FEDERAL UNIVERSITY OF PARÁ
INSTITUTE OF TECHNOLOGY
GRADUATE PROGRAM IN ELECTRICAL ENGINEERING

3D GEOMETRIC RECONSTRUCTION OF CIVIL INFRASTRUCTURES WITH NEURAL RADIANCE FIELDS

THIAGO FIGUEIRÓ RIBEIRO

DM 08/2025

UFPA / ITEC / PPGEE
Guamá University Campus
Belém-Pará-Brazil
2025

FEDERAL UNIVERSITY OF PARÁ
INSTITUTE OF TECHNOLOGY
GRADUATE PROGRAM IN ELECTRICAL ENGINEERING

THIAGO FIGUEIRÓ RIBEIRO

3D GEOMETRIC RECONSTRUCTION OF CIVIL INFRASTRUCTURES WITH NEURAL RADIANCE FIELDS

Master dissertation submitted to the Examining Board of the Graduate Program in Electrical Engineering from the Federal University of Pará to obtain the Master's Degree in Electrical Engineering, Area of Concentration in Applied Computing.

DM 08/2025

UFPA / ITEC / PPGEE
Guamá University Campus
Belém-Pará-Brazil
2025

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

"3D GEOMETRIC RECONSTRUCTION OF CIVIL INFRASTRUCTURES WITH NEURAL RADIANCE FIELDS"

AUTOR: **THIAGO FIGUEIRO RIBEIRO**

DISSERTAÇÃO DE MESTRADO SUBMETIDA À BANCA EXAMINADORA APROVADA PELO COLEGIADO DO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA, SENDO JULGADA ADEQUADA PARA A OBTENÇÃO DO GRAU DE MESTRE EM ENGENHARIA ELÉTRICA NA ÁREA DE COMPUTAÇÃO APLICADA.

APROVADA EM: 04/04/2025

BANCA EXAMINADORA:

Prof. Dr. João Crisóstomo Weyl Albuquerque Costa
(Orientador – PPGEE/ITEC/UFPA)

Prof.ª Dr.ª Jasmine Prsicyla Leite de Araújo
(Avaliadora Interna – PPGEE/ITEC/UFPA)

Dr. Moisés Felipe Mello da Silva
(Avaliador Interno – PPGEE/LOS ALAMOS)

Prof. Dr. Adam Dreyton Ferreira dos Santos
(Avaliador Externo – UNIFESSPA)

VISTO:

Prof. Dr. Diego Lisboa Cardoso

(Coordenador do PPGEE/ITEC/UFPA)

Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD
Sistema de Bibliotecas da Universidade Federal do Pará
Gerada automaticamente pelo módulo Ficat, mediante os dados fornecidos pelo(a) autor(a)

R4843 Ribeiro, Thiago Figueiró.
3d geometric reconstruction of civil infrastructures with Neural Radiance Fields / Thiago Figueiró Ribeiro. — 2025.
82 f. : il. color.

Orientador(a): Prof. Dr. João Crisóstomo Weyl Albuquerque Costa
Coorientador(a): Prof. Dr. Moisés Felipe Mello da Silva
Dissertação (Mestrado) - Universidade Federal do Pará, Instituto de Tecnologia, Programa de Pós-Graduação em Engenharia Elétrica, Belém, 2025.

1. Neural radiance fields. 2. Escaneamento 3D. 3. Núvens de ponto 3D. 4. Renderização volumétrica. 5. SHM. I. Título.

CDD 006.37

To Julía.
And to my past self.

Acknowledgements

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.

“I’m in love with my lust. Burning angel wings to dust. I wish I had your angel tonight.”

Nightwish - Wish I had an angel

Abstract

The use of noncontact sensing technologies for structural health monitoring (SHM) has significantly broadened the scope of tools available for precise measurement and analysis in engineering and scientific contexts. They address several limitations of the conventional contact-based sensors and at times outperforming them while easier survey, more convenient to install and often low-cost. Digital Twins — dynamic, data-driven virtual replicas of physical structures — have further revolutionized SHM by integrating real-time sensor data with predictive analytics and computational modeling. LiDAR and photogrammetry technology are leveraged to build high-fidelity 3D reconstruction models which can be used to create building information models and digital twins for civil structures. Recent deep learning advancements marked a paradigm shift in several areas, including 3D reconstruction. One particular approach is using Neural Radiance Fields, a deep-learning-based methodology capable of producing high-fidelity 3D models from sparse image datasets, such as those captured using standard consumer-grade cameras or smartphones. NeRF is capable of generating dense point clouds comparable to those generated by Multiview Stereo photogrammetry and terrestrial laser scanning. However, there is a gap in literature addressing the quantitative capabilities of NeRF-based 3D scanning of bridges. This work evaluates the performance of NeRF 3D reconstructions of real-world bridges against SFM/MVS photogrammetry-based models and ground truth data generated via LiDAR. We demonstrate NeRF's feasibility for large-scale structural assessments, with key insights into its performance under varying data availability, impacting both LOA (Level of Accuracy) metrics and error measurements. NeRF offers higher LOA and lower standard error, mean average error and when compared to Photogrammetry. 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, budget-constrained, specialized-equipment-scarce scenarios. This approach offers an efficient, cost-effective, and accurate solution to address the increasing demands for infrastructure monitoring within the Architecture, Engineering and Construction industry, thereby reducing economic and technical barriers to adoption.

Key-words: Neural radiance fields, 3D scanning, 3D Point clouds, Volumetric rendering, Building information modeling, SHM

Resumo

O uso de tecnologias de sensoriamento sem contato para o monitoramento da saúde estrutural (SHM) ampliou significativamente o conjunto de ferramentas disponíveis para medições e análises precisas em contextos científicos e de engenharia. Essas tecnologias superam várias limitações dos sensores convencionais de contato, muitas vezes apresentando desempenho superior, com levantamentos mais ágeis, instalação mais conveniente e frequentemente custo inferior. Gêmeos Digitais — réplicas virtuais dinâmicas e orientadas por dados de estruturas físicas — revolucionaram ainda mais o SHM ao integrar dados de sensores em tempo real com análises preditivas e modelagem computacional. Tecnologias como LiDAR e fotogrametria são utilizadas para criar modelos 3D de alta fidelidade, que podem servir de base para modelos de informações de construções (BIM) e gêmeos digitais de estruturas civis. Avanços recentes em aprendizado profundo marcaram uma mudança de paradigma em diversas áreas, incluindo a reconstrução 3D. Uma abordagem promissora é o uso de Neural Radiance Fields (NeRF), uma metodologia baseada em aprendizado profundo capaz de produzir modelos 3D de alta fidelidade a partir de conjuntos esparsos de imagens, como fotos capturadas por câmeras convencionais ou smartphones. NeRF é capaz de gerar nuvens de pontos densas, comparáveis às produzidas por fotogrametria Multiview Stereo (MVS) e laser scanner terrestre. No entanto, há uma lacuna na literatura quanto à avaliação quantitativa das capacidades do NeRF para escaneamento 3D de pontes. Este trabalho avalia o desempenho das reconstruções 3D de pontes reais usando NeRF em comparação com modelos baseados em fotogrametria SFM/MVS e dados de referência gerados por LiDAR. Demonstramos a viabilidade do NeRF para avaliações estruturais em larga escala, com insights interessantes sobre seu desempenho sob diferentes disponibilidades de dados, impactando tanto as métricas de Nível de Precisão (LOA) quanto as medições estatísticas de erro. Modelos NeRF possuem maior LOA e valores menores de erro médio, desvio padrão e erro quadrático médio quando comparados com modelos de fotogrametria. No geral, o NeRF se mostrou um método mais robusto e preciso, especialmente ao equilibrar a disponibilidade de dados com a qualidade da reconstrução, posicionando-se como a escolha preferida para modelagem 3D em cenários de dados limitados, restrições orçamentárias e escassez de equipamentos especializados. Essa abordagem oferece uma solução eficiente, econômica e precisa para atender à crescente demanda por monitoramento de infraestrutura no setor de Arquitetura, Engenharia e Construção, reduzindo assim as barreiras econômicas e técnicas à adoção dessa tecnologia.

Palavras-chave: Neural radiance fields, Escaneamento 3D, Núvens de ponto 3D, Renderização volumétrica, SHM

Contents

List of figures

List of tables

List of abbreviations and acronyms

3D-GAN	3D Generative Adversarial Network
3D-R2N2	3D Recurrent Reconstruction Neural Network
ASPRS	American Society for Photogrammetry and Remote Sensing
6-DOF	Six degrees of freedom
AEC	Architecture, Engineering & Construction
ALPMS	Activity-Level Progress Monitoring System
AR	Augmented reality
BIM	Building Information Modeling
BMS	Bridge Management System
BrIM	Bridge Information Model
CHP	Cultural heritage preservation
CT	Computed tomography
Full HD	Full High Definition
FOV	Field of view
FPS	Frames per second
GNSS	Global Navigation Satellite System
ICP	Iterative Closest Point
Instant-NGP	Instant Neural Graphics Primitives
Instant-NSR	Instant Neural Surface Reconstruction
KNN	K-Nearest Neighbors
LiDAR	Light Detection and Ranging
LoA	Level of Accuracy
LPIPS	Learned Perceptual Image Patch Similarity

MLS	Mobile Laser Scanner
MLP	Multilayer Perceptron
MP	Mega Pixel
MSE	Mean Squared Error
MVS	Multi View Stereo
NCC	Normalized Cross-Correlation
NeRFs	Neural Radiance Fields
NURBS	Non-Uniform Rational B-Splines
PDF	Probability Density Function
PPS	Points per second
PSNR	Peak Signal-to-Noise Ratio
px	Pixel
RADAM	Retified Adaptive Moment Estimation
RaDAR	Radio Detection and Ranging
ReLU	Rectified Linear Unit
RGB	Red, Blue, Green
RLE	Run-Length-Encoded
RMSE	Root Mean Square Error
RMSD	Root Mean Square Deviation
SDF	Signed Distance Fields
SLAM	Simultaneous Localization and Mapping
SNR	Signal-to-Noise Ratio
SfM	Structure from Motion
SHM	Structural Health Monitoring
SIFT	Scale Invariant Feature Transform
SSIM	Structural Similarity Index Measure

TLS	Terrestrial Laser Scanning
ToF	Time-of-Flight
UAV	Unmanned Aerial Vehicles
UHD	Ultra High Definition
USP	University of São Paulo
USIBD	United States Institute of Building Documentation
VR	Virtual Reality
Voxel	Volumetric Pixel

1 Introduction

The extraction of geometric data from civil structures is a cornerstone of precise structural assessment and modeling. Whether for design optimization, numerical simulations, or operational inspections, advanced 3D scanning and measurement technologies have become integral across the architecture, engineering, and construction (AEC) industries. The accelerating degradation of critical infrastructure, exacerbated by the intensifying impacts of climate change, highlights the growing demand for rapid, efficient, and cost-effective structural assessment methodologies (CHALMERS, 2014). Such assessments are pivotal in ensuring public safety, optimizing maintenance strategies, and extending the operational lifespan of infrastructure assets.

1.1 Context

Recent advancements in non-contact sensing technologies for structural health monitoring (SHM) have significantly broadened the scope of tools available for structural evaluation. Modern methods employ an array of imaging devices, including cameras, smartphones, unmanned aerial vehicles (UAVs), satellites, and ultrasonic sensors, to gather precise data for structural assessments (KAARTINEN; DUNPHY; SADHU, 2022). Many of these technologies hinge on the generation of 3D point clouds—a spatial dataset comprising discrete points that accurately capture the geometry and surface attributes of structures (LEE; KIM, 2022; JO; SOHN; LIM, 2020). Beyond their utility as visual representations, point clouds enable engineers to construct 3D models that facilitate the detection of critical discrepancies between as-built and design specifications. This capability is instrumental in evaluating construction quality, monitoring structural deformation, and identifying deviations from intended configurations (WANG; KIM, 2019). Moreover, point cloud data has proven indispensable in applications such as tracking construction progress (KARAMOOZIAN et al., 2023; ALWYN et al., 2024), verifying geometric accuracy (MIRZAEI et al., 2023), and detecting early-stage damage, including cracks, in buildings and civil infrastructure (ALAMDARI; EBRAHIMKHANLOU, 2024).

In addition, accurate geometry serves as the foundation for constructing digital twins[1] increasingly employed in the Architecture, Engineering & Construction (AEC) industry for predictive maintenance, operational optimization, and life-cycle management. For digital twins to function effectively and accurately reflect their physical counterpart, a robust 3D reconstruction model is essential since the fidelity of its simulation relies on the accuracy, precision, and overall quality of the model (BOJE et al., 2020). Research in different fields highlight the importance of accurate digital twin models for applications ranging from manufactory industry (LIU et al., 2021b), to aviation (XIONG; WANG, 2022) and healthcare (WRIGHT; DAVIDSON, 2020).

[1] virtual replicas of physical assets

Among the array of non-contact geometry extraction techniques, terrestrial laser scanning (TLS) has emerged as one of the most prevalent. TLS is primarily based on Light Detection and Ranging (LiDAR) technology, which utilizes laser pulses to measure distances between the scanning device and surrounding objects. Operating on principles similar to Radar (Radio Detection and Ranging) but employing laser light instead of radio waves, LiDAR provides high-precision distance measurements essential for generating detailed 3D models of complex structures (CARTER et al., 2012). However, the advanced technical components of LiDAR systems—such as high-grade optics and precise rotational mechanisms—contribute to significant costs ranging between $18,908 and exceeding $100,000[2]. In contrast, consumer-grade LiDAR devices have emerged as cost-effective alternatives, with commercially available options priced at a fraction of this expense, typically under $1,000[3].

1.2 Problem and motivations

Problem statement

Despite the high accuracy and utility of LiDAR, its substantial cost and technical complexity restrict its use for routine inspections, particularly for smaller-scale projects or budget-constrained organizations.

Alternative approach

To overcome these limitations, recent research has explored alternative approaches for 3D geometry extraction. A particularly promising avenue is the application of Neural Radiance Fields (NeRF), a machine learning-based methodology capable of producing high-fidelity 3D models from sparse image datasets, such as those captured using standard consumer-grade cameras or smartphones. REMONDINO et al. (2023) (REMONDINO et al., 2023) demonstrated NeRF's potential for generalized 3D reconstruction, while HUANG et al. (2024) (HUANG; TIAN; CHEN, 2024) successfully applied NeRF to model individual trees in 3D with exceptional accuracy. Although extensively utilized in domains like computer graphics (LI et al., 2024; LOPES; LALONDE; CHARETTE, 2024), cultural heritage preservation and documentation (CROCE et al., 2024; CROCE et al., 2024), robotics and autonomous systems (SHEN et al., 2023; MARZA et al., 2023), environmental monitoring (CHIAPPINI et al., 2024) and precision agriculture (HU et al., 2024), NeRF's application to civil infrastructure assessment remains underexplored. Xu et al. (2024) (XU et al., 2024) used a Multi-tiling NeRF approach for general urban mapping using using large-scale aerial datasets and compared their results to MVS pipelines against LiDAR reference data. Kim et al. (2024) (KIM; CHA, 2024) recently conducted

[2] The Ouster OS1-64, as reported by (AARON et al., 2023), is priced at $18,908. The Leica RTC360 retails at $91,086.65 as of September 2024.

[3] Consumer-grade LiDAR devices like the Livox Mid-40 retail at $599, though they achieve lower accuracy levels, as noted by (ARTEAGA; SCOTT; BOEHM, 2019).

a study on the 3D mapping of segmented cracks using NeRF-generated models of laboratory-scale bridges. Also recently Cui et al. (2024) (CUI et al., 2024) introduced NeRFusion, a model to provide an efficient and accurate 3D reconstruction technique for architectural structures with advantages in visual fidelity, detail preservation, and training efficiency for single large buildings and building complexes.

Research questions

1. Feasibility: Can NeRF generate complete, visually coherent 3D models of full-scale bridges using only sparse, consumer-grade images?
2. Accuracy: How does NeRF's geometric fidelity compares to LiDAR ground truth and SfM-MVS Photogrammetry-based reconstructions?
3. Data-Efficiency: How does NeRF perform when trained on sparse image datasets captured with low-cost, consumer-grade cameras in bridge reconstruction scenarios

Motivation

While recent advances in NeRF demonstrate usefulness in urban mapping and reconstruction of buildings, the literature has an interesting gap for civil infrastructure, particularly for bridges. Moreover, there is a lack of quantitative analysis regarding geometric accuracy. This work directly address this gap by evaluating the performance of NeRF 3D reconstructions of real-world bridges against SFM/MVS photogrammetry-based models and ground truth data generated via LiDAR and establishing quantitative benchmarks for its measurement fidelity in structural contexts.

1.3 Objectives

This study investigates the application of NeRF for the 3D reconstruction of civil infrastructures, particularly bridges. Unlike prior studies, this study explicitly benchmark NeRF-based models of real bridges against datasets acquired via terrestrial LiDAR and photogrammetry techniques, widely regarded as industry standards for geometric fidelity. By training NeRF on a constrained set of images captured using consumer-grade cameras, its evaluated its ability to replicate structural geometries that are essential for engineering assessments. Ultimately, this study seek to introduce a scalable, cost-effective, and high-accuracy solution to address the increasing demands for infrastructure monitoring within the AEC industry, thereby reducing economic and technical barriers to adoption.

Specific objectives

1. Investigate the applicability of NeRF for 3D reconstruction of large-scale civil infrastructure, particularly at bridges, addressing the current literature gap in real-world structural applications.

2. Benchmark NeRF's geometric accuracy against high-resolution LiDAR-based ground truth data and industry-used SfM-MVS Photogrammetry models to quantify discrepancies in key metrics.

3. Evaluate NeRF's performance under resource constraints, including training on sparse image sets captured with low-cost, consumer-grade cameras.

1.4 Contributions

The key contribution of this dissertation is tightening the gap in the literature regarding the application of NeRF for 3D reconstruction of civil infrastructure, particularly bridges. It is shown that NeRF produce models with lower statistical errors, higher Level of Accuracy, more complete models with fewer input images.

- **Lower statistical errors**: NeRF-based point clouds show reduced statistical differences in Gaussian Mean (μ), Standard Deviation, and Root Mean Square Error (RMSE) relative to LiDAR ground truth. Outperforms SfM-MVS Photogrammetry in these metrics, indicating enhanced precision and alignment with ground truth.

- **Higher Level of Accuracy (LOA)**: Achieves a superior LOA metric compared to SfM-MVS, validating NeRF's robustness for structural modeling applications.

- **Higher reconstruction completeness**: NeRF reconstructions maintain higher detailed, more complete reconstructions when compared to Photogrammetry models with fewer input pictures, especially in larger structures with finer important structural elements.

These findings hold significant relevance for structural engineering and monitoring, especially in data-scarce, budget-limited or other situations with restricted data acquisition where accurate reconstruction is essential. This dissertation is organized as follows: Chapter 2 reviews the literature on traditional and novel 3D-reconstruction techniques, highlighting their main applications. Chapter 3 focuses on Neural Radiance Fields (NeRF), providing an in-depth explanation of its underlying principles and operation. Chapter 4 describes the framework developed in this work for generating 3D models with NeRF. Chapter 5 introduces the civil structures selected for testing the framework. Chapter 6 discusses the experimental results. Chapter 7 presents the conclusions of this study and outlines directions for future research.

2 Literature review

In this chapter are discussed applications of noncontact sensors for SHM and the applications of digital twins in SHM, then the main reconstruction techniques used in this work: LiDAR, Photogrammetry and Neural Radiance fields. In the end traditional volume rendering techniques are brefly reviewed.

2.1 Structural health monitoring

Traditional SHM methodologies employ global responses (such as vibrational characteristics) and localized mechanical responses (such as strains) to evaluate structural integrity (SONY; LAVENTURE; SADHU, 2019). Vibration-based SHM approaches are particularly effective in evaluating global health state of structures, performing rapid risk assessment and proactive hazard mitigation. These techniques predominantly rely on acceleration measurements, which require the installation of either contact-based or noncontact sensors collecting high-fidelity data.

Contact-based sensing methods rely on sensors that are directly attached to the structure to measure dynamic responses such as acceleration, velocity, displacement, or inclination. However, these systems present significant economic and practical limitations. For instance, the contact-based wired sensors involve time-consuming and labor-intensive installation processes and demand substantial maintenance to ensure long-term monitoring.

Noncontact sensors, such as Time of Flight (ToF) Sensors, digital and high-speed cameras, UAVs and smartphones, address several limitations of the conventional contact-based sensors and at times outperform over the latter. They are easier and convenient, and allow faster installation and offer data acquisition that is more reliable with high-resolution temporal and spatial information of the structures. Moreover, they are less labor intensive and often cost-effective[1]. Such smart sensing techniques are integrated with visual and mobile monitoring systems.

Applications of LiDAR for SHM range from analyzing 3D objects, structural geometries, deformations, crack information, and visualization using a dense 3D dataset with finer resolution and precision (KAARTINEN; DUNPHY; SADHU, 2022). Despite its benefits, LiDAR operation is often costly, static and time consuming since a sensor has to stay at a place for several minutes to execute a survey. Unmanned Aerial Vehicles (UAVs) based techniques are becoming more common due the easiness of surveying with this tool. Duque et al. (2018) (DUQUE; SEO; WACKER, 2018) monitored glued-laminated timber arch bridge in South Dakota using a DJI

[1] (CHEN et al., 2019) compared the cost and time efficiency of an UAV survey against TLS-based data aquisition on the Boyne Viaduct Bridge, ocated in Drogheda, Ireland. While data accuracy was better in the TLS-based model, UAV equipment costs were less than 3% of TLS'. Moreover, time on site during the UAV survey was 1/3 of the time spent on the TLS survey.

Phantom 4 drone UAV combined with a damage quantification protocol. A DJI Inspire 1 Pro UAV equipped with infrared thermography capabilities of the FLIR Vue Pro was used to observe damages of concrete bridge decks in London, Ontario, Canada (OMAR; NEHDI, 2017). Potenza et al. (2020) presented a quantitative evaluation of defect extension using a Aibot X6 UAV several railway bridges belonging to the Italian National Network (POTENZA et al., 2020). The use of UAVs have revolutionized data collection by enabling rapid, high-resolution structural surveys while minimizing operational costs and human risk.

Concurrently, the proliferation of smartphone technologies presented an opportunity for SHM applications. In recent years the utilization of smartphones in SHM has attracted increasing attention owing to its unique features. Shrestha et al. (2018) (SHRESTHA; DANG; WANG, 2018) introduced the possibility of using multiple iPhones for vibration measurement and compared their results using vertical high-quality servo-velocity VSE-15D sensors. iPhone's 6 were used to assess the structural displacement from the movement of a laser spot on projected in a conveniently placed plate in a 1/28th model of the Zhuanghe Construction Bridge, in Zhuanghe/China (ZHAO et al., 2016). Pan et al. (2023) proposed a vision-based structural vibration measurement in term of structural displacement using an ensemble model for tracking structural motions (PAN et al., 2023). The proposed deep learning method was verified in a lab-scale two-story steel structure by recoding videos with an iPhone Xs Max.

2.2 Digital twins in SHM

Digital twin technology refers to the process of mapping an object in the real world to create a highly accurate digital representation and reflecting the information collected in the real world onto the digital model to make predictions based on the simulation, analysis, and visualization of the natural world (XU et al., 2023). Digital twins for for SHM purposes involves simultaneous data exchange among the physical and virtual twins (SUN et al., 2025). The physical twin represents the tangible, real-world asset with sensors and devices to collect real-time operational and environmental data. The virtual twin is its digital counterpart, created using advanced modelling and simulation techniques to replicate the physical asset's behaviour, characteristics, and performance. There are two main approaches of creating digital replicas of civil infrastructures: physics-based model-driven and a measurement-based data-driven method.

- **Physics-based digital twin:** Physics-based techniques can detect damage and evaluate its criticality, conduct simulations that enable assessing load capacity and the remaining useful life of the asset, and increase the efficiency of health monitoring. Finite Element Modelling (FEM), Finite Difference Method (FDM) and Computational Fluid Dynamics (CFDs) are commonly adopted for physics-based model-driven methods (SUN et al., 2025).
- **Data-driven digital twin:** The data-driven measurement-based methods include the deployment of machine learning algorithms, time series analysis, data mining, and genetic

algorithms; they can also detect anomalies, identify trends and future predictions (in asset behaviour), and quantify uncertainties.

Applications of digital twins

Torzoni et al. (2023) (TORZONI et al., 2024) states that the concept of digital twins has been applied to several fields including additive manufacturing (PHUA; DAVIES; DELANEY, 2022), smart cities (JANS-SINGH et al., 2020), urban sustainability (TZACHOR et al., 2022) and spacecraft health management (YE et al., 2020) for operational monitoring, control, and decision support.

Digital twins have been used with great success in SHM. Ye et al (2019) (YE et al., 2019) used both physics-based FEM and data-driven approaches to investigate their monitoring capabilities in the Railway bridges in Staffordshire, UK. The 3D FE model was constructed incorporating solid, shell and rebar elements, with time-dependent concrete properties, staged construction and torsional effects due to the skewed bridge geometry. The FE model predictions were verified by fibre optic sensors installed on the bridge measuring strain in different location along the girders and different balance loads [2]. Meanwhile, the data-driven approach focused on real-time statistical modeling of sensor data. Statistical modeling was then used to perform short-term forecasting and to detect anomalies in the data. This integrated approach can be leveraged into building more accurate and complete digital twins.

