Preprints.org

Article

Comparative Assessment of NeRF and Photogrammetry in Digital Heritage: Impact of Varying Image Conditions on 3D Reconstruction

Valeria Croce , Dario Billi , Gabriella Caroti [*] , Andrea Piemonte , Livio De Luca , Philippe Véron

Article

Comparative Assessment of NeRF and Photogrammetry in Digital Heritage: Impact of Varying Image Conditions on 3D Reconstruction

Valeria Croce [1], Dario Billi [2], Gabriella Caroti [2,*], Andrea Piemonte [2], Livio De Luca [3] and Philippe Véron [1]

[1] LISPEN EA 7515, Arts et Métiers Institute of Technology, 13100 Aix-en-Provence, France; valeria.croce@ensam.eu; philippe.veron@ensam.eu

[2] Department of Civil and Industrial Engineering, ASTRO Laboratory, University of Pisa, L.go Lucio Lazzarino, 56122 Pisa, Italy; dario.billi@studenti.unipi.it; andrea.piemonte@unipi.it

[3] UMR MAP 3495 CNRS/MC, Campus CNRS Joseph-Aiguier, 13402 Marseille, France; livio.deluca@map.cnrs.fr

* Correspondence: gabriella.caroti@unipi.it

Abstract: While Neural Radiance Fields (NeRF) are gaining increasing interest in various domains as innovative methods for novel view synthesis and image-based reconstruction, their potential application in the realm of Cultural Heritage remains unexplored. Purpose of this paper is to assess the effectiveness of applying NeRF to sets of images of digital heritage objects and sites. The study's findings demonstrate that NeRF could be valuable when used in combination with or as a comparison to other well-established techniques such as photogrammetry, to expand the possibilities of preserving and presenting heritage assets with enhanced visual fidelity and accuracy. Particularly, NeRF show promising results in improving the rendering of translucent and reflective surfaces, objects with homogeneous textures, and elements with intricate details. In addition, we demonstrate that, when considering the same set of input images (with known camera poses), reducing the image quality or the number of images results in significantly less information loss with NeRF compared to photogrammetry. This suggests that NeRF is preferentially suited for scenarios involving sparse information or low-quality photos or videos, which could be especially valuable in risky or challenging situations.

Keywords: NeRF; Cultural Heritage; 3D reconstruction; photogrammetry; 3D surveying

1. Introduction

The process of reconstructing and digitally documenting heritage artifacts and scenes holds great significance in studying, enhancing and preserving tangible Cultural Heritage (CH). This significance is demonstrated by the extensive digitization campaigns undertaken, the remarkable advancements in active and passive sensor-based survey techniques, and the various established methodologies for three-dimensional reconstruction [1]. Artificial Intelligence, Virtual Reality and Extended Reality technologies are currently being extensively applied to three-dimensional data for a wide range of tasks related to the large-scale analysis [2], valorization [3,4] and communication [5,6] of cultural heritage objects. These technologies are designed to cater the needs of both experts and the general public, for a series of applications: preservation of CH by creation of digital replicas [7]; support of restoration and conservation activities [8,9]; education and tourism, cultural promotion (to make CH more accessible, e.g., by virtual tours, interactive exhibits and educational materials) [10].

In this context, the rendering of material and surface characteristics of existing heritage objects and sites, faithfully representing actual shape and color of the latter, is of paramount significance. Image-based reconstruction techniques are studied for this purpose, and Structure-from-Motion (SfM) and Multi-View-Stereo (MVS) photogrammetry is considered as a consolidated methodology in this field, currently complementing active survey methods as Light Detection and Ranging (LiDAR) [11].

Photogrammetry is employed at multiple levels and scales of documentation to produce point clouds, textured 3D meshes, orthophotos, and to extract metric information, through control points. These several outputs can be harnessed for further semantic enrichment, in-depth analysis, precise measurements, animations as well as immersive visualizations [12].

With recent advancements in the realm of Artificial Intelligence applied to digital representation [5,13,14], and following Mildenhall's pioneering work [15], the introduction of Neural Radiance Fields (NeRF) might appear a promising alternative to photogrammetry for 3D reconstruction from images. NeRF rely on a fully connected neural network to generate novel views of 3D objects and scene based on a series of overlapping images with known camera poses. Initially designed for the specific task of novel view synthesis [16], NeRF optimize an underlying continuous volumetric scene function that allows the generation of neural renderings.

NeRF models are photo-realistic [16] and they have achieved visual quality for the reproduction of real shapes and appearance of a 3D scene. However, the interest in applying NeRFs compared to other well-established technologies like photogrammetry is still not yet recognized to its full potential. Even though the starting point is still a series of overlapping images, NeRF use neural networks to create the so-called radiance fields, instead of relying on the reconstruction of geometrical relations between image and 3D world space [17]. The output, provided in the form of a neural rendering, can be turned into common, consolidated 3D objects as point clouds or meshes, e.g., via the marching cubes algorithm [18].

1.1. Background and motivation

3D reconstruction methods have constituted a major endeavor within the Cultural Heritage domain. They can be broadly divided into contact and non-contact sensors.

Among non-contact (passive) sensors, photogrammetry relies on capturing and analyzing data from photographs and does not involve actively emitting any form of energy or signals to gather information about the objects or the environment being studied. It appears to date as a consolidated technique for three-dimensional (3D) models reconstruction of features or topographies from overlaying two-dimensional photographs taken from various locations and points of views [19].

Common photogrammetric software as MICMAC, RealityCapture and Agisoft Metashape enable the reconstruction of 3D data, in the form of point clouds, polygonal meshes, textured meshes. Ortho-rectified imagery and bi-dimensional maps can even be derived from a photogrammetric model, and properly geo-referenced based on the combination with traditional topographic surveying methods [11]. Photogrammetry is widely applied in many domains ranging from industrial design [20] to archeology [21], architecture [22] and agriculture [23], geology [24] etc.. However, photogrammetric reconstruction techniques face many common limitations in case of objects exhibiting challenging optical properties (as absorptivity, intense reflectivity or extensive scattering) [25], variations in lighting conditions as shadows, glare or inconsistent lighting, uniform or repetitive textures [11], and complex shapes or geometries [26].

In this context, NeRFs emerge as a cutting-edge technology that holds great promise for addressing some of these inherent limitations [16]. They rely on a supervised neural network model, the Multi-Layer Perceptron (MLP), i.e., a feedforward (non-convolutional) deep network that consists of fully connected neurons with a nonlinear kind of activation function.

The process essentially consists of three phases (Figure 1):

- First, as in photogrammetry, images are oriented in 3D space.
- Then, the sampled points, characterized by their three spatial and by the viewing direction, are processed by the MLP, resulting in a color and volume density information as output.
- Finally, volumetric rendering techniques as ray tracing [15] or cone tracing [27] are used to synthesize the output information from the previous phase to produce new views (novel view synthesis).

