3D reconstruction methods in industrial settings: a comparative study for COLMAP, NeRF and 3D Gaussian Splatting

Zeno Sambugaro[1,†], Lorenzo Orlandi[2,*,†] and Nicola Conci[3]

DISI, University of Trento, via Sommarive, 5, Povo, 38123, Italy

Abstract

3D rendering techniques have undergone a rapid evolution with the emergence of novel and advanced methodologies, redefining the boundaries of realism and computational efficiency. This study explores recent advancements in the field, comparing established approaches like photogrammetry with software such as COLMAP against the new frontiers opened by emerging view synthesis approaches like Neural Radiance Fields (NeRF), and 3D Gaussian Splatting. In this paper, we present a comprehensive comparison of the described methods tailored for industrial applications, where the data acquisition is generally conducted by human operators employing handheld devices.

Keywords

Photogrammetry, NeRF, Gaussian Splatting, 3D Reconstruction

1. Introduction

In recent years, the advancement of 3D reconstruction technologies has opened new avenues in the documentation and analysis of urban landscapes, such as working, industrial and archaeological sites. Among these, photogrammetry has long been established as the baseline for precise, high-resolution mapping and modeling. However, recent advent of Artificial Intelligence (AI) in the 3D field, thanks to the introduction of Neural Radiance Fields (NeRF) [1] and more recently 3D Gaussian Splatting [2] techniques presents a novel paradigm, potentially overcoming some of the inherent limitations faced by traditional methods. This paper aims to provide a comprehensive comparison between these cutting-edge techniques, focusing on their application in the industrial context of excavation sites. Excavation sites present unique challenges for 3D reconstruction due to their dynamic nature and intricate details. Operators data collection methods must adapt to ensure fidelity in reconstructing occluded regions. Integrating geo-spatial data with 3D reconstructions aids utility companies in locating subsurface infrastructure accurately. This enhances worksite planning, management, and reduces the risk of accidental damage during future excavation.

Moreover, the use of geo-referenced data in the 3D reconstruction enables the development of augmented reality (AR) technologies, thus offering an additional layer of information, enhancing operational safety and efficiency. By overlaying digital models onto the physical world, operators can gain real-time insights, further preventing the accidental severing of critical infrastructure.

This paper investigates the strengths and limitations of photogrammetry, NeRF, and 3D Gaussian Splatting in excavations, where geographical positioning data is essential. Utilizing datasets of images and precise coordinates, it aims to measure each method's efficacy, particularly where traditional photogrammetry is not accurate. Implicit methods like NeRF show promise in rendering complex scenes with diverse surface properties, while 3D Gaussian splatting provide very accurate estimation of the surfaces and fine structures, areas challenging for conventional methods.

NeRF-based methods have recently proven to be a valuable alternative to traditional photogrammetry in the field of image-based 3D reconstruction. This innovation is especially significant for the challenging scenarios of excavation sites, where the accuracy and detail of 3D models are crucial. This research is motivated by the potential of NeRF to enhance the precision and reliability of reconstructions in such complex scenario. By comparing NeRF with traditional photogrammetry across varied scenes, this study aims to comprehensively assess their performances in capturing intricate details, surface textures, and overall geometric accuracy. Our goal is to evaluate the suitability of NeRF techniques to be adopted in real-world applications with a particular focus on excavation sites.

Ital-IA 2024: 4th National Conference on Artificial Intelligence, organized by CINI, May 29-30, 2024, Naples, Italy

[*] Corresponding author.

[†] These authors contributed equally.

zeno.sambugaro@unitn.it (Z. Sambugaro); lorenzo.orlandi@unitn.it (L. Orlandi); nicola.conci@unitn.it (N. Conci)

0000-0002-4541-4528 (Z. Sambugaro); 0000-0002-8376-043X (L. Orlandi); 0000-0002-7858-0928 (N. Conci)