Shim et al. (2019) (SHIM; KANG; DANG, 2019) proposed a concept of digital twin models for the maintenance system for cable-stayed and suspension bridges using FEM. They suggest building lightweight FE models compared to BIM-based counterparts to be linked to machine-learning-based maintenance systems. This model and the additional parameter information of specific elements of the bridge are then accumulated into a database. Finally, researchers can use standardized models from this database for training of machine learning algorithms for automatic damage inspection.

Jayasinghe et al. (2024) (JAYASINGHE et al., 2024) introduced innovative methods leveraging Neural Networks as surrogates for real-time FEM of digital twins of the Dalrymple Bay Coal Terminal at Hay port, Queensland, Australia. An ANN was trained from data coming from sensors installed on critical elements of the port structure to output their 3D displacement patterns for an FE model. This trained ANN model generates fast-inference (within 15ms), high-accuracy (testing error in the order of 10^{-5}) results.

Mohammad et al. (2023) (MOHAMMADI et al., 2023) integrated a TLS-based bridge information model with a Decision Support System on the cable-stayed Werrington Bridge, in New South Wales, Australia. A TLS-based scan is used to build a Bridge Information Model (BrIM) which is input into the Bridge Management System (BMS) integrated with a multi-

[2] Dynamic responses during a train passage event.

criteria Decision Support System to score a “Priority Rating Condition Index” and provide more objective decisions for optimal budget allocation and remedial planning. Through the proposed framework, the BMS has been greatly improved not only by the BrIM-oriented geometrical and non-geometrical information, but also by the proposed priority ranking and decision support systems.

Hosamo et al. (2022) (HOSAMO; HOSAMO, 2022) reviewed digital twinning of bridges using TLS in applications using machine learning, BrIM, BMS and general 3D modeling. They assess the initiatives to integrate digital twins for bridges is in the early stages, lagging behind in research and applications compared to digital twins of other science fields. To that end, 3D reconstruction techniques can be leveraged to build highly detailed models that can facilitate the creation of digital twins for bridges.

Explicit and implicit representations

Explicit and implicit representations of 3D reconstruction fundamentally differ in their approach to modeling and rendering scenes. Explicit representations, such as voxel grids (WU et al., 2014), mesh-based methods (NGUYEN; CHOI, 2019), and point clouds (ROLIN et al., 2019), directly encode geometric structures into discrete formats. These methods provide a clear, visualizable structure that can be readily manipulated and rendered but often require significant storage, processing power for high-resolution models and intensive post-processing because of missing connectivity information. Most mesh-based methods do not perform post-processing, but they often require a deformable template mesh or represent geometry as a collection of 3D patches which leads to self-intersections and non-watertight meshes (NIEMEYER et al., 2019).

In contrast, implicit representations encode geometry indirectly as a mathematical function, such as Signed Distance Fields (SDFs) (ZYGUN, 2024) or weights of a neural network (MILDENHALL et al., 2020). Highly flexible, capable of representing complex, continuous surfaces at any resolution, implicit representations are often compact in storage. However, they require additional computation to render explicit structures or surfaces and may face challenges in interpretability and direct manipulation.

2.3 Light detection and ranging

Laser-based ranging, profiling, and scanning technologies have been active research topics since the 1960s. Two principal techniques are commonly employed to measure object distances, as illustrated in Figure 1. The first method, Time-of-Flight (ToF), involves emitting a short, intense laser pulse and measuring the time it takes to reflect back from the target surface. By leveraging the constant speed of light, distance accuracy is determined by the precision of the time measurement (SHAN, 2018). The second method calculates distance by measuring the phase difference between the emitted laser signal and the reflected signal. Modern LiDAR

systems employ scanning mechanisms that sample a sequence of measured distances, capturing both vertical and azimuthal angles. These sampled points are digitally and explicitly recorded to form a high-resolution 3D representation of the scanned environment. The resulting point clouds enable precise geometric modeling, critical for applications such as structural health monitoring, deformation analysis, and construction progress tracking.

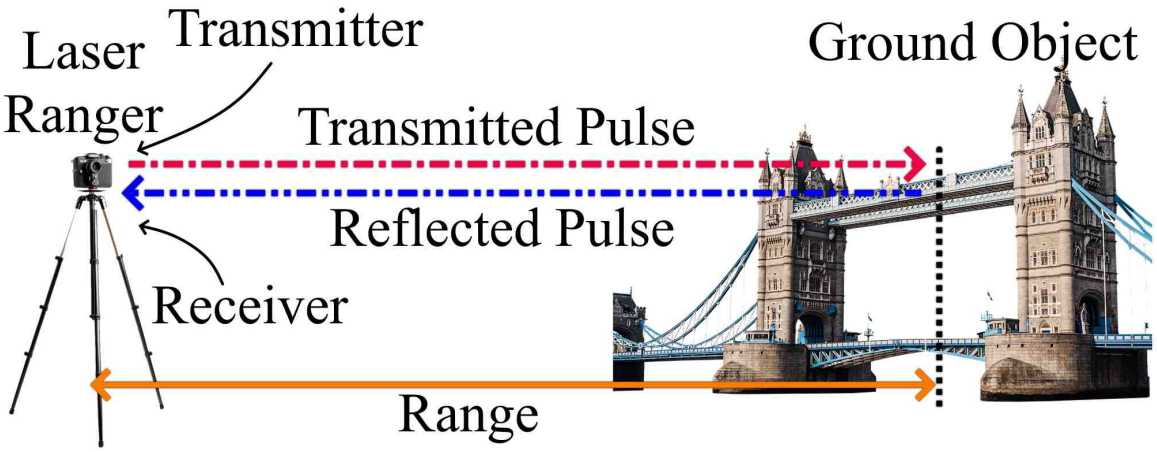

Figure 1 – Operation of a laser rangefinder that is using the timed pulse or ToF method. Adapted from SHAN et al. (2018) (SHAN, 2018)

Applications of 3D reconstructions using LiDAR

The use of laser scanners such as LiDAR is widespread in Civil Engineering as it enables precise data collection for infrastructure planning, monitoring, and maintenance in diverse areas such as terrain mapping and topographic surveys, SHM and construction site management.

Terrain mapping and topographic surveys

Liu et al. (2024) (LIU et al., 2024) introduced Point2Building, a model that reconstructs buldings as 3D polynomial meshes from airborne LiDAR point clouds. The model consists of two parts: a vertex module for sequential vertex generation and a subsequent face module to that connects vertices to mesh faces. When combined with geometric plausibility checks and rejection sampling, the novel model consistently outperforms both a classical optimization-based and a recent learning-based baseline across a range of performance metrics.

Huang et al (2022) (HUANG et al., 2022) presented an automatic approach for reconstructing 3D building models from large-scale airborne LiDAR point clouds. The paper uses an hypothesis-and-selection-based framework for reconstructing general piecewise-planar objects from a set of planar segments and vertical planes inferred from the LiDAR point cloud. They assess this is sufficient for reconstructing most urban buildings, however they note that in current implementation the model could not handle curved surfaces present in several types of architecture buildings.

Structural health monitoring

Terrestrial Laser Scanning (TLS) allow engineers to detect deformations, cracks, and other defects in bridges, dams, and high-rise buildings with sub-millimeter accuracy (TENG et al., 2022). Andreas et al. (2019) (ANDREAS et al., 2019) used a Terrestrial Laser Scanner to accurately survey the alleged[3] tilting of the Menara Saidah high-rise and other Jakarta-based buildings. Interestingly, survey results show that all the buildings investigated have zero tilting even on the top-most floors.

Xu et al. (2019) (XU; YANG, 2019) proposed the use of Signal-to-Noise Ratio (SNR) distribution with the parameter σ of Gaussian filtering method for automatic crack identification based on TLS-based 3D point cloud data. The method proposed could have a capability of noise immunity, which could help solve the texture problem and extract the correct cracks in tunnel-based environments. However, incorrect σ selection can lead to an over-smooth effect, undermining important features in the image, such as cracks whose size is very small in the crack width direction.

Xu et al. (2019) (XU et al., 2019) proposed a methodology leveraging TLS with Non-Uniform Rational B-Splines (NURBS[4]) technology to monitor the deformation of the Changheba earth-rock dam during different periods. Compared to planar mesh-based models, NURBS produced greater surface precision due its curve-like nature, especially if a higher number of adjacent points are used in expense of processing time. Using this model a 0.0976m deformation was observed in some parts of the damn when data from October 2016 was compared to April 2017, showing NURBS potential for point-cloud-based reconstructions.

A Leica RTC360 terrestrial laser scanner was used for health assessment of The Cloister of the Cathedral of Tarragona (Spain) (BUILL et al., 2020). From cross sections was possible to verify the verticality of the walls and observe anomalies such as bulging of certain walls using different reference planes for comparison.

A Trimble GX 200 TLS was used for deformation measurement of Haci Piri Mosque minaret[5], Istanbul (SELBESOGLU; BAKIRMAN; GOKBAYRAK, 2016). This particular minaret's tilt can be observed by naked eye so constant monitoring and documentation of deformation processes of the structure is vital. The offsets from the vertical axes for the cylindrical part, the balcony part and the upper parts of the minaret were found as 10.14 cm, 13.97 cm and 16.51 cm, respectively.

[3] The Menara Saidah, located in southern part of Jakarta city, was suddenly abandoned after a rumor of tilting on the building was rapidly spread out. Cracks and unleveling was observed at its parking yard supposedly caused by land subsidence. Nobody would dare to enter the building anymore and so the owner experienced financial loss from building rentals, events, etc.

[4] NURBS is a mathematical model using basis splines (B-splines) that is commonly used in computer graphics for representing curves and surfaces.

[5] Minaret, in Islamic religious architecture, is the tower from which the faithful are called to prayer five times each day by a muezzin, or crier. Such a tower is always connected with a mosque and has one or more balconies or open galleries (The Editors of Encyclopaedia Britannica, 2023).

The Z+F Imager 5010c laser scanner was used in combination with IR Thermography as a diagnostic tool for the St. Augustine Monumental Compound in Cosenza, Italy (COSTANZO et al., 2015). Their model detected anomalies by combining reflectance, geometric and thermal features originating from cracks, degradation of materials, hidden elements, misalignment. As such, their methodology not only reconstructed the 3D model, but was also leveraged to detect the anomalies and to analyze the vulnerability of the structural elements.

A FARO Focus S120 phase shift laser scanner was utilized to assess and monitor surface change in earthen architecture at Çatalhöyük, Turkey (LERCARI, 2019). By leveraging multiscale model-to-model cloud comparison to assess surface changes, the authors observed the material loss of walls between 2014 and 2016 at the architecture site. This study has showed M3C3 method is successful in quantifying surface materials loss in mud brick walls with millimeter-level accuracy, providing conservators and site managers with a powerful tool to detect patterns of change and calculate their significance as a preventative measure.

Construction site management

LiDARs are used to provide real-time, high-precision as-built models, improve project efficiency and reduce errors in design implementation (ALIZADEHSALEHI; YITMEN, 2023). Wang et al. (2015) (WANG et al., 2015) proposed a construction quality control system integrating BIM, LiDAR-based scanning, digital twins and extended reality technologies to work collaboratively. This system orchestrated the analysis, management and visualization of construction progress data, information and reports. It has the potential to reduce manpower required for progress status capture, automation of progress control progress, real-time checking of progress and quality and high quality visualization at the cost of increased costs of sensors, software licenses and trained employees to implement the system.

More recently, Kim et al. (2024) (KIM; LEE; KIM, 2024) assessed the work progress using Laser Scanners in a framework for construction management. Their models compares the project's BIM-based point cloud and the LiDAR-based point clouds then extract parts which work has been completed. Finally, a bill of quantity of work is calculated and progresses measurement is assessed. Their results suggest progress measurement using laser scanning and BIM contribute to faster and better progress measurement compared to manual evaluation.

2.4 Photogrammetry

Photogrammetry facilitates the creation of 3D models of objects, structures, or environments by analyzing 2D images captured from multiple perspectives. The process begins with feature detection and matching across images using advanced algorithms such as Scale Invariant Feature Transform (SIFT) (LOWE, 1999), as depicted in Figure 2. Subsequently, Structure-from-Motion (SfM) techniques (SCHONBERGER; FRAHM, 2016) estimate the position and

orientation of the camera for each image. These camera poses enable triangulation, which calculates the 3D coordinates of the matched feature points to produce an initial sparse point cloud. Modern drones encode the azimuth and position of the camera in their metadata, which can assist in the position estimation.

To enhance point cloud alignment, the Iterative Closest Point (ICP) algorithm (BESL; MCKAY, 1992) is frequently applied. Following this, Multi-View Stereo (MVS) techniques (GOESELE; CURLESS; SEITZ, 2006) are employed to generate a dense point cloud. For each image, depth maps are derived by triangulating matching features across multiple views, representing the distance of each pixel from the camera. These depth maps are then aligned and merged into a comprehensive dataset. Since every pixel in the depth map corresponds to a 3D point with a known distance from the camera, this approach effectively produces a dense and detailed explicit 3D point cloud, supporting high-precision modeling and visualization tasks.

Applications of 3D reconstructions using photogrammetry

Cultural heritage preservation (CHP)

CHP manages the protection, restoration, and management of cultural assets to safeguard their historical, artistic, and social value for future generations. Photogrammetry assist in the accurate 3D reconstruction of cultural heritage artifacts, aiding in preservation and study. It is an effective technique in documenting artifacts of varying sizes and complexities, from small relics to large architectural structures.

Calisi et al (2023) (CALISI; BOTTA; CANNATA, 2023) used the integration between laser scanning with a Z+F IMAGER 5010 sensor and photogrammetry techniques to scan Rocca Farnese in Capodimonte. Using different drone cameras and georeferenced-high-resolution-photos, their reconstructions generated point clouds with overlapping errors in the order of 0.001m. This highlights photogrammetry capabilities when high-resolution (4000x2250 px to 5280x3956 px), georeferenced photos are provided.

Croce et al (2024) (CROCE et al., 2021) proposed a pipeline for photogrammetry based point clouds to integrate to a Heritage-Building Information Modeling system. The photogrammetry-based point cloud is used as input to a random-forest-based semantic segmentation model which is then used to reconstruct parametric geometries to be used by Heritage BIM platforms. The approach is then tested on the laser scanned point clouds of the Pisa Charterhouse, in Italy. This technique can be employed to update, retrieve and archive heritage information.

Apollonio et al. (2021) (APOLLONIO et al., 2021) proposed a Photogrammetry-Based workflow for 3D reconstruction of museums assets which was then compared to LiDAR-based ground truth from a Faro Focus X 130 TLS. Phone-based photographs are used into the photogrammetry reconstruction pipeline, then the mesh is output as an Unity 3D model and compared to a TLS-based model, with measured mean errors within 1.03mm and standard deviation within

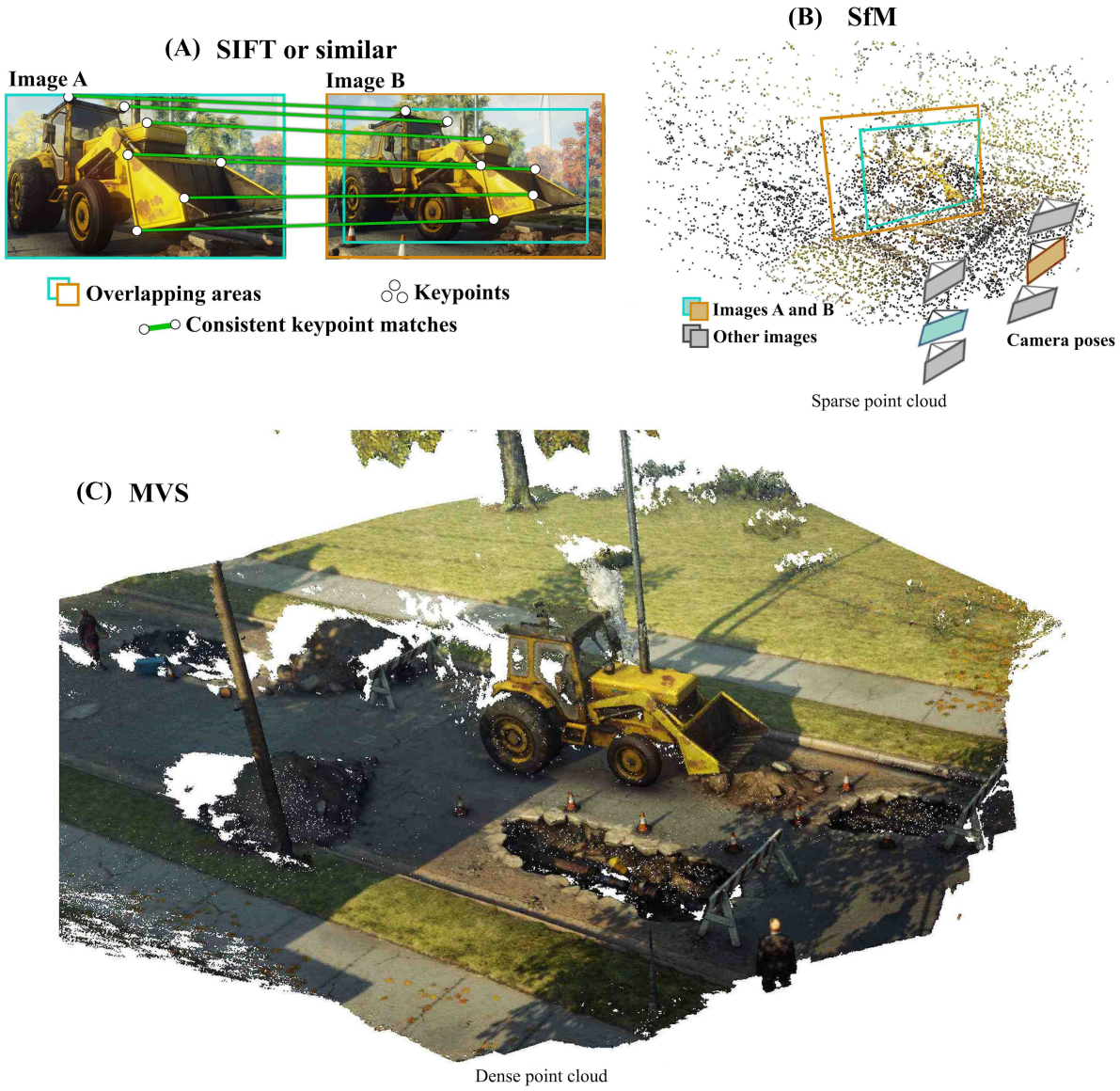

Figure 2 – The three key stages in a SfM-MVS workflow illustrated on two hypothetical images of a construction truck: (A) keypoint identification and matching (e.g. SIFT), (B) SfM with camera parameters and a sparse point cloud as output and (C) the densified point cloud following MVS.

1.231mm. Their results demonstrate that devices and workflow allow to get high-quality 3D models that can be used for the management, enhancement and monitoring the state of preservation of cultural heritage.

Civil infrastructures

Vien et al. (2020) (VIEN et al., 2020) proposes a strategy of UAV-assisted photogrammetry with finite element analyses to determine the structural integrity of floating high-density polyethylene membrane cover at the Western Treatment Plant anaerobic wastewater lagoons in Melbourne, Australia. The photogrammetry model is used to build a digital elevation model of

two deformed membrane specimens in a controlled laboratory setting, then preprocessed using a smoothing interpolation with different smoothing parameters and applied as displacement loads for FE analysis to predict its displacement and strain field. Such methodology could be used to assess strain and displacement in other civil-related scenarios.

Jung et al. (2024) (JUNG; MERKLE; REITERER, 2024) explore the possibility of automating the UAV camera pose generation for bridges using voxel-based camera pose placement. These poses are generated based on a voxelized representation of the bridge with the size of the voxels calculated from quality requirements and camera specifications. This method was able to generate high-resolution images necessary for digital damage inspection and photogrammetry-based 3D reconstruction to overcome time and labor-consuming manual inspection.

2.5 Neural radiance fields

In contrast to conventional LiDAR-based approaches, Neural Radiance Fields (NeRF) are considered an implicit representation, encoding the 3D light field of a given scene into the weights of a neural network (YU et al., 2021; BARRON et al., 2023) (Figure 3). This method represents a static scene as a continuous 5D function that outputs the radiance emitted in each direction, (θ, ϕ), from any point in the 3D coordinate system, (x, y, z). Ray samples are generated for each pixel using a classical pinhole camera model and conventional ray tracing[6] techniques. For each ray sample, the model learns an RGB color value and a density parameter, σ, which serves as a differential opacity factor, determining how much radiance is accumulated along a ray passing through a specific 3D location. Subsequently, volume rendering algorithms are applied to all samples to reconstruct the final pixel color of an image rendered from a given camera angle (MILDENHALL et al., 2020; NIEMEYER et al., 2019).

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, SfM algorithms (SCHONBERGER; FRAHM, 2016) are used. They use triangulation techniques to estimate the three-dimensional coordinates and angles from each input photo.

[6] There are 2 main ways to model ray tracing: **Forward ray tracing** involves simulating light particles (photons) emanating from a light source towards objects in a scene. While this method provides highly accurate color determination for each object, it is computationally intensive (SNELLENBURG et al., 2010). **Backward ray tracing** enhances efficiency by originating rays from the viewer's perspective. This approach creates eye rays starting at the camera or viewpoint, passing through the view-plane into the scene (XING et al., 2017). The first object intersected by these rays determines visibility, allowing for efficient rendering of visible elements.

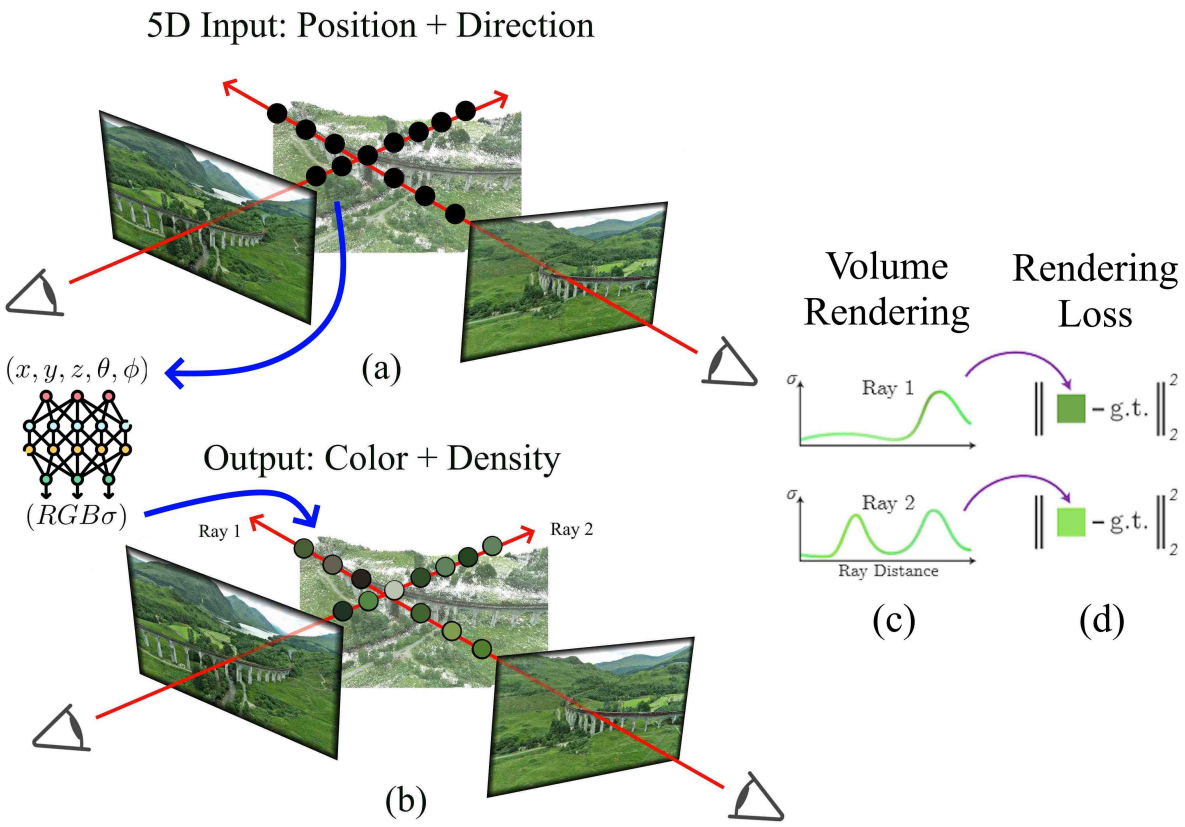

Figure 3 – An overview of the neural radiance field scene representation and differentiable rendering procedure. Images are synthesized by sampling 5D coordinates (location and viewing direction) along camera rays (a), feeding those locations into an MLP to produce a color and volume density (b), and 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).

Applications of 3D reconstructions using neural radiance fields

REMONDINO et al. (2023) (REMONDINO et al., 2023) demonstrated NeRF's potential for generalized 3D reconstruction by comparing their reconstrutions traditional photogrammetry, evaluating the strengths and weaknesses of NeRFs and provide insights into their applicability to different real-life scenarios. Quantitative analysis was evaluated with geometric accuracy (RMSE, Gaussian Mean, Standard Deviation) and the number of required images (comparing models with progressively fewer input images). Their work show that NeRFs exhibit superior performance over photogrammetry in terms of non-collaborative objects with texture-less, reflective, and refractive surfaces. Nevertheless, photogrammetry has outperforms NeRFs in cases where the object's surface possesses cooperative texture.