In NeRF's formulation, the scene is intended as a continuous 5D function that associates to every point in space the corresponding radiance emitted in each direction. The function maps spatial

coordinates within the scene (x, y, z) and two viewing angles, azimuthal and polar, which specify the viewing direction (ϑ, φ), to output the volume density σ and a RGB color dependent on the viewing direction. The volume is intended as a differential opacity indicative of the radiance accumulated by a ray passing through each point. Considering the Multi-Layer Perceptron F_Θ, the following formulation applies:
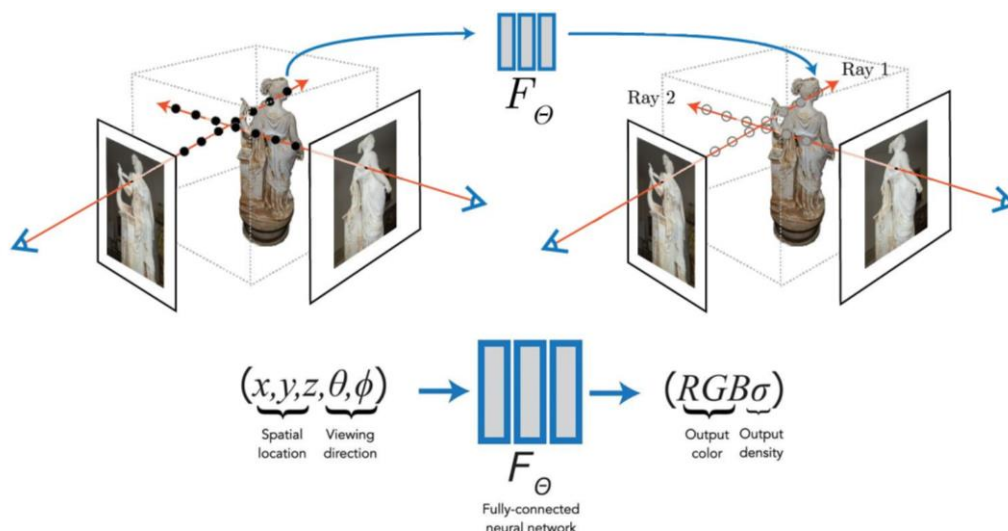
$$F_\Theta : (x, d) \rightarrow (c, \sigma)$$

(1)

Where **x** = (x, y, z) are the in-scene coordinates, the direction is expressed as a 3D Cartesian unit vector **d** (ϑ, φ); **c** = (r, g, b) represents the color values; σ indicates the volume density. Although σ is independent of the viewing direction, **c** depends on both viewing direction and in-scene coordinates.

The theoretical formulation of NeRF has the following consequences: (i) the 3D scene is represented as neural rendering and not - at least initially - as a point cloud or mesh; (ii) the representation of the scene is view-dependent, which results in a more realistic variation in color and illumination, both with respect to lighting conditions and handling of reflective surfaces.

Since their introduction in 2020, NeRFs have seen numerous implementations and have been extended to various fields of research, including urban mapping [28,29], robotics [30,31], autonomous driving [32], simulation of climate effects [33], industrial design [34], human pose estimation [35]. Mazzacca et al. [36] have proposed several methods such as noise level, geometric accuracy, and the number of required images (i.e., image baselines) to evaluate NeRF-based 3D reconstructions.

In the specific field of cultural heritage, the first experiments began with the work by Condorelli and Rinaudo [37], and they have continued over the past three years, incorporating the developments and improvements that NeRFs have undergone since their earliest version. Interest in the application of NeRF to the Cultural Heritage sector is on the rise, as evidenced by a recent session at the CIPA 2023 Conference on Documenting, Understanding, Preserving Cultural Heritage, entirely dedicated to the theme 'AI and NeRF for 3D reconstruction' [36,38–41].

Figure 1. NeRF training overview, adapted from the original NeRF paper by Mildenhall et al. [15].

1.2. Research aim

The aim of this work is to assess the intrinsic advantages or limitations of NeRF compared to photogrammetry and the possible benefits of integrating the two methods for digital 3D reconstruction of cultural heritage objects. Considering a same set of input images with known camera poses, we aim at comparing the two techniques in terms of quality and consistency of the results, handling of challenging scenes (e.g.: objects with reflective, metallic or translucent surfaces), realistic renderings of the detected objects, processing time and impact of image resolution and numbering on the accuracy and fidelity of the 3D reconstruction.
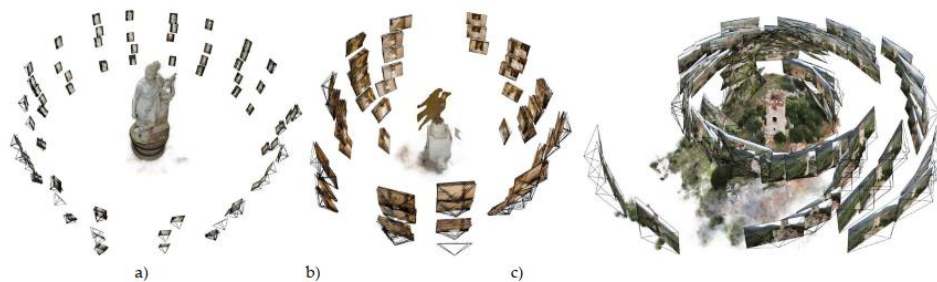
4

2. Case studies

The analyzed case studies (Figures 2 and 3) are related to datasets that are typically difficult to handle for photogrammetry, due to different peculiarities of the objects' surfaces, e.g. very fine details or specific material characteristics (homogeneity, roughness, brilliance, chromaticity, luminosity and hardness). Some of these objects feature non-Lambertian surfaces, meaning that they do not reflect light uniformly in all directions and exhibit angle-dependent reflectance properties.

We consider objects of various scales, ranging from small museum artifacts to monumental statues and even urban contexts. In this contribution, we present the results of three specific case studies, anticipating that other cases are still being implemented at the time of writing. The pilot case study chosen to test the approach is the statue of the Terpsichore, the masterpiece of the Italian artist, sculptural exponent of Neoclassicism, Antonio Canova (Figure 2a).

The statue represents the muse of dance and choral singing Terpsichore, as can be deduced from the lyre placed on the high pedestal. The work has intricate details, such as the strings and soundbox of the lyre or the drapery of the robe, and has specific material characteristics that are difficult to reproduce, such as the homogeneity of the color and the white and pink veining of its material. The original dataset consists of 233 images.

Figure 2. The three case studies analyzed: (a) Terpsichore by Antonio Canova, (b) eagle-shaped lectern and (c) Caprona Tower.

Figure 3. Image acquisition pattern displayed for the neural renderings of: (a) the Terpsichore, (b) the eagle-shaped lectern and (c) the Caprona Tower.

The second case study considered is that of the eagle-shaped lectern (254 images), that presents specular reflections and other variations in appearance based on the direction of incoming light and the viewing angle, due to the metallic, shiny surface of the bronze (Figure 2b).