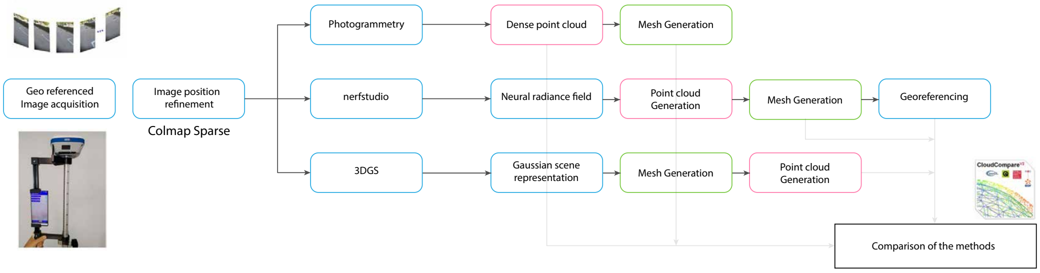

Figure 1: Overview of the proposed methodology

2. Background

3D reconstruction is crucial in fields like construction, excavation, and worksite management. Employing multi-view reconstruction techniques, scenes are captured from various angles using 2D images. This enables detailed monitoring of the project progress and provides the ability to virtually navigate sites, both during and after completion, utilizing geo-referencing and virtual reality. Among the most common photogrammetric solutions for 3D view reconstruction, we focus on COLMAP [3] for its open-access policy and continual improvements. COLMAP enables the conversion of 2D images into comprehensive 3D models, including point clouds and textured meshes, enabling advanced spatial analyses. However, the application of photogrammetric reconstruction encounters several challenges, particularly when dealing with objects characterized by complex optical properties such as high absorbency, reflectivity, or scattering. These methods can also suffer from variance in lighting conditions, including shadows, glare, or inconsistent illumination, as well as by surfaces with uniform or repetitive textures and complex shapes or geometries.

NeRF-based technologies offer cutting-edge solutions to overcome limitations in scene representation by resembling the scene with particles characterized by density and color. This study compares two neural radiance-based techniques, Nerfacto (a variation of InstantNGP [4] in Nerfstudio [5]) and SuGaR [6] (a variation of 3D Gaussian Splatting [2]), against traditional photogrammetry methods.

Neural Radiance Fields Neural Radiance Fields (NeRF) have emerged as a significant advancement in the field of 3D scene reconstruction. The scene is represented with a novel 5D function. This function correlates each spatial point (x, y, z) with the radiance emitted in any direction, defined by azimuthal and polar angles (θ, ϕ). The outcome, characterized by volume density σ and RGB color values c, varies with the viewing direction. This relation is formulated through the Multi-Layer Perceptron (MLP) F_θ, expressed as:

$$F_\theta : (\mathbf{x}, \mathbf{d}) \rightarrow (\mathbf{c}, \sigma) \tag{1}$$

where $\mathbf{x} = (x, y, z)$ denotes the coordinates within the scene, and $\mathbf{d}\,(\theta, \phi)$ represents the 3D Cartesian unit vector indicating the direction. The color $\mathbf{c} = (r, g, b)$ shifts with the viewing angle, while σ, denoting volume density, remains invariant. The usage of neural volume rendering pipelines, over traditional point clouds or meshes, enable the modeling of variations in color and illumination. InstantNGP [4], short for Instant Neural Graphics Primitives, is a variant that enhances NeRF's framework to expedite scene reconstruction significantly. By refining the neural network's architecture and computations, InstantNGP facilitates quicker achievement of high-quality results, positioning it as a viable option for real-time applications.

NeRFStudio introduces an innovative platform, leveraging the Nerfacto model, to streamline NeRF-based model creation and manipulation. Nerfacto integrates insights from very recent research, including MipNeRF-360 [8], Instant-NGP [4], and Ref-NeRF [7], focusing on optimizing camera views and sampling processes.

3D Gaussian Splatting for Real-Time Radiance Field Rendering 3D Gaussian Splatting [2], a novel approach to scene representation, contrasts with neural fields by optimizing an explicit point-based scene model. Each point in this representation is associated with various attributes: a position $p \in \mathbb{R}^3$, opacity $o \in [0, 1]$, third-degree spherical harmonics (SH) coefficients $k \in \mathbb{R}^{16}$, 3D scale $s \in \mathbb{R}^3$, and 3D rotation $R \in SO(3)$ represented by 4D quaternions $q \in \mathbb{R}^4$. Rendering to the image plane involves accumulating the color c_{GS} from correctly-sorted points using the equation:

$$c_{GS} = \sum_{j=1}^{N_p} c_j \alpha_j \tau_i \quad \text{where} \quad \tau_i = \prod_{i=1}^{j-1} (1 - \alpha_i) \tag{2}$$

with c_j determined by SH coefficients k and α_j calculated from the projected 2D Gaussian with covariance $\Sigma' = JM\Sigma M^T J^T$, incorporating per-point opacity o, viewing transformation M, and Jacobian J of the affine approximation of the projective transformation. The 3D covariance matrix Σ ensures positive semi-definiteness through the scale matrix $S = \text{diag}(s_1, s_2, s_3)$ and rotation R, following $\Sigma = RSS^T R^T$.