NeRFs in computer vision

Recent advances in NeRFs have enhanced 3D reconstruction for AR/VR, but real-world applications remain constrained by both inefficiencies in acceleration and scalability. Li et al. (2024) (LI et al., 2024) proposes Fusion-3D, a framework capable of real-time ($\geq$ 30FPS) rendering and near-instant (fewer than 2 seconds) scene reconstruction by using optimizations to all stages of the NeRF pipeline and using a multi-chip architecture for handling large-scale scenes.

Lopes et al. (2024) (LOPES; LALONDE; CHARETTE, 2024) proposed an approach to leverage NeRF to infer material transformations in complex scenes by learning a material mapping function. By optimizing NeRF representations from paired observations of a scene in its original and transformed states, the method accurately models material changes such as varnishing or painting. This learned transformation function can then be applied to new scenes with different geometries and materials, enabling realistic appearance modification. The framework enhances material editing capabilities in graphics applications, offering a flexible and scalable solution for predicting material changes across diverse scenes.

Cultural heritage preservation (CHP)

Croce et al (2024) (CROCE et al., 2024) leverages NeRF and photogrammetry for 3D reconstruction of the Terpsichore from Dodona Statue. Results show that NeRFs outperform photogrammetry in preserving completeness and material details with fewer input data or lower resolution[7], making them valuable for large-scale mapping in emergency scenarios. However, NeRFs exhibit higher noise sensitivity. The study suggests integrating NeRFs with photogrammetry to overcome their respective limitations for this type of surveying.

The same author further assess the flexibility of NeRF reconstructions for CHP objects with different dimensions and geometries (CROCE et al., 2024), such as pavement laying patterns from Piaza dei Priori, the Spanish Mill of the Orbetello Lagoon and the 19th century stove and mirror and the façade of Palazzo Boileau in Pisa. In their work, NeRF models of historic origin were compared to photogrammetry, showing that while NeRF reconstructions were more complete, they had higher noise and lower texture quality.

Robotics and autonomous systems

3D reconstruction models enable robots to interact with dynamic environments by accurately modeling changes over time. NeRFs provide robots with detailed 3D scene understanding, improving navigation and manipulation tasks. Shen et al. (2023) (SHEN et al., 2023) leverages the use of NeRF for 3D reconstrution models for robot grasp inference. In his work, a robot uses

[7] Initially, 233 images with a resolution of 2752x4128 pixels were used, followed by 233 images downscaled by a factor of 3, then 116 images at full resolution, and finally 166 images downscaled by a factor of 3

a selfie-stick to capture RGB photos of a given scene. Then, dense features are extracted from these images to distill them into a feature field and to train a NeRF model. Finally, when the user query the system via natural language, it generates a heatmap and infer 6-DOF grasps on novel objects.

Marza et al. (2023) (MARZA et al., 2023) introduced AutoNeRF, a framework for automatically generating 3D models of a scene by training NeRFs from data collected by autonomous agents[8] . The data collecting occurs using automated agents that explore an unseen environment efficiently and use the experience to build an implicit map representation autonomously. Then, local policies such as area-to-clean can be attributed to each ambient.

Environmental monitoring and precision agriculture

Precision Agriculture is the use of technology and data-driven techniques to optimize farming practices, improving efficiency, yield, and sustainability. 3D models of natural habitats help track biodiversity and monitor ecosystem changes. Accurate reconstruction of plant phenotypes (traits such as growth, development, tolerance) plays a key role in optimizing sustainable farming practices in the field of Precision Agriculture.

Huang et al. (2024) (HUANG; TIAN; CHEN, 2024) successfully applied NeRF to reconstruct individual trees with different canopy structures. Then compared their results to SfM/MVS Photogrammetry and laser scanning methods. It was noted that for trees with denser foliages, the quality of the reconstructed NeRF model was superior although noisy in comparison to Photogrammetry. NeRF canopy metrics (Witdh, height, area, volume) were also more accurate.

Chiappini et al. (2024) (CHIAPPINI et al., 2024) assessed the accuracy of urban olive 3D models originated from an iPhone 14 Pro LiDAR, Photogrammetry and NeRF (Luma.AI based-model) against a Mobile Laser Scanner (MLS) model as ground truth. Total height, canopy base height and canopy volume measurements were extracted and compared. Overall, NeRF had the highest RMSE and Standard Deviation when compared to the other models.

Hu et al. (2024) (HU et al., 2024) used NeRF (Instant-NGP[9] and Instant-NSR[10]) for view synthesis and 3D reconstruction of crop and plant models for plant phenotyping. In this context, NeRF generate competitive reconstructions against leading compercial 3D MVS-based reconstruction. Results indicate that Instant-NGP is suitable for representing plants with uneven surfaces such as litchi. Instant-NRS is more suitable for reconstructing smooth and continuous plant surfaces such as watermelon and grape.

[8] Such as cleaning robots.
[9] Volume density-based NeRF model
[10] Signed Distance Fields-based NeRF model

Civil construction and structural health monitoring (SHM)

SHM is the process of evaluating and monitor the condition of structures using sensors and data analysis to detect damage or deterioration. Xu et al. (2024) (XU et al., 2024) used a Multi-tiling NeRF approach into reconstruction of structures using large-scale aerial datasets over the city of Dortmund and Bordeaux and compared their results to MVS pipelines against LiDAR reference data. Instead of training all the scene at once, the model decomposes a large-frame image into multiple tiled images with different camera models, allowing these small frame images to be fed into the training process as needed for specific locations without a loss of accuracy. NeRF return more complete and detailed results when compared to traditional techniques, while having higher statistical error when compared to SfM/MVS pipelines.

Pal et al. (2024) (PAL et al., 2024) proposed a framework for measuring the progress of civil construction activities through an Activity-Level Progress Monitoring System (ALPMS). This system uses photographs from the construction site and a building information model (BIM) and returns the completion percentage necessary for adjusting project schedules. It leverages novel views [11] generated from NeRF models to compare to planned BIM with deep learning-based semantic segmentation for progress reasoning, and estimates completion percentages. This information is transferred to a BIM for 3D visualization of the progress of the construction.

A recent and promising paper of Cui et al. (2024) (CUI et al., 2024) introduces NeRFusion, a novel model aiming to facilitate modeling of expansive building structures in large-scale scenes. By integrating recent NeRF-related advances into a fast-trainning, fast-inferring model, it produces higher PSNR and SSIM while being 15 times faster than state-of-the-art models. While this paper focus on architectural buildings, it could be leveraged to reconstruct bridge structures, thus this work was cited as suggestion for future works in this dissertation.

Comparison of features between reconstruction techniques

Table 1 further compares the main features of LiDAR-and photogrammetry-based point cloud acquisition and the potential advantages of NeRFs. LiDAR has the best results in accuracy and surface penetration, making it the best option for structural assessments requiring millimeter-level precision or environments with occlusion such as vegetation-covered bridges. However, the high equipment and deployment cost and limited color information restrict its usefulness in projects with budget constraints. Photogrammetry offers moderate accuracy at a lower cost, returning good results even using commonly available consumer cameras such as phone cameras. However, it relies on good light conditions and its unable of penetrating surfaces which can generate problems in shadowed or texture-poor bridge environments. NeRF is as a promising alternative with high color fidelity and ease of use, generating photorealistic 3D reconstructions

[11] In the context of NeRF, a novel view is a 2D image rendered from a previously unseen camera viewpoint, using a trained NeRF model that volumetrically integrates color and density along rays.

with moderate geometric accuracy from the same photos from photogrammetry. Yet, its high computational costs and processing time from neural network training can limit real-world scalability for large-scale infrastructure projects.

Feature	**LiDAR**	**Photogrammetry**	**NeRF**
Accuracy	High	Moderate	Moderate
Surface penetration	High	None	None
Light dependence	Low	High	High
Cost equipment	High	Moderate	Low
Computational costs	Low	High	High
Color information	Low	Moderate	High
Ease of use	Moderate	Moderate	High
Processing Time	Low	Moderate	High

Table 1 – Comparison of LiDAR, Photogrammetry, and NeRF based on various features.

2.6 Volumetric rendering

Volumetric rendering is a set of techniques that construct a 3D image from a series of 2D cross-sectional images (XU; SUN; LIANG, 2021). It is used in a range of applications, from medical imaging to gaming graphics (ZHANG; EAGLESON; PETERS, 2011; BECHER et al., 2024). In Computed tomography (CT) imaging, for instance, raw data is first acquired by projecting X-rays through the body from multiple different angles (PALAMADAI; YU, 2025). Sensors then measure the X-ray absorption as they pass through different tissues. The data collected from these projections is then used to generate a series of 2D cross-sectional images that represent the internal structures of the measured body from multiple different depths. Volumetric datasets are expressed by volume pixel (voxel) (IEEE, 2015). Volume rendering convey more information than surface rendering images at the cost of increased algorithm complexity and increased rendering times (KAUFMAN; MUELLER, 2005).

Traditional volume rendering

The history of computer vision is marked by the progressive evolution of visual representation. In 1965, Larry Roberts[12] discussed the possibilities of extracting 3D geometrical information from 2D perspective views of blocks (HUANG, 1996). By the 70s to mid 80s, surface rendering was extensively used for medical visualization (HERMAN; LIU, 1979; ARTZY, 1979; UDUPA, 1983; HERMAN, 1986; HOHNE; BERNSTEIN, 1986; LORENSEN; CLINE, 1987). Traditional surface rendering techniques worked well for clearly defined tissues such as high-density bone structure, whereas struggled with reconstruction of blurry, noisy and weakly

[12] Today known as the father of Computer Vision

defined surfaces such as thin tissues or structures with low-density-contrast to their surroundings (SRAMEK, 2006).

Then in the late 1980s, Levoy (1988) (LEVOY, 1988) and Drebin et al. (1988) (DREBIN; CARPENTER; HANRAHAN, 1988) introduced volume rendering, a new method for visualizing 3D data that improved upon older surface rendering techniques. Volume rendering solved the problem above by using transfer functions that encode data values such as density to colors and opacity values in order to create visuals directly from the raw data. This made it possible to visualize complex or ambiguous shapes that surface rendering couldn't.

Since its inception, volume rendering found applications in several fields such as medicine (GEE et al., 2003; ZHANG; EAGLESON; PETERS, 2011), archaeology (BERNARDES; MADEIRA; MARTINS, 2020), geology (CHAVES et al., 2011), computer graphics (WESTERMANN; ERTL, 1998; WEILER et al., 2000)

Traditional volume rendering techniques

There are four main 3D volume rendering algorithms in the literature: Ray casting or raymarching, splatting, shear-warp and 3D texture slicing (MCREYNOLDS; BLYTHE, 2005).

1. **Raymarching** is a technique for rendering implicit surfaces using geometric distance (HADJI-KYRIACOU; ARANDJELOVIĆ, 2021). Given a signed distance function (SDF) returning the distance to an object from a given point in space, sphere-tracing marches along a ray towards the closest intersection point of the surface iteratively.

2. The **Splatting** approach represents a volume as an array of reconstruction voxel kernels, which are classified to have colors and opacities based on the transfer functions applied. Each of these kernels leaves a footprint or splat on the screen, and the composite of all splats yields the final image (HUANG et al., 2002).

3. **Shear-warp** can be considered a hybrid between image-order algorithms, such as ray-casting, and object-order algorithms, such as splatting. In shear-warp, the volume is rendered by a simultaneous traversal of Run-Length-Encoded (RLE) voxel and pixel runs, where opaque pixels and transparent voxels are efficiently skipped during these traversals (SWEENEY; MUELLER, 2002).

4. **The 3D texture slicing** principle is similar to the shear-warp algorithm. It involves storing a set of three rectilinear volume data sets, and using them as three perpendicular stacks of object aligned texture slices (BENASSAROU et al., 2005). Slices are taken through the volume orthogonal to each of the principal axes and the resulting information for each slice is represented as a 2d texture that is then pasted onto a square polygon of the same size (Figure 4).

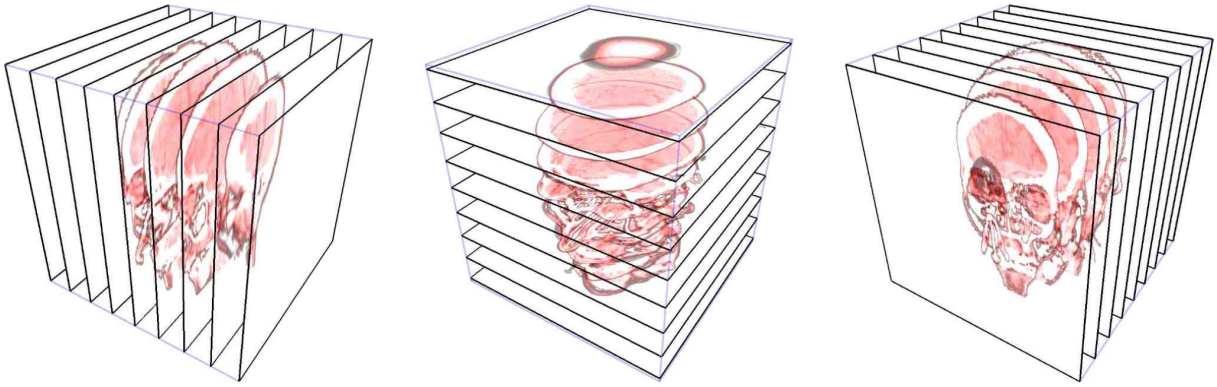

Figure 4 – Object-aligned slice stacks with 2d texture mapping. Adapted from (BENASSAROU et al., 2005).

Deep learning applied to volumetric rendering

The recent deep learning advancements marked a paradigm shift in several areas of study, volume rendering being one of them (WEIß, 2022). Neural networks have seen more and more applications in volume visualization and reconstruction, with several articles in recent years. Choy et al. (CHOY et al., 2016) introduces the 3D Recurrent Reconstruction Neural Network (3D-R2N2), a network which inputs a set of object images from arbitrary views and outputs the reconstructed object in form of a 3D occupancy grid. Wu et al. (WU et al., 2017) proposed the 3D Generative Adversarial Network (3D-GAN), a framework combining volumetric convolutional networks and adversarial training to generate 3D objects from latent space. Tatarchenko et al. (TATARCHENKO; DOSOVITSKIY; BROX, 2017) proposed an octree-based convolutional decoder that generates memory-efficient 3D shapes by predicting structure and occupancy, avoiding cubic complexity. It enables high-resolution outputs for autoencoders, single-image reconstruction, and scene generation, surpassing voxel grids in scalability and detail. Lombardi et al. (LOMBARDI et al., 2019) propose a volumetric neural approach using an encoder-decoder network and differentiable ray-marching, trained on multi-view images without explicit reconstruction.

Neural representations leverage the power of neural networks to provide a more complete, continuous representation of the studied object. Instead of explicitly representing each data sample as mesh or a point, neural representation encodes the data inside the weights of that particular neural network. The rendering process must be differentiable to be able to infer features about the data from images or reconstruct certain properties. Differentiability allows the optimization of parameters using gradient-based optimizers (WEIß, 2022).

3 Fundamentals of Neural Radiance Fields

The NeRF architecture is discussed in this chapter, detailing their rendering processes and optimization strategies, as described by Mildenhall et al. (MILDENHALL et al., 2020).

3.1 MLP architecture

The NeRF architecture employs a multilayer perceptron (MLP) consisting of nine fully connected layers, each with 256 neurons and ReLU activation functions. A skip connection is incorporated into the fifth layer, where positional encoding (VASWANI et al., 2023) is introduced to enhance the network's ability to capture fine spatial details (Figure 5). The ninth layer outputs the density parameter, σ. An additional tenth layer, with 128 neurons and ReLU activation, processes the remaining 256-dimensional feature vector. A final sigmoid-activated output layer predicts the emitted RGB radiance at position x, as observed along a ray with direction d (MILDENHALL et al., 2020).

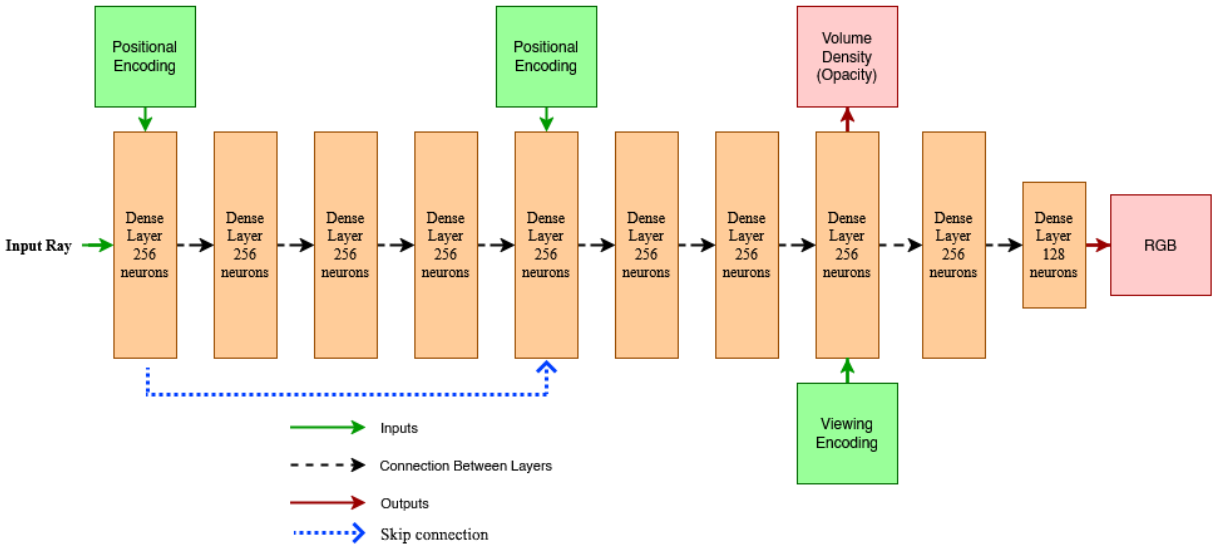

Figure 5 – Visualization of the NeRF's fully-connected network architecture.

3.2 Volume rendering

(MILDENHALL et al., 2020) models the 5D neural radiance field to represent a scene as the volume density and directional emitted radiance at any point in space. The color of any ray passing through the scene is rendered using principles from classical volume rendering (KAJIYA; HERZEN, 1984).

The volume density $\sigma(\mathbf{x})$ is interpreted as the differential probability of a ray terminating at an infinitesimal particle at location $\mathbf{x}$. The expected color $C(\mathbf{r})$ of camera ray $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$ with near and far bounds t_n and t_f is:

$$C(\mathbf{r}) = \int_{t_n}^{t_f} T(t)\sigma(\mathbf{r}(t))c(\mathbf{r}(t), d)dt, \quad T(t) = \exp\left(-\int_{t_n}^{t} \sigma(\mathbf{r}(s))ds\right). \tag{3.1}$$

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. Rendering a view from a continuous neural radiance field requires estimating this integral $C(\mathbf{r})$ for a camera ray traced through each pixel of the desired virtual camera.

This continuous integral can be numerically estimated using quadrature. Deterministic quadrature, which is typically used for rendering discretized voxel grids, would effectively limit the representation's resolution because the MLP would only be queried at a fixed discrete set of locations. Instead, a stratified sampling approach is used where its partitioned $[t_n, t_f]$ into N evenly-spaced bins and then draw one sample uniformly at random from within each bin:

$$t_i \sim \mathcal{U}\left[t_n + \frac{i-1}{N}(t_f - t_n), t_n + \frac{i}{N}(t_f - t_n)\right]. \tag{3.2}$$

Although a discrete set of samples is used to estimate the integral, stratified sampling enables representing a continuous scene representation because it results in the MLP being evaluated at continuous positions over the course of optimization. These samples are used to estimate $C(\mathbf{r})$ with the quadrature rule discussed in the volume rendering review by Max (MAX, 1995):

$$\hat{C}(\mathbf{r}) = \sum_{i=1}^{N} T_i(1 - \exp(-\sigma_i\delta_i))c_i, \quad T_i = \exp\left(-\sum_{j=1}^{i-1} \sigma_j\delta_j\right), \tag{3.3}$$

where $\delta_i = t_{i+1} - t_i$ is the distance between adjacent samples. This function for calculating $\hat{C}(\mathbf{r})$ from the set of (c_i, σ_i) values is trivially differentiable and reduces to traditional alpha compositing with alpha values $\alpha_i = 1 - \exp(-\sigma_i\delta_i)$.

3.3 Positional encoding

Positional encoding[1] is a method used to inject spatial information into neural networks, enabling them to effectively model high-frequency variations and complex spatial structures. It maps low-dimensional inputs, such as spatial coordinates (x, y, z) or viewing directions (θ, ϕ), into a higher-dimensional feature space using a series of sinusoidal functions. This technique was introduced in the context of Transformers (VASWANI et al., 2023) and has found applications in Large Language Models (KAZEMNEJAD et al., 2023), audio processing (PEPINO; RIERA;

[1] Positional encoding was inspired by Fourier feature mapping (TANCIK et al., 2020), a technique used to enable neural networks, especially MLPs, to better capture high-frequency signals in data, particularly in low-dimensional domains such as image and graphics processing

FERRER, 2022) and graph-structured data (CANTüRK et al., 2024). It was also adapted for use in Neural Radiance Fields (MILDENHALL et al., 2020), where capturing fine-grained spatial details is essential.

The positional encoding function is defined as:

$$\text{PE}(p) = \left[\sin(2^k \pi p), \cos(2^k \pi p)\right]_{k=0}^{L-1} \tag{3.4}$$

where p represents the input coordinate (e.g., x, y, z), and L determines the number of frequency bands used in the encoding (VASWANI et al., 2023). The canonical NeRF uses $L = 10$ frequencies for spatial coordinates (x, y, z) and $L = 4$ for viewing directions (θ, ϕ).

For a single input p, the positional encoding expands it into a $2L$-dimensional feature vector:

$$\text{PE}(p) \in R^{2L}. \tag{3.5}$$

For 3D spatial coordinates (x, y, z), the encoded representation is computed as:

$$\text{PE}(x, y, z) = \left[\text{PE}(x), \text{PE}(y), \text{PE}(z)\right], \tag{3.6}$$

resulting in a $6L$-dimensional feature vector for each 3D point. Similarly, for viewing directions (θ, ϕ) the encoded representation is:

$$\text{PE}(\theta, \phi) = \left[\text{PE}(\theta), \text{PE}(\phi)\right], \tag{3.7}$$

Positional encoding is crucial in overcoming the limitations of standard neural networks, which struggle to model high-frequency signals due to the inherent smoothness of activation functions like ReLU. By incorporating Fourier features, the network gains the capacity to represent fine details and sharp edges in data, significantly improving the quality of reconstructions and representations in tasks like 3D rendering and image synthesis.

Why the positional/viewing encoding on layer 1 and 4? (MILDENHALL et al., 2020) noted that the basic implementation of optimizing a neural radiance field representation for a complex scene does not converge to a sufficiently high resolution representation and is inefficient in the required number of samples per camera ray. They addressed these issues by transforming input 5D coordinates with a positional encoding that enables the MLP to represent higher frequency functions. Simply applying a Fourier feature mapping to the input (x, y) points before passing them to the network allows for rapid convergence.

3.4 Hierarchical volume sampling

NeRF's rendering strategy of densely evaluating the neural radiance field network at N query points along each camera ray is inefficient: free space and occluded regions that do not

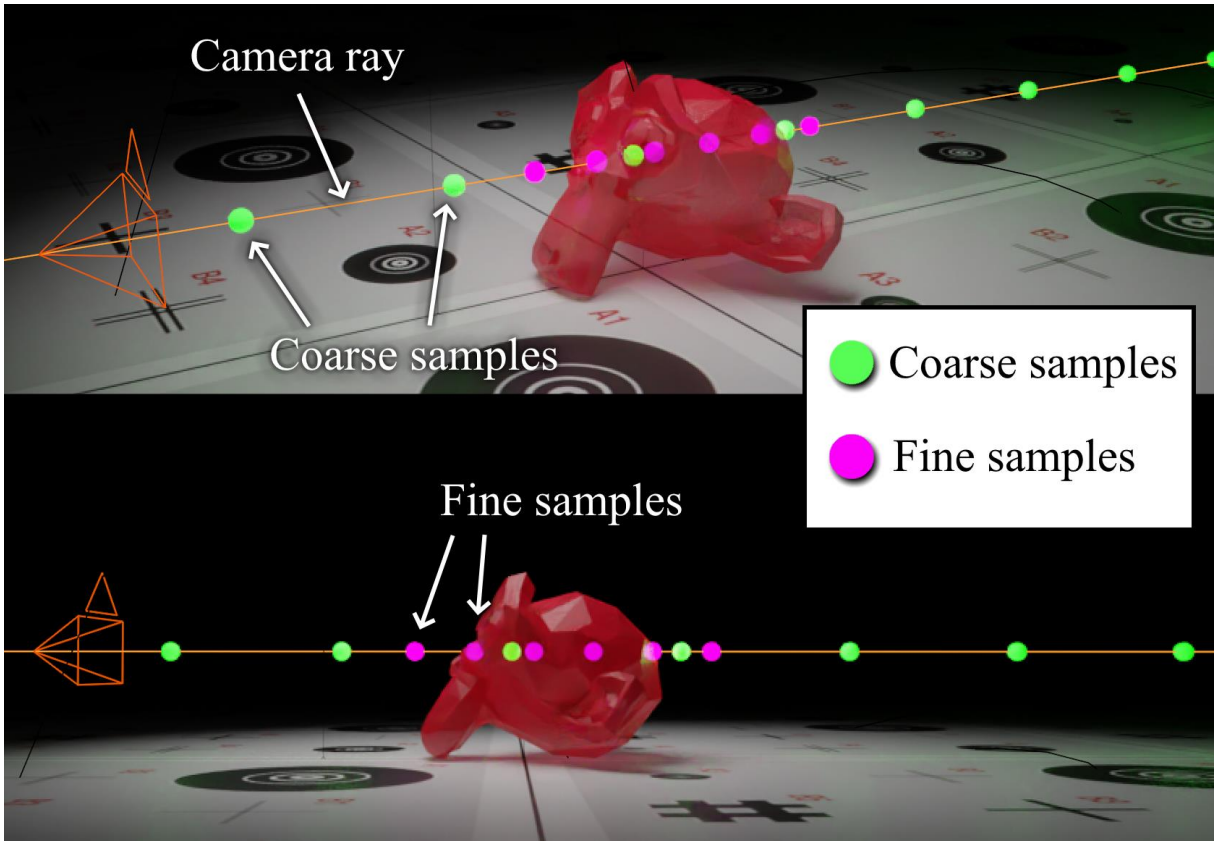

Figure 6 – Visual representation of NeRF's hierarchical sampling. First a set of coarse points (in green) are uniformly sampled along a particular ray using a coarse network to estimate general scene density. This initial density prediction is then used to form a probability distribution that guides the fine network to finely sample additional points (in pink) in regions that likely contain important details.

contribute to the rendered image are still sampled repeatedly. An hierarchical representation is proposed by Mildenhall et al. (2020)(MILDENHALL et al., 2020) that increases rendering efficiency by allocating samples proportionally to their expected effect on the final rendering.