Finally, the Caprona Tower dataset, composed of 124 images acquired by drone, presents a wider context scale comprising the natural environment surrounding the tower, composed of trees and shrubs (Figure 2c). The photogrammetric reconstruction of this study object, together with the historical remarks on the tower, were previously documented in reference [44].
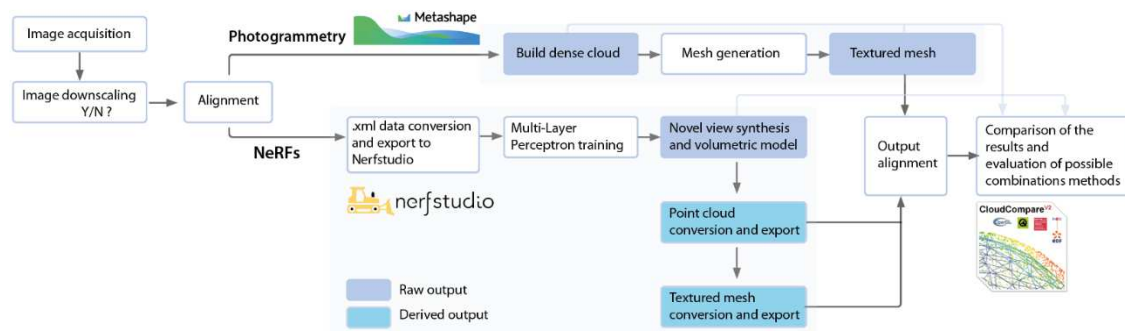
3. Methodology

This present study aims to test the validity and potential advantages of using NeRF in comparison to well-established image-based reconstruction techniques, with photogrammetry being the foremost among them.

For a direct comparison between NeRF and photogrammetry, this research employs the same set of synthetic images.

These images depict objects with non-Lambertian surfaces or composed of very fine materials. In this dataset, following standard photogrammetric practices, the object of study is captured through a sequence of successive shots, ensuring more than 70% overlap.

Over these initial datasets, two distinct methodologies are employed (Figure 4):

- **Photogrammetric Procedure.** This involved estimating camera orientation parameters for sparse point cloud construction, generating a dense point cloud, creating a mesh, and extracting textures. The software used for this task is Agisoft Metashape, and the alignment, dense point cloud, mesh and texture generation phases are run in *high quality* mode.

- **NeRF-Based Reconstruction.** This method requires the camera pose estimate to be known. With this input, a Multi-Layer Perceptron is trained for novel view synthesis, and the neural rendering (volumetric model) is generated. For the latter part, the NeRFstudio Application Programming Interface by Tancik et al. [42]. By default, this application applies a scaling factor to the images to reduce their dimensions and expedite the training process (*downscaling*).

Figure 4. Overview of the proposed methodology.

As the alignment procedure is common to both processes and is executed following photogrammetry principles, an initial assumption is made about the dataset: the images used for reconstructing the 3D scene can be aligned, which is to say that the estimation of camera poses is feasible using the input images.

In both cases, camera orientation parameters are calculated using the Agisoft Metashape software. The camera data file, in .xml format, exported from Metashape, is then imported into NeRFstudio [42], serving as initial data for the 3D scene reconstruction. It is important to note that for this step to work, all images are required to be of the same sensor type.

By way of comparison, since the output of photogrammetry cannot be directly confronted with neural rendering, an additional conversion phase is introduced. NeRFstudio allows the export of NeRF to point clouds and meshes via the *ns-export* function. The marching cubes algorithm [18] and the Poisson surface reconstruction method [43] are used for this purpose, while texture coordinates are derived via the *xatlas* library (github.com/mworchel/xatlaspython).

Referring to known measurements from topographic surveys, the point clouds obtained through photogrammetry and NeRF are individually scaled and aligned. Subsequently, they are processed to derive a cloud-to-cloud comparison using the CloudCompare software (CloudCompare 2.12.4, 2023). Various configurations -of both input image sizes and quantities- are considered in this phase to

determine whether reducing the resolution or the number of images leads to a decrease in the quality of the 3D reconstruction.

For the pilot case of the Terpsichore statue, considering that the production process of volumetric models using NeRF includes, by default, a downscaling factor, i.e. a reduction of the image size by factor 3 (Figures 5 and 6) and even considering that a lower number of images results in a reduction of the training times, 4 different dataset configurations are tested, as follows:

1. 233 photos, no downscale;
2. 233 photos with downscale of factor 3 (3x);
3. Reduced dataset of 116 photos (~50% of the input dataset) with no downscale;
4. Reduced dataset of 116 photos with downscale 3x.

It has to be noted that, in combinations 3 and 4, the reduction in dataset size is accomplished by systematically removing images one by one.

a)   b)

Figure 5. Comparison of image sizes between initial dataset (a, 2752x4128 pixels) and reduced dataset, downscaled by factor 3 (to the left, 344x516 pixels).
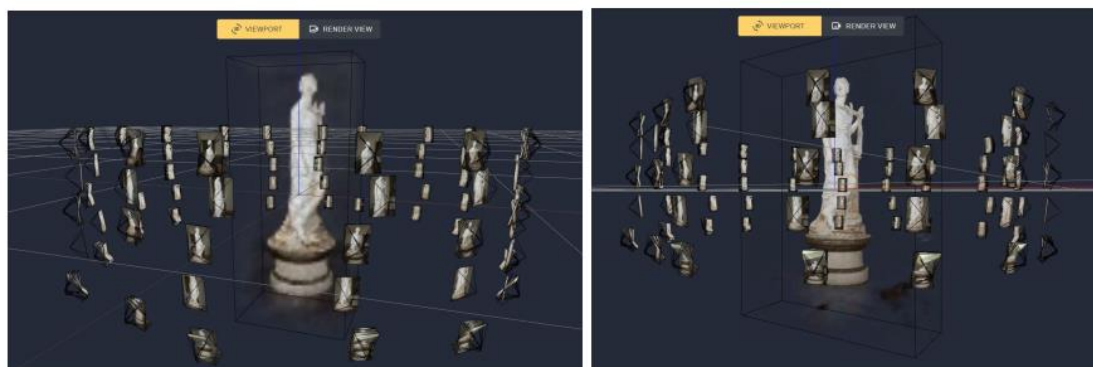

Figure 6. Comparison of image resolution between initial dataset (a, 2752x4128 pixels) and reduced dataset, downscaled by factor 3 (b, 344x516 pixels).

4. Results

For each of the 4 combinations considered, we proceeded by drawing up synoptic tables showing the comparison, respectively, between the point cloud and the mesh acquired from photogrammetry, and between the neural rendering, the point cloud and the mesh extracted from Nerfstudio (Figures 7, 8, 9).

The point cloud extracted by NeRFstudio and the photogrammetric point cloud are later aligned on CloudCompare and compared with each other (Figure 10), to derive an analysis of deviation between minimum values (in blue) and maximum values (in red).

