2023 IMEKO TC-4 International Conference on
Metrology for Archaeology and Cultural Heritage
Rome, Italy, October 19-21, 2023

Ponte San Lorenzo, a case study for the comparison of image-based survey tools. NeRF as an alternative to photogrammetry

Maurizio Perticarini[1], Andrea Giordano[2]

[1] *Università degli Studi di Padova, Via Marzolo, 9 - 35131 Padova, andrea.giordano@unipd.it*
[2] *Università degli Studi di Padova, Via Marzolo, 9 - 35131 Padova, maurizio.perticarini@unipd.it*

Abstract – **The aim of the research is to experiment with the new possibilities offered by artificial intelligence applied to surveying. Ponte San Lorenzo, a perfectly preserved archaeological artefact under Via Riviera dei Ponti Romani in Padua, one of the case studies of the 2018 project of the University of Padua entitled: "PD-Invisible: PaDova INnovative VISions - visualizations and Imaginings Behind the city Learning", was the object of an accurate integrated image and range-based survey and made it possible to compare the different 3D acquisition methodologies of the cultural heritage. The Bridge is now digitally rendered through the new system developed by Nvidia called Instant Nerf and the new web-based software Luma AI which exploit machine learning and Volume Ray Marching for the creation of virtual twins. The same database of photographs taken for photogrammetry was partially used as a training database for the algorithm. The research aims to evaluate the results of the new survey system and to compare them with the results of the methods used in the past.**

Keywords: Nerf; Cultural Heritage; Photogrammetry; Artificial Intelligence; Machine Learning

I. INTRODUCTION

The PD-Invisible: PaDova INnovative VISions - visualizations and Imaginings Behind the city Learning project was an initiative of the University of Padua aimed at revealing and enhancing the remarkable Roman archaeological heritage present in the city of Padua, which has remained entirely or partially hidden due to changes in the urban environment over the centuries. The main objective is to highlight the numerous manufacts from the Roman age present in the urban context, with particular attention to the Roman bridges located below street level in the Riviera Tito Livio and Riviera dei Ponti Romani: the Ponte Altinate (at the crossroads Via Altinate) and the Ponte San Lorenzo (at the crossroads with Via San Francesco) [1].

The project is based on rigorous archival research in order to acquire all the information necessary for the subsequent virtual reconstruction of the missing parts of the artefacts. The integrated survey based on both a range-based technique such as laser scanning and an image-based technique such as photogrammetry served to acquire the greatest amount of information for the virtual reconstruction of the artefacts. Particular attention has been given to Ponte San Lorenzo, which is perfectly preserved and easily accessible. The two surveys, which both provided excellent results, were analysed and compared, both to carry out the scaling of the photogrammetric survey and to identify the strengths and weaknesses of the two different acquisition techniques.

This article describes a new data acquisition method based on artificial intelligence: the NeRF (Neural Radiance Fields) algorithm, an innovative method for the reconstruction of three-dimensional scenes and the generation of photorealistic images. The same dataset of images previously used for photogrammetry was subjected to this new system and the results obtained were analysed and compared. The NeRF represents a significant step forward in image synthesis and 3D modelling based on neural networks, it was introduced in 2020 by Ben Mildenhall together with a group of researchers at Google Research and is based on the idea of modelling the scene as a function of density of radiance, which represents the amount of light emitted at each point in space. Using a neural network, the NeRF learns this function and can then generate new views of the scene or place virtual objects within it. The training process requires a dataset consisting of several views of a scene from different angles. For each view, information about the colour and direction of light is captured. A significant advantage of NeRF is its ability to handle the representation of translucent or reflective objects. Because NeRF models light at every point in space, it can capture the effects of translucency and reflection, making reconstructions more realistic than traditional methods. Another strong point are the calculation times of the rendering phase, which are reduced to fractions of a second, as the rendering method used by the algorithm is no longer based on raytracing, but on the ray marching volume, which with the new hardware developed by Nvidia (such as RTX GPUs) is much more manageable than in the past. A platform already available on the web that uses this type of algorithm is Luma AI, which is easy to understand and use. It allows both the training phase and the rendering phase with the help of an external server and has quite reliable output functions. With the increasingly frequent implementation of AI within software for the management and disclosure of complex three-dimensional objects, this acquisition method is very valid and could easily replace the most common systems based on photogrammetry.