Instead of just using a single network to represent the scene, two networks are simultaneously optimized: one "coarse" $(N_c = 64)$ and one "fine" $(N_f = 192)$. First a set of N_c locations are sampled using stratified sampling, and evaluate the "coarse" network at these locations as described in equations (3.2) and (3.3). Given the output of this "coarse" network, it is then produced a more informed sampling of points along each ray where samples are biased towards the relevant parts of the volume (Figure 6). To do this, first the alpha composited color is rewritten from the coarse network $\hat{C}_c(\mathbf{r})$ in equation (3.3) as a weighted sum of all sampled colors c_i along the ray:

$$\hat{C}_c(\mathbf{r}) = \sum_{i=1}^{N_c} w_i c_i, \qquad w_i = T_i(1 - \exp(-\sigma_i \delta_i)). \tag{3.8}$$

Normalizing these weights as $\hat{w}_i = w_i / \sum_{j=1}^{N_c} w_j$ produces a piecewise-constant Probability Density Function (PDF) along the ray. A second set of N_f locations is sampled from this

distribution using inverse transform sampling, evaluate the "fine" network at the union of the first and second set of samples, and compute the final rendered color of the ray $\hat{C}_f(\mathbf{r})$ using the equation (3.3) but using all $N_c + N_f$ samples. This procedure allocates more samples to regions expected to contain visible content. This addresses a similar goal as importance sampling, but the sampled values are used as a non-uniform discretization of the whole integration domain rather than treating each sample as an independent probabilistic estimate of the entire integral.

3.5 Loss function

During training of deep learning models, a loss function is used to optimize the model's parameters (TERVEN et al., 2024). It measures the difference between the predicted and expected outputs of the model. The objective of training is to minimize this difference as the training iterations occur. Given the vast array of deep learning applications, different loss functions can be applied to different kind of problems. For instance, using 0-1 loss, perceptron loss, logarithmic loss, sigmoid cross entropy loss, and others for classification problems and Absolute loss, Huber Loss, Log-cosh loss and others in regression problems (WANG et al., 2020). In image models, the loss function is computed by comparing the colors of a synthesized image with the ground-truth image (ZHANG et al., 2022). Traditional image models rely on the L2 norm, also known as the Euclidean Norm, for image reconstruction (ZHAO et al., 2017).

In the context of NeRF training, the Loss function defined at Equation 3.9 was used to measure the difference between a predicted image $\hat{C}_p(r)$ and a ground truth image $C(r)$. The loss function of the NeRF is the total squared error between the rendered and true pixel colors for both the coarse $(N_c = 64)$ and fine $(N_f = 192)$ renderings:

$$L = \sum_{r \in R} \left[\|\hat{C}_c(r) - C(r)\|_2^2 + \|\hat{C}_f(r) - C(r)\|_2^2 \right]. \tag{3.9}$$

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.

3.6 Training metrics

Peak signal-to-noise ratio (PSNR)

PSNR is the golden standard parameter for quality evaluation of a reconstructed image. It compares the predicted image $\hat{C}_p(r)$ with the ground truth image $C(r)$ in terms of the Mean Squared Error (MSE):

$$\text{MSE} = \frac{1}{N_{images}} \sum_{i}^{N_{images}} (\hat{C}_p(i) - C(i))^2 \tag{3.10}$$

Where N_{images} is tne number of images being compared. Then, PSNR is computed as:

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\text{MAX}^2}{\text{MSE}} \right) \tag{3.11}$$

where MAX is the maximum possible pixel value (e.g., 255 for an 8-bit image). A higher PSNR value corresponds to a lower MSE and indicates better image quality. However, PSNR is not sensitive to minor adjustments such as slight changes in brightness, contrast, or saturation because it is normalized relative to the maximum pixel value. This normalization reduces its responsiveness to subtle variations in these parameters (RABBY; ZHANG, 2024). PSNR is typically measured in decibels (dB) and higher values represent a higher quality of reconstructed image.

Structural similarity index measure (SSIM)

The SSIM was first introduced by Wang et al. (2004)[2] (WANG et al., 2004) as a tool for quality assessment that compares local patterns of pixel intensities that have been normalized for luminance and contrast. SSIM is defined by:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{3.12}$$

where μ_x, μ_y, σ_x and σ_y are the means and standard deviations of both the original and reference images respectively, C_1 and C_2 are constants given by $C_1 = (K_1L)^2$, $C_2 = (K_2L)^2$, where L is the dynamic range for the sample data $L = 255$ for an 8-bit gray-level image, and $K_1 \ll 1$ and $K_2 \gg 1$ are two scalar constants. (NASR; ALRAHMAWY; TOLBA, 2017). SSIM ranges from 0 (No structural similarity between images) to 1 (two identical images).

Learned perceptual image patch similarity (LPIPS)

Traditional image assessing parameters such as PSNR and SSIM fail to address many nuances of human perception as they assume pixel-wise independence, whereas deep perceptual similarity metrics compute distances between image features extracted by deep neural networks (DOSOVITSKIY; BROX, 2016). Proposed by Zhang et al. (2018) (ZHANG et al., 2018), the LPIPS (Figure 7) leverage deep neural networks trained on datasets with human perceptual judgments[3] on geometric distortions[4] from uncompressed images to infer the perceptual similarity

[2] This paper was groundbreaking to say the least. As of 26/02/2025 Google Scholar states it has received 61.096 citations!

[3] "Given 2 distorted images, for instance, which is more similar to the ground truth reference image?"

[4] Photometric distortions, random noise, blurring, spatial shifts and corruptions, and compression artifacts.

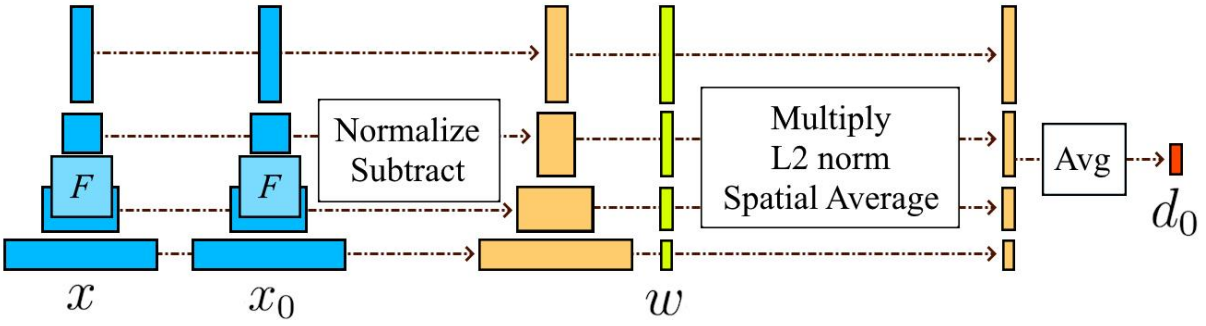

Figure 7 – LPIPS architecture. Given two pictures x and x_0 deep embeddings are calculated to compute the distance d_0 representing their perceptual similarity. Adapted from Zhang et al. (2018) (ZHANG et al., 2018)

between two images. LPIPS ranges from 0 (no perceptual difference) to 1 (Large perceptual differences, poor reconstruction quality).

4 Framework for NeRF-based Point Cloud Generation

This work follows the general framework illustrated in Figure 8: First data acquisition (photos or videos) is performed using consumer-grade RGB cameras (smartphones or drones), followed by general image pre-processing, camera-pose estimation via triangulation, NeRF network training for point-cloud extraction, and post-processing of the resulting point cloud. This chapter aims to present main NeRF steps.

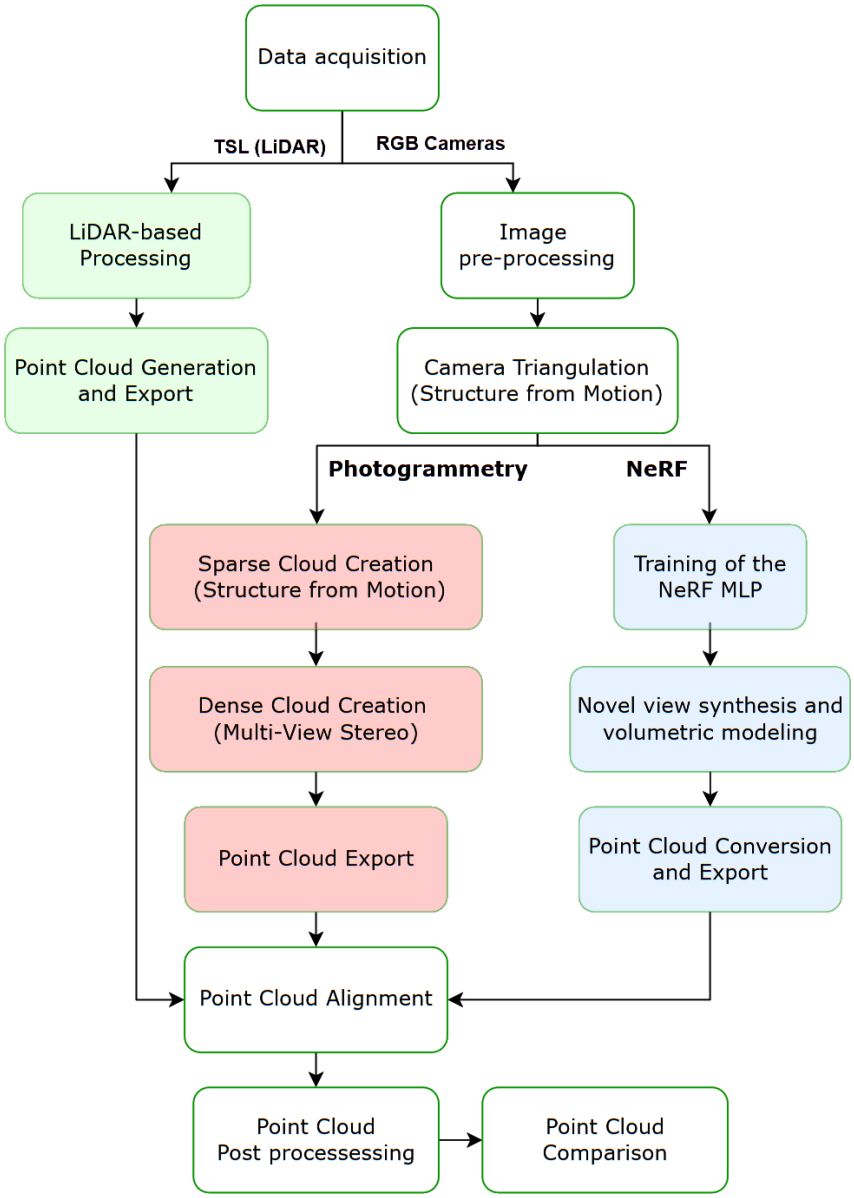

Figure 8 – General Methodology for Generating NeRF-based point clouds of Critical Structures.

4.1 Data acquisition

For comparison purposes, two main techniques are leveraged in this work: Terrestrial Laser Scanners (TLS) for the LiDAR approach and RGB sensors for the Photogrammetry and NeRF approach.

4.1.1 RGB data

This study uses consumer-based RGB-cameras to obtain input data for the Photogrammetry and NeRF pipeline. Thus, it is recommended the use of cameras capable of at least FullHD capture for this part, mainly phone cameras or airborne drone cameras if available. Also, phone-based cameras have often small sensor sizes, which can limit light capture and introduce noise. Photon noise (shot noise) occurs in low-light conditions due to limited photons reaching the sensor. To avoid noise-related issues, it is worth capturing pictures of well illuminated structures. Moreover, lossless formats such as PNG or RAW, although large and noise-prone, should be used when producing pictures. Lossy picture formats such as JPEG should be avoided as they discard high-frequency details causing blocking artifacts and aliasing.

RGB cameras

(MILDENHALL et al., 2019) suggests using pictures taken from a grid-like pattern in front of the study object. They must have at most 64 pixels of disparity in order to maintain perceptual quality within Learned Perceptual Image Patch Similarity (LPIPS) metric (ZHANG et al., 2018). To achieve a comprehensive representation of the studied object, it is advisable to capture continuous images from multiple azimuthal angles, ideally following a helical pattern (Figure 9) going from lower to higher relative height.

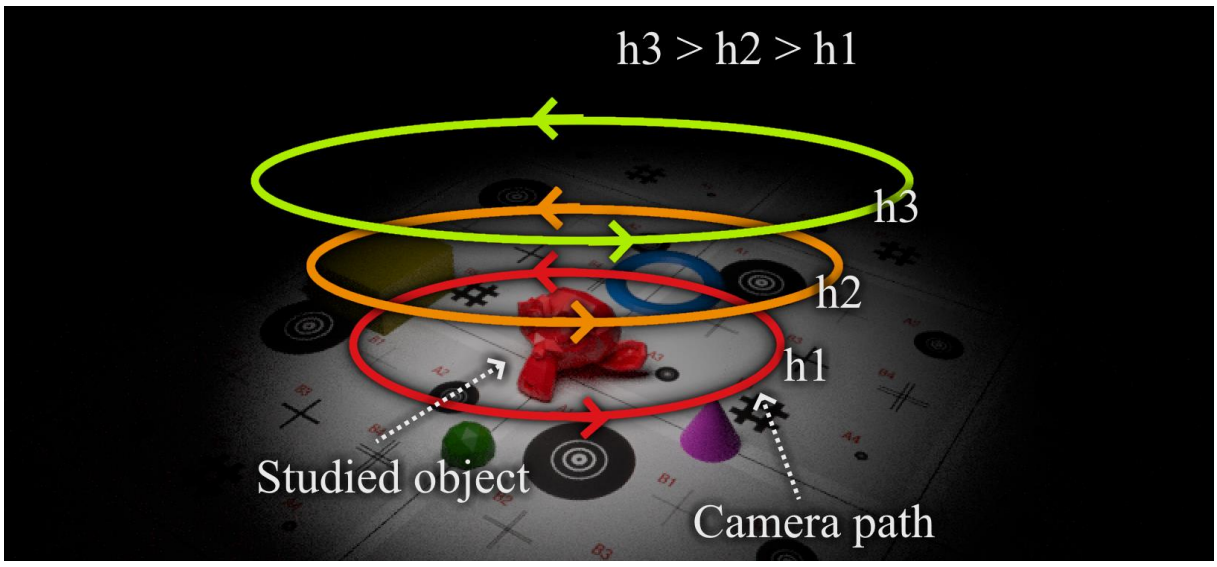

Figure 9 – Suggested helical pattern for image capturing.

Image pre-processing

It would be ideal if ultra high-resolution pictures, 100MP (11.608 x 8.708 px) for instance, could be used for photographic capture. However often high-resolution images are too memory demanding of one's hardware, so images should be downscaled. Also, images with blur, overexposed (too bright) or underexposed (too dark) should be avoided since they generate errors in the final reconstruction. Motion blur introduces distortions that degrade feature detection and extraction. It occurs because of slow shutter speeds or camera motion. Overexposed pictures often have saturated highlights while underexposed pictures often exhibit photon noise.

In this work, all photos were taken at or downscaled to FullHD (1920 x 1080 px) resolution. Blurred, overexposed or underexposed pictures were discarded from further processing, although they have small impact if used in minimal quantities since NeRF neural network would learn to ignore them when compared to better taken pictures.

Camera triangulation

Accurate camera poses (extrinsic parameters) are required as input to both the photogrammetry and NeRF pipelines. In simulated environments, such as software generated scenes, the task of obtaining extrinsic and intrinsic camera parameters of each photo is trivial. However, for phone-based photos this becomes challenging. Thus, it is necessary the use of an triangulation algorithm applied on these RGB pictures to obtain their locations in a coordinate system. Usually the Structure from Motion algorithm is used for this task, as one of its outputs is the cameras coordinates from input images. A suite called COLMAP (SCHONBERGER; FRAHM, 2016) is commonly used in the literature of this task for triangulation before the NeRF pipeline.

4.1.1.1 NeRF training

Here are discussed the steps of the NeRF MLP training adopted in this work, the novel view synthesis from the training model and the extraction of point-cloud data.

Model architecture

An architecture similar to the canonical NeRF architecture (shown on Figure 5) was adopted in this work: A base MLP configuration with 8 layers, each processing 256 channels with a ReLU activation function, followed by 2 layers processing 128 channels also activated by a ReLU function. Skip connection is applied to layer 4 to help with gradient propagation. Two embedding dimensions are set: a positional encoding of 10 frequencies on layer 1 and a viewing direction encoding of 4 frequencies on layer 9. The Rectified Adaptive Moment Estimation (RADAM[1]) optimizer (LIU et al., 2021a) is used for learning rate optimization with a learning rate of 5e-4.

[1] Rectified Adam, or RAdam, is a variant of the Adam stochastic optimizer that introduces a term to rectify the variance of the adaptive learning rate. It seeks to tackle the bad convergence problem suffered by Adam. The

Why the skip connection on layer 4? (PARK et al., 2019) observed that without skip connections a plateau in training loss ocurred past 4 layers. However, with skip connections the training loss continues to decrease although with diminishing returns past 12 layers.

Loss function

For training purposes, the Loss function (Equation 3.9) was used with 64 as the number of coarse Samples and 192 as the number of fine samples, as used in the canonical NeRF.

Training schedule

Based on the observations in this work, it is recommended an absolute minimum of 30k iterations for reconstruction if only the rough outline is desired of bigger structures in the scene, such as the main deck, pier or bridge towers. However, if finer and smaller parts of the structure is of interest, a higher number of iterations is necessary, from 100k to 300k iterations in some cases. For instance, the cable-stays in the Octavio Frias Bridge (Figure 36) appear after 25 thousand iterations. In this work, all models were trained with 100k iterations within Nerfstudio suite (TANCIK et al., 2023). Each model took approximately 3 hours to train using the hardware mentioned in Table 2.

It is expected that the NeRF depth estimation improves progressively over the training iterations, as depicted in Figure 10. At the beginning of the training, at early iterations, the reconstructed scene exhibits significant artifacts, with blurry and less defined structural elements of the bridge. Then, as the training advances, the model begins to refine the scene geometry, reducing distortions but still presenting inconsistencies in certain regions. By iteration 20131, the depth representation becomes more accurate, capturing finer structural details such as vertical support beams with greater clarity. Thus, one may correctly infer that if a higher LoA is necessary, the NeRF model should be trained for more iterations to achieve higher precision in their reconstruction. All processing through training made in this study was conducted in a computer described in table 2.

Component	**Specification**
CPU	AMD Ryzen 7 5700G
GPU	Galax RTX 4060 8GB
RAM Memory	Juhor 16GB DDR4 2666MHz
Solid State Drive	Kingston SA2000M8/500G

Table 2 – Hardware Specifications

authors argue that the root cause of this behaviour is that the adaptive learning rate has undesirably large variance in the early stage of model training, due to the limited amount of training samples being used. Thus, to reduce such variance, it is better to use smaller learning rates in the first few epochs of training - which justifies the warmup heuristic.

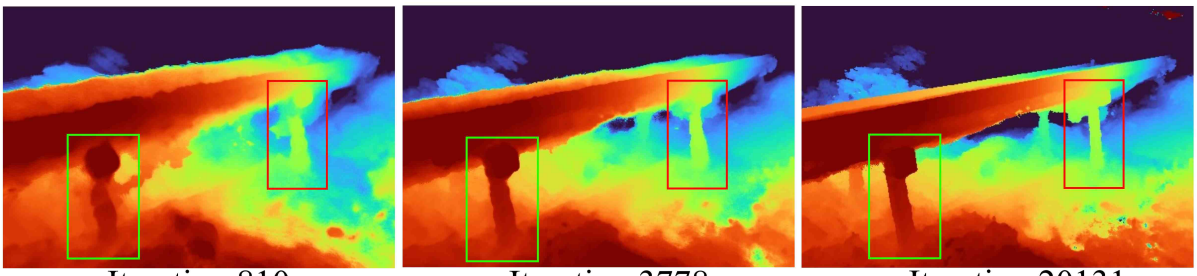

Figure 10 – Comparative visualization of depth estimations across three progressive training iterations.

Novel view synthesis and volumetric modeling

At this point, the trained NeRF model is ready for final novel view synthesis. The objective here is to transform the volume density profile along one ray into a pixel value. This process is repeated for every pixel in the image to obtain the rendered image. To assess the contribution from all points along a ray, the classical rendering equation[2] is used. For NeRF models, the rendering equation can be simplified to a sum since a finite number of samples is used along the ray. The sum of all RGB values from all points along a ray are summed, as shown at the equation (3.3). As mentioned before, the probability of a ray stopping at any point when going from the camera to the object is considered: The higher the opacity of that particular point, the higher the probability of that particular ray stopping at that point and the higher the probability that the RGB value of that point having an impact on the rendered pixel.

Point cloud extraction and export

Extracting explicit geometry information in the shape of a point cloud from an implicit volumetric representation of a scene from a trained NeRF model is done by sampling a batch of rays from different viewing angles. Using these rays, the NeRF model infers their RGB color and opacity for every sampled 3D coordinate point. Since the camera origins and directions of every ray are known, the distance/depth of every sampled point to the camera origin are also known from the camera extrinsics. The distance/depth from the camera's origin to the sample point is encoded into 3D coordinates in space.

$$P = \{x_i \mid \sigma(x_i) > \sigma_{threshold} \text{ for } x_i \in \mathbb{R}^3\}, \tag{4.1}$$

where $\sigma(x_i)$ is the predicted density at point (x_i) and $\sigma_{threshold}$ is a predefined opacity threshold used to extract surface points. Points below certain $\sigma_{threshold}$ (often 0.9) or outside a bounding box are discarded. The remaining valid points are gradually collected, with the process repeating until an arbitrary number of high quality points have been gathered. Finally, the accumulated 3D

[2] The rendering equation was first introduced by the groundbreaking paper from Immel et al. (1986) (IMMEL; COHEN; GREENBERG, 1986). It is an integral equation in which the equilibrium radiance leaving a point is given as the sum of emitted plus reflected radiance under a geometric optics approximation.

points and their color data can be exported into a standardized point cloud format for further processing.

An advantage of implicit representations is the continuous nature of the reconstructed model. Therefore, given sufficient memory resources, point clouds containing an arbitrary number of points can be generated from any region of the model. In this work, it was established that the generated NeRF point clouds should contain between 2 million points (such as USP footbridge) to 10 million points (such as Octavio Frias Bridge) before post-processing.

4.2 Point cloud post processing

The resulting point clouds, while dense, are often in need of post processing due to diverse factors. Outliers, mixed points, holes, and other defects may be generated in the target point cloud due to the external environment, the discreteness of the scanning technique and occlusion of objects (ZHAO et al., 2021). Decompressed point clouds can suffer of quantization distortions, which lower the level of detail of these reconstructions (FAN et al., 2022). These defects degrade the quality of the point cloud and affect the subsequent reconstruction accuracy of the three-dimensional surface. Thus these issues must be addressed by processing/filter the point clouds at this moment, removing spurious points outside the scope of the object under study, denoising[3] where needed, then scaling, aligning and registering for further comparison. However, the extracted point cloud's coordinates do not conform to any real-world unit system. Therefore, the cloud must be rescaled to the appropriate measurement units. Point removal and rough scaling was mainly made using Blender software (COMMUNITY, 2018), while alignment and registering was made using Cloud Compare suite (CLOUD. . . , 2025).

Point cloud scaling and alignment

NeRF and SfM-MVS Photogrammetry point clouds are often generated in a coordinate system that differs from the metric system (meters). Time-of-flight sensors usually generate points scaled in milimeters or centimeters (which can be considered ground truth for our purposes) while image-based point clouds are within their own local coordinate system. Point clouds cannot compared with different scales, so an scaling process is needed, where all points of an arbitrary-coordinate-system is brought to the metric coordinates. Two approaches can be considered:

1. If a ground truth point cloud (often LiDAR-based) in the metric system is available, one can scale all points of the desired point cloud to the first one.

[3] One interesting suggestion comes from Reuter et al (2005) (REUTER et al., 2005). Their paper describes a technique involving the projection of point-clouds into a continuous surface to minimize a local least-square error. This filter out high frequency noise normally present in raw data.