Building upon the principles of 3D Gaussian Splatting, Surface Gaussian Approximation for Rendering (SuGaR) [6] leverages Gaussian functions to model object surfaces within a scene, achieving precision in handling occlusions and detailed surface texturing through Gaussian "splats" projected onto a volume grid. Each splat influences the volume's density and color, based on its spatial location and Gaussian distribution, described mathematically as:

$$G(\mathbf{x};,\Sigma) = \frac{1}{(2\pi)^{\frac{3}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x}-)^T\Sigma^{-1}(\mathbf{x}-)\right) \quad (3)$$

In Eq. (3) $\mathbf{x}$ denotes a point in space, the mean location (center of the splat), and Σ the covariance matrix shaping the Gaussian distribution. SuGaR's method for accumulating multiple splats across a scene constructs a volumetric representation capturing density and color information, enabling a precise shading and depth rendering.

3. Methodology

This study aims to evaluate the effectiveness and potential benefits of Neural Radiance Fields (NeRF) against traditional image-based reconstruction techniques, particularly photogrammetry, in the context of augmented/virtual reality applications. Our focus is on challenging outdoor scenarios. We include excavation sites and playground objects, which are characterized by unbounded environments and non-Lambertian surfaces. To facilitate a direct comparison, the same dataset of images, captured with geo-referencing, is utilized across all reconstruction methods. This standardized approach ensures that differences in the reconstruction quality and efficiency can be attributed solely to the methodologies rather than due to a bad alignment.

3.1. Dataset acquisition

The datasets are collected using a system comprised of two devices: a smartphone and an RTK-GNSS spatially calibrated as can be seen from [9]. These devices ensure highly accurate pose information for all the collected scenes. The study aims to analyze industrial applications, therefore, as scenarios, we have selected some simple playground games mixed with real excavation scenarios where the reconstruction is more challenging. Our datasets consist of 7 playground scenarios and 3 excavation scenarios.

Acquisition Process. The dataset has been acquired following the standard procedure that an operator would follow when working in a given site. The trajectory reflects a rotation around the object, maintaining the capture at eye level. During acquisition, the frame rate is set at 5 frames per second with a resolution of 1280 x 720. The accuracy of the geopose data is always less than 3 cm in traslation and less than 1 degree for each acquired image. We maintain a uniform velocity during acquisition, so that the number of images for each scenario depends on the length of the trajectory. The playground dataset comprises approximately 200 images, while the excavation dataset contains around 500 images, which reflects longer trajectories.

3.2. Methodologies Employed

Three distinct reconstruction methodologies were applied to the captured datasets; an overview is shown in Figure 1:

1. **Photogrammetry:** The classical photogrammetric procedure involves estimating camera orientation parameters for sparse point cloud construction, generating a dense point cloud; mesh creation and texture extraction complete the reconstruction process. For this purpose we used COLMAP, with all phases conducted in high-quality mode to ensure maximum detail and accuracy.
2. **NeRF-Based Reconstruction:** The training of Neural Radiance Field reconstruction requires known camera poses as input. We use *nerfstudio* [5], and in particular "nerfacto", a model strongly based on InstantNGP [4], used for its fast training and inference. We then extract the dense point clouds and textured mesh from nerfstudio's API; in particular, for mesh extraction we exploited Poisson reconstruction.
3. **Gaussian Splatting (SuGaR):** Similarly to NeRF this method requires known camera poses as input. This explicit model is then trained to approximate the radiance field of the scene. The training of SuGaR involves more than one step. The training starts with 7k iterations of normal 3D Gaussian Splatting and 7k iterations of SuGaR finetuning to extract a more precise geometry.