II. THE CASE STUDY AND THE INTEGRATED SURVEY

Ponte San Lorenzo is situated beneath street level, at the junction of Via San Francesco and Riviera Tito Livio. It can be accessed through an underpass, previously open on both sides of the Riviera but now accessible only from the eastern side (Piazza della Tomba di Antenore). The bridge boasts three remarkably preserved arches, derived from polycentric arches with varying spans and curvature. Apart from the arches, other notable features visible include the armillae on the south-facing side, the abutments at the tip and tail ends, the piles with rostra (visible only on the southern side), and a single armilla on the north-facing side. The comprehensive survey of the bridge was conducted utilizing laser scanning and photogrammetry techniques. The laser scanning survey, which did not require uniform lighting, was carried out without any preparatory stage of the environment. It served as a supporting element for the photogrammetric survey, providing precise spatial coordinates and ensuring accurate sizing based on image-based surveying. For photogrammetry, a lighting study phase was conducted to ensure even lighting across the entire surface being captured. Given the limited dimensions and inadequate lighting conditions, spotlights were positioned in a schematic manner, directed towards the bridge's surfaces characterized by sharp edges and deep undercuts. The spotlights were placed at the centre of each arch, perpendicular to the vault portions that needed illumination, thereby creating a bright and uniform environment. Photographs were captured following a regular grid pattern to ensure proper framing of each individual point on the arches. A reflex camera was used without a tripod, while one of the mobile spotlights was directed to ensure lighting consistency. This approach allowed for the creation of a point cloud with a high density of points, enabling the detection of even the most intricate components such as undercuts. The described method was employed for both the survey of the better-preserved southern elevation and the single armilla on the northern elevation. However, certain challenges were encountered in the less accessible areas. For instance, the intact elevation of the bridge was partially obstructed by architectural elements such as piers and abutments, which were shielded by slender protective railings. Additionally, the central and westbound armillae were located in close proximity to the supporting beams and pillars of the road above. Special efforts were made to capture the central armilla, which held particular significance due to the Roman inscription engraved at the top of the arch, despite the limited space for adequate detection. Upon completion of the photography phase, all the collected data were processed using Agisoft Metashape software, specifically designed for generating point clouds from photographs (Fig. 1).

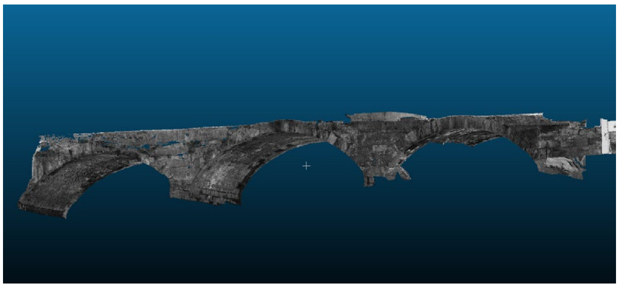

Fig. 1. Point cloud of Ponte San Lorenzo obtained with Metashape Software

III. THE NEURAL RADIENCE FIELD

A few sets of photographs used for photogrammetry formed the data set for the new Nerf-based system.[2]

The Neural Radience Field (NeRF) is a neural network designed to generate novel perspectives of intricate three-dimensional scenes using a partial set of 2D images. By interpolating the input images available in the dataset, the network creates a 3D scene, addressing the problem known as Novel View Synthesis. It synthesizes an image from a specific viewpoint (camera target) by leveraging a series of source images, each captured from its own unique perspective. To optimize rendering times, Nerf employs several optimization techniques, including Multi-layer Perceptron (MLP) to avoid redundant density calculations for specific points, positional coding, and hierarchical sampling. Early NeRF models were capable of rendering detailed scenes without artifacts within minutes, but the training process still required hours. However, Instant NGP (Instant neural graphics primitives) Instant NeRF, developed by Nvidia, significantly reduces rendering time by implementing a technique called "multi-resolution hash grid coding." This optimization is designed to efficiently run on the independent system of NVIDIA RTX GPUs, eliminating the need for complex machine learning operations on external servers. Instant NeRF also introduces a faster input coding method, utilizing a single continuous 5D coordinate that encompasses spatial position (x, y, z) and observation direction (θ, φ). The output of this coding method includes volume density and spatial position-dependent emitted radiance (RGBσ). Understanding the importance of the viewpoint direction in an image is crucial. In the case of a completely