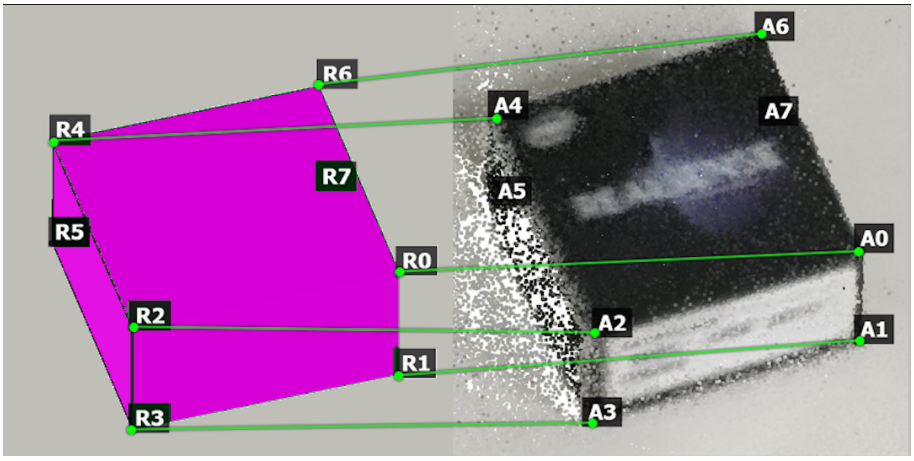

Figure 11 – Mesh with correct measurement scale with their vertices selected (Left) and the point cloud of the object to be transformed in scale with their equivalent vertices also selected (Right).

2. If no ground-truth model is available, one can scale the point cloud using a reference object of known X,Y and Z dimensions within the scene. This approach has been successfully used in the work of (HUANG et al., 2022), where several 30×30×0.3cm conveniently placed control points were used to align their point clouds.

For the second case, an object of known measurements such as a box near the studied object can be used to help scaling the cloud to the same size of the known object. At first a 3D mesh of the object of known measurements is created with the correct scale. Then equivalent vertices are selected from the object in the cloud and in the mesh (Figure 11). Finally, a transformation is computed by minimizing the Root Mean Square Deviation (RMSD) between the matched points, allowing the point cloud to be rescaled to the reference mesh's scale. It is recommended using objects with a similar size to the studied object in order to lower the scaling error of imperfect point picking.

Point cloud registration

The alignment process all but roughly approximates clouds of different origins to a reference cloud. However, fine adjustment is needed in order for both clouds accurately overlay each other. Registration is a more precise process where one point cloud is finely adjusted to match another through iterative methods such as Iterative Closest Point (ICP).

The classical ICP algorithm iteratively obtains the optimal transformation matrix using least squares and uses a point-to-point metric for registration (YANG et al., 2025). Given a source point cloud p and a target point cloud q with 3D coordinates (p_x, p_y, p_z) and (q_x, q_y, q_z), respectively. the goal is to find an optimal transformation $T = (R, t)$, where R is a rotation matrix and t is a translation vector, such that the distance between corresponding points is minimized. The objective function for ICP is defined as:

$$E(R,t) = \sum_{i=1}^{n_p} \|q_i - (Rp_i + t)\|^2, \tag{4.2}$$

where n_p is the number of points between two point clouds being compared and $\|\cdot\|$ denotes the Euclidean norm. One should register the cloud of lower confidence into the cloud of higher confidence. In this work, photogrammetry and NeRF clouds were registered into LiDAR clouds which are considered as ground truth. This option is used to compare the quality and average distance to a cloud obtained by a common method in the industry. However, when there is a direct photogrammetry/NeRF comparison without a LiDAR cloud, the NeRF cloud was registered into the photogrammetry cloud.

5 Experimental framework

In this chapter, we present the bridges selected for the LiDAR–Photogrammetry–NeRF comparison, outline the imaging sensors employed in this study, and conclude with a discussion of the performance metrics used for assessment.

5.1 Real-world bridges

In this work is focused on four bridges, one of a smaller scale, two medium scale bridges and finally a much larger scale bridge. For the two first bridges, the Tucunduba river bridge and the Taurizinho River bridge drone-based videos were obtained to reconstruct their models. For the last two bridges, the USP footbridge and the Octavio Frias Cable-stayed bridge LiDAR-based point cloud was also available, which was used as a ground-truth model for comparison. First some history and photos of the studied structures are exhibited, then it is shown how the data for the study was collected.

5.1.1 Tucunduba river bridge

The Federal University of Para (UFPA), Guamá Campus, is crossed by the 19m wide Tucunduba River. There is a car bridge connecting both sides of the campus (Figure 12). Therefore, the bridge is subjected to real-life operational and environmental conditions, allowing one to verify the applicability of NeRFs for structures outside the near-ideal laboratory environment.

For the Tucunduba river bridge, the data acquisition was made with a DJI Mini 2 drone, which was flown over the bridge and recorded a 4k video. 300 equally spaced frames were taken

Figure 12 – Side view of the Tucunduba river bridge in the UFPA campus

from this video, downsized to Full HD and served as input data for both photogrammetry-based and NeRF-based dense point cloud creation (Figure 13). Since a LiDAR-based point cloud ground truth is not available for this specific structure, the distance between the NeRF-based cloud was compared directly to the photogrammetry-based cloud using the least square plane local modeling.

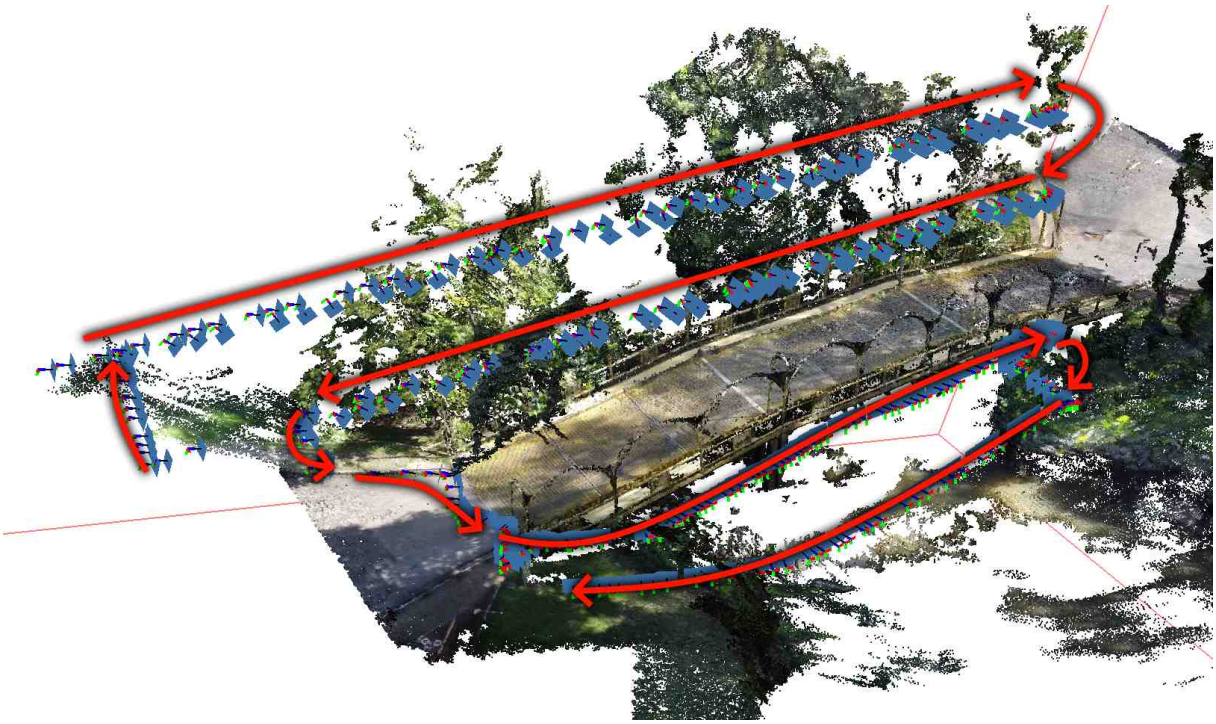

Figure 13 – Camera angles of the photos taken by the drone over the Tucunduba river bridge.

5.1.2 Tauarizinho river bridge

The Tauarizinho river bridge is a 122m wide concrete bridge located over the Tauarizinho River, Marabá/PA (Figure 14). It connects the city of Marabá and São Domingos do Araguaia. For this bridge, the data acquisition was made with a DJI Mini 2 drone, which was flown over the bridge and recorded a 4k video. 200 equally spaced frames were taken from this video, downsized to Full HD and served as input data for both photogrammetry-based and NeRF-based dense point cloud creation (Figure 15). Since a LiDAR-based point cloud ground truth was not available for this specific structure, the distance between the NeRF-based cloud was compared directly to the photogrammetry-based cloud using the least square plane local modeling.

5.1.3 USP footbridge

The University of São Paulo (USP) Bridge is a pedestrian footbridge spanning the Tejo stream, strategically positioned between the Mining Engineering and Civil Engineering buildings on the USP-SP campus in São Paulo, Brazil. The bridge features a deck with a 12-meter span, constructed from cross-sectional timber elements measuring 7 cm × 16 cm, which

Figure 14 – Side view of the Taurarizinho river bridge

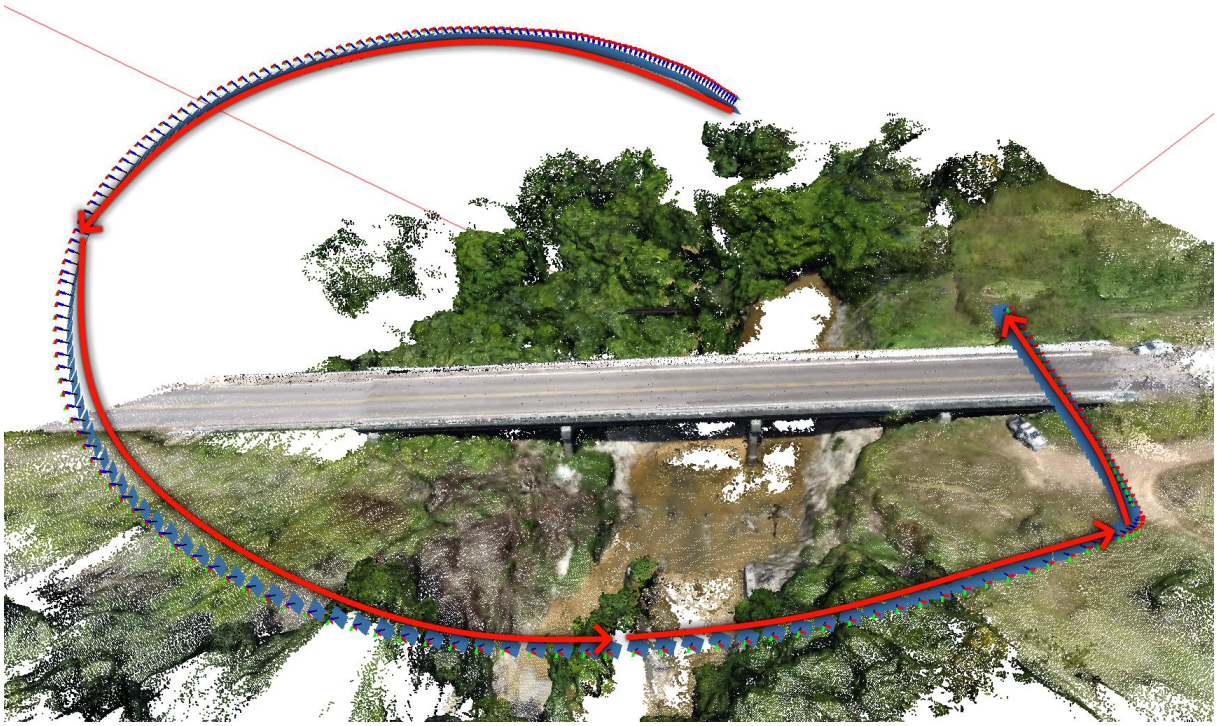

Figure 15 – Camera angles of the photos taken by the drone over the Taurizinho River Bridge.

are transversely prestressed and overlaid with reinforced concrete. The prestressing system incorporates a metallic diverter positioned beneath the deck at mid-span, anchored by tendons at both deck ends. The superstructure is supported by reinforced concrete blocks repurposed from a pre-existing structure on the site. Additionally, on the Mining Engineering side access, a precast reinforced concrete slab approximately 2 meters in length spans the gap between the support

pillars and the abutment (Laboratório de Sistemas Estruturais Ltda., 2006).

Figure 16 – The USP Bridge over the Tejo stream, São Paulo, Brazil

A total of 348 photographs of the USP footbridge were captured using a phone Poco X3 Pro from various azimuths and used as input for photogrammetry and NeRF-based point cloud generation. All images were taken at a resolution of 960 × 1280 pixels. Figure 17 illustrates the camera angles corresponding to each image. The generated point clouds were evaluated against a LiDAR-based point cloud produced using a Leica RTC360, a high-performance 3D laser scanner widely utilized in construction and architectural applications. The RTC360 point cloud served as the ground truth, providing a reliable benchmark for assessing accuracy and precision. The comparison employed K-Nearest Neighbors for local error modeling. To further analyze error behavior relative to the ground truth, the experiment was iteratively performed with reduced photo subsets, generating point clouds using 75% (261 photos), 50% (174 photos), 25% (87 photos), 10% (35 photos), and 5% (17 photos) of the original dataset.

5.1.4 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 (Pedro Manuel Calas; MURIETA, 2015). 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 Nações 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

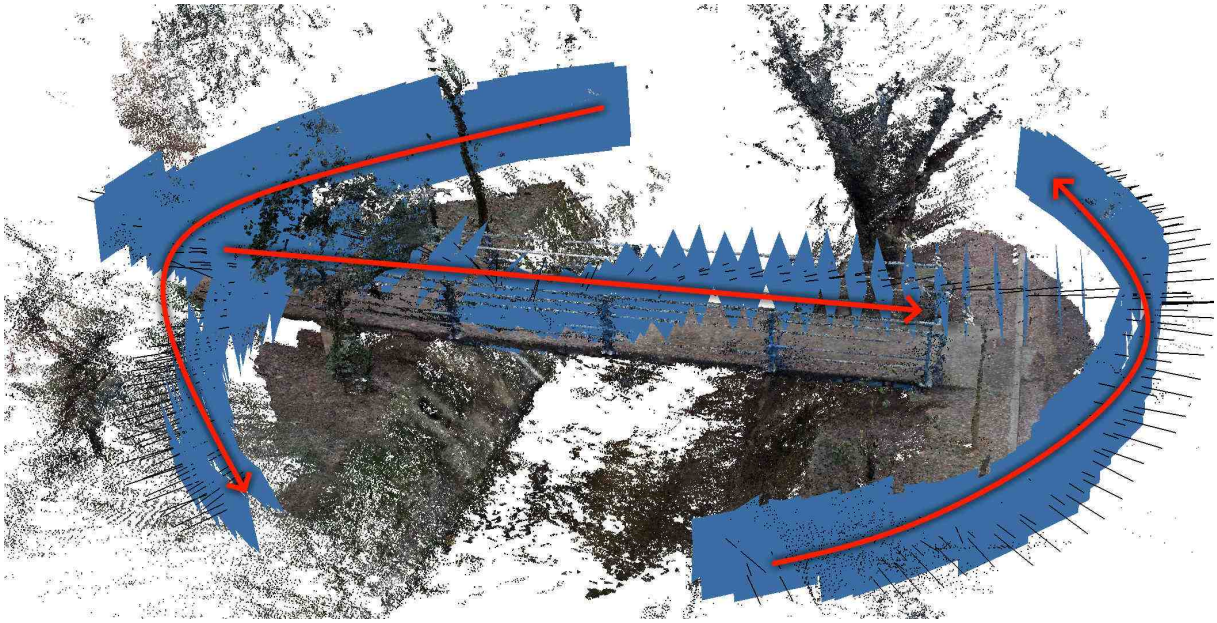

Figure 17 – Camera angles and positions of the photos taken over the USP Bridge.

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 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.

Figure 18 – The Octavio Frias de Oliveira Cable-stayed Bridge crossing the Pinheiros River

For the Octavio Frias Bridge, 500 equally spaced frames were selected from a YouTube

video (DRONE 4K, 2022) filmed by a DJI Mini 2 drone. Figure 19 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, point cloud models are created 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).

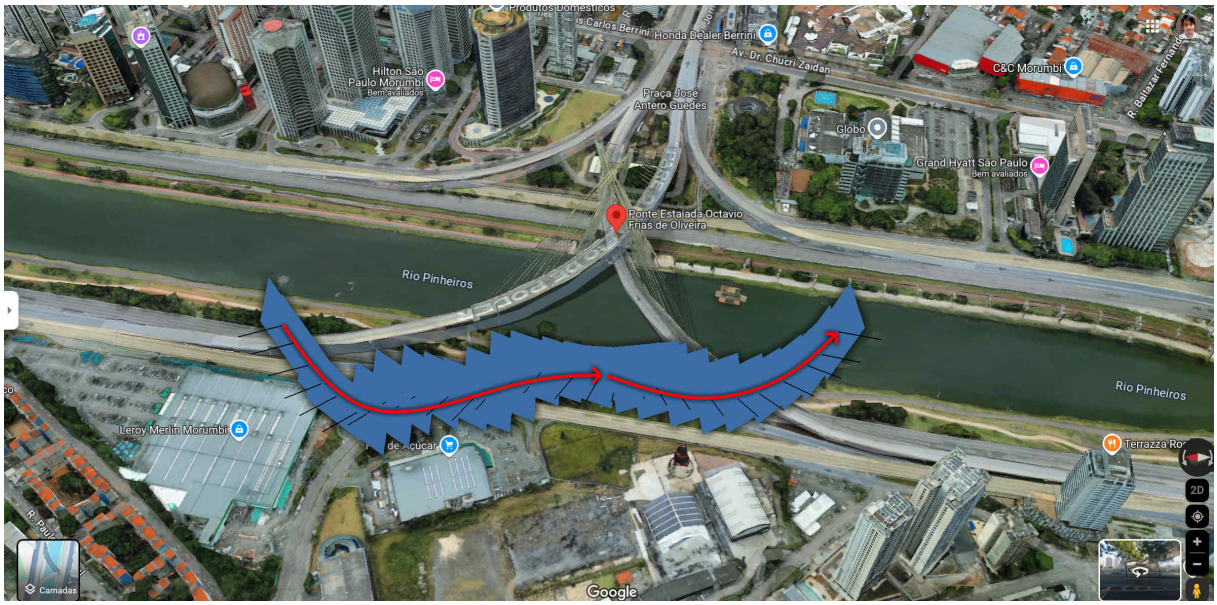

Figure 19 – Camera angles of the photos taken by the drone over the Octavio Frias Bridge. Source: Google Maps

5.2 Imaging sensors

Leica RTC360 LiDAR

The Leica RTC360 terrestrial laser scanner (Figure 20) was utilized to generate the ground truth point clouds, serving as a reference standard for comparison with point clouds derived from NeRF and photogrammetry-based methods. Its operational range of up to 130 meters under standard conditions renders it well-suited for monitoring extensive structures, including bridges, high-rise buildings, and industrial facilities. The device achieves a positional accuracy of 3mm at 10 meters, meeting the precision requirements of SHM applications. In particular, its application in structural monitoring has been demonstrated in studies involving bridges (ZHOU et al., 2024) and salt mines (ŚWIERCZYŃSKA; KURDEK; JANKOWSKA, 2024). This LiDAR was used in the USP footbridge and the Octavio Frias Cable-Stayed Bridge.

Xiaomi Poco X3 Pro

Released in March 2021, The Xiaomi Poco X3 Pro is a consumer grade cellphone with a 48 MP, a f/1.8 aperture, a 1/2.0" sensor size, a 0.8µm pixel size and a Phase Detection Autofocus function. It was used for the data acquisition of the USP footbridge on the default out-of-the-factory configuration.

DJI Mini 2 drone

Launched in March 2022, the DJI Mini 2 is a consumer grade drone. It is equipped with a 12 MP, a f/2.8 aperture and a 1/2.3” sensor size camera. It was used for the data acquisition of the Tucunduba river bridge, the Tauarizinho river bridge and the Octavio Frias Bridge.

Figure 20 – The Leica RTC360

5.3 Performance metrics

In this work there were chosen 2 main parameter groups for performance assessment: Statistical Cloud-to-Cloud parameters and industry-used Level of Accuracy (LoA) parameters.

Statistical parameters

When comparing distances between point clouds, several techniques can be employed, including *nearest neighbor* and *nearest neighbor with local modeling*. The nearest neighbor approach computes the distance between a 3D point in one cloud and the closest Euclidean point in the other cloud. The Euclidean distance $d(\mathbf{p}, \mathbf{q})$ between two 3D points $\mathbf{p}$ and $\mathbf{q}$ with coordinates (p_x, p_y, p_z) and (q_x, q_y, q_z), respectively. is calculated as:

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2 + (p_z - q_z)^2} \tag{5.1}$$

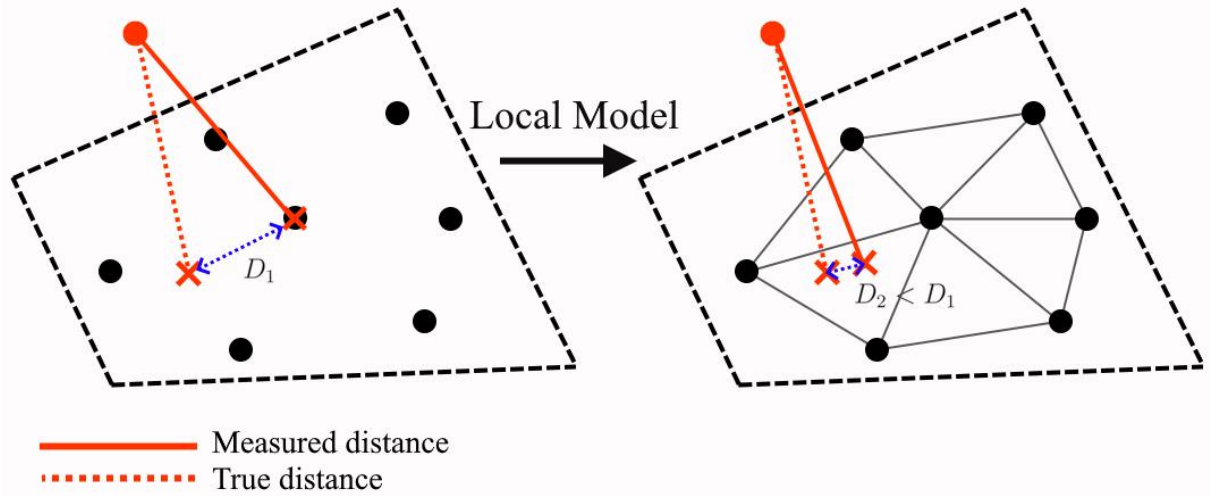

Figure 21 – Distance to the nearest neighbor and to the nearest neighbor with local modeling.

However, as point clouds represent discrete approximations of continuous surfaces, this method may yield inaccuracies relative to the true continuous geometry (Figure 21). To address this limitation, local modeling techniques can be applied. Rather than relying on a single point, the K closest points—known as *K-Nearest Neighbors (KNN)*—are selected, and a spatial plane is estimated using linear regression. The equation of the plane is:

$$Ax + By + Cz = D \tag{5.2}$$

The coefficients A, B, C, and D are obtained by solving the least squares minimization problem:

$$\min_{A,B,C,D} \sum_{i=1}^{K} (Ax_i + By_i + Cz_i - D)^2 \tag{5.3}$$

Once the plane is estimated, the perpendicular distance (height) between the original point and the derived plane is calculated as:

$$d_\perp = \frac{|Ap_x + Bp_y + Cp_z + D|}{\sqrt{A^2 + B^2 + C^2}} \tag{5.4}$$

This approach introduces a sense of continuity to the comparison but may produce errors in regions where the surface is genuinely discontinuous.

Following the methodology of ELKHRACHY et al. (2020) and SHAN et al. (2023) (ELKHRACHY, 2020; SHAN et al., 2023), this paper evaluates model accuracy using the *Gaussian Mean Difference (μ)*, the *Root Mean Square Error (RMSE)* of the calculated differences, and the *Standard Deviation* to quantify data noise.

The Mean Difference (μ) is given by:

$$\mu = \frac{1}{n_p} \sum_{i=1}^{n_p} d_{\perp i} \tag{5.5}$$

where n_p is the number of points between two point clouds being compared, $d_{\perp i}$ is the height between a particular point and the derived plane of the reference point.

The RMSE is defined as:

$$\text{RMSE} = \sqrt{\frac{1}{n_p} \sum_{i=1}^{n_p} d_{\perp i}^2} \tag{5.6}$$

The Standard Deviation, a metric to quantify noise, is calculated as:

$$\text{Standard Deviation} = \sqrt{\frac{1}{n_p} \sum_{i=1}^{n_p} (d_{\perp i} - \mu)^2} \tag{5.7}$$

Level of accuracy (LOA)

The Level of Accuracy (LOA) is a standardized scale used to define the precision requirements of a model relative to a point cloud (USIBD, 2016). This scale, developed by the United States Institute of Building Documentation (USIBD), provides an industry-wide standard for specifying accuracy levels. The LOA framework consists of five increments of precision, ranging from LOA10 (lowest precision) to LOA50 (highest precision). Each level corresponds to a maximum allowable error margin:

$$\text{LOA} = \{\text{LOA10, LOA20, LOA30, LOA40, LOA50}\}, \tag{5.8}$$

where each increment specifies increasing levels of precision. In this study, three LOA levels are evaluated for the smaller structure, the USP footbridge (table 3).

