Chapter 9
Neural Radiance Fields for Low-cost, High-resolution 3D Scanning of Critical Infrastructures

Thiago F. Ribeiro, Felipe S. Brandão, Victor H. R. Cardoso, João C. W. A. Costa, Túlio N. Bittencourt, and Moisés F. Silva

Abstract 3D reconstruction of civil and mechanical structures is vital for structural assessment and health management. Often, 3D scan data are integrated into finite element analysis (FEA) to refine geometry, evaluate models, and measure displacement, stress, and loading effects. This process offers a cost-effective way to capture surface geometry and create a digital structure representation, with accuracy heavily dependent on the 3D scanning technology used. Terrestrial laser scanning is frequently employed for non-contact estimation of accurate, high-fidelity geometry data to predict a structure's remaining service life. However, high costs and budget limitations restrict its use in many applications. Multi-View Stereo (MVS) based photogrammetry has many advantages, such as high resolution and the ability to capture complex surfaces. However, it also has drawbacks such as heavy computational loads, high sensitivity to conditions and image quality, and the need for specialized software. The photogrammetric process can be time-consuming, especially with many images. Moreover, these methods generate large data volumes, which can be prohibitive for continuous data acquisition in real-time applications. In counterpart, Neural Radiance Fields (NeRF) based 3D reconstruction models are capable of producing photo-realistic renderings of complex geometries with significantly reduced memory overhead and storage requirements, as it creates an implicit 3D representation of the structure (i.e., the geometry is implicitly embedded into the weights of a neural network). This can be achieved by combining recent advances in inverse neural rendering and volumetric representation of scenes to create a generative model inherently capable of representing the 3D light field from a scene from a small and sparse set of images. In this work, we demonstrate the capabilities of using NeRF-based point cloud extraction to evaluate critical elements of civil infrastructures from sparse sets of smartphone images. We compare this to traditional laser scanning and photogrammetry techniques. We propose a non-contact workflow aimed at providing rapid, extremely low-cost, high-fidelity, and accurate 3D scanning for fast assessment. Our findings show the potential of NeRF-estimated dense point clouds as a viable alternative to MVS photogrammetry and terrestrial laser scanning.This approach offers an efficient, cost-effective method for precise measurement and analysis in engineering and scientific contexts.

Keywords Neural radiance fields · 3D scanning · 3D Point clouds · Volumetric rendering · Building information modeling

Introduction

Extracting geometric data from civil and mechanical structures is fundamental for accurate structural assessment and modeling. Whether it is applied to design, numerical simulations, or operational inspections, 3D scanning and measurement technologies are indispensable tools across the architecture, engineering, and construction (AEC) sectors. With the ongoing

Thiago F. Ribeiro · Victor H. R. Cardoso · João C. W. A. Costa
Applied Electromagnetism Laboratory, Federal University of Pará, Belém, PA, Brazil
e-mail: thiago.figueiro.ribeiro@gmail.com; victorcard@ufpa.br; jweyl@ufpa.br

Felipe S. Brandão · Túlio N. Bittencourt
Department of Structural and Geotechnical Engineering, University of São Paulo, São Paulo 05508-900, Brazil
e-mail: fbrandao2k@usp.br; tbitten@usp.br

Moisés F. Silv
Los Alamos National Laboratory, Los Alamos, New Mexico 87544, USA
e-mail: cien.felipe@gmail.com

Janko Slavic et al. (eds.), *Computer Vision & Laser Vibrometry, Vol. 6*, Conference Proceedings
of the Society for Experimental Mechanics Series, https://doi.org/10.13052/97887-438-0151-1_9

deterioration of key infrastructures due to the escalating effects of climate change, there is an increasing need for structural assessments that are rapid, efficient, and cost-effective [1]. These assessments are vital not only to ensure safety but also to optimize maintenance efforts and prolong the lifespan of infrastructure assets.

The rapid evolution of non-contact sensing technologies for structural health monitoring (SHM) has expanded the use of cameras, smartphones, unmanned aerial vehicles (UAVs), satellites, ultrasonic devices, and other imaging sensors for conducting structural assessments [2]. A significant number of these technologies rely on the generation of 3D point clouds, which are collections of spatial data points that describe the shape and surface characteristics of structures [3, 4]. The versatility of point cloud data extends far beyond visual representation; it allows engineers to derive 3D models that highlight discrepancies between projected models and as-built structures, facilitating critical evaluations of construction quality, structural deformation, and deviations from original design specifications [5]. Additionally, point cloud data has become essential for monitoring construction progress, assessing geometry quality, and detecting early-stage damage, such as cracks, in buildings and civil infrastructures.

Among the numerous techniques for non-contact geometry extraction, terrestrial laser scanning (TLS) stands out as one of the most widely used. It is predominantly based on Light Detection and Ranging (LiDAR) technology, which uses pulses of laser light to measure distances between the scanner and surrounding objects. LiDAR operates similarly to Radar (Radio Detection and Ranging), but instead of using radio waves, it relies on discrete pulses of laser light [6]. This allows for high-precision distance measurements, which are crucial for producing accurate 3D models of complex structures. However, the advanced technological complexity of LiDAR systems, which includes high-quality optical components and precise rotational mechanisms, results in considerable costs. The Ouster OS1-64 used by AARON et al. (2023) [7] retails at €17,424.00 as of September 2024. Consumer-grade LiDARs such as the Livox Mid-40, which retails for $599 as of September 2024, trade higher accuracy for lower prices as noted by ARTEAGA et al. (2019) [8].

Despite their effectiveness, the high cost and technical challenges associated with LiDAR limit its accessibility for routine inspections, particularly for smaller projects or organizations with limited budgets. To address these cost-related barriers, recent research has explored alternative methods for 3D geometry extraction. One promising approach involves the use of Neural Radiance Fields (NeRF), a machine learning-based technique that generates high-fidelity 3D models from sparse sets of images, such as those captured by standard consumer-grade cameras, including smartphones. REMONDINO et al. (2023) [9] has shown a generalized capacity of 3D reconstruction using NeRF. HUANG et al. (2024) [10] recently demonstrated the successful application of NeRF in reconstructing individual trees in 3D with remarkable accuracy. Although NeRF has primarily been applied in fields like computer graphics and environmental modeling, its potential for civil and mechanical infrastructure assessment remains largely unexplored. As far as the authors knowledge goes this is the first study and application of NeRFs with a focus on civil infrastructures. We aim to demonstrate that NeRF, trained on a limited set of photos taken with inexpensive cameras, can generate 3D models that rival those produced by more expensive and complex systems like LiDAR and photogrammetry. By conducting a detailed comparison with alternative methods, such as stabilished LiDAR systems and traditional photogrammetry, we seek to provide a comprehensive evaluation of NeRF's viability for large-scale structural assessments. The ultimate goal is to offer a more accessible, scalable, and accurate approach to infrastructure monitoring, one that addresses the growing demands of the AEC industry while reducing costs and technological barriers.