Figure 7. Neural rendering with training in progress (a) and training completed (b).
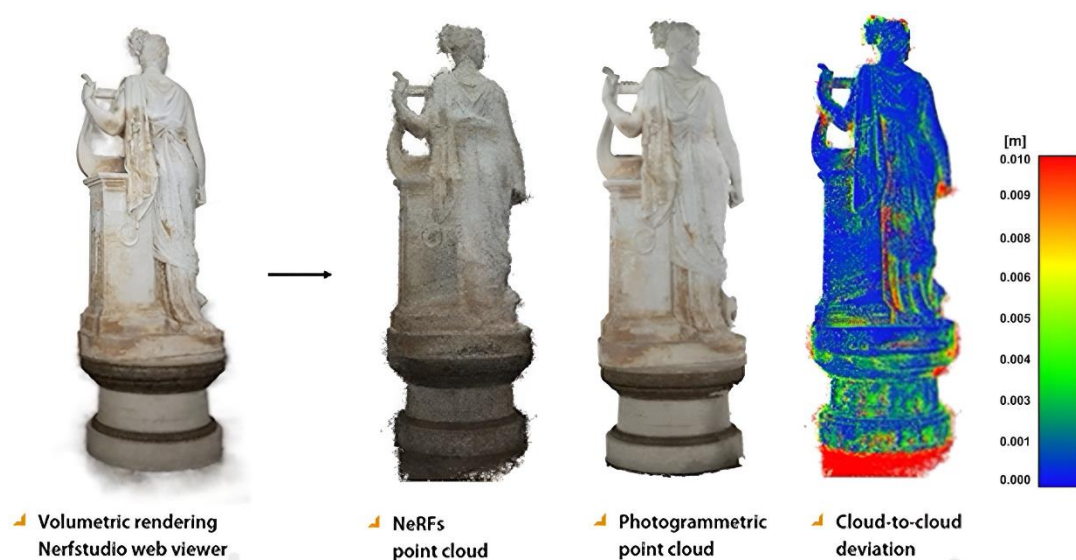

Figure 8. Neural renderings of the Terpsichore statue from different points of view, case 1.
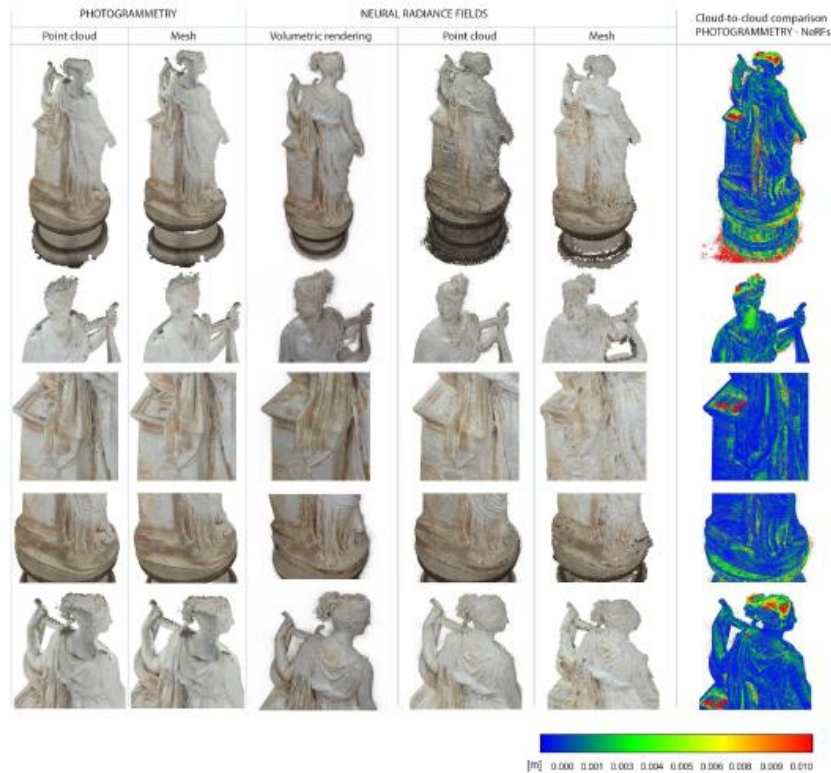
Figure 9. Several details of the neural renderings of the Terpsichore statue, case 4.
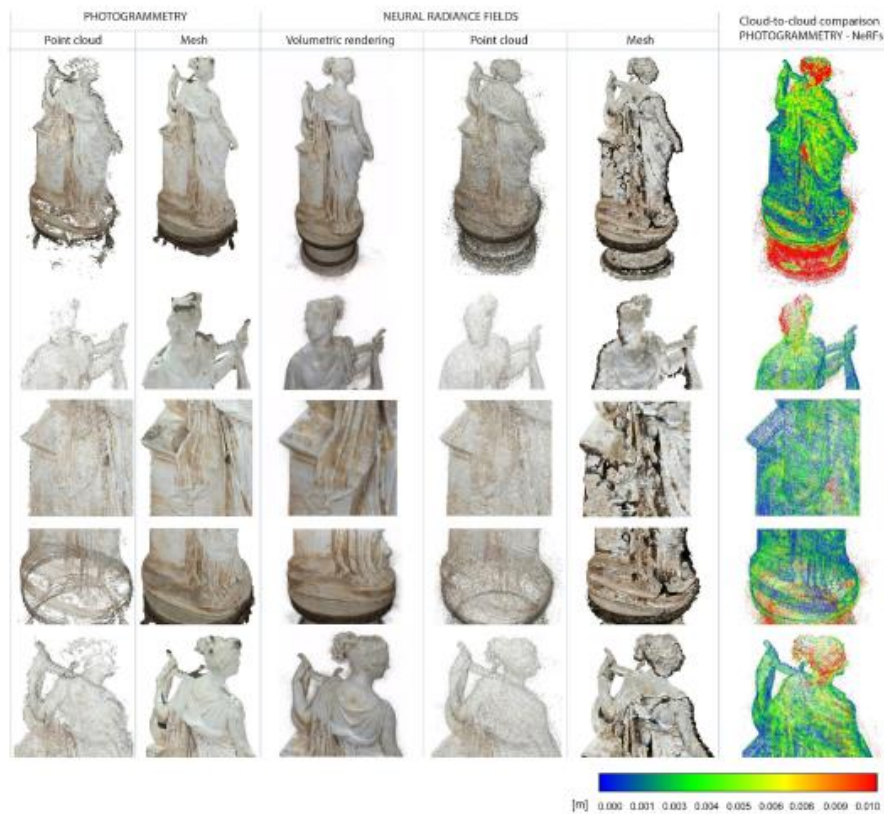
The complete synoptic tables for all 4 datasets are presented, respectively, in Figures 11, 12, 13 and 14. First, each synoptic tables illustrates a general overview of the whole result, respectively: of point cloud and mesh from photogrammetry; of neural rendering, extracted point cloud and extracted mesh from NeRF; of the cloud-to-cloud comparison between NeRF and photogrammetry. In addition, detailed, zoomed-in views of specific areas where the differences among the various tests are most pronounced are presented for each result. These areas of focus include the head, pedestal, base, and the rear of the statue, with specific attention to details of the *lira.*
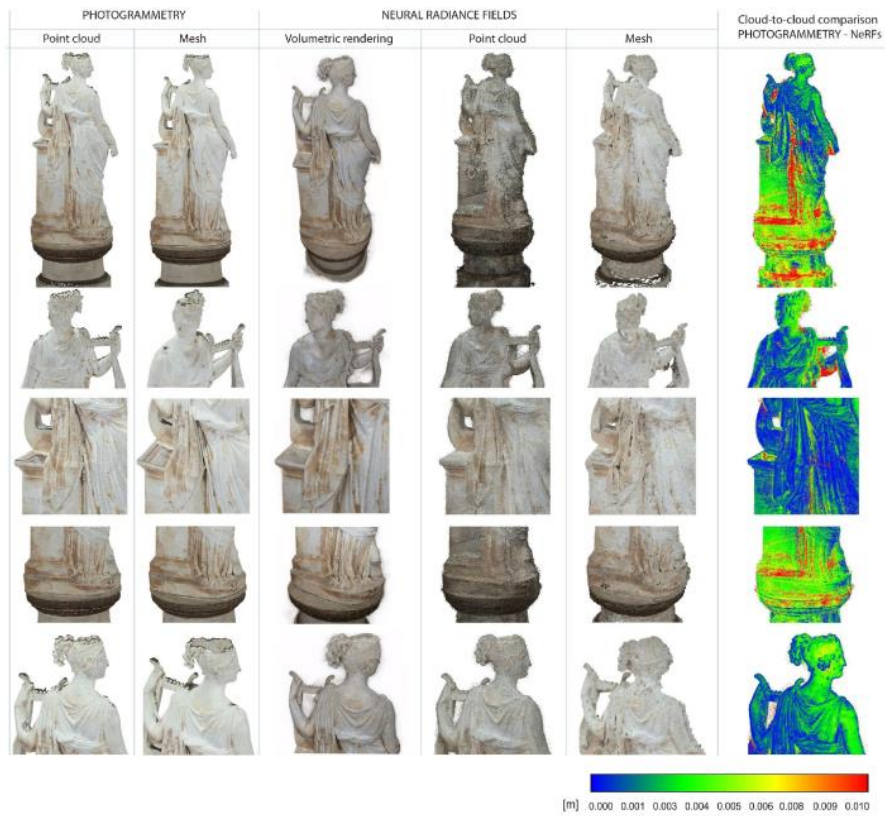