LOA	**Maximum Allowable Error Margin**
LOA30	15 mm
LOA40	5 mm
LOA50	1 mm

Table 3 – Maximum Allowable Error Margins for LOAs for the USP footbridge

Given the larger size and distance of the Octavio Frias Bridge, which is approximately 100 times greater than that of the USP Bridge, the error margins are scaled proportionally. For each LOA level, the thresholds are adjusted as follows:

$$\text{Scaled LOA Threshold} = \text{Original LOA Threshold} \times 100, \tag{5.9}$$

LOA x 100	**Maximum Allowable Error Margin**
LOA30 $\times$ 100	$1500\,\text{mm}$
LOA40 $\times$ 100	$500\,\text{mm}$
LOA50 $\times$ 100	$100\,\text{mm}$

Table 4 – Maximum Allowable Error Margins for LOAs for the Octavio Frias Bridge

leading to the adjustments shown on table 4.

The deviations between the reconstruction models and the ground truth are assessed using these adjusted LOA thresholds for both the USP Bridge and the Octavio Frias Bridge. This analysis allows for a consistent evaluation of accuracy across structures of different scales.

6 Results and discussion

In this chapter, the methodology from the previous chapter is put to work on real-world structures. First in the the Tucunduba river bridge and the Tauarizinho river bridge. For these two bridges, reconstruction models were created with progressively fewer photographs with Photogrammetry and NeRF and compared in terms of reconstruction quality and completeness.

Then reconstruction models for the the USP footbridge and the Octavio Frias Bridge are created, first with all input photos used, then with progressively fewer photos. These last two models are then compared to LiDAR-based reconstruction point clouds using the KNN local modeling with 6 neighbors. The performances of the SfM-MVS Photogrammetry and NeRFs models of these two structures are measured in terms of different levels of LOA and Gaussian mean, Standard Deviation and RMSE. Interestingly, in these scenarios NeRF outperform Photogrammetry in both LoA and statistical terms.

6.1 Tucunduba river bridge

Both techniques were able to successfully reconstruct the Tucunduba Bridge (Figure 22), capturing elements such as railings and pavement patterns with sharp edges. Photogrammetry was able to reconstruct the deck of the bridge in a more complete manner whereas NeRF managed to get finer details on the railings. The comparison picture further exibits the lack of points on the deck with the red spot in the middle of the figure. The statistical comparison for this bridge are: GM = 0.052665m, Standard Deviation = 0.062362m, RMSE = 0.062362m. NeRF model delivers highly accurate geometry for the 19m bridge, with mean absolute errors of just 5.3 cm and an RMSE of 6.2 cm ($\approx$0.33% of total span), indicating largely unbiased, randomly distributed residuals. This sub–0.5% error comfortably meets typical tolerances for structural assessment and as-built documentation.

Figure 23 shows photogrammetry and NeRF reconstruction models when available data is progressively reduced, ranging from 100% down to 33%. NeRF reconstructions maintain a relatively coherent overall shape, but begin to show smoother surfaces and less crisp detail in the finer elements, such as the railings, as the photo count decreases. Photogrammetry, on the other hand, tends to produce sharper edges and textures when enough images are available, but suffers from increasingly noticeable gaps and missing patches in the geometry with fewer images, especially around complex or thin structures. At 33% of image input NeRF may appear more continuous but slightly “soft”. while photogrammetry retains sharper regions where sufficient views exist but struggles to fill in areas with limited coverage (Figure 24). This suggests that NeRF can provide more globally consistent results under sparse data conditions, whereas photogrammetry often excels in detail when adequate coverage is present but degrades more

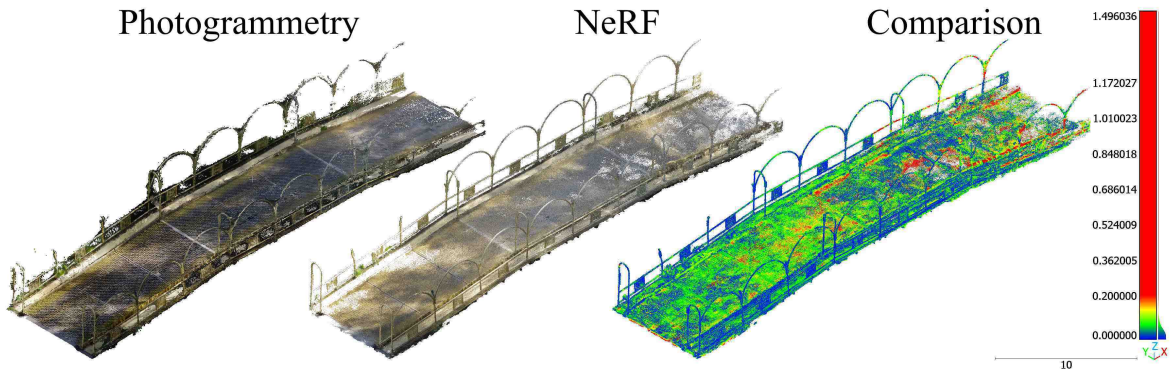

Figure 22 – RGB Photogrammetry-based, NeRF-based point clouds of the Tucunduba river bridge and a comparison between each model using Nearest Neighbors (KNN = 6) local modeling.

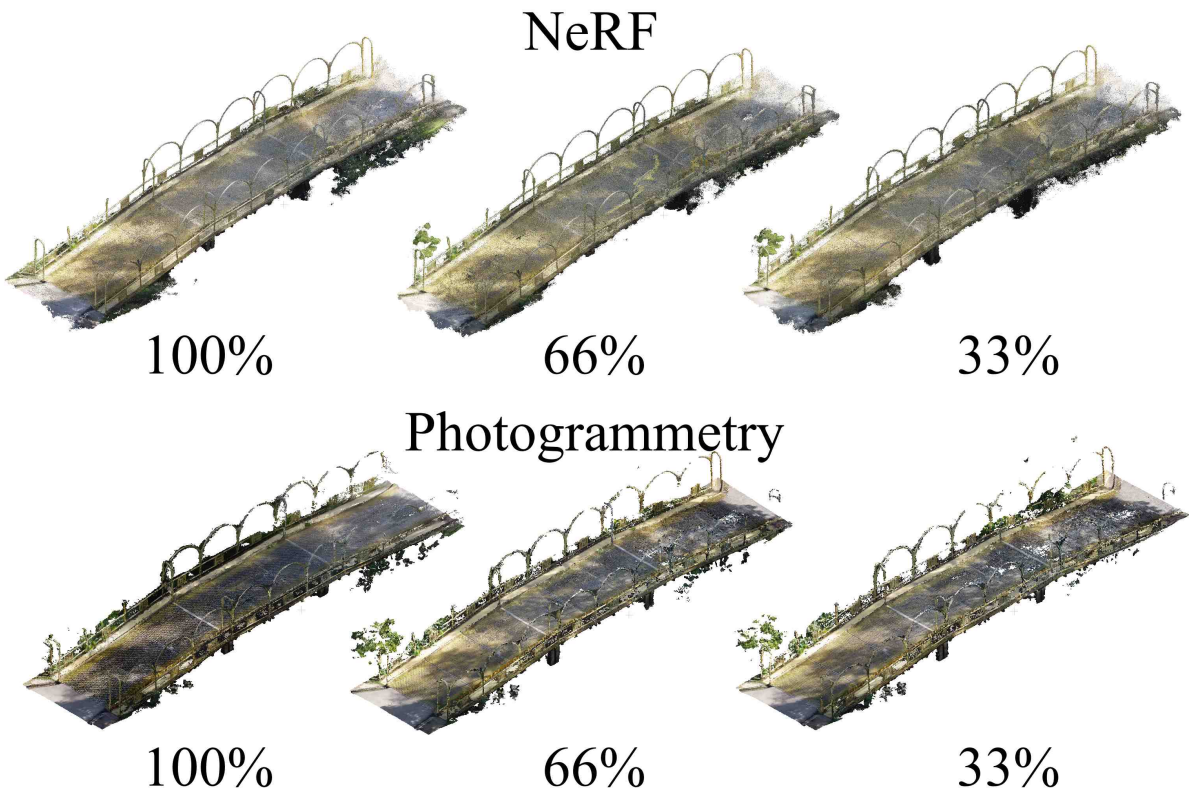

Figure 23 – RGB Photogrammetry-based and NeRF-based point clouds of the Tucunduba river bridge with progressively fewer pictures.

noticeably with reduced image sets.

Figure 25 exhibits novel view reconstructions and depth estimations in different points of the training of the Tucunduba river bridge model, namely iteration 5558, 11987 and 2500. It is observed that a lower-quality initial representations gradually refine into a sharper, more detailed image overtime, especially in the deck and curved arches on the side of the bridge. The depth estimations become more refined as well. The bottom row exhibits the behavior of the Loss and PNSR over training iterations for both training and test data. After 60k iterations, the training

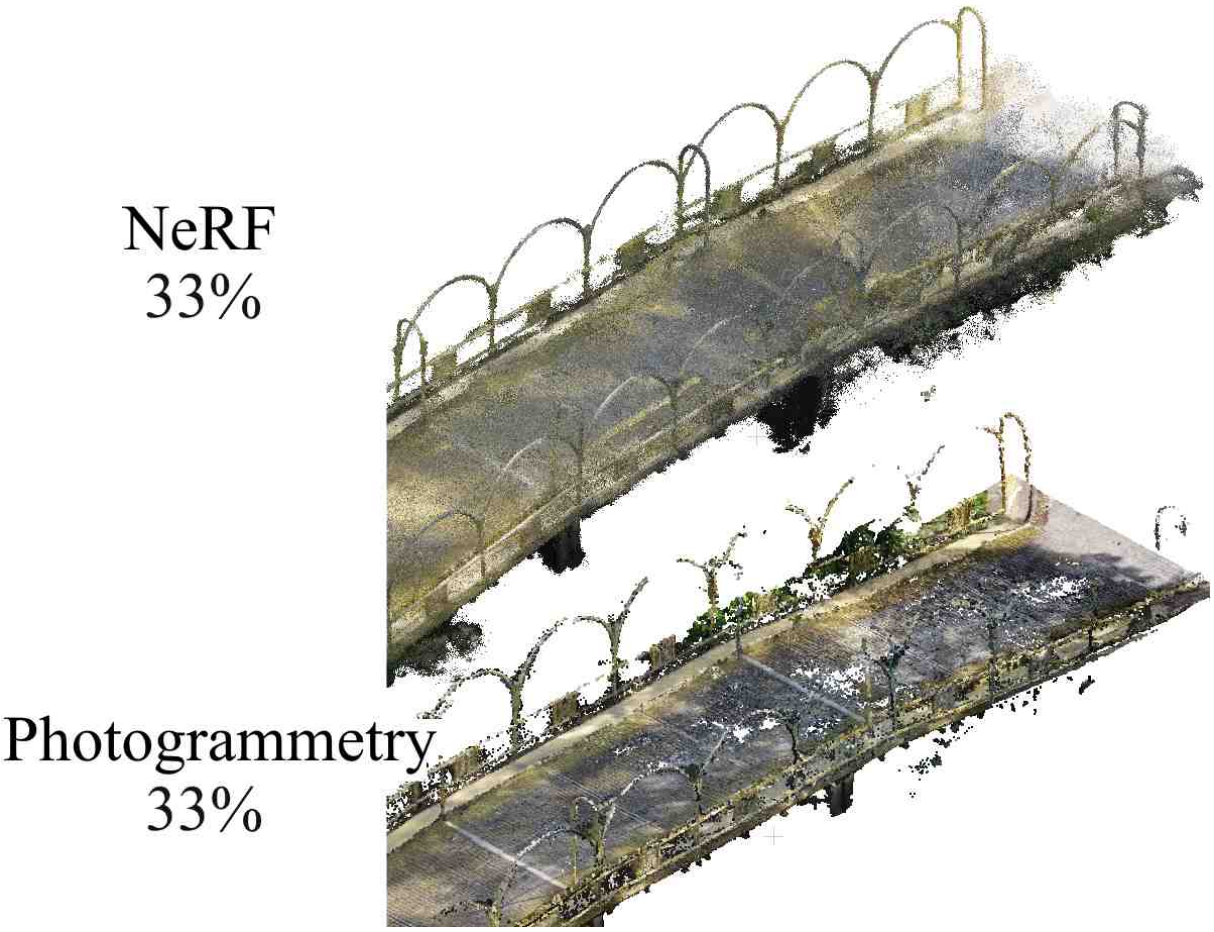

Figure 24 – Photogrammetry and NeRF point clouds of the pillars Tucunduba river bridge with 33% of input pictures.

loss gets as low as 0.017 while the test loss stagnates at 0.031. The training PSNR increasses to 18.29 while test PSNR stagnates at 15.11. SSIM value behaves erratically in the training, going from 0.374 to 0.307 while having variations as high as 0.3 in between. LPIPS also has high frequency behavior. While it starts at 0.937 and ends at a higher 0.948, it has an overall downward trend during training.

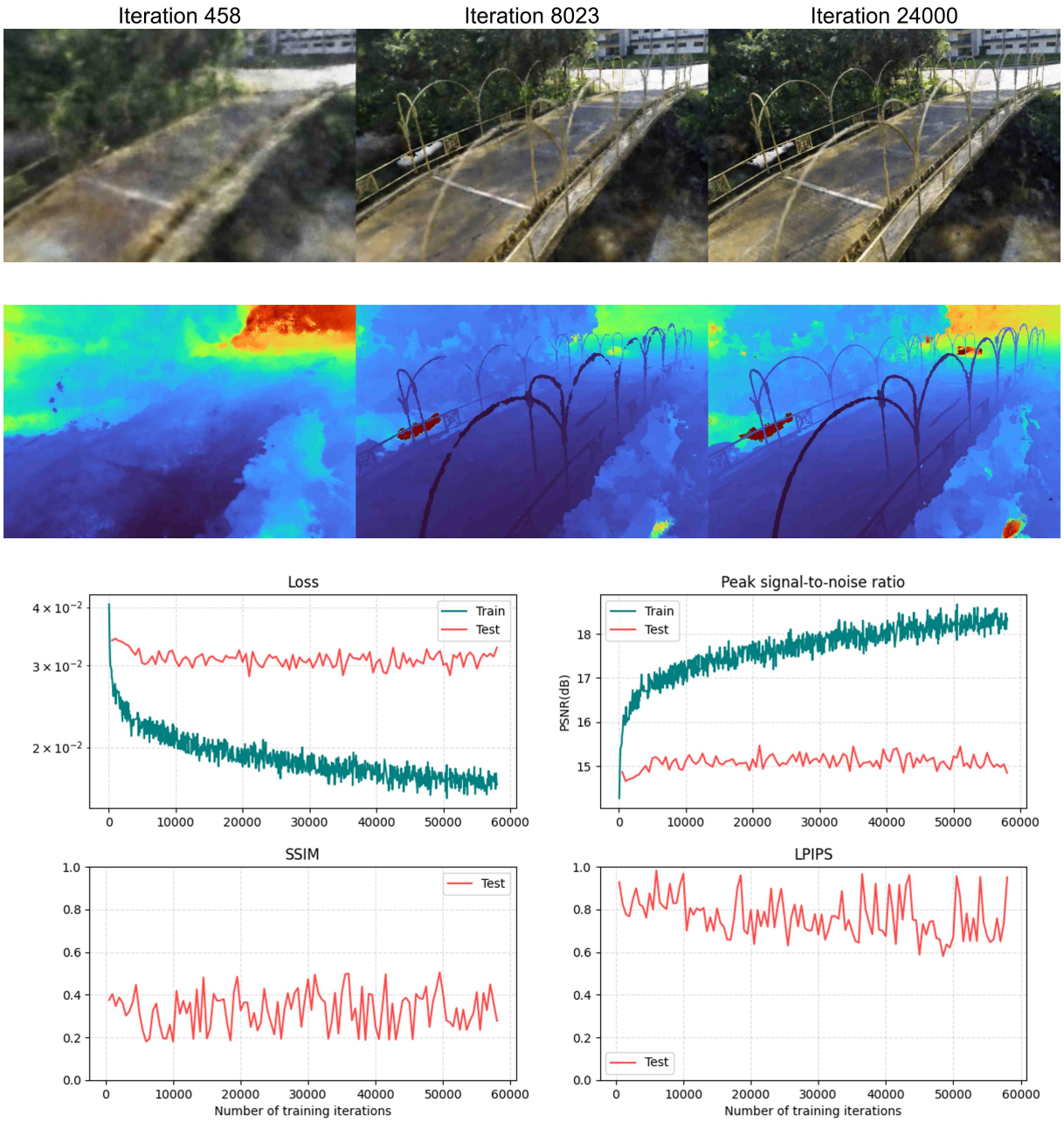

Figure 25 – First row: Novel view reconstructions of the Tucunduba Bridge over progressing iterations. Second Row: Depth maps of the model on that specific iterations. Third Row: Loss and PSNR. Forth Row: SSIM and LPIPS over training iterations.

6.2 Tauarizinho river bridge

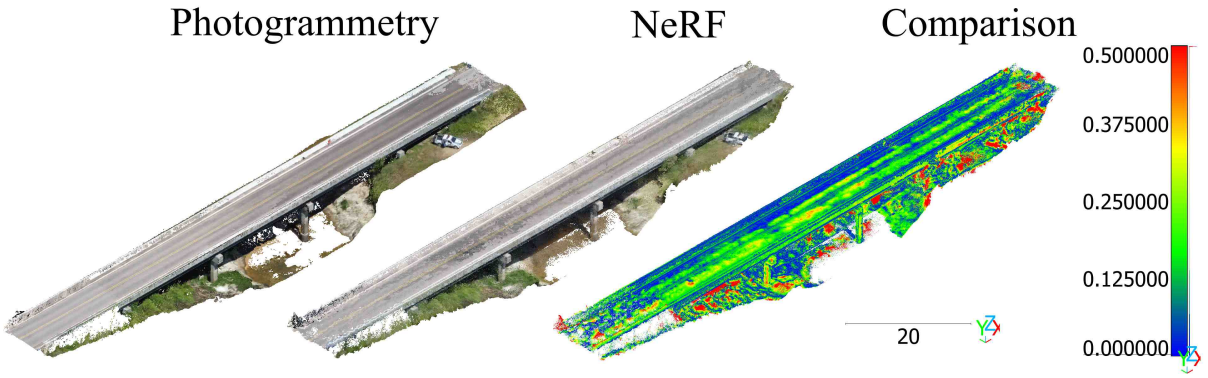

Figure 26 – RGB Photogrammetry-based, NeRF-based point clouds of the Tauarizinho river bridge and a comparison between each model using Nearest Neighbors (KNN = 6) local modeling.

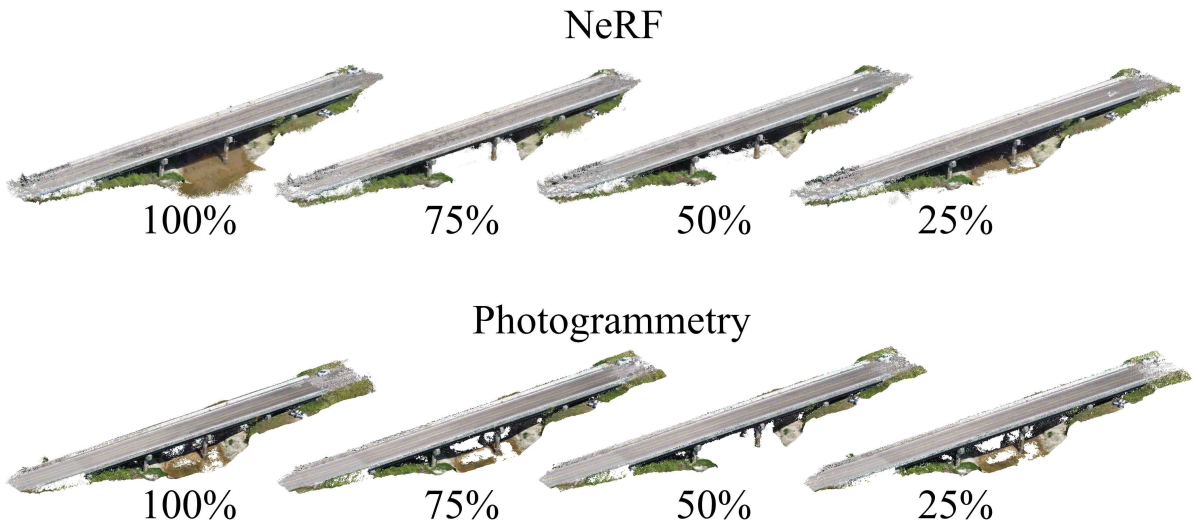

Figure 27 – RGB Photogrammetry-based and NeRF-based point clouds of the Tauarizinho river bridge with progressively fewer pictures.

Both techniques were able to sucessfully reconstruct the Tauarizinho river bridge (Figure 26). The NeRF has more visible noise in the deck of the bridge in comparison to the Photogrammetry model. The comparison picture shows the further exibits the lack of points on the deck with the red spot in the middle of the Figure. The statistical comparison for this bridge are: GM = 0.118324m, Standard Deviation = 0.163613m, RMSE = 0.201915m. With an average absolute error of just 0.12m and an RMSE of 0.20m on a 122m bridge, the reconstruction model delivers sub-0.2% geometric accuracy, well within the tolerances required for most structural-engineering assessments. The standard deviation of 0.16m shows that the bulk of pointwise deviations lie within 16cm, while the gap between GM and RMSE signals a small number of larger outliers. Overall, achieving mean errors under 12cm on a full-scale bridge demonstrates a high-fidelity reconstruction model.

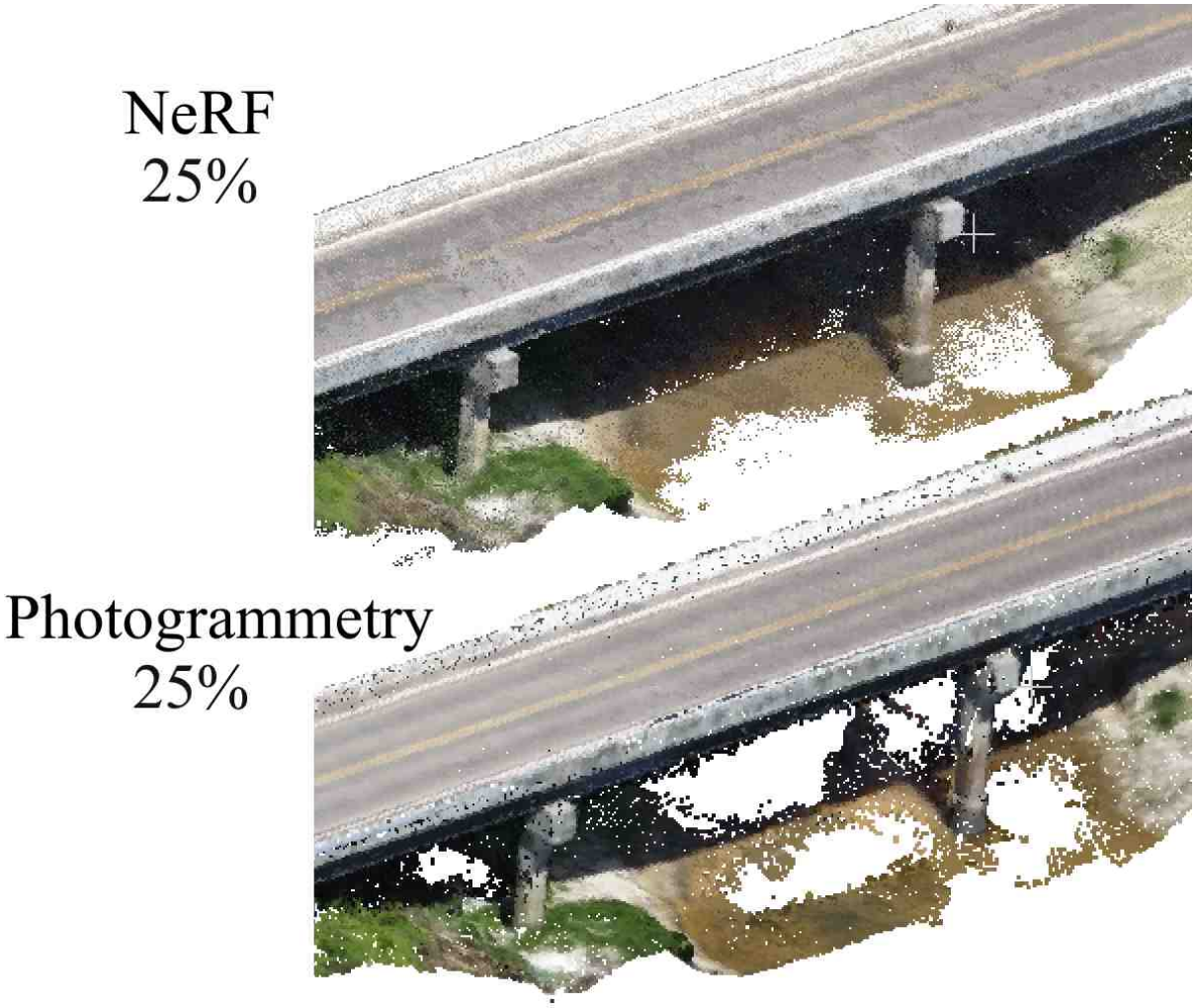

Figure 28 – Photogrammetry and NeRF point clouds of the pillars Tauarizinho river bridge with 25% of input pictures.