The acquisition of our dataset incorporated georeferencing, so as to simplify the alignment process for the reconstructions. The only exception is NeRF, as an

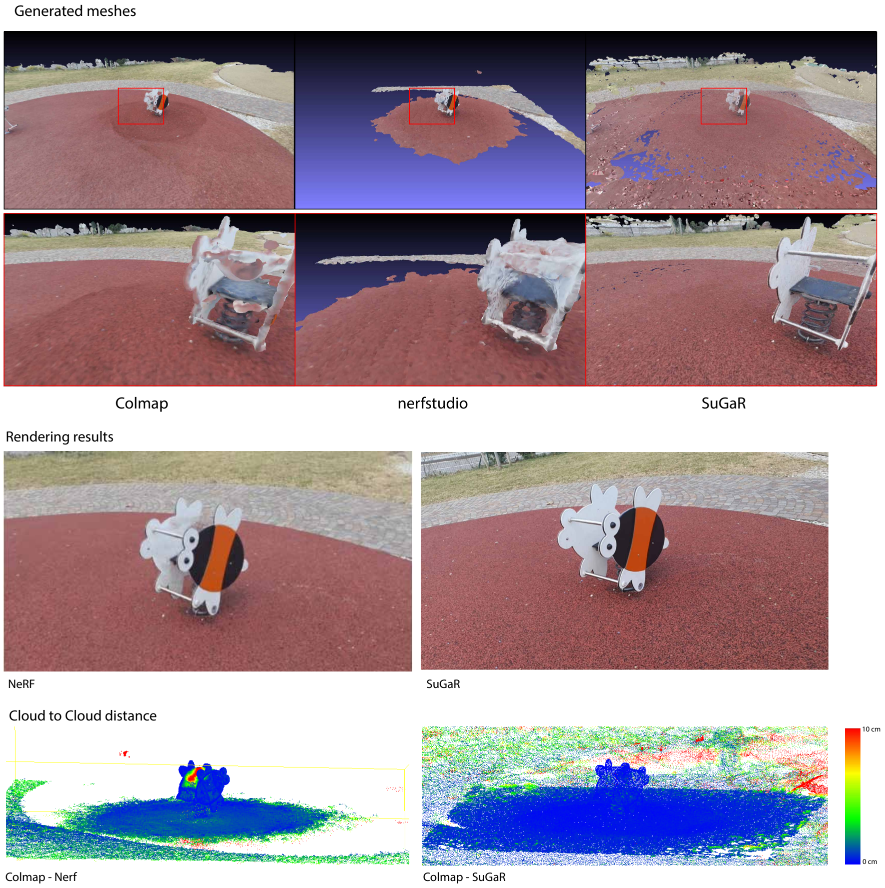

Figure 2: Comparison of the mesh obtained with the proposed methodologies on the playgrounds dataset. Other scenes can be found at: **https://zenos4mbu.github.io/photogrammetry_nerf.github.io/**

implicit framework this model normalizes its coordinates between -1 and 1. This aspect of NeRF requires an additional step to calibrate the model, to incorporate scale and translation derived from the geo-referenced input to ensure accurate alignment. For the dataset to be used in training, we first need to estimate the camera parameters from the input images. This estimation is necessary because the neural network requires knowledge of both the camera's positions and the corresponding images to accurately generate the scene representation. To achieve this, we utilized COLMAP, a known software for its application of Structure from Motion (SfM) techniques [3], for estimating three-dimensional structures from two-dimensional image sequences.

To facilitate comparison, given that outputs from photogrammetry are not directly comparable with those from neural fields or Gaussian splatting, we incorporate an additional conversion phase. NeRFstudio provides functionality to convert NeRF outputs into point clouds and meshes. The point clouds are easily exported since the neural representation can be inspected at any 3D point. For the meshes this conversion employs the *marching cubes* algorithm and the Poisson surface reconstruction method. In the SuGaR framework the mesh extraction phase it also done through *marching cubes* or Possian surface reconstruction. In this case the reconstruction is enhanced thanks to the precise estimation of the normals of the sampled points. To obtain an accuracy metric we derive a cloud-to-cloud comparison using the CloudCompare software.

