 2011. *Advances in Visual Computing: 7th International Symposium, ISVC 2011, Las Vegas, NV, USA, September 26-28, 2011. Proceedings, Part I.* 6938, Springer.

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. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, 565–570.

Deng, B., Barron, J. T., Srinivasan, P. P., 2020. JaxNeRF: an efficient JAX implementation of NeRF.

Dou, J., Qin, J., Jin, Z., Li, Z., 2018. Knowledge graph based on domain ontology and natural language processing technology for Chinese intangible cultural heritage. *Journal of Visual Languages & Computing*, 48, 19–28.

Felicetti, A., Paolanti, M., Zingaretti, P., Pierdicca, R., Malinverni, E. S., 2021. Mo. Se.: Mosaic image segmentation based on deep cascading learning. *Virtual Archaeology Review*, 12(24), 25–38.

Fiorucci, M., Khoroshiltseva, M., Pontil, M., Traviglia, A., Del Bue, A., James, S., 2020. Machine learning for cultural heritage: A survey. *Pattern Recognition Letters*, 133, 102–108.

Gao, K., Gao, Y., He, H., Lu, D., Xu, L., Li, J., 2022. Nerf: Neural radiance field in 3d vision, a comprehensive review.

Garbin, S. J., Kowalski, M., Johnson, M., Shotton, J., Valentin, J., 2021. Fastnerf: High-fidelity neural rendering at 200fps.

Grilli, E., Remondino, F., 2020. Machine learning generalisation across different 3D architectural heritage. *ISPRS International Journal of Geo-Information*, 9(6), 379.

Hurtut, T., Gousseau, Y., Cheriet, F., Schmitt, F., 2011. Artistic line-drawings retrieval based on the pictorial content. *Journal on Computing and Cultural Heritage (JOCCH)*, 4(1), 1–23.

Lindell, D. B., Martel, J. N. P., Wetzstein, G., 2021. Autoint: Automatic integration for fast neural volume rendering.

Liu, L., Gu, J., Lin, K. Z., Chua, T.-S., Theobalt, C., 2021. Neural sparse voxel fields.

Lorensen, W. E., Cline, H. E., 1987. Marching Cubes: A High Resolution 3D Surface Construction Algorithm. *SIGGRAPH Comput. Graph.*, 21(4), 163–169. https://doi.org/10.1145/37402.37422.

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

Machidon, O.-M., Tavčar, A., Gams, M., Duguleană, M., 2020. CulturalERICA: A conversational agent improving the exploration of European cultural heritage. *Journal of Cultural Heritage*, 41, 152–165.

Malinverni, E. S., Pierdicca, R., Paolanti, M., Martini, M., Morbidoni, C., Matrone, F., Lingua, A., 2019. Deep learning for semantic segmentation of 3D point cloud. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*.

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. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7210–7219.

Matrone, F., Lingua, A., Pierdicca, R., Malinverni, E., Paolanti, M., Grilli, E., Remondino, F., Murtiyoso, A., Landes, T., 2020. A benchmark for large-scale heritage point cloud semantic segmentation. *XXIV ISPRS Congress*, 43, 1419–1426.

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., Ng, R., 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1), 99–106.

Müller, T., Evans, A., Schied, C., Keller, A., 2022. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. *ACM Trans. Graph.*, 41(4), 102:1–102:15. https://doi.org/10.1145/3528223.3530127.

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

Pierdicca, R., Paolanti, M., 2022. GeoAI: a review of artificial intelligence approaches for the interpretation of complex geomatics data. *Geoscientific Instrumentation, Methods and Data Systems*, 11(1), 195–218.

Pierdicca, R., Paolanti, M., Matrone, F., Martini, M., Morbidoni, C., Malinverni, E. S., Frontoni, E., Lingua, A. M., 2020. Point cloud semantic segmentation using a deep learning framework for cultural heritage. *Remote Sensing*, 12(6), 1005.

Qi, C. R., Yi, L., Su, H., Guibas, L. J., 2017. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30.

Reiser, C., Peng, S., Liao, Y., Geiger, A., 2021. Kilonerf: Speeding up neural radiance fields with thousands of tiny mlps.

Remondino, F., El-Hakim, S., 2006. Image-based 3D modelling: a review. *The photogrammetric record*, 21(115), 269–291.

Remondino, F., Rizzi, A., 2010. Reality-based 3D documentation of natural and cultural heritage sites—techniques, problems, and examples. *Applied Geomatics*, 2(3), 85–100.

Rowley, R., 2021. An evaluation of Image-Based Modelling for metrically recording cultural heritage subjects suitably to enable further use in geomatics, geoinformatics, and digital humanities. PhD thesis, Bournemouth University.

Schonberger, J. L., Frahm, J.-M., 2016a. Structure-from-motion revisited. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4104–4113.

Schönberger, J. L., Frahm, J.-M., 2016b. Structure-from-motion revisited. *Conference on Computer Vision and Pattern Recognition (CVPR)*.

Shen, T., Gao, J., Yin, K., Liu, M.-Y., Fidler, S., 2021. Deep marching tetrahedra: a hybrid representation for high-resolution 3d shape synthesis. *Advances in Neural Information Processing Systems (NeurIPS)*.

Verbin, D., Hedman, P., Mildenhall, B., Zickler, T., Barron, J. T., Srinivasan, P. P., 2021. Ref-nerf: Structured view-dependent appearance for neural radiance fields.

Vicini, D., Speierer, S., Jakob, W., 2022. Differentiable signed distance function rendering. *ACM Transactions on Graphics (TOG)*, 41(4), 1–18.

Wang, Y., Sun, Y., Liu, Z., Sarma, S. E., Bronstein, M. M., Solomon, J. M., 2019. Dynamic graph cnn for learning on point clouds. *Acm Transactions On Graphics (tog)*, 38(5), 1–12.

Wu, L., Lee, J. Y., Bhattad, A., Wang, Y., Forsyth, D., 2022. Diver: Real-time and accurate neural radiance fields with deterministic integration for volume rendering.

Wysocki, O., Grilli, E., Hoegner, L., Stilla, U., 2023. Combining visibility analysis and deep learning for refinement of semantic 3D building models by conflict classification. *arXiv preprint arXiv:2303.05998*.

Yu, A., Li, R., Tancik, M., Li, H., Ng, R., Kanazawa, A., 2021. Plenoctrees for real-time rendering of neural radiance fields.