

Neural Radiance Fields for Processing Drone Images

Subjects: **Imaging Science & Photographic Technology**

Contributor: Luigi Barazzetti , Mattia Previtali , Lorenzo Cantini , Annunziata Maria Oteri

Reconstruction of 3D scenes starting from images is a research topic in which scholars have been developing new methods since the beginning of the 20th century. The approach receiving more attention so far is the one based on the geometric consideration of image geometry and image-matching technique for radiometric consistency. Neural Radiance Fields (NeRFs) are an emerging technology that may provide novel outcomes in the heritage documentation field. NeRFs are also relatively new compared to traditional photogrammetric processing, and improvements are indeed expected with future research work.

drone

NeRFs

1. Introduction

Nowadays, 3D modeling from drone images has reached significant maturity and a high automation level. Digital photogrammetry coupled with computer vision methods (e.g., Structure from Motion/SfM) can automate most phases of the reconstruction workflow [1]. In addition, flight planning software also allow users to automate the data acquisition phase, which can be remotely planned before reaching the site [2].

Digital Recording of Ruins for Restoration and Conservation

The versatility of advanced digital recording tools proves to be invaluable in the field of conservation. Restoring and conserving historical buildings necessitates a precise understanding of their geometrical features. Digital recording techniques, such as laser scanning and digital photogrammetry, offer a very high level of detail, providing geometrically accurate information. This data can be utilized to analyze the historic building based on its geometrical proportions and to facilitate further investigations, including assessing the material composition and state of conservation.

The successful preparation of various conservation actions relies on creating a comprehensive recording dossier, instrumental for analyzing the proportions and architectural heritage. This includes verticality deviations or other deformations, understanding the construction logic, identifying materials, and assessing their conservation conditions.

Digital recording techniques offer the advantage of flexible 2D and 3D representations utilized throughout the conservation process, from preliminary studies and intervention identification to devising enhancement strategies

to ensure ongoing use and accessibility of the buildings.

2. NeRFs for Processing Drone Images

2.1. Preliminary Considerations

Reconstruction of 3D scenes starting from images is a research topic in which scholars have been developing new methods since the beginning of the 20th century. The approach receiving more attention so far is the one based on the geometric consideration of image geometry and image-matching technique for radiometric consistency [1][3][4][5]. New approaches are emerging in the field based on implicit representations. In those methods, a reconstruction model can be trained from a set of 2D images, and the trained model can learn implicit 3D shapes and textures [6]. In particular, Neural Radiance Field (NeRF) models are gaining popularity recently. NeRF models use volume rendering with implicit neural scene representation by using Multi-Layer Perceptrons (MLPs) to learn the geometry and lighting of a 3D scene starting from 2D images. The first paper introducing the concept of NeRF was published by Mildenhall et al. [7] in 2020, and to date (June 2023) it has received more than 3000 citations (source: Google Scholar), proving the increasing interest in this topic. Providing a literature review of this topic is out of the scope of this paper, and the reader is referred to recent review papers [8][9][10].

- NeRF models can be trained using only multi-view images of a scene. Unlike other neural representations, NeRF models require only images and poses to learn a scene;
- NeRF models are photo-realistic;
- NeRF models require low memory and perform computations faster.

In its basic form, a NeRF model represents three-dimensional scenes as a radiance field approximated by a neural network. The radiance field describes color and volume density for every point and every viewing direction in the scene. NeRF models are self-supervised. This means that they do not need extra data (like 3D or depth information) for the 3D reconstruction: the input of a NeRF neural network are camera poses (i.e., spatial location and viewing directions) of a set of 2D images of a scene. The output is the volume density of the scene and the view emitted radiance (RGB color) in any direction and for each scene location. The function mapping camera poses to a radiance field (color and volume density) are approximated by one or more Multi-Layer Perceptrons (MLPs). The prediction of the volume density (i.e., the scene's content) is constrained as independent concerning the viewing direction, while the color is set as dependent on both view direction and scene position. In its standard formulation, NeRF models are designed in two stages. The first stage takes as input the camera poses, passes them to a Multi-Layer Perceptron, and outputs volume density and a high-dimensional feature vector. In the second stage, the feature vector is concatenated with the viewing direction, using them as input for a further MLP, which outputs colors. Then, starting from a trained NeRF, new synthetic views can be created using this workflow:

- For each pixel to be created in the new synthetic image, camera rays and some sampling points are generated into the scene;
- Local color and density are computed for each sampling point by using the view direction, the sampling location, and the trained NeRF;
- The computed colors and densities are the input to produce the synthetic image using volume rendering.