3.3. Comparative Analysis Framework

The comparative analysis between these methods focuses on the following key metrics: (i) **Accuracy and Detail Resolution**, to evaluate the fidelity of the reconstructed models to the original scenes, and (ii) **Processing Time**, to assess the efficiency of each methodology in terms

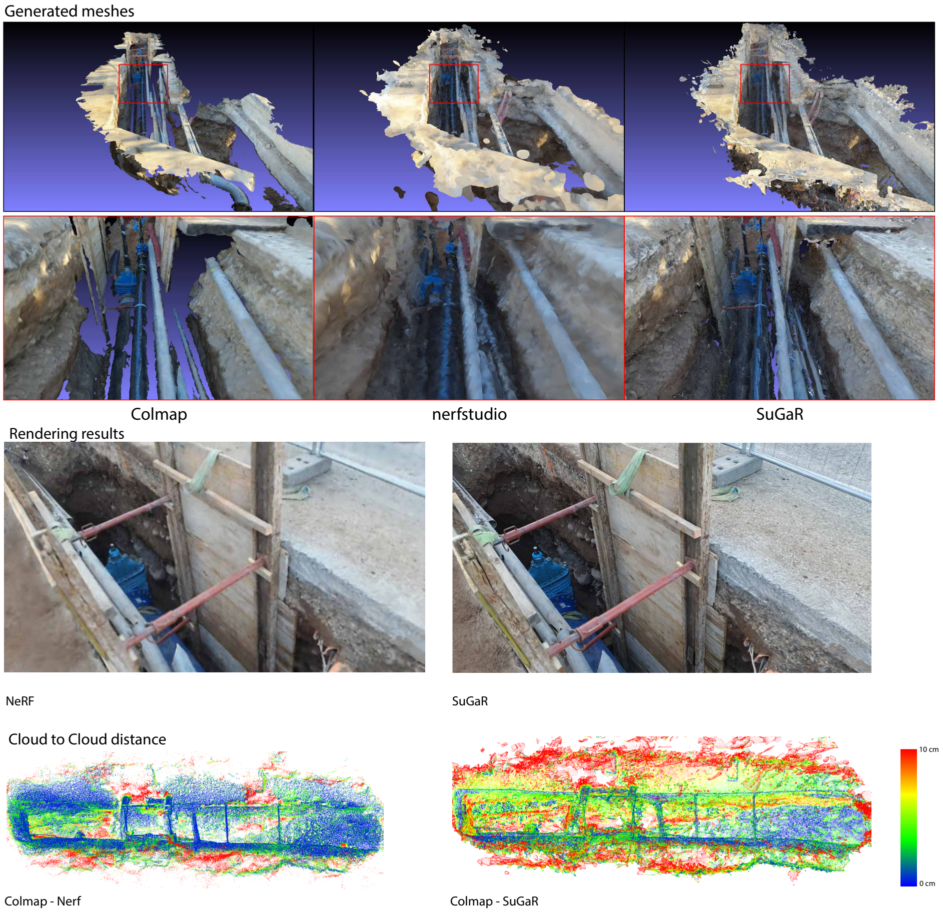

Figure 3: Comparison of the cloud to cloud distance of the proposed methodologies on the excavation sites dataset. Other scenes can be found at: **https://zenos4mbu.github.io/photogrammetry_nerf.github.io/**

of computational resources and time required for reconstruction. To compare the level of fidelity of the reconstructed models we propose using the point clouds generated by the studied methods. In this way we can obtain a quantitative metric. To be more specific we measure the cloud to cloud deviation of the methods based on radiance fields with respect to the reconstruction using classical photogrammetry. This measure is an absolute value, which doesn't tell which method is performing better; it only informs about the deviation from one reconstruction to the other. Therefore, we also show the rendering results in order to see the performances in graphical terms, in Figure 3. In addition to this quantitative result we also propose a qualitative comparison of the resulting meshes, comparing the proposed methodologies in Figure 2.

4. Discussion

We show a comparison of NeRF-based techniques against traditional photogrammetry utilizing COLMAP. All models are trained on an NVIDIA RTX 3090 GPU. The assessment focuses on their effectiveness in view synthesis and 3D reconstruction, particularly in expansive, unbounded environments. The results of our analysis highlights that the three methodologies produce high quality point clouds, with very close results especially in the fine structures of the 3D scene, as illustrated in in Figure 2. Notably, NeRF's output shows a denser point cloud around high-frequency scene features but has gaps in smoother regions. The radiance field rendering results show that the quality of the reconstructed views is really high and is very difficult to say if nerfstudio or SuGaR presents the best result. However, the comparison illustrated in Figure 2 highlights a failure case of nerfstudio, with a red area within the scene's object of interest indicating