Lambertian surface (an ideal "opaque" surface like terracotta), knowing the viewpoint is not essential since objects with similar characteristics exhibit nearly the same colour, regardless of the viewpoint. However, when dealing with reflective surfaces (e.g., perfect mirrors), each ray of light is reflected in a single direction. Therefore, the formula involving volume density and spatial position-dependent emitted radiance becomes necessary. Volume rendering enables the creation of a 2D projection from a 3D dataset. Each camera position corresponds to an RGB value and density "σ" for every voxel the camera beam passes through. The reverse process involves generating a 3D object from multiple 2D images captured from different angles and perspectives. The system predicts the depth and density of the objects based on this data. The Coldmap model, which combines 3D photogrammetry algorithms with artificial intelligence, organizes the information from each photographic shot, generating a volumetric cloud that accurately represents the complete object. By synthesizing views and analysing the 5D coordinates along camera beams, traditional volumetric rendering techniques can be applied to project colors and output densities onto an explorable model. Volumetric rendering considers not only specific object data but also the importance of acquisition rays that penetrate inside the objects. By operating at a volumetric level and utilizing multiple simultaneous shots, this approach eliminates the need for global illumination calculations, where secondary rays simulate ambient lighting and create a realistic spatial configuration [3][4][5][6][7][8][9].
To conclude, the latest RTX graphics cards elevate the concluding phases of volumetric rendering while providing assistance during the initial stage of machine learning for acquiring intricate visual information through an "image-based" approach. These advancements facilitate the concurrent production of renderings for a three-dimensional environment (Fig. 2).

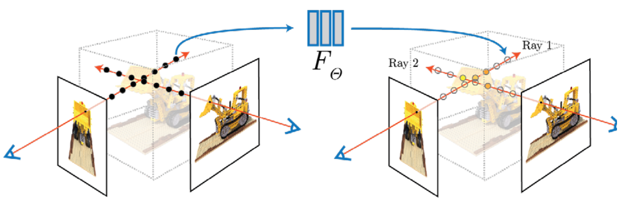

Fig. 2. Working diagram of the Volume Ray Marching. Rendering system at the base of the Nerf. Mildenhall et al 2020 (arXiv:2003.08934)

IV. RESULTS WITH INSTANT NGP

NERF works differently from photogrammetry, so the entire database of photographs used for the photogrammetric survey has not been used. Using the entire database of shots would have led to an inaccurate result and above all to very long training times. In fact, Instant NGP works well with a small number of photographs. The best results obtained were found in the elaboration of the rostrum and the central armilla characterized by the epigraph (Fig. 3, 4). From the software you can extract the mesh together with the vertex color, vertex normals and face ids attributes. As can be seen from fig. 5 the mesh has no texture but the vertex color parameter can be easily converted to a raster texture.