NeRFs are an emerging technology that may provide novel outcomes in the heritage documentation field. A review of possible applications in digital recording and a comparison between NeRFs and traditional photogrammetric approaches was recently proposed in [11][12][13][14]. Previous work was also carried out by [15]. The author recognized the advantages in the case of complex objects with occlusions or irregular shapes, such as vegetated areas (one of the main problems faced in this work). The results also seem promising for homogenous surfaces or transparent/shining objects, which are difficult to model with photogrammetric methods. The metric accuracy of the achieved results is still questionable, especially when compared to photogrammetric or laser scanning point clouds. However, recent neural surface reconstruction methods seem very powerful [16].

Images and videos acquired from drones can also be processed using NeRFs, and previous work was carried out by [17]. The growing interest in such novel technology is demonstrated by the number of examples generated with Luma AI, accessible at <https://lumalabs.ai/featured> (accessed on 1 August 2023).

2.2. Castel Mancapane

2.2.1. Processing with Nerfstudio

The first test with NeRFs was carried out with the images oriented in Metashape and Nerfstudio (<https://docs.nerf.studio/en/latest/> (accessed on 1 August 2023)). Exterior, interior, and calibration parameters were exported from the software and processed using Nerfacto. Processing was very rapid (a few minutes, faster than the generation of the point clouds and mesh with Metashape), even considering the significant number of images in the project (549). Default parameters were used in the processing. An image of the viewer after training is shown in **Figure 1**, whereas some frames of the rendered video were illustrated at the beginning of the manuscript.

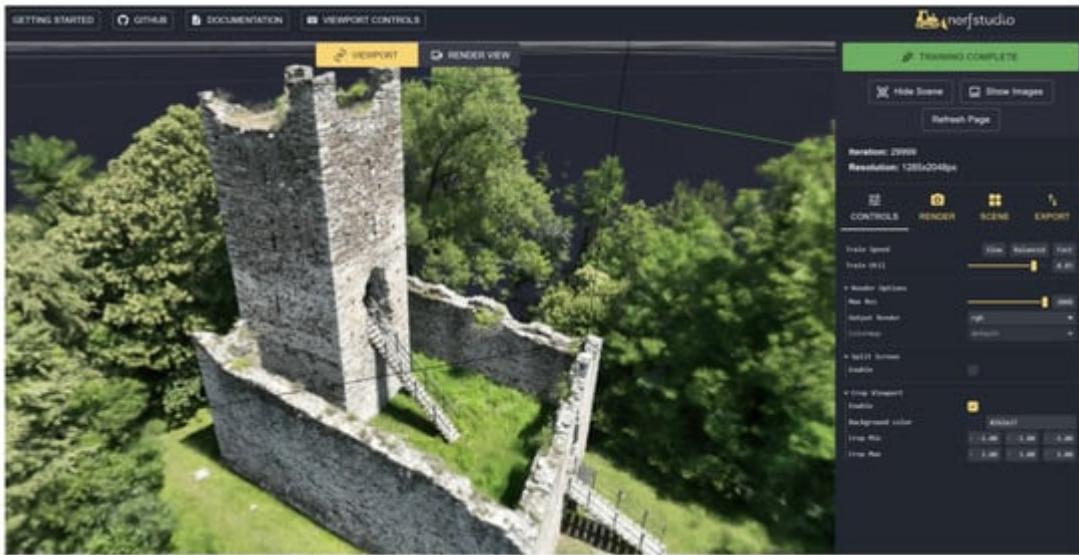


Figure 1. Castel Mancapane processed using Nerfstudio.

Rendering a video is very straightforward, and the quality of the results was excellent. In this sense, NeRFs provided excellent results for Castel Mancapane. Creating a camera path is also relatively simple, making NeRFs an alternative to traditional approaches for rendering based on point clouds or meshes.

Nerfstudio also allows for exporting the geometry. The point cloud was then exported, setting the number of points to 20 million and using the default setting provided by Nerfstudio. A visual comparison with the point cloud generated by Metashape is shown in **Figure 2**, which comprises 90 million points. The results achieved with Metashape are more complete and less noisy. However, processing was carried out using the default set of options, and the authors know that results could be improved. NeRFs are also relatively new compared to traditional photogrammetric processing, and improvements are indeed expected with future research work.



Figure 2. Point clouds produced with Nerfstudio (**left**) and Metashape (**right**) using the same set of images.