a high cloud-to-cloud distance. This issue not only produce a discrepancy in the point cloud representation but also results in blurring within the targeted region of the neural reconstruction.
Considering the extensive usage of meshes in VR and AR applications, for their simplicity and low memory footprint, we present a comparison of the meshes produced with the three methodologies. In Figure 2 we show the obtained meshes also showing a detail of the reconstruction in the region of the 3D scene with finer details. As depicted in Figure 2, there's a noticeable variance in detail and texture among the outputs. The COLMAP mesh, while being consistent, falls short on representing thin structures. In contrast, the NeRF mesh shows greater detail but presents some holes. The SuGaR mesh stands out for its superior detail, accurately capturing complex structures where others falter, thanks to its precise normal calculations. Another point to consider is the difference in accuracy between the two scenarios we have examined. The playground scene is easier and, in fact, has better results compared to the case of excavations. The complexity of the excavation scenario reduces the performance in reconstruction, especially with the SuGaR and NeRF method. It is noticeable in the figure 3 that there are many artifacts on the road surface visible on the Cloud to Cloud distance analysis, especially in the case of SuGaR, and there are also many holes, especially in the excavation bottom. Finally, we analyze the processing time for each method. Regarding this aspect, there is no difference between SuGaR and COLMAP. Instead, the best performance is observed with InstantNGP, which takes about a quarter of the time compared to the other methods. Additional materials regarding to our analysis, they can be accessed through this link [1].

5. Conclusions

In this paper we provide a comparative analysis of Neural radiance fields based reconstruction methods and classical photogrammetry for unbounded scenarios. We show results in playgrounds and excavations sites, to access the performances in easy and complex scenarios. In our set-up, photogrammetry has provided superior reliability in complex scenes, especially on the excavation sites. Proving also better results in modeling completely flat area which in the NeRF methods presents some artifacts. Although training/reconstruction times are generally not the main concern in the reconstruction of working areas, some application might benefit from fast reconstruction times. In this aspect nerfstudio provided the best speed in the reconstruction, requiring just 15 minutes for the training of a scene. An important aspect that needs to be analyzed is the reliance of the current rendering pipelines for virtual and augmented reality on meshes representations. This advantages the classical photogrammetry since its final goal is to obtain a mesh representation. In contrast, neural rendering technologies focus primarily on view synthesis, offering an alternative that eliminates the need for mesh generation. SuGaR and more in general 3D Gaussian Splatting techniques produce an explicit representation that allow for the splatting of Gaussians in the same way traditional methods splat triangle. This feature enable SuGaR to render the scene in real time, making it possible to use it into existing pipelines. In the future, we see 3D Gaussian Splatting to be a potential replacement for for meshes representations, especially in scenarios requiring the realistic reconstruction of complex environment.

This research is supported by the project DIMOTY, funded by the Autonomous Province of Trento under the LP6/99 framework

References

[1] Mildenhall, Ben, et al. "Nerf: Representing scenes as neural radiance fields for view synthesis." Communications of the ACM 65.1 (2021): 99-106.

[2] Kerbl, Bernhard, et al. "3d gaussian splatting for real-time radiance field rendering." ACM Transactions on Graphics 42.4 (2023): 1-14.

[3] Schonberger et al. "Structure-from-motion revisited." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[4] Müller et al. "Instant neural graphics primitives with a multiresolution hash encoding." ACM transactions on graphics (TOG) 41.4 (2022): 1-15.

[5] Tancik et al. "Nerfstudio: A Modular Framework for Neural Radiance Field Development." ACM SIGGRAPH 2023.

[6] Chen et al., C. "SuGaR: Pre-training 3D Visual Representations for Robotics." arXiv preprint arXiv:2404.01491, 2024.

[7] Verbin et al. "Ref-nerf: Structured view-dependent appearance for neural radiance fields." 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

[8] Barron et al.. "Mip-nerf 360: Unbounded anti-aliased neural radiance fields." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5470-5479, 2022.

[9] Lorenzo O., Kevin D., et al. "Spatial-Temporal Calibration for Outdoor Location-Based Augmented Reality'. IEEE Sensor Journal (2024): "accepted for publication".

[1] **https://zenos4mbu.github.io/photogrammetry_nerf.github.io/**