Figure 10. Scheme of the comparison between NeRF and photogrammetric point cloud.

Figure 11. Analysis of the deviation between NeRF and photogrammetric processing results. Dataset 1: 233 images, no downscale.

Figure 12. Analysis of the deviation between NeRF and photogrammetric processing results. Dataset 2: 233 images with 3x downscale.
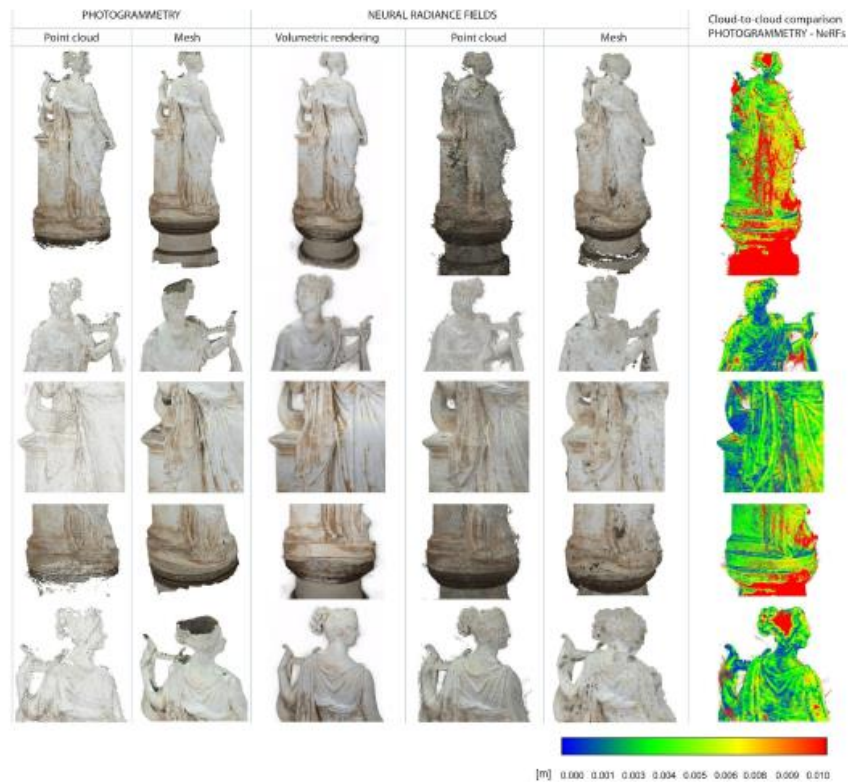
Figure 13. Analysis of the deviation between NeRF and photogrammetric processing results. Dataset 3: 116 images with no downscale.

Figure 14. Analysis of the deviation between NeRF and photogrammetric processing results. Dataset 4: 116 images with 3x downscale.

In detail, Figures 11–14 reveal that as image resolution and the number of available images decrease (provided image alignment remains feasible), NeRFs exhibit significantly less quantitative information loss compared to photogrammetry. This phenomenon becomes particularly pronounced when examining specific details such as the statue's head (present in only a few initial dataset images) and the upper and lower sections of the pedestal. Additionally, a noticeable enhancement in rendering very fine lyre details is observed when transitioning from case 1 to case 4.

The deviation analysis shows a general range from 0 mm (blue points) to a maximum value of 10 mm (red points). In the first dataset (Figure 11), the maximum deviation occurs at the top of Terpsichore's head and at the upper and lower sections of the pedestal. Moreover, it can be noted that the point cloud obtained from NeRF exhibits more noise, with deviations spread across all surfaces. This is particularly noticeable where the deviation is minimal (1 mm in blue), but there appear isolated points with medium (4 to 7 mm in green-yellow) and high (1 cm in red) deviations.

In the second dataset (Figure 12), the maximum deviation of 1 cm is found in the Terpsichore head and the lower section of the pedestal. Compared to the previous case, deviations between 1 and 4 mm are present throughout the entire statue, while the lyre and the pedestal maintain 1 mm deviations.

For the third dataset (Figure 13), maximum deviations of 1 cm shift to the lower part of the pedestal, concentrated between the square of the pedestal and the drapery.The deviations in the Terpsichore head decrease to an average of 3 to 8 mm.

For the last case (Figure 14), higher deviations, reaching the value of 1 cm, are observed in the head of the statue, as well as in the lower part of the pedestal: indeed, these parts appear to be entirely missing in the photogrammetric result. Meanwhile, average deviations of 4 to 5 mm are found along the entire drapery. The consideration on the observed noise in the NeRF point cloud, as anticipated for the first case, is relevant for all the other cases mentioned above. This can also be appreciated by a visual comparison between the volumetric rendering and the photogrammetric mesh.

Overall, these results on the pilot case study suggest that:

- Compared to photogrammetry, NeRF may offer the ability to handle reduced image data or reduced resolution of the images, with lower quantitative information loss. For the 3rd and 4th cases analyzed, indeed, NeRF capture details, such as the head and lower pedestal, that are absent in the photogrammetric output. This is true, however, if the reconstruction of camera poses is possible over the *reduced* datasets;
- NeRF neural renderings more faithfully reproduce the statue's material texture compared to the textured mesh obtained through photogrammetry.
- However, NeRF are more prone to noise and, for higher-resolution datasets, they may encounter challenges in capturing specific fine details compared to photogrammetry.