3d Reconstruction Techniques

Lidar

Laser-based ranging, profiling, and scanning have been subjects of research since the 1960s. Two primary techniques are employed to measure the distance to an object (Figure 1). The first method involves precisely measuring the Time-of-Flight (ToF) of a short but intense laser pulse emitted from the instrument, which is then reflected back from the object's surface. Given the precise knowledge of the speed of light, the accuracy or resolution of the distance measurement is determined by the precision of the time measurement [11]. In the second method, distance is determined by accurately measuring the phase difference between the emitted signal and the signal received after reflecting off the object. Modern LiDAR systems scan objects by sampling a series of measured distances, capturing both vertical and azimuthal angles. Each sampled point is digitally recorded and stored, creating a detailed 3D representation of the scene.

Photogrammetry

Photogrammetry is a technique that involves creating a three-dimensional model of an object, structure, or environment by analyzing two-dimensional images taken from multiple angles. The process starts with detecting and matching features

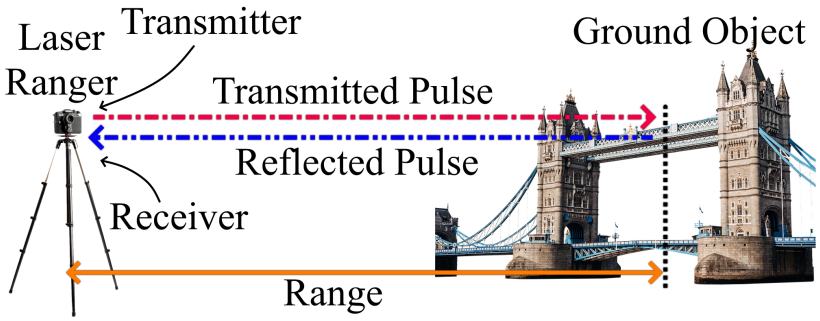

Fig. 1 Operation of a laser rangefinder that is using the timed pulse or TOF method. Adapted from SHAN et al. (2018) [11]

between the images using algorithms like Scale Invariant Feature Transform (SIFT) [12] (Figure 2). Next, Structure-from-Motion (SfM) techniques [13] are applied to estimate the camera's position and orientation for each image. Using these camera poses, the 3D coordinates of the matched feature points are determined through triangulation, producing an initial sparse point cloud. To improve the alignment of point clouds, the Iterative Closest Point (ICP) algorithm [14] is often used. Afterward, Multi-View Stereo (MVS) techniques [15] are applied to produce a dense point cloud. For each photo, a depth map is generated by triangulating matching features from multiple images, which represents the distance of each pixel from the camera. These depth maps are subsequently aligned and merged. Since every pixel in the depth map corresponds to a 3D point, and its distance to the camera is known, this approach effectively generates a dense 3D point cloud.

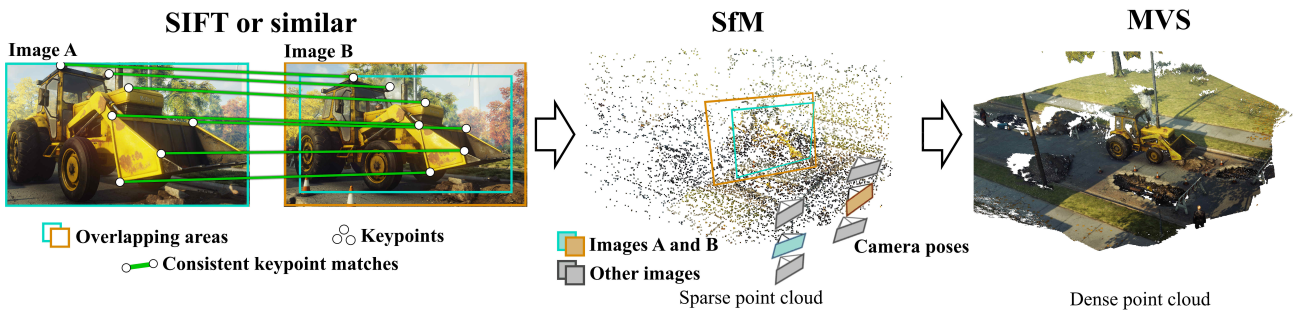

Fig. 2 The three key stages in a SfM-MVS workflow illustrated on two hypothetical images of a construction truck: (1) keypoint identification and matching (e.g. SIFT), (2) SfM with camera parameters and a sparse point cloud as output and (3) the densified point cloud following MVS. Adapted from IGLHAUT et al. (2018) [16]

NeRF

In contrast to the conventional LiDAR-based approaches, NeRFs are considered an implicit representation, as it stores the 3D lightfield of a given scene into the weights of a neural network [17, 18]. In this method, a static scene is represented as a continuous 5D function that outputs the radiance emitted in each direction, (θ, ϕ), and each point in the 3D coordinate system, (x, y, z). Ray samples are created from each pixel using a classical pinhole camera model using conventional ray tracing. Thus, the overall goal is, for each ray sample, to learn an RGB color and a density value σ at the 3D location, acting as a differential opacity, controlling how much radiance is accumulated by a ray passing through that point (x, y, z). Later, volume rendering algorithms can be applied over all samples to reconstruct the final pixel color of an image rendered from a given camera angle [19, 20]. The NeRF architecture first processes the input with a multilayer perceptron (MLP) with 9 fully connected layers (ReLU activated, 256 neurons each) with skip connections implemented in the fifth and ninth layers where positional information [21] is integrated (Figure 4). The ninth layer outputs σ. A tenth, 128 neurons-wide, ReLU layer processes the remaining 256-dimensional feature vector. A final Sigmoid activated layer outputs the emitted RGB radiance at position x, as viewed by a ray with direction d [19].