2.2.2. A Few Considerations on (Some) Other NeRFs Implementations

The same dataset was processed using Luma AI (<https://lumalabs.ai/>) (accessed on 1 August 2023) and Instant-npg (<https://github.com/NVlabs/instant-npg> (accessed on 1 August 2023)). Luma AI provides a fully automated workflow through a website in which images can be loaded as a single zip file. As the maximum size is 5 gigabytes, the images were resized by 50%. Instant-npg can be downloaded from the provided website, and we used COLMAP for image orientation. Additionally, in this case, compressed images were used in the workflow, knowing that image resizing affects the final results.

Rendering a video was straightforward with both solutions. The visual quality was excellent, especially for areas with vegetation, which is always difficult to model from a set of images and photogrammetric techniques. The background was also visible in the videos, including elements far from the castle. This interesting result could significantly improve traditional rendering from point clouds or meshes. A couple of sample frames from both videos are shown in **Figure 3**.



Figure 3. Two frames from the rendered videos using Lumia AI (**top**) and Instant-npg (**bottom**).

2.3. Rocca di Vogogna

A second test with NeRF was carried out for the Rocca di Vogogna dataset. This test aimed to compare results obtained with traditional image-matching techniques, those obtained with NeRF, and a common benchmark represented by a terrestrial laser scanning point cloud. A complete survey of the area for the Rocca di Vogogna is available. It was carried out using a Faro Focus X130 and a Leica BLK360.

For this case study, the comparison is between Metashape and Nerfstudio. The same images were used for both reconstructions: 390 oblique and convergent images. The set of normal images was discarded since it determined some issues with the NeRF training. Default parameters were used in the processing workflows with both software. Once the training in Nerfstudio was completed, the point cloud was exported, setting the target points to 20 million. The point cloud obtained with Metashape has about 35 million points and was then subsampled to 20 million for further comparison.

As previously noticed, also in the case of the Rocca di Vogogna dataset, a visual comparison shows a NeRF point cloud that is more noisy and less detailed than the one obtained with Metashape (**Figure 4**).



Figure 4. Point clouds produced with Nerfstudio (**left**) and Metashape (**right**) using the same set of images for the Vogogna dataset.

During the comparison with the terrestrial laser scanning data, the absolute distance between the point cloud is considered the test metric, and the laser scanning point cloud is considered the reference one. The computation of the distance between point clouds was performed by using the software CloudCompare (<https://www.danielgm.net/cc/> (accessed on 1 August 2023)). Results of the comparison for image matching and NeRF are reported in **Figure 5**. The color scale is the same in both figures.

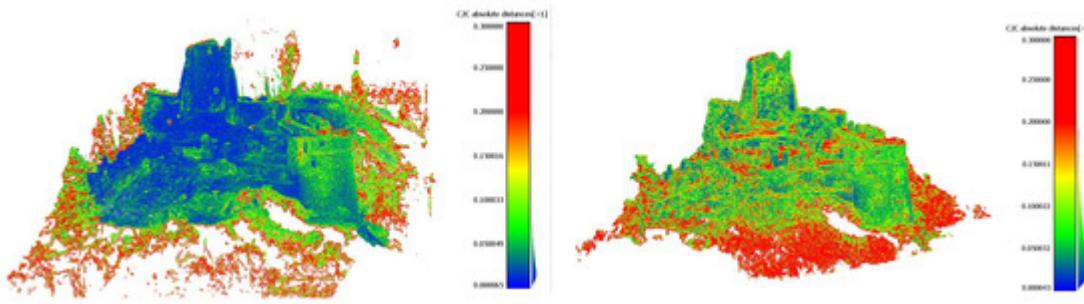


Figure 5. Comparison with TLS data: absolute distance. False color colorized point cloud (**left**) for image matching—Metashape (**left**) and NeRF—Nerfstudio (**right**).

No systematic error or discrepancy can be visualized in comparing both point clouds. However, some more significant problems with the NeRFs solution can be observed in correspondence of vegetated areas, which are not detectable visually.

As shown in **Figure 5**, the discrepancies between image matching and laser scanning data are in the order of 2–3 cm. In particular, 50% of the discrepancies are below 3.6 cm.

On the other hand (**Figure 5-right**), discrepancies for the NeRF reconstruction are in the order of magnitude of 4–5 cm, and 50% of the discrepancies are below 4.9 cm. In addition, the distribution of discrepancies for the image matching shows a lower dispersion of this dataset than the NeRF results, as it can be observed both in the density and the cumulative distribution of discrepancies for the two datasets (**Figure 6**). This is probably related to the higher noise in the NeRF solution.