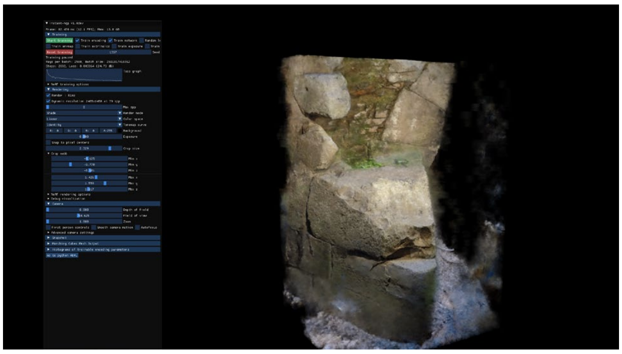

Fig. 3. Rostrum of the San Lorenzo Bridge rendered in Instant NGP

Fig. 4. Central armilla with epigraph rendered in Instant NGP

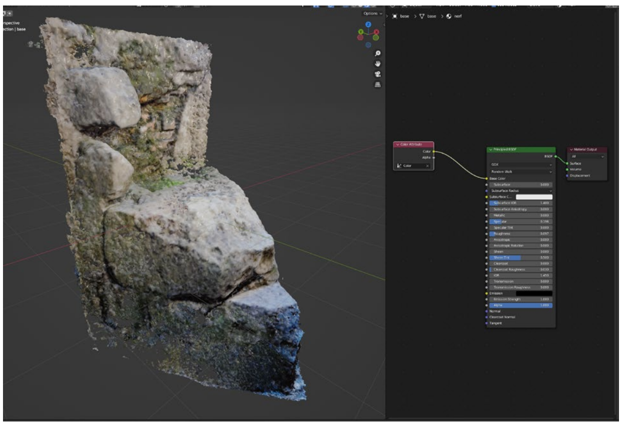

Fig. 5. The mesh obtained from the export from Instant NGP and the shader created in Blender 3D by means of vertex color

V. LUMA AI

Recently, Nerf systems are becoming increasingly popular, and the use of such systems in image set training methodologies to generate metrically accurate and photorealistic three-dimensional scenery in material rendering is growing. The most promising results were obtained by Nvidia with Instant NGP and by Berkeley students in the KAIR lab at Berkeley AI Research (BAIR) in October 2022 with NERF Studio. Now LumaAI, developed by a group of independent researchers in collaboration with the Berkeley Artificial Intelligence Research Lab and funded from NVIDIA's NVentures, General Catalyst, Matrix Partners, South Park Commons, and RFC's Andreas Klinger, constitutes a valid platform for experimenting with the potential of this new method.

Luma Ai was used for the same case study involving the digitization of an archaeological artifact. In addition to having developed an APP for iOS devices that allows to carry out the survey using the smartphone camera, it offers a web platform where you can upload photographs or a video of the surveyed object to subsequently generate a representation of it in 3D.

VI. RESULTS WITH LUMA AI

Also, in this case not all the photographs from the photogrammetric database were used. In fact, the LUMA AI platform has an upload limit of 5 gb and obviously, as in the case of Instant NGP, the use of a large database would have generated a confused model and greatly increased training times. It was therefore decided to use photographs taken at a not too close distance of the entire southern elevation of the bridge. The shots were previously optimized in photoshop, cutting out only the parts relating to the bridge. In this way we obtained (albeit partially) a result free of elements that could disturb the training phase. After about fifteen minutes, necessary for the NeRF training process via server, the software provided access to the integrated viewer which shows a preview of the processed 3D object. As in the smartphone app, the web viewer shows the object in three ways: a video in which a virtual camera moves smoothly between the points of view of the shots taken, a 3D viewer of the object and a viewer of the whole scene 3D. Once the virtual twin has been created, it is possible to access the output which allows to obtain the model in the formats .gtlf, .usdz, .obj and in the .ply format for the entire scene. Luma Ai also provides API for scene and model export directly compatible with Unreal Engine 5.1 and 5.2. Through a plugin it is easy to import the file in a special ".luma" format into the software. This allows to have the entire 3D setting and the object on which the survey is focused within the Unreal content browser. The model consists of a complex mesh which, as in Instant NGP, is equipped with the vertex color parameter. Inside Unreal there is a panel relating to the management of ".luma" files which allows cropping of the area of interest with respect to the entire environment. Furthermore, thanks to the new Nanite algorithm present from Unreal version 5.0 onwards, this type of assets, made up of many polygons, can be used - especially in applied games - without carrying out preliminary optimization operations. (Fig. 6) [10][11][12][13][14].