NeRFs perform the reconstruction of color and opacity of ray samples using an MLP to represent the lightfield of the scene, which regresses a 5D coordinate-space (x, y, z, θ, ϕ) to the appropriate σ and view-dependent color $C = (r, g, b)$,

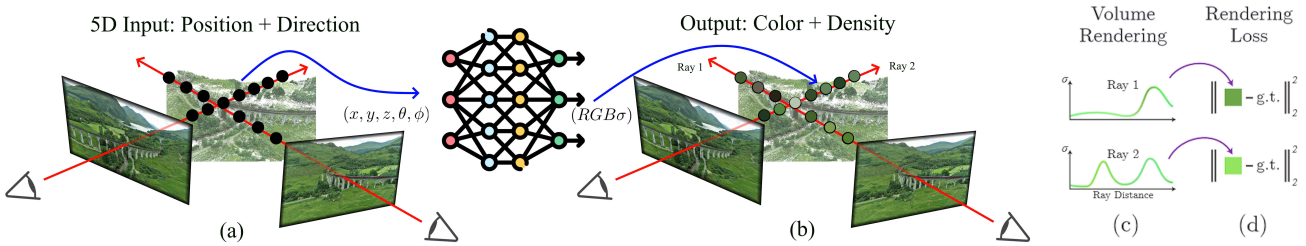

Fig. 3 An overview of the neural radiance field scene representation and differentiable rendering procedure. Images are synthesized by sampling 5D coordinates (location and viewing direction) along camera rays (a), feeding those locations into an MLP to produce a color and volume density (b), and using volume rendering techniques to composite these values into an image (c). This rendering function is differentiable, so optimization of the scene can be done by representation by minimizing the residual between synthesized and ground truth observed images (d). Adapted from MILDENHALL et al. (2020) [19]

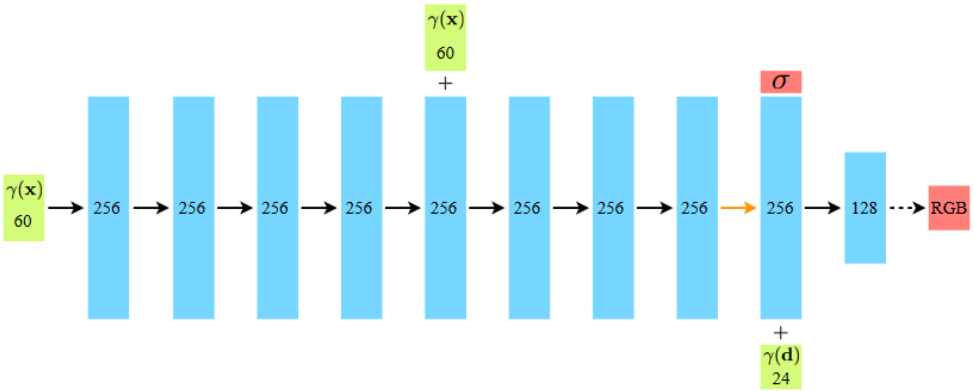

Fig. 4 Visualization of the NeRF's fully-connected network architecture. Adapted from MILDENHALL et al. (2020) [19]

$$C \approx \sum_{i=1}^{N} T_i \alpha_i c_i, \tag{1}$$

where N represents the number of samples along a given ray, T_i is the accumulated transmittance from the camera to the i-th sample point, α_i represents the opacity at the i-th point, which is related to the volume density σ by $\alpha_i = 1 - \exp(-\sigma_i \delta_i)$, and c_i represents the color of the i-th point, expressed as an RGB vector. In this context, a high volume density σ_i will increase the opacity α_i, making the color c_i of that point more significant in the final rendered color C.

The formula states that the accumulated density between the near bound t_n and t will decrease the importance of whatever lies between t and the far bound. The function $T(t)$ denotes the accumulated transmittance along the ray from t_n to t, i.e., the probability that the ray travels from t_n to t without hitting any other particle [19]. In other words, the opaquer surfaces between the near bound and position t, the less light will go through that space, and the lower the accumulated transmittance be at position t are measured. Therefore, the resulting color for this ray on the image, C, is a linear integral over the scene's domain of the color function weighted by the density and accumulated transmittance [22]. The loss function of the NeRF is the total squared error between the rendered and true pixel colors for both the coarse ($N = 64$) and fine ($N = 192$) renderings:

$$L = \sum_{r \in R} \left[\|\hat{C}_c(r) - C(r)\|_2^2 + \|\hat{C}_f(r) - C(r)\|_2^2 \right]. \tag{2}$$

NeRFs are trained on batches of rays rather than entire images. R represents the set of rays in each training batch. For each ray r in a batch, $C(r)$ denotes the ground truth RGB color, while $\hat{C}_c(r)$ and $\hat{C}_f(r)$ represent the predicted RGB colors from the coarse and fine models, respectively. The training aims minimizing the difference between these predicted colors and the ground truth, ensuring that both coarse and fine models learn to accurately reconstruct the scene. The input data for the training of a NeRF must include the spatial location of each camera, a 5D coordinate (x, y, z, θ, ϕ) tensor. In computer-generated images, this process is simpler since you already have this information in order to generate the image. In contrast,

to generate the input data for real-world photos, we use Structure from Motion (SfM) algorithms [13]. They use triangulation techniques to estimate the three-dimensional coordinates and angles from each input photo.

Performance Metrics

Cloud to Cloud Comparison

When comparing distances between cloud points, different techniques can be used, such as nearest neighbor and nearest neighbor with local modeling. The first measures the distance between a 3d point to the nearest euclidean point of the other cloud. However, since point clouds are discrete models, this distance can be inaccurate when compared to the true continuous reality. This problem can be mitigated when using local modeling. Instead of using a single point and measuring distance, the K closest points (K Nearest Neighbors or KNN) are selected and a spatial plane is calculated from these points using linear regression. Then, the height of this plane and the initial point is measured. This height will be the distance using KNN. This technique gives a continuity to the compared model, but introduces errors if the model is indeed discontinuous in a region. Similarly to Elkhrachy (2020) [23], this paper considers the Mean value and Root Mean Square (RMS) of all differences to assess the accuracy of the model, while the Standard Deviation (Std. Dev.) was used as an indicator of data noise.

Level of Accuracy (LOA)

The Level of Accuracy (LOA) is a scale that is used to determine how precise a model should be to the point cloud [24]. The LOA scale was developed by the United States Institute of Building Documentation (USIBD) to standardize specifications industry-wide. The LOA standard is structured in five increments of ten, beginning with LOA10 (lower precision) and extending through LOA50 (higher precision). We will examine LOA30 (with an error margin of up to 15 mm), LOA40 (up to 5 mm), and LOA50 (up to 1 mm) for the smaller structure (USP Bridge). Since the distance from the camera to the Octavio Frias Bridge is approximately 100 times greater than that to the USP Bridge, we anticipate that the LoA will be affected by the same factor. Therefore, we will use LOA30 x 100 (with an error margin of up to 1500 mm), LOA40 x 100 (up to 500 mm), and LOA50 x 100 (up to 100 mm) to evaluate the errors between the reconstruction models of this bridge and the ground truth.