Figure 27 shows photogrammetry and NeRF reconstruction models when available data is progressively reduced, ranging from 100% down to 25%. When 100% of photographs is used both NeRF and photogrammetry generate high-fidelity reconstructions, capturing the overall bridge geometry. For 75% and below, NeRF maintains a relatively consistent global shape and fills in the scene without large holes, but begins to exhibit a softer appearance. This aligns with the understanding that NeRF is a neural network, thus is able to infer accurate data even with no direct training data. Photogrammetry meanwhile continues to deliver sharper detail and clearer texturing where sufficient input data exist, but it suffers from incomplete regions and noticeable gaps in the pillars under the bridge with fewer input pictures (Picture 28).

Figure 29 exhibits novel view reconstructions and depth estimations in different points of the training of the Tauarizinho river bridge model, namely iteration 752, 4030 and 20374. It is observed that a lower-quality initial representations gradually refine into a sharper, more detailed image overtime, especially in pillars below the bridge and car deck above. The depth estimations become more refined as well. The bottom row exhibits the behavior of the Loss and

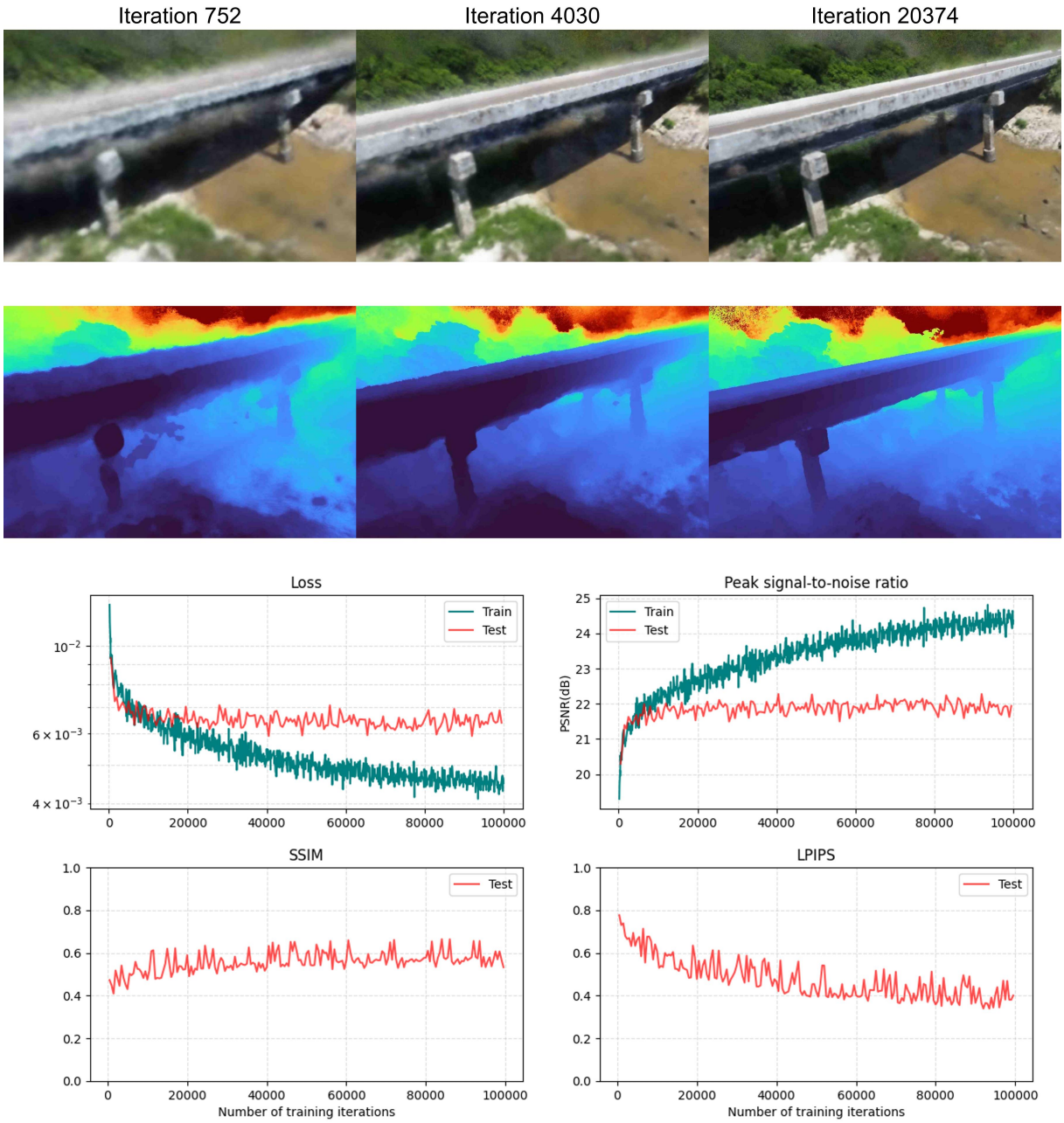

Figure 29 – First row: Novel view reconstructions of the Tauarizinho Bridge over progressing iterations. Second Row: Depth maps of the model on that specific iterations. Third Row: Loss and PSNR. Forth Row: SSIM and LPIPS over training iterations.

PNSR over training iterations for both training and test data. After 100k iterations, the training loss gets as low as 0.0044 while the test loss stagnates at 0.0064. The training PSNR increasses to 24.64 while test PSNR stagnates at 21.92. SSIM has an overall upward trend going from 0.467 to 0.563. LPIPS has an downward trend from 0.0779 to 0.410.

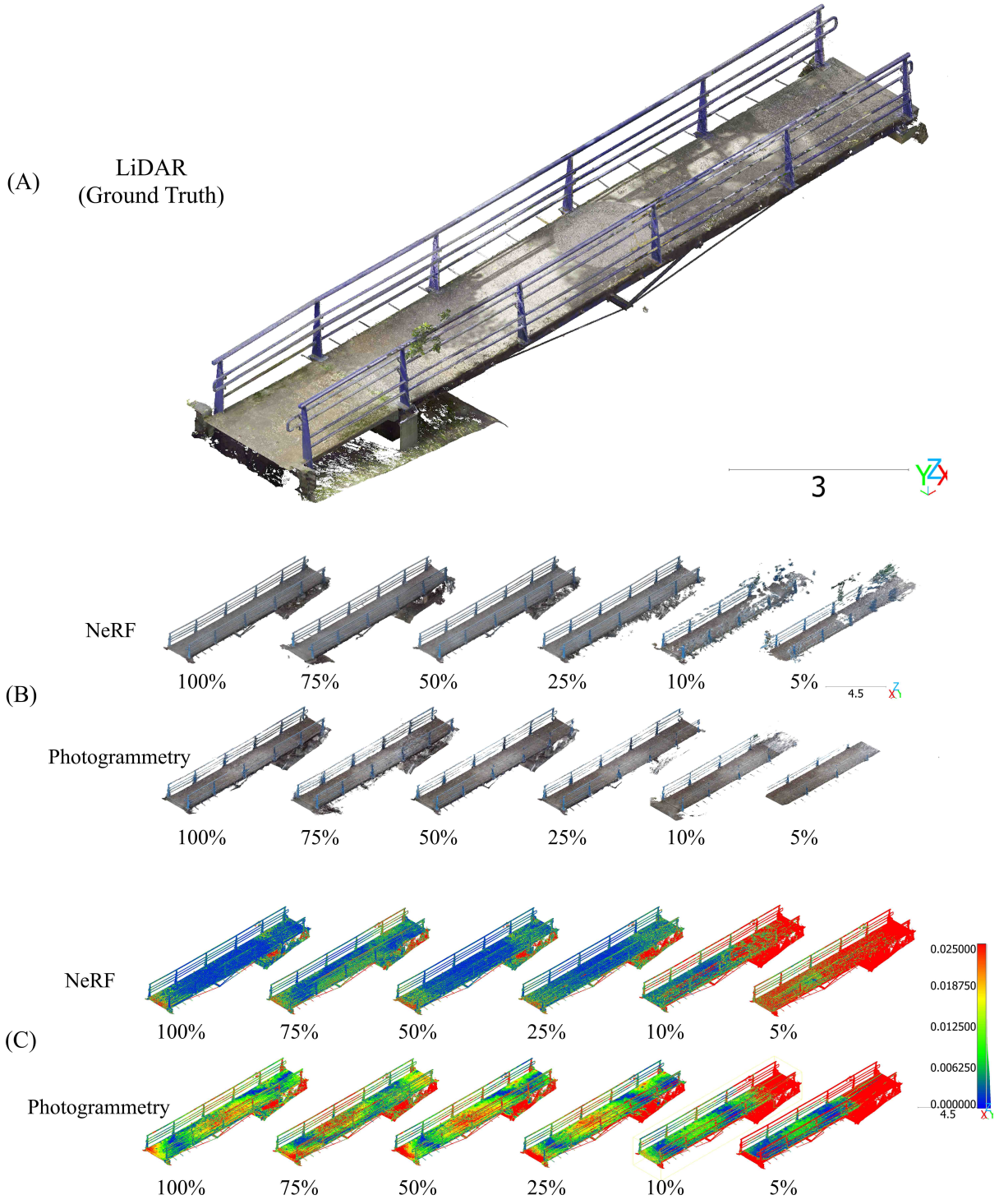

Figure 30 – (A) LiDAR based point cloud from the USP Bridge. (B) RGB NeRF-based and Photogrammetry-based point clouds created utilizing progressively fewer pictures of the USP Bridge. (C) Models compared to the LiDAR-based ground truth using Nearest Neighbors (KNN = 6) local modeling using progressively fewer pictures of the USP Bridge.

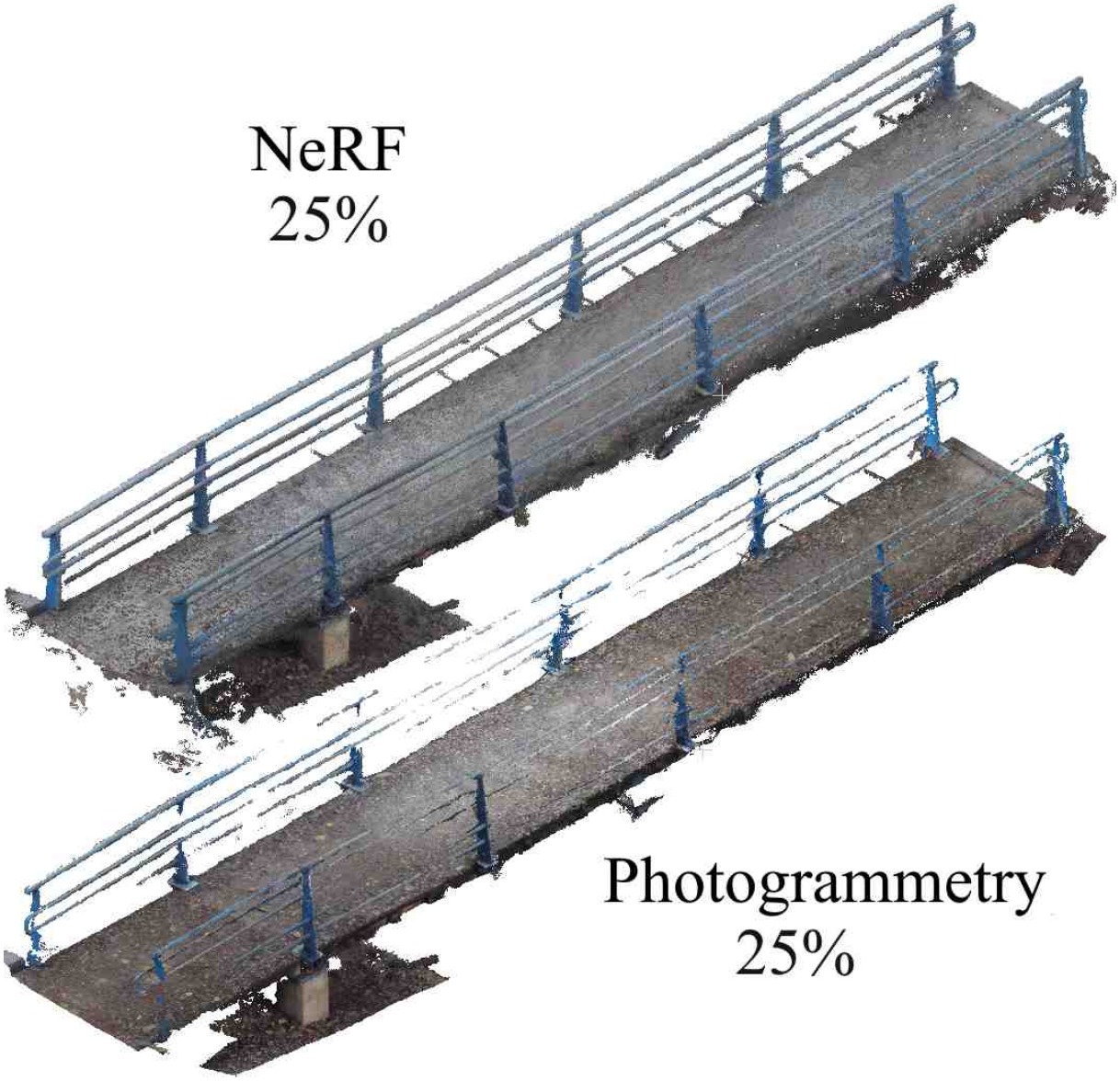

Figure 31 – Photogrammetry and NeRF point clouds of the USP footbridge with 25% of input pictures.

6.3 USP footbridge

Figure 30 (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 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 (Figure 31). 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. This outcome aligns with the findings of Larrivée et al. (BISSON-LARRIVÉE; LEMOINE, 2022), who highlight the influence of camera localization

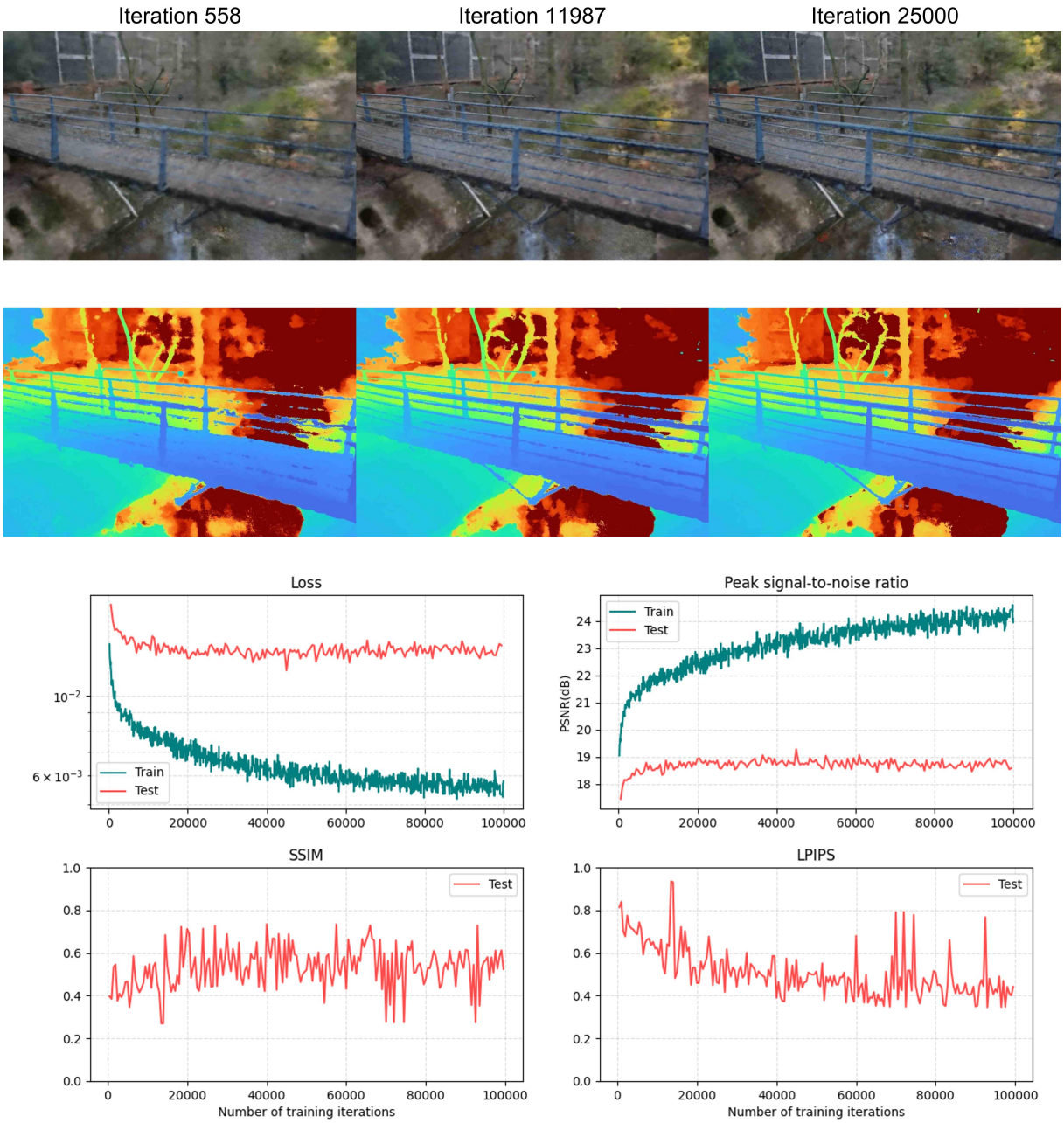

Figure 32 – First row: Novel view reconstructions of the USP footbridge over progressing iterations. Second Row: Depth maps of the model on that specific iterations. Third Row: Loss and PSNR. Forth Row: SSIM and LPIPS over training iterations.

and the angular disparity between input images on photogrammetry results. Specifically, they demonstrate that Photogrammetry-based 3D reconstructions become increasingly incomplete as the number of input images decreases and the angular separation between them increases.

Figure 30 (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

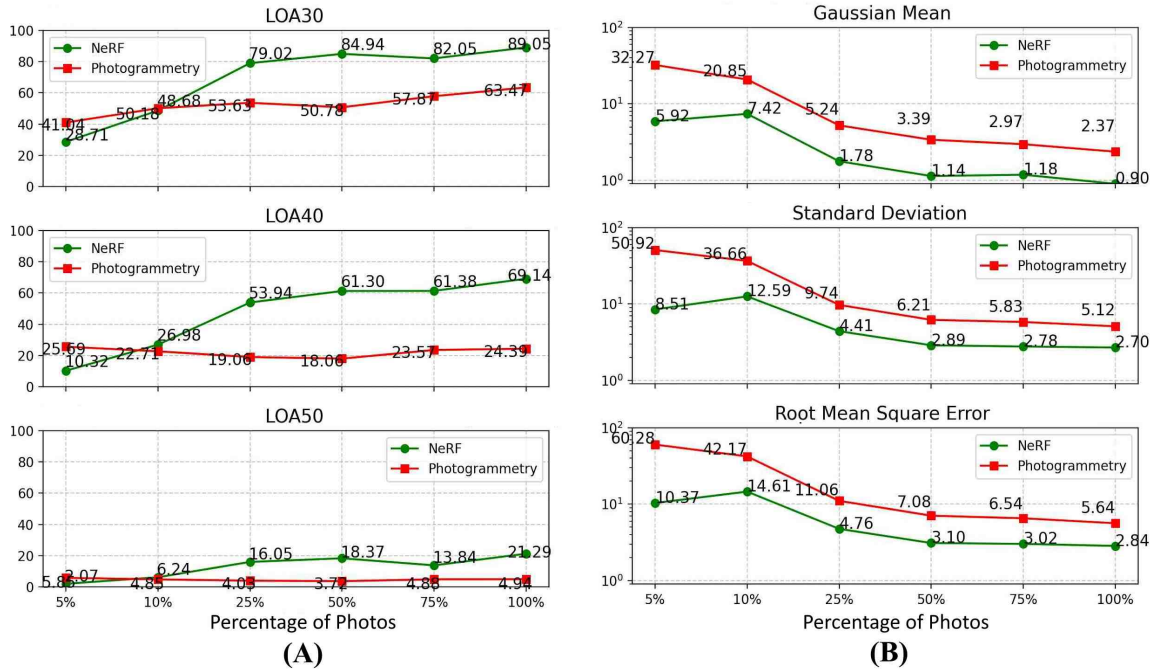

Figure 33 – (A) LOA30, 40 and 50 by percentage of used photos when creating NeRF and Photogrammetry models of the USP footbridge (Higher is better). (B) Gaussian mean, standard Deviation and Root Mean Square Error by percentage of used photos when creating NeRF and Photogrammetry models of the USP Bridge (Lower is better).

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.

When comparing Levels of Accuracy, NeRF consistently outperforms photogrammetry at all percentages of photos, except for 5% (Figure 33 (A)). 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 (Figure 33 (B)). With 100% of photos, NeRF's RMSE is 2.85 cm, significantly lower than photogrammetry's 5.64 cm. The NeRF reconstruction model delivers exceptionally accurate geometry over the full 12m bridge span, with an average deviation under 1 cm and a worst-case spread under 3 cm. In relative terms, this error amounts to less than 0.1% of the structure's length, placing the model well within the tolerance requirements for structural health monitoring, deformation tracking, and precision inspection. The low standard deviation further confirms that errors are not only small on average but also tightly clustered, ensuring that the

reconstruction is both precise and reliable for critical engineering applications. 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.

Figure 32 exhibits novel view reconstructions and depth estimations in different points of the training of the USP footbridge model, namely iteration 5558, 11987 and 2500. It is observed that a lower-quality initial representations gradually refine into a sharper, more detailed image overtime, especially in the handrails of the bridge. The depth estimations become more refined as well. The bottom row exhibits the behavior of the Loss and PNSR over training iterations for both training and test data. After 100k iterations, the training loss gets as low as 0.0055 while the test loss stagnates at 0.013. The training PSNR increasses to 20.2 while test PSNR stagnates at 18.71. SSIM values start as low as 0.396 and then present a high-frequency behavior throughout the training, finishing at 0.582. LPIPS valures start at 0.831 and then finish at 0.431 while presenting peaks as high as 0.2 in between.

6.4 Cable-stayed Octavio Frias bridge

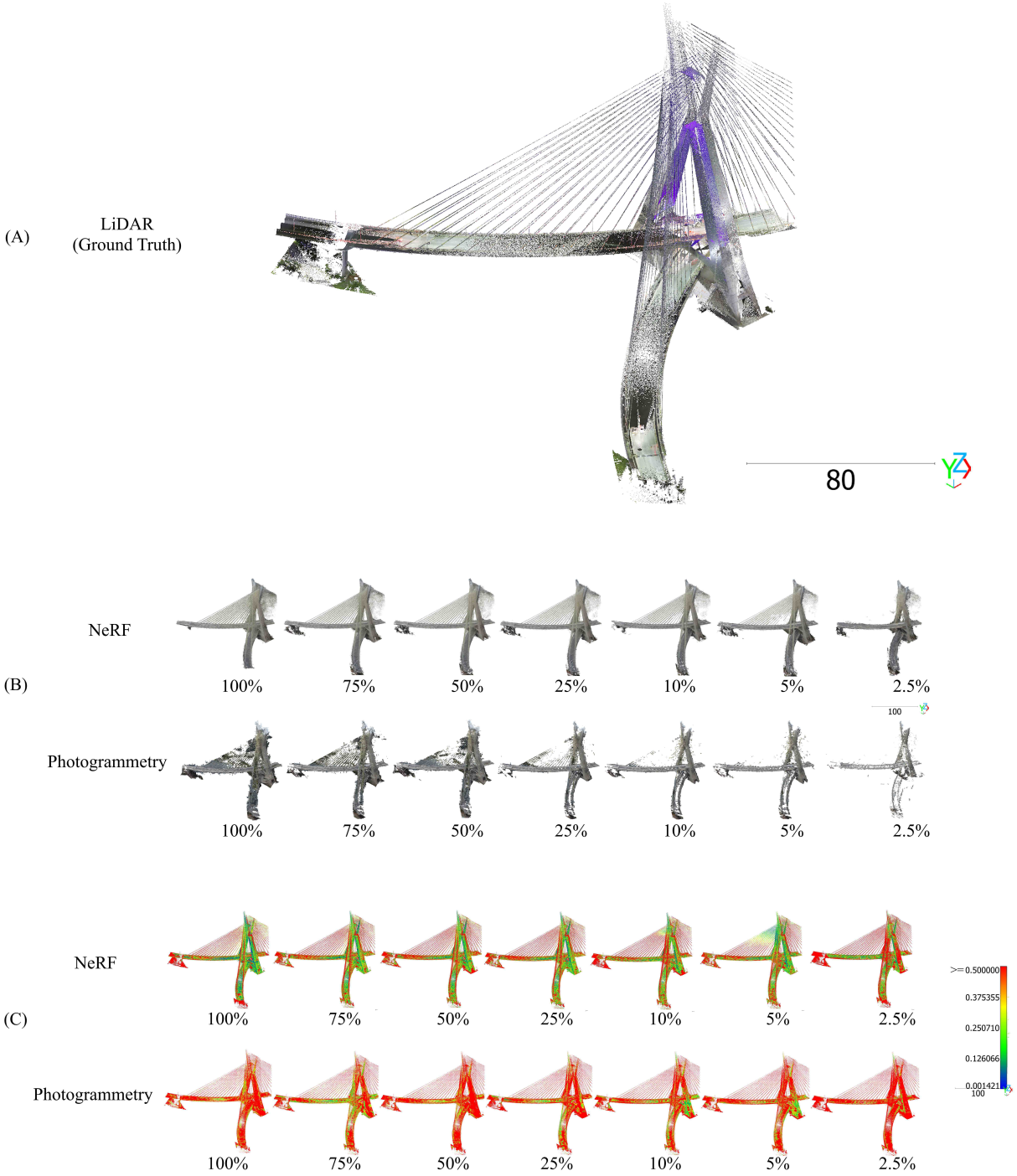

Figure 34 – (A) LiDAR based point cloud from the Octavio Frias Bridge. (B) RGB NeRF-based and Photogrammetry-based point clouds created utilizing progressively fewer pictures of the Octavio Frias Bridge. (C) Models compared to the LiDAR-based ground truth using Nearest Neighbors (KNN = 6) local modeling using progressively fewer pictures of the Octavio Frias Bridge

Figure 34 (B) shows a completeness comparison between NeRF and photogrammetry reconstructions as the amount of data is progressively reduced. Both methods display reconstruc-

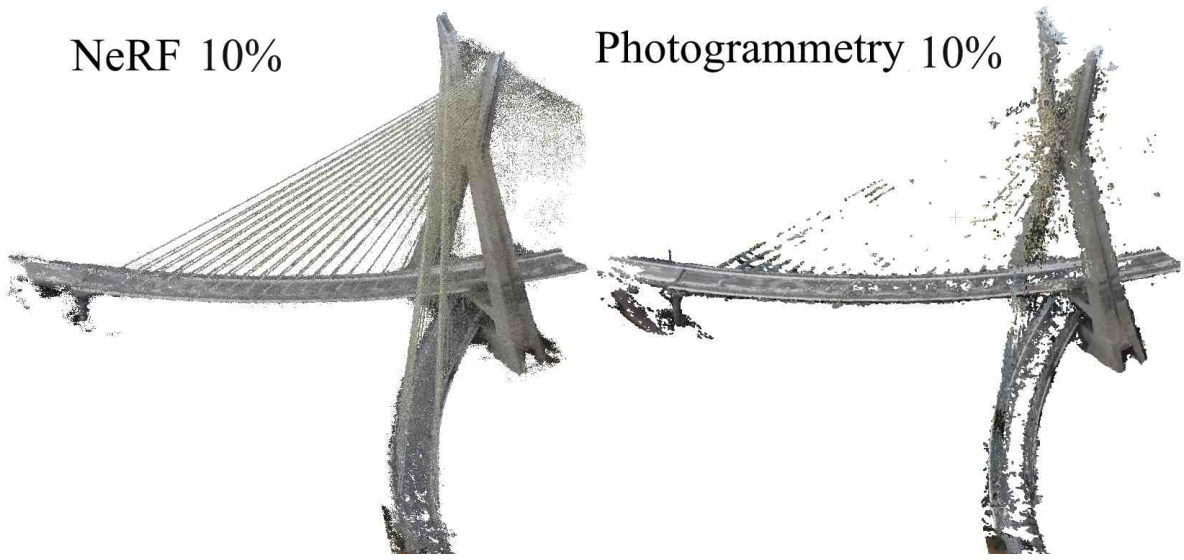

Figure 35 – Photogrammetry and NeRF point clouds of the Octavio Frias Bridge with 10% of input pictures.

tion models of the Octavio Frias Bridge at different levels of data 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% (Figure 35) 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 34 (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 37 (A)), 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.