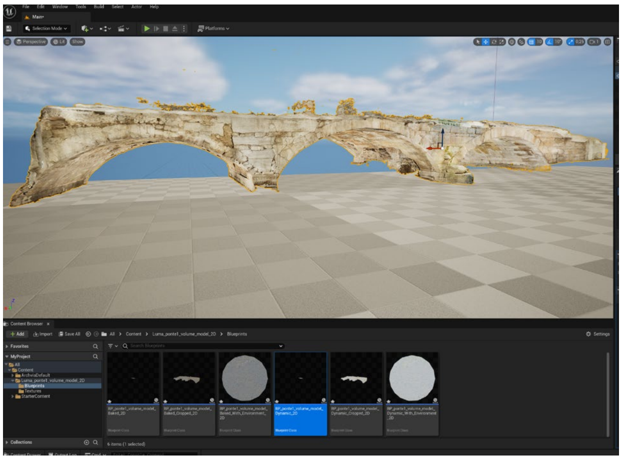

Fig. 5. Importing the ".luma" file into Unreal Engine 5

VII. CONCLUSIONS

In conclusion, the use of machine learning represents a valid contemporary alternative to traditional photogrammetry methods in the context of data acquisition for cultural heritage. This approach allows efficient results but the survey approach must necessarily be different from the photogrammetric one. The study highlighted how it is not possible to use the same number of shots used in the photogrammetric survey, as training times are considerably extended; above all the shots must respect a precise order and the photographs taken at a medium distance and not too close are much more effective. The acquisition technique is moving towards video acquisition and no longer based on frames, precisely because Nerf is facilitated by the continuity of the shot.

The meshes obtained are difficult to manage from the point of view of modification and optimization, but the system is designed to intervene as little as possible on the models obtained. In fact, the models have the features of real ready-to-use assets - today with their defects and imperfections - which are gradually improving with the algorithm updates.

The Nerf therefore, although in an embryonic phase, would seem to offer a wide range of users the possibility of creating high-quality interactive and engaging virtual experiences more and more rapidly. It offers promising opportunities for applications in fields such as immersive experiences, applied games for the dissemination of cultural heritage, in augmented reality or virtual simulation applications, opening new perspectives for the creation of immersive and realistic digital content.

REFERENCES

[1] Carraro F., Monego M., Callegaro C., Mazzariol A., Perticarini M., Menin A., Achilli V., Bonetto J., Giordano A. (2019). The 3d Survey of the Roman Bridge of San Lorenzo in Padova (Italy): A Comparison between SFM and TLS Methodologies Applied to the Arch Structure. International Archives of The Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 42, p. 255-262, ISSN: 1682-1750, doi: 10.5194/isprs-archives-XLII-2-W15-255-2019

Palestini C, Basso A, Perticarini M. (2022). Machine Learning As an Altenative to 3D Photomodelling Employed in Architectural Survey and Automatic Design Modelling. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLVIII-2/W1-2022, p. 191-197, ISSN: 2194-9034, doi: https://doi.org/10.5194/isprs-archives-XLVIII-2-W1-2022-191-2022

[2] Lombardi, S. et al.: Neural volumes: Learning dynamic renderable volumes from images. Association for Computing Machinery Trans Graph 38, New York, NY, USA (2019).

[3] Tewari, A. et al.: State of the Art on Neural Rendering, Computer Graphics Forum, John Wiley & Sons Ltd, Chichester United Kingdom (2020).

[4] Park, K. et al.: Nerfies: Deformable Neural Radiance Fields. Proceedings of the IEEE/CVF International Conference on Computer Vision (2021).

[5] Martin-Brualla, R. et al.: NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections. Computing Research Repository (2020).

[6] Mildenhall, B. et al.: NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. Computing Research Repository (2020).

[7] Chen, A. et al.: MVSNeRF: Fast Generalizable Radiance Field Reconstruction from Multi-View Stereo. Computing Research Repository (2021).

[8] Li, K., Rolff, T., Schmidt, S., Bacher, R., Frintrop, S., Leemans, W., & Steinicke, F.:(2022). Immersive Neural Graphics Primitives. arXiv preprint arXiv:2211.13494

[9] Hong, K., Wang, H. & Yuan, B.: Inspection-Nerf: Rendering Multi-Type Local Images for Dam Surface Inspection Task Using Climbing Robot and Neural Radiance Field. Buildings 13, Switzerland (2023).

[10] Müller, T., Evans, A., Schied, C. & Keller, A.: Instant neural graphics primitives with a multiresolution hash encoding. ACM Trans Graph 41, Association for Computing Machinery, New York, USA (2022).

[11] Li, Z., Niklaus, S., Snavely, N. & Wang, O.: Neural Scene Flow Fields for Space-Time View Synthesis of Dynamic Scenes. arXiv preprint arXiv:2011.13084v3 (2020).

[12] Kurz, A., Neff, T., Lv, Z., Zollhöfer, M. & Steinberger, M.: Supplementary Material for ‘AdaNeRF: Adaptive Sampling for Real-time Rendering of Neural Radiance Fields’. arXiv e-prints (2022).

[13] Park, K. et al.: Nerfies: Deformable Neural Radiance Fields. arXiv e-prints (2020).