Objects of Study

The USP Footbridge

The University of São Paulo (USP) Bridge is a footbridge over the Tejo stream, located between the Mining Engineering and Civil Engineering buildings on the USP-SP campus, São Paulo, Brazil. The bridge has a deck with a span of approximately 12m made up of cross-section timber pieces measuring 7 cm x 16 cm, transversely prestressed, covered with reinforced concrete, and a prestressing system consisting of a metal diverter that rests under the deck in the middle of the span and tendons that in turn fix the diverter and are anchored at the ends of the deck. The superstructure is supported by reinforced concrete blocks that were recovered from a pre-existing structure on the site, and at the access on the Minas side there is a precast reinforced concrete slab approximately 2 meters long allocated to span the gap between the support pillars and the abutment [25].

The Cable-stayed Octavio Frias Bridge

The Octavio Frias de Oliveira Cable-stayed Bridge is in the Rela Parque Roadway Complex and cross the Pinheiros River at the end of Jornalist Robert Marinho Avenue, in São Paulo, Brazil [26]. It is constituted by two curved ramps giving direct access between the Journalist Roberto Marinho Avenue and the expressway of the Nações Unidas Avenue. The cable-stayed decks present a curvature in plan with a constant radius, with 140-meter spans to cross the Nacoes Unidas Avenue and the CPTM trains, and 150-meter spans across the Pinheiros River. Inaugurated in May 2008, it stands out for its unique ”X” shape, with two curved decks that intersect around a single concrete tower standing 138 meters high. The bridge is named after Octavio Frias de Oliveira, a prominent Brazilian businessman and former owner of the newspaper Folha de S.Paulo. This structure stands out for its innovative design, which serves not only as a crucial connection between Marginal Pinheiros

Fig. 5 The USP Bridge over the Tejo stream, São Paulo, Brazil

Fig. 6 The Octavio Frias de Oliveira Cable-stayed Bridge crossing the Pinheiros River

and the surrounding regions but also as a symbol of modernity and architectural excellence in São Paulo. At night, the bridge is often illuminated in vibrant colors, making it a key landmark in the city's skyline.

Experiments

We captured 348 photos of the USP Bridge from various azimuths and used them as input for photogrammetry and NeRF-based point cloud generation. All photos were taken at a resolution of 960 x 1280 px. Figure 7 shows the camera angles for each image. The resulting point clouds were compared to a LiDAR-based point cloud generated by a Leica RTC360, a high-performance 3D laser scanner widely employed in industries such as construction and architecture. The RTC360 point cloud was considered the ground truth, providing a reliable benchmark for accuracy and precision in the comparison. This comparison was performed using K-Nearest Neighbors (K = 6) local modeling. To investigate the behavior of the errors between our models and the ground truth, we repeated the experiment with progressively fewer photos, generating point

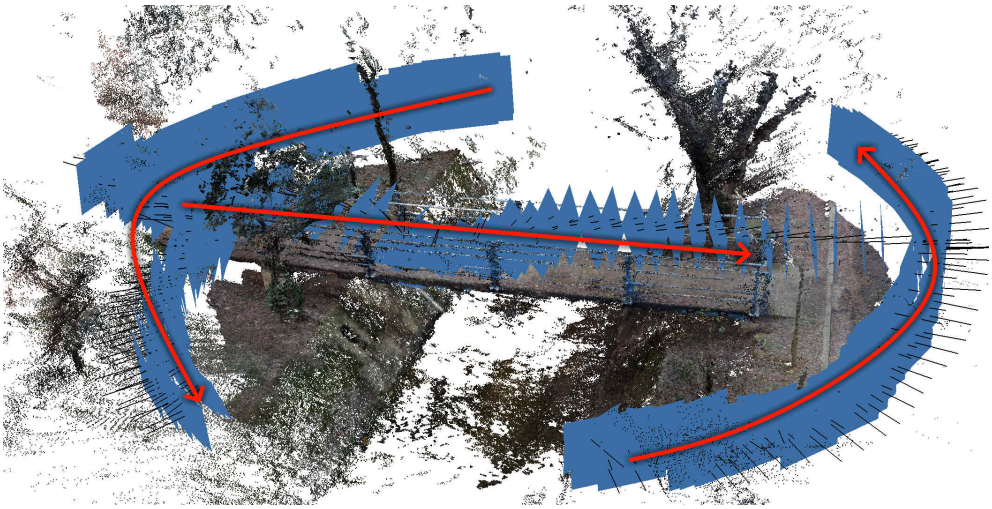

Fig. 7 Camera angles and positions of the photos taken over the USP Bridge.

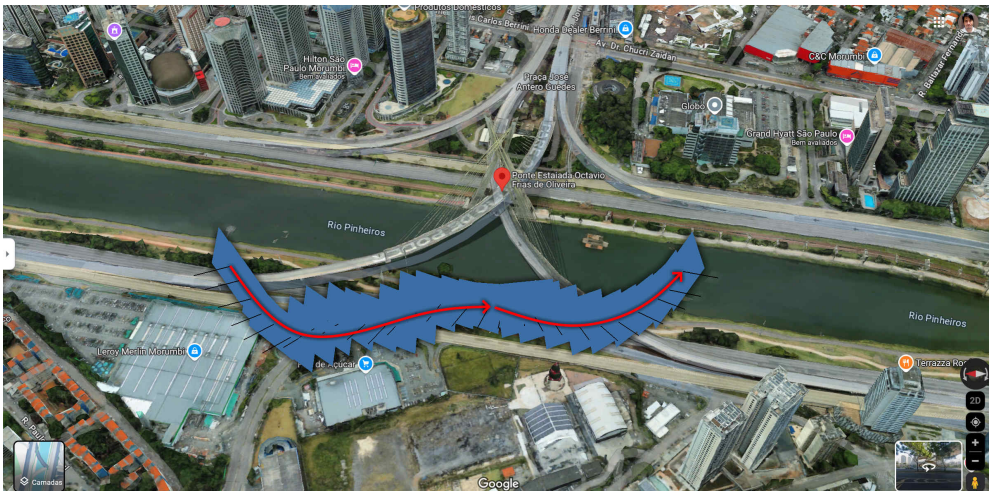

Fig. 8 Camera angles of the photos taken by the drone over the Octavio Frias Bridge. Source: Google Maps

clouds from 75% (261 photos), 50% (174 photos), 25% (87 photos), 10% (35 photos), and 5% (17 photos) of the original dataset.