To underscore the stark contrast in texture rendering between photogrammetry and NeRF, the example of the eagle-shaped lectern vividly demonstrates the concept of view-dependency achieved through NeRF. In the series of images provided as input, the illumination on the eagle's chest varies dramatically. Examining the result obtained via photogrammetry, such variation in appearance cannot be observed. NeRF, in this instance, excels in providing color changes and reflections that faithfully mimic the actual bronze surface, creating a highly realistic representation: as in Figure 15, the illumination on the eagle's chest varies dramatically. The images below, taken from different viewpoints clearly show how the rendering of the raptor's metallic surface changes depending on the view direction.

Figure 15. From left to right: the right side of the eagle's chest is bathed in light, while in the second image, the reflective part shifts towards the left. In the third image, the light centers and descends lower on the sculpture.

In stark contrast, the photogrammetric model appears quite matter in terms of texture (Figures 16 and 17). It does not convey the material's true nature, making it difficult to distinguish whether the surface is made of bronze or wood, illustrating the limitations of photogrammetry in capturing the nuanced material properties of the subject.

Figure 16. Photogrammetry and NeRF comparison results of the eagle-shaped lectern.

Figure 17. Photogrammetry and NeRF comparison results of the eagle-shaped lectern.

Concerning the case study of the Caprona Tower, when comparing the results obtained through NeRF with those produced by photogrammetry, a striking disparity becomes evident in terms of representing the vegetation surrounding the Tower (Figures 18 and 19). Notably, there are low bushes scattered about, and certain parts of the mesh model exhibit holes or pronounced sharpness, which are rendered exceptionally well-defined in the volumetric representation. This discrepancy is particularly noticeable when observing the tree on the left side of the tower and some of the lower bushes.

Figure 18. Photogrammetry and NeRF comparison results for the Caprona Tower.

Figure 19. Photogrammetry and NeRF comparison results for the Caprona Tower.

However, it is essential to acknowledge that while NeRF excels in capturing the volumetric details of natural elements, such as the aforementioned vegetation, it may fall short in representing the finer details of the Tower's masonry texture. In contrast, photogrammetry excels in capturing the intricate texture of the Tower's facades.

Given this disparity in strengths and weaknesses, a possible solution may involve integrating the two models. Combining the precision of NeRF in representing natural elements with the texture detail captured by photogrammetry could lead to a more comprehensive and accurate overall representation of the scene. This fusion of techniques could strike a balance between volumetric accuracy and textural fidelity, offering a more holistic view of the subject matter.

5. Discussion

Considering the outcomes of our investigation, the discussion comparing NeRF and photogrammetry can be framed around several key factors, including model description, data processing time and possible conversion to other forms of representation. In any case, a shared characteristic between these two techniques is represented by the reconstruction of internal and

external camera orientation parameters. As a result, ensuring that the model maintains metric correctness and accurate scaling necessitates processes like georeferencing or the utilization of known measurements. The main differences between the two methods are explained below.

Model Description. The comparisons between these two techniques reveal that, with respect to characterizing the shape of models, NeRF introduces a unique capability. It allows the appearance of materials to be contingent on the observer's perspective, an attribute that proves exceptionally powerful when dealing with surfaces that are not opaque, but rather transparent or reflective, and when handling textures that are uniformly consistent. These are challenging scenarios for standard photogrammetry to replicate. Moreover, the results on the pilot case of the Tersicore dataset demonstrate that NeRF models restitute finer details and parts as the number or the resolution of the input images decreases.

Representation. A notable difference in the model output between NeRF and photogrammetry lies in their representation: NeRF generates a continuous volumetric model, while photogrammetry produces a dense cloud or mesh-based model with discrete textured surfaces. This contrast in representation can significantly influence the suitability of each method for specific use cases, although many techniques for extracting meshes or point clouds from neural rendering are being proposed [42].

Data Processing Times. When it comes to the time required for processing data, photogrammetry typically demands a longer duration to generate a textured mesh from the same set of images. In contrast, NeRF training, although not instantaneous, is relatively swift, usually completing in approximately 30 minutes to an hour, contingent on the initial image file size. Furthermore, NeRF offers the capability to explore the model even before training reaches 100%. However, it is important to note that NeRF faces difficulties when processing very high-resolution images.

Sensitivity to reduction in number and resolution of images and trade-off with noise. NeRF is generally less sensitive to a number reduction of input images or to image downscaling than photogrammetry. In other words, NeRF models can maintain a more consistent level of detail in the 3D reconstruction, even with a smaller image dataset. However, it's important to note that as the number and resolution of images decrease, the NeRF point cloud can become noisier.

Conversion to Other Representations and Exporting Extension. A distinction between these methods pertains to their openness. Standard photogrammetry, particularly when based on Structure-from-Motion (SfM), adopts an open system approach. This allows for relatively straightforward conversion to alternative forms of representation and the exportation of models to various file formats. On the other hand, NeRF operates as a closed system, making the conversion to alternative representations and the export to different file extensions still challenging.

In conclusion, it must be noted that, as for the case of the Caprona Tower case study, a combination of NeRF and photogrammetry might offer a complementary approach, leveraging the strengths of each method for more comprehensive and versatile 3D modeling solutions. This hybrid approach could potentially yield superior results by addressing the limitations of each technique and opening up new possibilities for diverse applications in fields such as computer vision, augmented reality, and environmental mapping.

6. Conclusions

In this paper, we provide a comparative evaluation of NeRF and photogrammetry for applications in the Cultural Heritage domain. Our processing results, based on the pilot case study of the Terpsichore statue, indicate that as the input data and image resolution decrease on the same initial set of images, NeRF exhibits better preservation of both completeness and material description compared to photogrammetry. Consequently, we recommend the use of NeRF for datasets with limited images or low resolution, provided that the camera poses for the initial image set are known. This approach holds great significance in scenarios requiring large-scale area mapping by aerial surveying, especially during emergency situations in which the swift acquisition of extensive survey data is critical, and resources and accessibility are constrained. Based on the results of the three case

studied considered, a comparison between photogrammetry and NeRF is conducted in Section 5, respectively in terms of: model description and representation, data processing times, sensitivity to reduction in number and resolution of images and trade-off with noise, conversion to other representations and exporting extension.

Despite being a more recent technique still under development compared to the well-established photogrammetry, NeRF display higher potential for describing material characteristics. This technology may find applications in enhancing the representation of view-dependent materials and objects with intricate details, especially in cases of limited number of input images. Furthermore, the rendering property of NeRF may appear particularly useful for materials featuring homogeneous textures.

Our future developments encompass extending these findings to other case studies to validate and expand the range of possibilities for NeRF applications in cultural heritage, architecture, and industrial design. Specifically, we are interested in exploring applications on planar surfaces such as building facades, objects with homogeneous textures like statues or bronze furnishings, and mechanical components constructed from steel or other materials with reflective surfaces.

The very recent applications for creating models categorized into classes through semantic labels (semantic segmentation) [45] or for the reconstruction of in-the-wild scenes [46] further suggest possible developments in this context. The exploration of NeRFs in Virtual and Augmented Reality applications, following the works by Deng et al. [47], is also the subject of ongoing work.