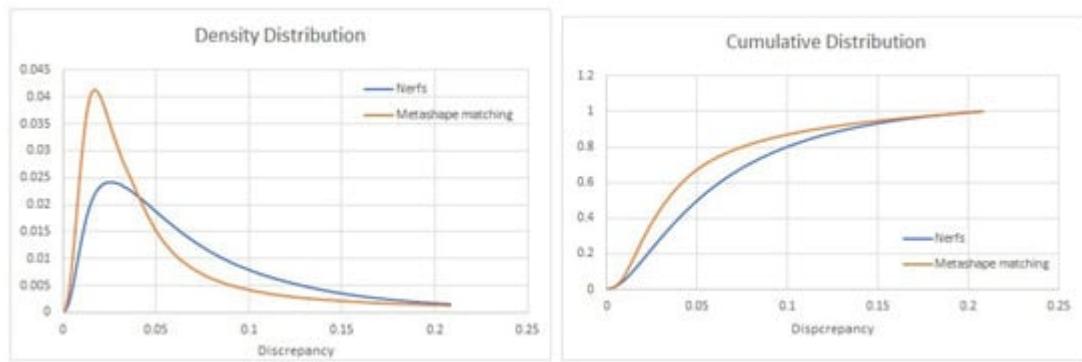


Figure 6. Density (**left**) and cumulative (**right**) distributions of the discrepancies for image matching with Metashape (orange) and the Nerfstudio solution (blue).

References

1. Remondino, F.; Nocerino, E.; Toschi, I.; Menna, F. A critical review of automated photogrammetric processing of large datasets. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* 2017, XLII-

2/W5, 591–599.

2. Hsieh, C.S.; Hsiao, D.H.; Lin, D.Y. Contour Mission Flight Planning of UAV for Photogrammetric in Hillside Areas. *Appl. Sci.* **2023**, *13*, 7666.
3. Luhmann, T. *Close Range Photogrammetry: Principles, Techniques and Applications*; Whittles Pub: Dunbeath, UK, 2006.
4. Hirschmuller, H. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *30*, 328–341.
5. Kraus, K. *Photogrammetry: Geometry from Images and Laser Scans*; Walter de Gruyter: Berlin/Heidelberg, Germany, 2011.
6. Liu, S.; Chen, W.; Li, T.; Li, H. Soft rasterizer: Differentiable rendering for unsupervised single-view mesh reconstruction. *arXiv* **2019**, arXiv:1901.05567.
7. Mildenhall, B.; Srinivasan, P.P.; Tancik, M.; Barron, J.T.; Ramamoorthi, R.; Ng, R. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. *Commun. ACM* **2020**, *65*, 96–106.
8. Gao, K.; Gao, Y.; He, H.; Lu, D.; Xu, L.; Li, J. NeRF: Neural Radiance Field in 3D Vision, A Comprehensive Review. *arXiv* **2023**, arXiv:2210.00379.
9. Debbagh, M. Neural Radiance Fields (NeRFs): A Review and Some Recent Developments. *arXiv* **2023**, arXiv:2305.00375.
10. Zhu, F.; Guo, S.; Song, L.; Xu, K.; Hu, J. Deep Review and Analysis of Recent NeRFs. *Apsipa Trans. Signal Inf. Process.* **2023**, *12*, e6.
11. Croce, V.; Caroti, G.; Luca, L.; Piemonte, A.; Véron, P. Neural radiance fields (nerf): Review and potential applications to digital cultural heritage. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2023**, *XLVIII-M-2-2023*, 453–460.
12. Balloni, E.; Gorgoglione, L.; Paolanti, M.; Mancini, A.; Pierdicca, R. Few shot photogrammetry: A comparison between nerf and mvs-sfm for the documentation of cultural heritage. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2023**, *XLVIII-M-2-2023*, 155–162.
13. Mazzacca, G.; Karami, A.; Rigon, S.; Farella, E.M.; Trybala, P.; Remondino, F. Nerf for heritage 3D reconstruction. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2023**, *XLVIII-M-2-2023*, 1051–1058.
14. 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. Spat. Inf. Sci.* **2023**, *XLVIII-M-2-2023*, 1113–1118.

15. 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. Spat. Inf. Sci.* 2021, **XLIII-B2-2021**, 565–570.
16. Li, Z.; Müller, T.; Evans, A.; Taylor, R.H.; Unberath, M.; Liu, M.Y.; Lin, C.H. Neuralangelo: High-Fidelity Neural Surface Reconstruction. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 20–22 June 2023.
17. Turki, H.; Ramanan, D.; Satyanarayanan, M. Mega-NeRF: Scalable Construction of Large-Scale NeRFs for Virtual Fly-Throughs. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 12922–12931.

Retrieved from <https://encyclopedia.pub/entry/history/show/109038>