For the Octavio Frias Bridge, we extracted 500 equally spaced frames from a YouTube video. Figure 8 shows the camera angles for each frame. The frames, originally in UHD 4K resolution, were downsampled to Full HD (1080p) and used as input for photogrammetry and NeRF-based point cloud generation. These models were then compared to the LiDAR-based point cloud, which was used as the ground truth. Additionally, we created point cloud models using progressively fewer frames, ranging from 100% of the original 500 frames down to 2.5%, to compare their accuracy against the ground truth.

The distance-errors between models and the LiDAR-Ground Truth are presented as color-coded point clouds, where the color bar indicates the magnitude of error from blue (lower error) to red (higher error).

Results

The USP Footbridge

Figure 9 (B) shows a comparison between NeRF and photogrammetry models of the USP Bridge. NeRF manages to reconstruct the bridge with high accuracy from 100% down to 25%, preserving structural and visual details. 25% some degradation starts occurring, especially in the handrails. The 10% model shows reduction on the density of the model, with gaps forming

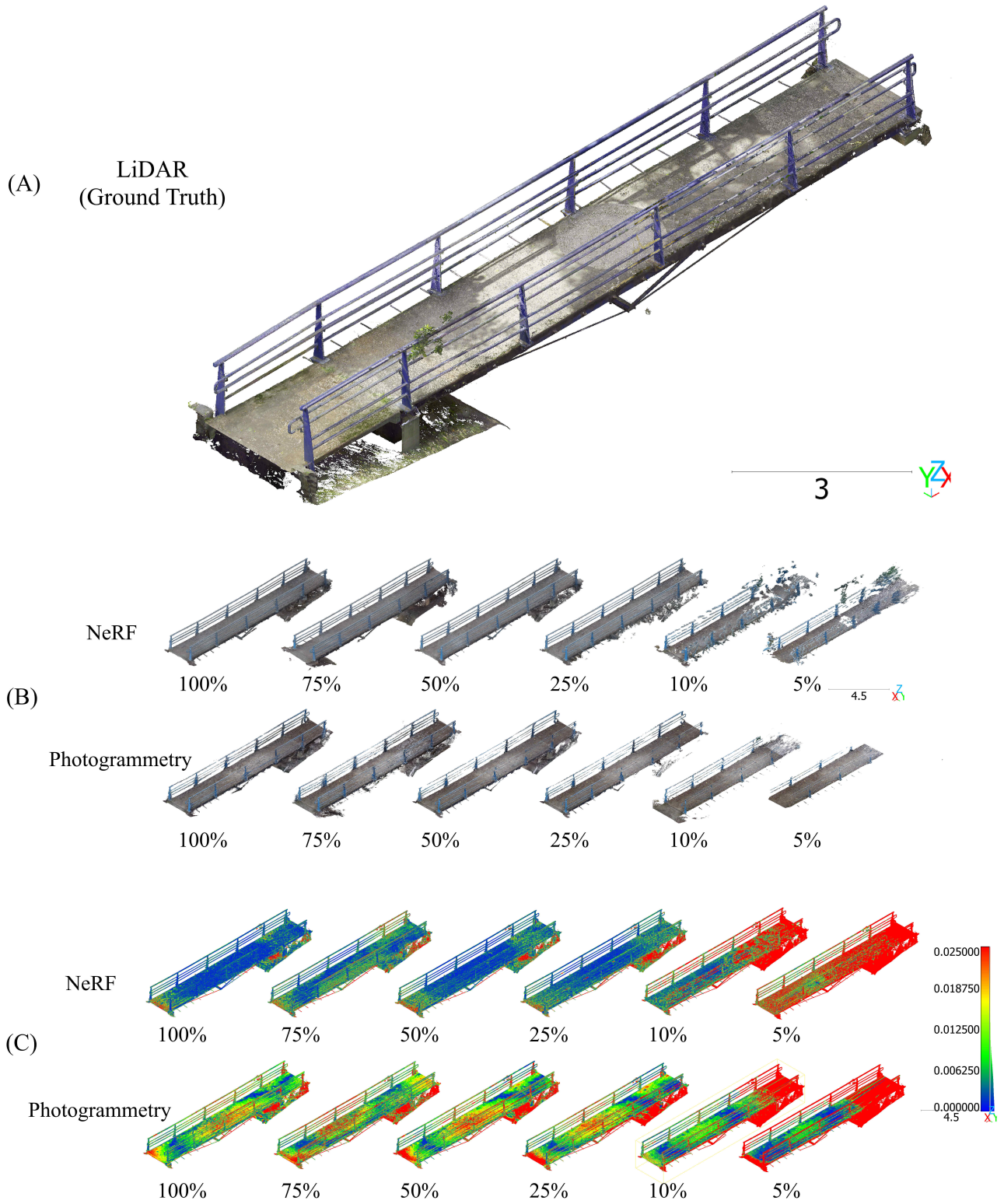

Fig. 9 (A) LiDAR based point cloud from the USP Bridge. (B) RGB NeRF-based and Photogrammetry-based point clouds created utilizing progressively fewer pictures of the USP Bridge. (C) Models compared to the LiDAR-based ground truth using Nearest Neighbors (KNN = 6) local modeling using progressively fewer pictures of the USP Bridge.

on the bridge deck. Finally the 5% model shows large portions of the structure missing. The photogrammetry model provides accurate reconstruction up to 50%, where some parts of the structure start disappearing. 25% shows gaps in the rails. 10% shows parts of the structure missing and 5% shows greatly reduced details on the rails, with only part of the bridge deck maintaining a coherent shape.

Figure 9 (C) shows photogrammetry and NeRF reconstruction models errors when compared to the LiDAR ground truth as the amount of available data is progressively reduced, ranging from 100% down to 5%. NeRF has a mostly blue representation up to 25%, showing a close proximity with the ground truth model. On the other hand, the photogrammetry demonstrates noticeable higher errors on the same range, with greener and red colors on the bridge deck. At 10% and especially 5% data availability, the photogrammetry reconstruction becomes significantly fragmented and incomplete. Despite this, unlike NeRF, at 5% the photogrammetry model maintains a more coherent representation.