References

1. Moyano, J.; Nieto-Julián, J.E.; Bienvenido-Huertas, D.; Marín-García, D. Validation of Close-Range Photogrammetry for Architectural and Archaeological Heritage: Analysis of Point Density and 3d Mesh Geometry. *Remote Sensing* **2020**, *12*, 1–23, doi:10.3390/rs12213571.

2. Wojtkowska, M.; Kedzierski, M.; Delis, P. Validation of Terrestrial Laser Scanning and Artificial Intelligence for Measuring Deformations of Cultural Heritage Structures. *Measurement: Journal of the International Measurement Confederation* **2021**, *167*, doi:10.1016/j.measurement.2020.108291.

3. Díaz-Rodríguez, N.; Pisoni, G. Accessible Cultural Heritage through Explainable Artificial Intelligence. New York, Association for Computing Machinerym 2020; pp. 317–324.

4. Škola, F.; Rizvić, S.; Cozza, M.; Barbieri, L.; Bruno, F.; Skarlatos, D.; Liarokapis, F. Virtual Reality with 360-Video Storytelling in Cultural Heritage: Study of Presence, Engagement, and Immersion. *Sensors (Switzerland)* **2020**, *20*, 1–17, doi:10.3390/s20205851.

5. Fiorucci, M.; Khoroshiltseva, M.; Pontil, M.; Traviglia, A.; Del Bue, A.; James, S. Machine Learning for Cultural Heritage: A Survey. *Pattern Recognition Letters* **2020**, *133*, 102–108, doi:10.1016/j.patrec.2020.02.017.

6. Pedersen, I.; Gale, N.; Mirza-Babaei, P.; Reid, S. More than Meets the Eye: The Benefits of Augmented Reality and Holographic Displays for Digital Cultural Heritage. *Journal on Computing and Cultural Heritage* **2017**, *10*, doi:10.1145/3051480.

7. Trunfio, M.; Lucia, M.D.; Campana, S.; Magnelli, A. Innovating the Cultural Heritage Museum Service Model through Virtual Reality and Augmented Reality: The Effects on the Overall Visitor Experience and Satisfaction. *Journal of Heritage Tourism* **2022**, *17*, 1–19, doi:10.1080/1743873X.2020.1850742.

8. Gros, A.; Guillem, A.; De Luca, L.; Baillieul, É.; Duvocelle, B.; Malavergne, O.; Leroux, L.; Zimmer, T. Faceting the Post-Disaster Built Heritage Reconstruction Process within the Digital Twin Framework for Notre-Dame de Paris. *Scientific Reports* **2023**, *13*, doi:10.1038/s41598-023-32504-9.

9. Croce, V.; Caroti, G.; De Luca, L.; Jacquot, K.; Piemonte, A.; Véron, P. From the Semantic Point Cloud to Heritage-Building Information Modeling: A Semiautomatic Approach Exploiting Machine Learning. *Remote Sensing* **2021**, *13*, 1–34, doi:10.3390/rs13030461.

10. Bekele, M.K.; Pierdicca, R.; Frontoni, E.; Malinverni, E.S.; Gain, J. A Survey of Augmented, Virtual, and Mixed Reality for Cultural Heritage. *Journal on Computing and Cultural Heritage* **2018**, *11*, doi:10.1145/3145534.

11. Bevilacqua, M. G.; Caroti, G.; Piemonte, A.; Terranova, A. A. Digital Technology and Mechatronic Systems for the Architectural 3D Metric Survey. *Intelligent Systems, Control and Automation: Science and Engineering* **2018**, *92*, 161–180, doi:10.1007/978-3-319-68646-2_7.

12. Rea, P.; Pelliccio, A.; Ottaviano, E.; Saccucci, M. The Heritage Management and Preservation Using the Mechatronic Survey. *International Journal of Architectural Heritage* **2017**, *11*, 1121–1132, doi:10.1080/15583058.2017.1338790.

13. Croce, V.; Caroti, G.; Piemonte, A.; De Luca, L.; Véron, P. H-BIM and Artificial Intelligence: Classification of Architectural Heritage for Semi-Automatic Scan-to-BIM Reconstruction. *Sensors* **2023**, *23*, 2497, doi:10.3390/s23052497.

14. Condorelli, F.; Rinaudo, F. Cultural Heritage Reconstruction from Historical Photographs and Videos. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2018**, *XLII–2*, 259–265, doi:10.5194/isprs-archives-XLII-2-259-2018.

15. Mildenhall, B.; Srinivasan, P.P.; Tancik, M.; Barron, J.T.; Ramamoorthi, R.; Ng, R. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. *arXiv:2003.08934* **2020**.

16. Gao, K.; Gao, Y.; He, H.; Lu, D.; Xu, L.; Li, J. NeRF: Neural Radiance Field in 3D Vision, A Comprehensive Review **2022**, arXiv:2210.00379.

17. Murtiyoso, A.; Grussenmeyer, P. Initial Assessment on the Use of State-of-the-Art NeRF Neural Network 3D Reconstruction for Heritage Documentation. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2023**, *XLVIII-M-2–2023*, 1113–1118, doi:10.5194/isprs-archives-XLVIII-M-2-2023-1113-2023.

18. Lorensen, W.E.; Cline, H.E. Marching Cubes: A High Resolution 3D Surface Construction Algorithm. In Proceedings of the Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques; Association for Computing Machinery: New York, NY, USA, **1987**; pp. 163–169.

19. Yastikli, N. Documentation of Cultural Heritage Using Digital Photogrammetry and Laser Scanning. *Journal of Cultural Heritage* **2007**, *8*, 423–427, doi:10.1016/j.culher.2007.06.003.

20. James, D.W.; Belblidia, F.; Eckermann, J.E.; Sienz, J. An Innovative Photogrammetry Color Segmentation Based Technique as an Alternative Approach to 3D Scanning for Reverse Engineering Design. *Computer-Aided Design and Applications* **2017**, *14*, 1–16, doi:10.1080/16864360.2016.1199751.

21. Fiz, J.I.; Martín, P.M.; Cuesta, R.; Subías, E.; Codina, D.; Cartes, A. Examples and Results of Aerial Photogrammetry in Archeology with UAV: Geometric Documentation, High Resolution Multispectral Analysis, Models and 3D Printing. *Drones* **2022**, *6*, doi:10.3390/drones6030059.

22. Caroti, G.; Piemonte, A. Integration of Laser Scanning and Photogrammetry in Architecture Survey. Open Issue in Geomatics and Attention to Details. *Communications in Computer and Information Science* **2020**, *1246*, 170–185, doi:10.1007/978-3-030-62800-0_14.

23. Guimarães, N.; Pádua, L.; Marques, P.; Silva, N.; Peres, E.; Sousa, J.J. Forestry Remote Sensing from Unmanned Aerial Vehicles: A Review Focusing on the Data, Processing and Potentialities. *Remote Sensing* **2020**, *12*, doi:10.3390/rs12061046.