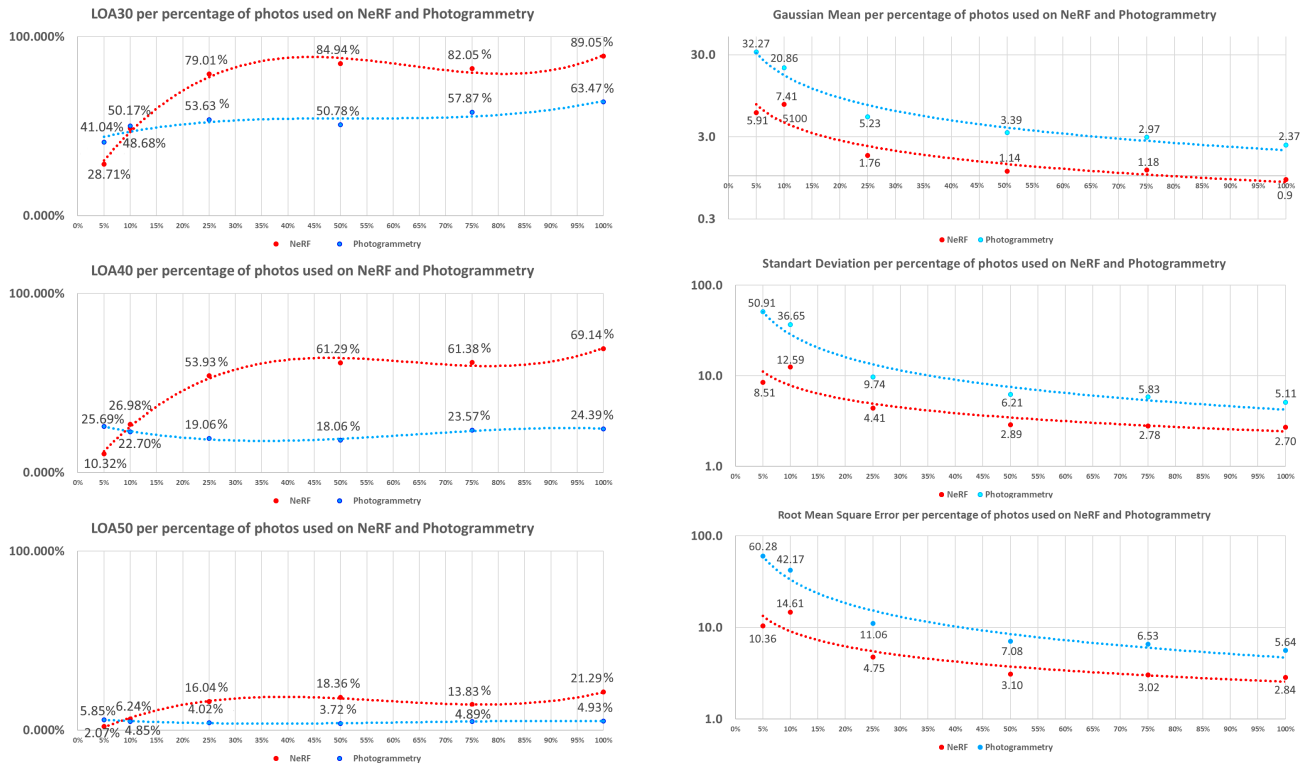

Fig. 10 Left column: LOA30, 40 and 50 by percentage of used photos when creating NeRF and Photogrammetry models (Higher is better). Right Column: Gaussian mean, Standart Deviation and Root Mean Square Error by percentage of used photos when creating NeRF and Photogrammetry models of the USP Bridge (Lower is better).

When comparing Levels of Accuracy, NeRF consistently outperforms photogrammetry at all percentages of photos, except for 5% (Figure 10). When using 100% of the photos, NeRF achieves an accuracy of 89.05% for LOA30 ($\leq$ 1.5 cm), whereas photogrammetry only reaches 63.47%. As the number of photos decreases, NeRF maintains a higher level of accuracy relative to photogrammetry. However, with as little as 5% of photos, photogrammetry performs better accuracy for LOA30, LOA40 and LOA50, compared to NeRF, but the gap narrows as the photo count decreases.

Moreover, NeRF maintains a lower Gaussian mean and RMSE compared to photogrammetry at higher percentages of photos, but both methods exhibit a sharp rise in error at lower photo percentages. With 100% of photos, NeRF's RMSE is 2.85 cm, significantly lower than photogrammetry's 5.64 cm. However, with only 5% of photos, NeRF's RMSE rises to 10.37 cm, while photogrammetry's skyrockets to 60.28 cm, indicating a much larger degradation in performance for photogrammetry.

The Cable-stayed Octavio Frias Bridge

Figure 11 (B) shows a completeness comparison between NeRF and photogrammetry reconstructions as the amount of data is progressively reduced. Both methods display reconstruction models of the Octavio Frias Bridge at different levels of data

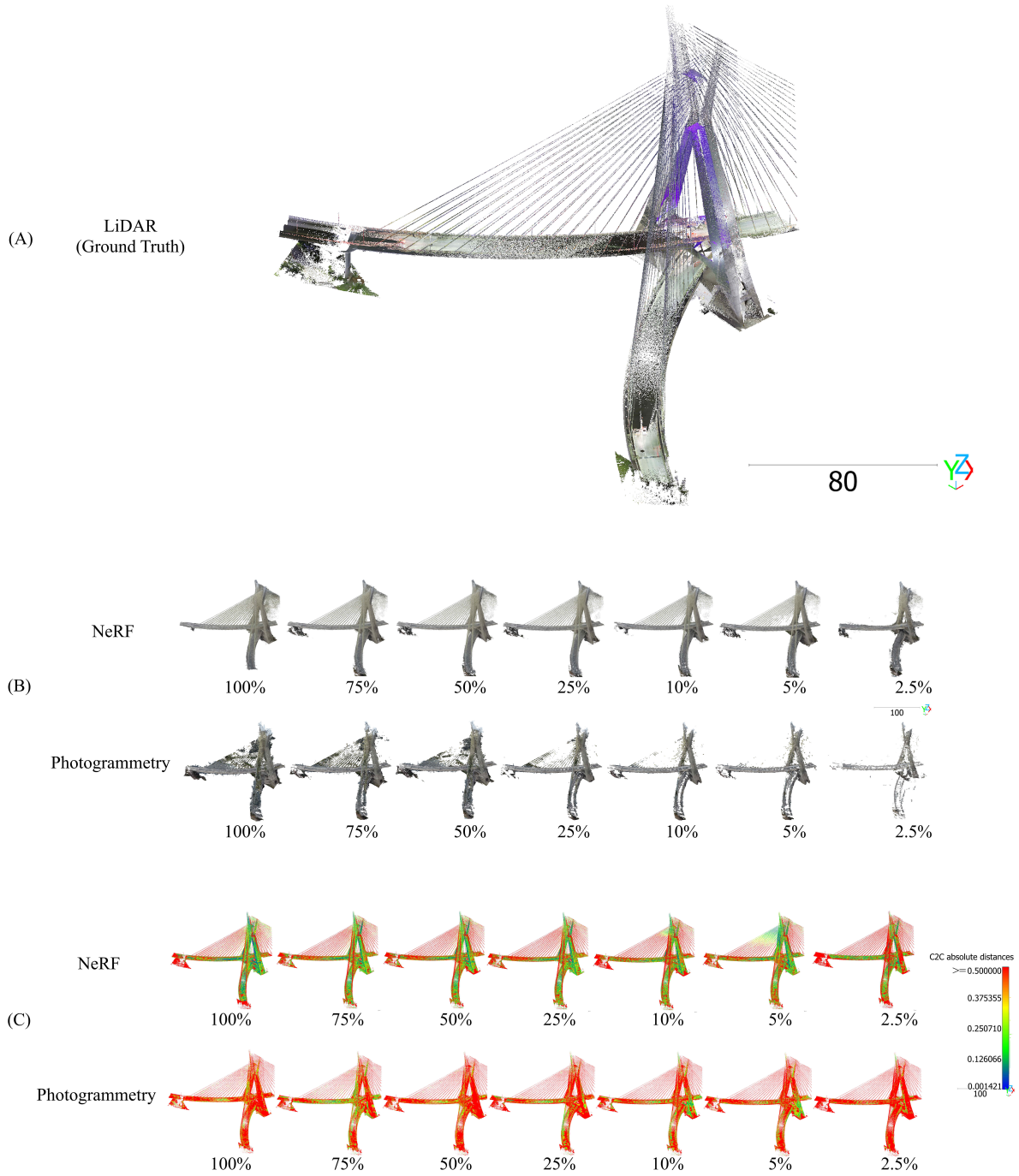

Fig. 11 (A) LiDAR based point cloud from the Octavio Frias Bridge. (B) RGB NeRF-based and Photogrammetry-based point clouds created utilizing progressively fewer pictures of the Octavio Frias Bridge. (C) Models compared to the LiDAR-based ground truth using Nearest Neighbors (KNN = 6) local modeling using progressively fewer pictures of the Octavio Frias Bridge

availability, ranging from 100% down to 5%. For NeRF, even at low percentages (2.5% and 5%), the model retains its structural integrity with minimal detail loss. However, photogrammetry exhibits significant degradation as the data percentage decreases. By the time only 10% or less of the data is available, the photogrammetry reconstruction becomes notably fragmented and incomplete, whereas NeRF maintains a more coherent representation of the bridge structure.

Figure 11 (C) presents a comparison between NeRF and photogrammetry based on error measurements across varying data availability levels. In the NeRF row, the reconstructions show relatively lower error, with most points in green and yellow, indicating smaller deviations from the reference model. Even at reduced data levels (2.5%), the NeRF model maintains lower errors, with fewer regions appearing in red. In contrast, the photogrammetry results exhibit a higher error distribution as the available data decreases. While the 100% and 75% models still maintain some accuracy (green and yellow regions), the 50%, 10%, 5%, and 2.5% models show a significant increase in red regions, indicating larger errors.

In terms of $LOA \times 100$ (Figure 12), NeRF scores fluctuate depending on the percentage of photos used, with higher percentages generally leading to higher $LOA \times 100$, except for the lowest error threshold ($\leq 10cm$), where the $LOA \times 100$ remains lower. The highest $LOA \times 100$ of 88.62% is observed when 5% of photos are used at $LOA30 \times 100$, but lower thresholds ($\leq 50cm$ and $\leq 10cm$) tend to perform better with a higher percentage of photos. For Photogrammetry, the trend is more consistent, with a drop in $LOA \times 100$ as fewer photos are used. Notably, Photogrammetry performs worse than NeRF across all $LOA \times 100$ thresholds, particularly for $LOA50 \times 100$ ($\leq 10cm$), where values remain below 3%, highlighting its difficulty in achieving low-error reconstructions.

NeRF displays more stability in terms of mean and error metrics as the percentage of photos decreases. Its lowest RMSE (1.14) occurs with 5% of photos, demonstrating better error tolerance than Photogrammetry at this level. Photogrammetry, however, shows higher error measures, particularly at lower photo percentages. When only 25% or 2.5% of photos are used, the RMSE increases sharply, reaching 4.55 and 4.66, respectively. This indicates that Photogrammetry struggles with fewer photos, becoming significantly less accurate compared to NeRF.

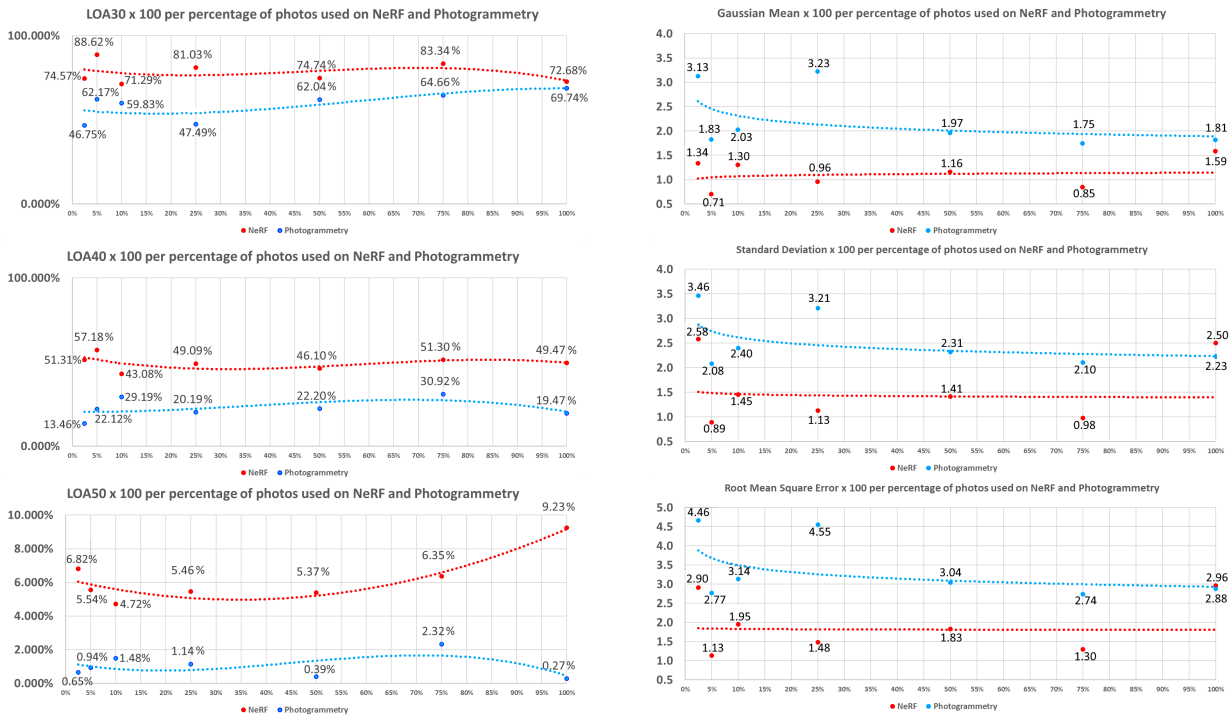

Fig. 12 Left column: LOA30 x 100, LOA40 x 100 and LOA50 x 100 by percentage of used photos when creating NeRF and Photogrammetry models (Higher is better). Right Column: Gaussian mean x 100, Standart Deviation x 100 and Root Mean Square Error x 100 by percentage of used photos when creating NeRF and Photogrammetry models of the Octavio Frias Bridge (Lower is better).