24. Haneberg, W.C. Using Close Range Terrestrial Digital Photogrammetry for 3-D Rock Slope Modeling and Discontinuity Mapping in the United States. *Bulletin of Engineering Geology and the Environment* **2008**, *67*, 457–469, doi:10.1007/s10064-008-0157-y.

25. Nicolae, C.; Nocerino, E.; Menna, F.; Remondino, F. Photogrammetry Applied to Problematic Artefacts. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2014**, XL-5, 451–456, https://doi.org/10.5194/isprsarchives-XL-5-451-2014.

26. Ippoliti, E.; Meschini, A.; Sicuranza, F. Digital Photogrammetry and Structure From Motion for Architectural Heritage: Comparison and Integration Between Procedures. In *Geospatial Intelligence: Concepts, Methodologies, Tools, and Applications* **2019**; Vol. 2, pp. 959–1018 ISBN 978-1-5225-8055-3.

27. Barron, J.T.; Mildenhall, B.; Tancik, M.; Hedman, P.; Martin-Brualla, R.; Srinivasan, P.P. Mip-NeRF: A Multiscale Representation for Anti-Aliasing Neural Radiance Fields. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV); IEEE: Montreal, QC, Canada, October 2021; pp. 5835–5844.

28. Derksen, D.; Izzo, D. Shadow Neural Radiance Fields for Multi-View Satellite Photogrammetry. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW); IEEE: Nashville, TN, USA, June 2021; pp. 1152–1161.

29. Semeraro, F.; Zhang, Y.; Wu, W.; Carroll, P. NeRF Applied to Satellite Imagery for Surface Reconstruction. *arXiv:2304.04133* **2023**.

30. Kerr, J.; Fu, L.; Huang, H.; Avigal, Y.; Tancik, M.; Ichnowski, J.; Kanazawa, A.; Goldberg, K. Evo-NeRF: Evolving NeRF for Sequential Robot Grasping of Transparent Objects. 2022.

31. Zhou, A.; Kim, M.J.; Wang, L.; Florence, P.; Finn, C. NeRF in the Palm of Your Hand: Corrective Augmentation for Robotics via Novel-View Synthesis *arXiv:2301.08556* **2023**.

32. Adamkiewicz, M.; Chen, T.; Caccavale, A.; Gardner, R.; Culbertson, P.; Bohg, J.; Schwager, M. Vision-Only Robot Navigation in a Neural Radiance World. *IEEE Robot. Autom. Lett.* **2022**, *7*, 4606–4613, doi:10.1109/LRA.2022.3150497.

33. Li, Y.; Lin, Z.-H.; Forsyth, D.; Huang, J.-B.; Wang, S. ClimateNeRF: Extreme Weather Synthesis in Neural Radiance Field. *arXiv:2211.13226* **2023**.

34. Mergy, A.; Lecuyer, G.; Derksen, D.; Izzo, D. Vision-Based Neural Scene Representations for Spacecraft. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW); IEEE: Nashville, TN, USA, June 2021; pp. 2002–2011.

35. Gafni, G.; Thies, J.; Zollhöfer, M.; Nießner, M. Dynamic Neural Radiance Fields for Monocular 4D Facial Avatar Reconstruction *arXiv:2012.03065* **2020**.

36.   Mazzacca, G.; Karami, A.; Rigon, S.; Farella, E.M.; Trybala, P.; Remondino, F. NeRF for Heritage 3D Reconstruction. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **2023**, *XLVIII-M-2–2023*, 1051–1058, doi:10.5194/isprs-archives-XLVIII-M-2-2023-1051-2023.

37.   Condorelli, F.; Rinaudo, F.; Salvadore, F.; Tagliaventi, S. A Comparison between 3D Reconstruction Using NERF Neural Networks and MVS Algorithms on Cultural Heritage Images. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2021**, *XLIII-B2-2021*, 565–570, doi:10.5194/isprs-archives-XLIII-B2-2021-565-2021.

38.   Croce, V.; Caroti, G.; De Luca, L.; Piemonte, A.; Véron, P. Neural Radiance Fields (NERF): Review and Potential Applications to Digital Cultural Heritage. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2023**, *XLVIII-M-2–2023*, 453–460, doi:10.5194/isprs-archives-XLVIII-M-2-2023-453-2023.

39.   Pansoni, S.; Tiribelli, S.; Paolanti, M.; Stefano, F.D.; Frontoni, E.; Malinverni, E.S.; finestra, C. a un sito esterno I. contenuti a cui indirizza il collegamento verranno aperti in una nuova; Giovanola, B. Artificial Intelligence and Cultural Heritage: Design and Assessment of an Ethical Framework. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2023**, *XLVIII-M-2–2023*, 1149–1155, doi:10.5194/isprs-archives-XLVIII-M-2-2023-1149-2023.

40.   Vandenabeele, L.; Häcki, M.; Pfister, M. Crowd-Sourced Surveying for Building Archaeology: The Potential of Structure From Motion (SFM) and Neural Radiance Fields (NERF). *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2023**, *XLVIII-M-2–2023*, 1599–1605, doi:10.5194/isprs-archives-XLVIII-M-2-2023-1599-2023.

41.   Balloni, E.; Gorgoglione, L.; Paolanti, M.; Mancini, A.; Pierdicca, R. Few Shot Photogrametry: A Comparison Between Nerf and Mvs-Sfm for the Documentation of Cultural Heritage. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2023**, *48M2*, 155–162, doi:10.5194/isprs-archives-XLVIII-M-2-2023-155-2023.

42.   Tancik, M.; Weber, E.; Ng, E.; Li, R.; Yi, B.; Kerr, J.; Wang, T.; Kristoffersen, A.; Austin, J.; Salahi, K.; et al. Nerfstudio: A Modular Framework for Neural Radiance Field Development 2023.

43.   Kazhdan, M.; Bolitho, M.; Hoppe, H. Poisson Surface Reconstruction. **2006**.

44.   Billi, D.; Rechichi, P.; Montalbano, G.; Croce, V. La Torre Degli Upezzinghi a Caprona: Analisi storico-archivistica e rilievo digitale per la documentazione dell'evoluzione temporale. In: Bevilacqua, M. G. and Ulivieri, D. (eds.) *Defensive Architecture of the Mediterranean, Proceedingsof the International Conference on Fortifications of the Mediterranean Coast FORTMED 2023* Pisa, 23, 24 and 25 March 2023 **2023**, XIII, 391–400.

45.   Zhi, S.; Laidlow, T.; Leutenegger, S.; Davison, A.J. In-Place Scene Labelling and Understanding with Implicit Scene Representation *arXiv:2103.15875* **2021**.

46.   Martin-Brualla, R.; Radwan, N.; Sajjadi, M.S.M.; Barron, J.T.; Dosovitskiy, A.; Duckworth, D. NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections *arXiv:2008.02268* **2021**.

47.   Deng, N.; He, Z.; Ye, J.; Duinkharjav, B.; Chakravarthula, P.; Yang, X.; Sun, Q. FoV-NeRF: Foveated Neural Radiance Fields for Virtual Reality. *IEEE Trans. Visual. Comput. Graphics* **2022**, 1–11, doi:10.1109/TVCG.2022.3203102.