Conclusion

We demonstrate clear differences in the performance of NeRF and photogrammetry as the percentage of input photos decreases, impacting both LOA (Level of Accuracy) metrics and error measurements. NeRF maintains a higher level of detail and exhibits lower errors as data decreases, showing greater resilience to data reduction compared to photogrammetry.

Although photogrammetry retains some structural coherence at very low data levels (such as 5%), this comes at the cost of significantly higher errors and loss of detail, making it less reliable for high-fidelity reconstruction. Overall, NeRF proves to be the more robust and accurate method, especially when balancing data availability with reconstruction quality, positioning it as the preferred choice for 3D modeling in data-limited scenarios.

Acknowledgments This paper was made with the support of the Fundação Amazônia de Amparo a Estudos e Pesquisas (Fapespa), Process 2023/592598, Grant agreement 050/2023, the Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp), Process 2022/10105-3 and CNPq.

References

1. Chalmers, P. "Climate change: Implications for buildings. key findings from the intergovernmental panel on climate change". Technical report, University of Cambridge (2014) Fifth Assessment Report.
2. Kaartinen, E., Dunphy, K., and Sadhu, A. "LiDAR-based structural health monitoring: Applications in civil infrastructure systems". *Sensors (Basel)*, 22(12):4610 (2022)
3. Lee, J. and Kim, R.E. "Noncontact dynamic displacements measurements for structural identification using a multi-channel lidar". *Struct. Contr. Health Monit.*, 29(11) (2022)
4. Jo, H.C., Sohn, H.G., and Lim, Y.M. "A LiDAR point cloud data-based method for evaluating strain on a curved steel plate subjected to lateral pressure". *Sensors (Basel)*, 20(3):721 (2020)
5. Wang, Q. and Kim, M.K. "Applications of 3D point cloud data in the construction industry: A fifteen-year review from 2004 to 2018". *Adv. Eng. Inform.*, 39:306–319 (2019)
6. Carter, J., Schmid, K., Waters, K., Betzhold, L., Hardley, B., Mataosky, R., et al. *LiDAR 101: An Introduction to LiDAR Technology, Data, and Applications*. National Oceanic and Atmospheric Administration (NOAA) Costal Services Center, Charleston, Silver Spring, MD (2012)
7. Aaron, J., Spielmann, R., McArdell, B.W., and Graf, C. "High-frequency 3D LiDAR measurements of a debris flow: A novel method to investigate the dynamics of full-scale events in the field". *Geophys. Res. Lett.*, 50(5) (2023)
8. Ortiz Arteaga, A., Scott, D., and Boehm, J. "Initial investigation of a low-cost automotive lidar system". *ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, XLII-2/W17:233–240 (2019)
9. Remondino, F., Karami, A., Yan, Z., Mazzacca, G., Rigon, S., and Qin, R. "A critical analysis of NeRF-based 3D reconstruction". *Remote Sens. (Basel)*, 15(14):3585 (2023)
10. Huang, H., Tian, G., and Chen, C. "Evaluating the point cloud of individual trees generated from images based on neural radiance fields (NeRF) method". *Remote Sens. (Basel)*, 16(6):967 (2024)
11. Shan, J. *Topographic laser ranging and scanning*. CRC Press, Second edition. — Boca Raton : Taylor & Francis, CRC Press, 2018. (2018)
12. Lowe, D.G. "Object recognition from local scale-invariant features". In *Proceedings of the Seventh IEEE International Conference on Computer Vision*. IEEE (1999)
13. Schonberger, J.L. and Frahm, J.M. "Structure-from-Motion revisited". In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE (2016)
14. Besl, P.J. and McKay, N.D. "A method for registration of 3-D shapes". *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(2):239–256 (1992)
15. Goesele, M., Curless, B., and Seitz, S.M. "Multi-View stereo revisited". In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2 (CVPR'06)*. IEEE (2006)
16. Iglhaut, J., Cabo, C., Puliti, S., Piermattei, L., O'Connor, J., and Rosette, J. "Structure from motion photogrammetry in forestry: A review". *Curr. For. Rep.*, 5(3):155–168 (2019)
17. Yu, A., Li, R., Tancik, M., Li, H., Ng, R., and Kanazawa, A. "PlenOctrees for real-time rendering of neural radiance fields" (2021)
18. Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P., and Hedman, P. "Zip-NeRF: Anti-aliased grid-based neural radiance fields" (2023)
19. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., and Ng, R. "NeRF: Representing scenes as neural radiance fields for view synthesis" (2020)
20. Niemeyer, M., Mescheder, L., Oechsle, M., and Geiger, A. "Differentiable volumetric rendering: Learning implicit 3D representations without 3D supervision" (2019)
21. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., and Polosukhin, I. "Attention is all you need" (2023)
22. Tagliasacchi, A. and Mildenhall, B. "Volume rendering digest (for nerf)" (2022)
23. Elkhrachy, I. "Modeling and visualization of three dimensional objects using low-cost terrestrial photogrammetry". *Int. J. Archit. Heritage: Conserv. Anal. Restor.*, 14(10):1456–1467 (2020)
24. USIBD. "Guide for usibd document c220tm: Level of accuracy (loa) specification for building documentation". Technical report (2016) Document C120TM, Version 2.0, URL: https://cdn.ymaws.com/www.nysapls.org/resource/resmgr/2019_conference/handouts/hale-g_bim_loa_guide_c120_v2.pdf.
25. Laboratório de Sistemas Estruturais Ltda. "Memorial descritivo da passarela de pedestres sobre o Córrego que separa os prédios da civil e da minas, no campus da USP-SP". Technical report (2006) Contrato DEE 77.
26. Pedro Manuel Calas and Sainz de Murieta, E., editors. *Multi-span large bridges*. CRC Press, London, England (2015)