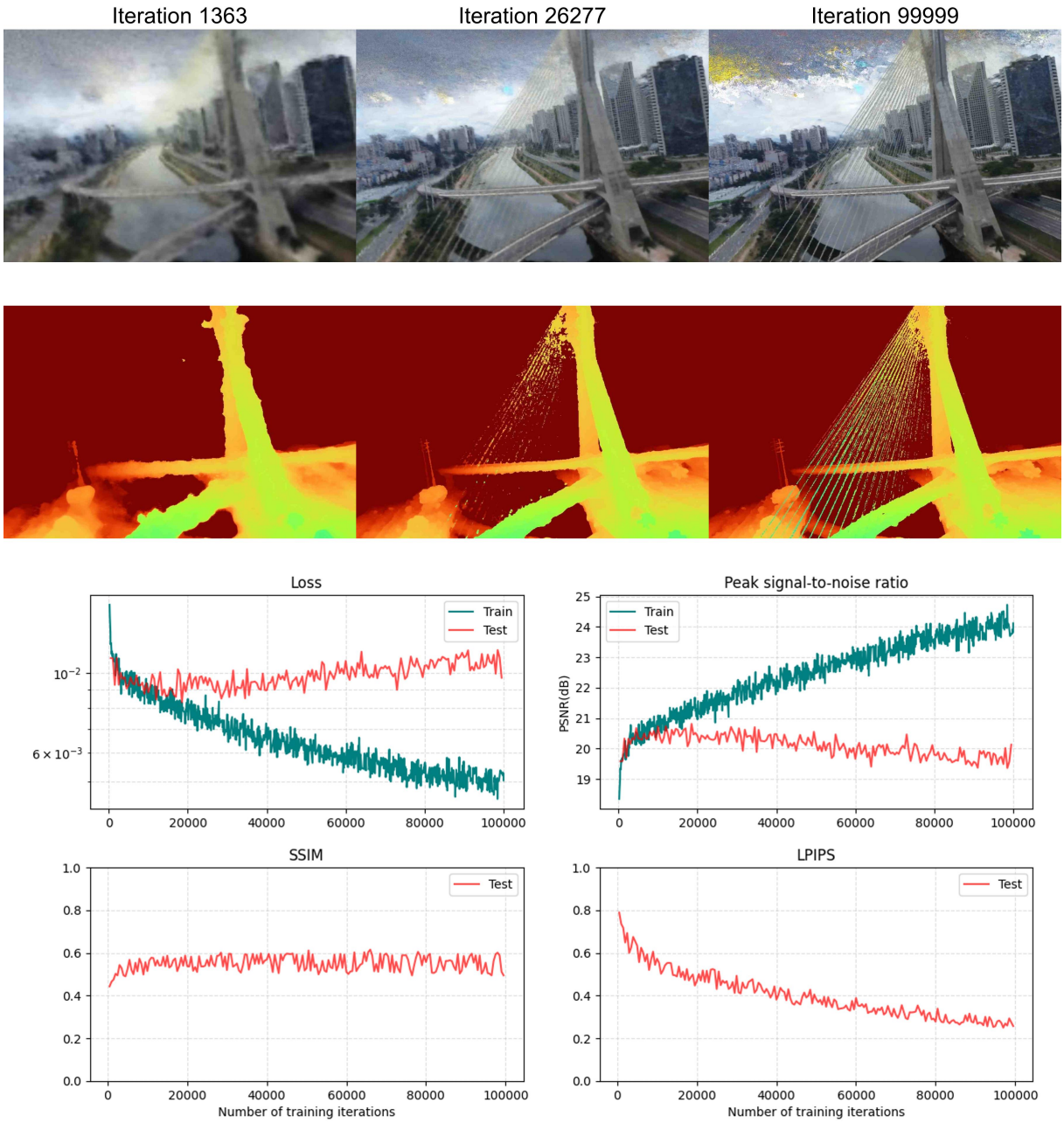

Figure 36 – First row: Novel view reconstructions of the Octavio Frias Cable-Stayed Bridge over progressing iterations. Second Row: Depth maps of the model on that specific iterations. Third Row: Loss and PSNR. Forth Row: SSIM and LPIPS over training iterations.

NeRF displays more stability in terms of mean and error metrics as the percentage of photos decreases (Figure 37 B)). Its lowest RMSE (1.14m) 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 4.55m and 4.66m, respectively. This indicates that Photogrammetry struggles with fewer photos, becoming significantly less accurate compared to NeRF. Although the 100% model achieves a low average error (GM $\approx$ 1.6m, just

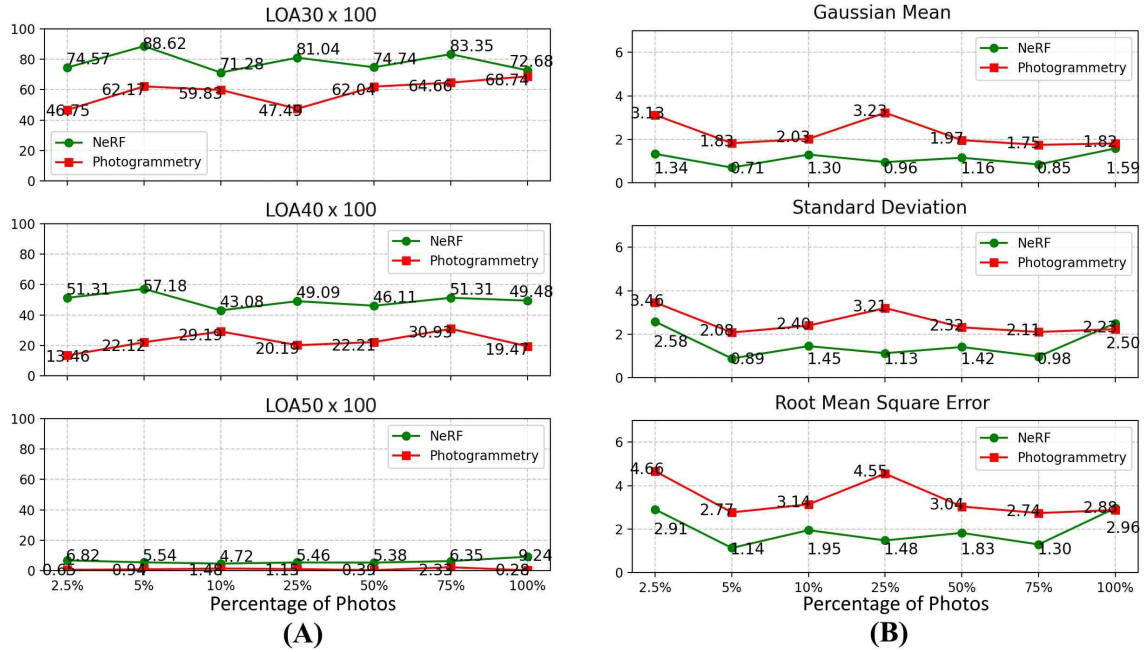

Figure 37 – (A) LOA30, 40 and 50 by percentage of used photos when creating NeRF and Photogrammetry models of the Octavio Frias Bridge (Higher is better). (B) Gaussian mean, standard Deviation and Root Mean Square Error by percentage of used photos when creating NeRF and Photogrammetry models of the USP Bridge (Lower is better).

over 1% of the 150m span), the higher standard deviation (2.5m) and RMSE (2.96m) expose sporadic large errors that could undermine precision-critical tasks like stress analysis or clearance verification.

Figure 36 exhibits novel view reconstructions and depth estimations in different points of the training of the Octavio Frias model, namely iteration 1363, 26277 and 99999. It is observed that a lower-quality initial representations can't reconstruct the finer cable-stays, while they start to appear after several training iterations. The depth estimations become more refined as well. The bottom row exhibits the behavior of the Loss and PNSR over training iterations for both training and test data. After 100k iterations, the training loss gets as low as 0.0049 while the test loss stagnates at 0.01. The training PSNR increasses to 24.18 while test PSNR stagnates at 19.72. SSIM present a value of 0.445 at the start of the training and then stagnates at 0.56 to its end. LPIPS however gets lower overtime, from 0.787 to 0.268 by the end.

6.5 Discussion

The results above highlight the comparative strengths and limitations of NeRF and photogrammetry for 3D reconstruction applications, with key insights into their performance under varying data availability.

- **Model reconstruction varying number of input images:** Overall, NeRF reconstructions maintain higher detailed, more complete reconstructions when compared to Photogrammetry models with fewer input pictures, especially in larger structures with finer important structural elements such as the cable-stays from the Octavio Frias bridge (Figure 18).

- **Level of Accuracy:** Overall, LoA metrics were higher (Better) for NeRF models with varying input images on real structures.

- **Statistical metrics:** NeRF's Gaussian mean, Standard Deviation and RMSE metrics were lower (Better) when compared to photogrammetry-based models.

- **Loss function:** Overall, the Loss exhibits a consistent decline during training, resulting in minimization of visual discrepancies across all models. Even after 60k and 100k training iterations he training loss continue declining while the test loss stagnates at a higher level. This shows that even further training doesn't make novel views better.

- **PSNR:** Training PSNR values steadily rise during training, indicating effective learning from low-quality representations towards clearer images and depth estimations. The test PSNR stagnates at a lower value in comparison.

- **SSIM and LPIPS:** Results show a clear negative correlation between these two parameters, with SSIM getting higher values while LPIPS gets lower overtime. Interestingly, models with highest instability (USP footbridge and Tucunduba Bridge) in these parameters had input pictures taken with farther distances in between them, which corroborates Zhang's et al. (2018) (ZHANG et al., 2018) assessment of LPIPS behavior previously mentioned. On the other hand, models with more stable SSIM tend to achieve a lower, more stable, LPIPS value.

Implications

The findings of this work have important implications, particularly in scenarios with limited data, where effective reconstruction and monitoring of structural integrity are critical. In addition, our results have practical implications for 3D reconstruction of bridges, particularly in resource-constrained environments where data collection may be limited by logistical or financial

constraints. NeRF's resilience to sparse data suggests its suitability for remote or emergency scenarios, such as post-disaster evaluations or inaccessible bridge sites. Its lower RMSE and superior LOA performance compared to SFM-MVS photogrammetry indicate a high reliability for creating accurate models with limited photographic data, reducing the need for extensive site surveys or specialized equipment.

A significant trend observed across both methods is the exponential increase in error metrics (RMSE and Gaussian mean) as data availability decreases. Although NeRF demonstrates a smaller downgrade in performance rate, both methods exhibit a sharp increase in errors at the 5% and 2.5% thresholds. This highlights the inherent challenges of data-sparse reconstruction and suggests that neither method can completely circumvent the limitations imposed by extremely low data availability.

7 Conclusions and future research

This study is set to explore the performance of NeRF-based 3D reconstruction models of civil structures, specifically bridges. NeRF models are compared to traditional photogrammetry and LiDAR in scenarios with limited input data. The key objectives were to assess how decreasing percentage of available photographs impact reconstruction quality and test these input-limited models to LiDAR ground truth. Demonstrated differences are shown in the performance of NeRF and photogrammetry as the percentage of input photos decreases, impacting both LOA metrics and error measurements.

Main conclusions

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. Below are the key contributions of this work:

- **NeRF efficiency in data acquisition**: This study demonstrates that NeRF can produce high-fidelity reconstruction models of bridges even with reduced numbers of input pictures, lowering costs and time for field applications.
- **Quantitative metrics validation**: This work provides quantitative metrics (RMSE, Gaussian statistics) comparing NeRF and photogrammetry in real-world settings, filling a gap in prior quantitative comparisons.
- **Training dynamics**: This study also demonstrated that prolonged training improves model reconstruction, especially of smaller elements such as cable-stays, guiding practitioners to optimize training epochs for these specific scenarios.
- **Suitability for complex scenarios**: The work shows NeRF as particularly effective for large, geometrically complex structures such as cable-stayed bridges. This can lead to adoption in civil engineering and infrastructure monitoring scenarios.

In summary, while NeRF shows promising advancements over photogrammetry under constrained data conditions, continued investigation into more varied environments and camera specifications is essential to fully validate its efficacy across different environments.

Directions for Future Work

- **Different network architectures**: The canonical NeRF MLP architecture was used in this work. Therefore a work to be attempted in the near future is the assessment of different NeRF MLP architectures to attest how the metrics in this work change. It is possible that with a more complex architecture a lower error can be achieved at cost of computational power and processing time. Previously mentioned work by Cui et al. (2024) (CUI et al., 2024) shows promise due its fast training and good PSNR with fewer training iterations than other NeRF-based models.

- **Different methodology for point cloud creation**: NeRF novel views can be leveraged to create unlimited inputs for the MVS process to create dense point clouds. These MVS-based depth maps from novel views can help create photogrammetry models where input data is limited. This NeRF-MVS[1] hybrid has the potential of providing point clouds with higher resolution compared to SFM-MVS photogrammetry.

- **Structure Scope**: The study focused on bridges, but there are other critical civil structures such as train networks, electrical power towers, dams that could have their geometry assessed with similar methodology.

- **Photograph resolutions**: The specific resolution of 1080p used in this study could be too small for proper reconstruction. Future research should expand on the range of resolutions for reconstructions, from very low resolutions (512 x 512px, for instance) up to standard 4k (3840 x 2160px).

- **Data Diversity**: The range of bridge scenarios examined with LiDAR-based ground truth was not exhaustive. Future research should expand on this issue to include a broader variety of LiDAR-based ground-truth models from bridges.

[1] Not to be confused with (CHEN et al., 2021), which leverages MVS-based depth maps *to* make Neural Radiance Fields models.

List of Publications

The main original works that support this dissertation are addressed in follow. In the time frame of this dissertation, 1 conference work was accepted and 1 paper is under review for publication.

International conference paper

Thiago Ribeiro; Felipe Brandão; Victor Cardoso; João Costa; Túlio Bittencourt; Moisés Silva, "**Neural Radiance Fields for Low-cost, High-resolution 3D Scanning of Critical Infrastructures**", XLIII International Modal Analysis Conference, 2025

Journal paper

Thiago Ribeiro; Felipe Brandão; Victor Cardoso; João Costa; Túlio Bittencourt; Moisés Silva, "**Benchmarking Neural Radiance Fields for 3D geometric reconstruction of bridges**", Engineering Structures, 2025, *Under review*

Bibliography

AARON, J. et al. 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.*, American Geophysical Union (AGU), v. 50, n. 5, mar. 2023. Cited on page 8.

ALAMDARI, A. G.; EBRAHIMKHANLOU, A. Multi-scale robotic scanning of surface cracks in concrete structures. In: GYEKENYESI, A. L. et al. (Ed.). *Nondestructive Characterization and Monitoring of Advanced Materials, Aerospace, Civil Infrastructure, and Transportation XVIII*. [S.l.]: SPIE, 2024. v. 12950, p. 129500M. DOI:10.1117/12.3010480. Cited on page 7.

ALIZADEHSALEHI, S.; YITMEN, I. Digital twin-based progress monitoring management model through reality capture to extended reality technologies (drx). *Smart and Sustainable Built Environment*, Emerald Publishing Limited, v. 12, n. 1, p. 200–236, Jan 2023. ISSN 2046-6099. DOI:10.1108/SASBE-01-2021-0016. Cited on page 17.

ALWYN, M. et al. Digital Twin Enabled Construction Process Monitoring. In: European Council on Computing in Construction (E3C3). *2024 EC(3) - 2024 European Conference on Computing in Construction*. Chania (Crète), Greece, 2024. DOI:10.35490/EC3.2024. Cited on page 7.

ANDREAS, H. et al. The investigation on high-rise building tilting from the issue of land subsidence in jakarta city. *MATEC Web Conf.*, v. 270, p. 06002, 2019. Cited on page 16.

APOLLONIO, F. I. et al. A photogrammetry-based workflow for the accurate 3d construction and visualization of museums assets. *Remote Sensing*, v. 13, n. 3, 2021. ISSN 2072-4292. DOI:10.3390/rs13030486. Cited on page 18.

ARTEAGA, A. O.; SCOTT, D.; BOEHM, J. Initial investigation of a low-cost automotive lidar system. *ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, Copernicus GmbH, XLII-2/W17, p. 233–240, nov. 2019. Cited on page 8.

ARTZY, E. Display of three-dimensional information in computed tomography. *Computer Graphics and Image Processing*, v. 9, n. 2, p. 196–198, 1979. ISSN 0146-664X. DOI:10.1016/0146-664X(79)90057-1. Cited on page 25.

BARRON, J. T. et al. Zip-NeRF: Anti-aliased grid-based neural radiance fields. 2023. Cited on page 20.

BECHER, M. et al. Your visualisations are going places: SciVis on gaming consoles. *J. Vis. (Tokyo)*, Springer Science and Business Media LLC, dez. 2024. Cited on page 25.

BENASSAROU, A. et al. MC slicing for volume rendering applications. In: *Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, (Lecture notes in computer science). p. 314–321. Cited 3 times on pages 0, 26, and 27.

BERNARDES, P.; MADEIRA, J.; MARTINS, M. *Traditional-based visualization methods for archaeological 3D data: An evaluation*. [S.l.]: The Eurographics Association, 2020. Cited on page 26.

BESL, P. J.; MCKAY, N. D. A method for registration of 3-D shapes. *IEEE Trans. Pattern Anal. Mach. Intell.*, Institute of Electrical and Electronics Engineers (IEEE), v. 14, n. 2, p. 239–256, fev. 1992. Cited on page 18.

BISSON-LARRIVÉE, A.; LEMOINE, J.-B. Photogrammetry and the impact of camera placement and angular intervals between images on model reconstruction. *Digit. Appl. Archaeol. Cult. Herit.*, Elsevier BV, v. 26, n. e00224, p. e00224, set. 2022. Cited on page 61.

BOJE, C. et al. Towards a semantic construction digital twin: Directions for future research. *Autom. Constr.*, Elsevier BV, v. 114, n. 103179, p. 103179, jun. 2020. Cited on page 7.

BUILL, F. et al. Terrestrial laser scanner for the formal assessment of a roman-medieval structure—the cloister of the cathedral of tarragona (spain). *Geosciences*, v. 10, n. 11, 2020. ISSN 2076-3263. DOI:10.3390/geosciences10110427. Cited on page 16.

CALISI, D.; BOTTA, S.; CANNATA, A. Integrated surveying, from laser scanning to uav systems, for detailed documentation of architectural and archeological heritage. *Drones*, v. 7, n. 9, 2023. ISSN 2504-446X. DOI:10.3390/drones7090568. Cited on page 18.

CANTüRK, S. et al. *Graph Positional and Structural Encoder*. 2024. Cited on page 30.

CARTER, J. et al. *LiDAR 101: An Introduction to LiDAR Technology, Data, and Applications*. Charleston, Silver Spring, MD: National Oceanic and Atmospheric Administration (NOAA) Costal Services Center, 2012. 76 p. Cited on page 8.

CHALMERS, P. *Climate Change: Implications for Buildings. Key Findings from the Intergovernmental Panel on Climate Change*. [S.l.], 2014. Fifth Assessment Report. Cited on page 7.

CHAVES, M. U. et al. Visualization of geological features using seismic volume rendering, RGB blending and geobody extraction. In: *12th International Congress of the Brazilian Geophysical Society & EXPOGEF, Rio de Janeiro, Brazil, 15–18 August 2011*. [S.l.]: Brazilian Geophysical Society, 2011. Cited on page 26.

CHEN, A. et al. *MVSNeRF: Fast Generalizable Radiance Field Reconstruction from Multi-View Stereo*. 2021. Cited on page 72.

CHEN, S. et al. Uav bridge inspection through evaluated 3d reconstructions. *Journal of Bridge Engineering*, v. 24, n. 4, p. 05019001, 2019. DOI:10.1061/(ASCE)BE.1943-5592.0001343. Cited on page 11.

CHIAPPINI, S. et al. Comparing the accuracy of 3D urban olive tree models detected by smartphone using LiDAR sensor, photogrammetry and NeRF: a case study of 'ascolana tenera' in italy. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, Copernicus GmbH, X-3-2024, p. 61–68, nov. 2024. Cited 2 times on pages 8 and 23.

CHOY, C. B. et al. *3D-R2N2: A Unified Approach for Single and Multi-view 3D Object Reconstruction*. 2016. Cited on page 27.

CLOUD Compare. [S.l.], 2025. GPL software. Disponível em: <http://www.cloudcompare.org/>. Cited on page 40.

COMMUNITY, B. O. *Blender - a 3D modelling and rendering package*. Stichting Blender Foundation, Amsterdam, 2018. Disponível em: <http://www.blender.org>. Cited on page 40.

COSTANZO, A. et al. Combined use of terrestrial laser scanning and ir thermography applied to a historical building. *Sensors*, v. 15, n. 1, p. 194–213, 2015. ISSN 1424-8220. DOI:10.3390/s150100194. Cited on page 17.

CROCE, V. et al. Comparative assessment of neural radiance fields and photogrammetry in digital heritage: Impact of varying image conditions on 3D reconstruction. MDPI AG, v. 16, n. 2, p. 301, jan. 2024. Cited 2 times on pages 8 and 22.

CROCE, V. et al. From the semantic point cloud to heritage-building information modeling: A semiautomatic approach exploiting machine learning. *Remote Sensing*, v. 13, n. 3, 2021. ISSN 2072-4292. DOI:10.3390/rs13030461. Cited on page 18.

CROCE, V. et al. Neural radiance fields (nerf) for multi-scale 3d modeling of cultural heritage artifacts. *ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, Copernicus GmbH, XLVIII-2/W4-2024, p. 165–171, fev. 2024. Cited 2 times on pages 8 and 22.

CUI, D. et al. 3d reconstruction of building structures incorporating neural radiation fields and geometric constraints. *Automation in Construction*, v. 165, p. 105517, 2024. ISSN 0926-5805. DOI:10.1016/j.autcon.2024.105517. Cited 3 times on pages 9, 24, and 72.

DOSOVITSKIY, A.; BROX, T. Generating images with perceptual similarity metrics based on deep networks. In: LEE, D. et al. (Ed.). *Advances in Neural Information Processing Systems*. [S.l.]: Curran Associates, Inc., 2016. v. 29. Cited on page 33.

DREBIN, R. A.; CARPENTER, L.; HANRAHAN, P. Volume rendering. *SIGGRAPH Comput. Graph.*, Association for Computing Machinery, New York, NY, USA, v. 22, n. 4, p. 65–74, jun. 1988. ISSN 0097-8930. DOI:10.1145/378456.378484. Cited on page 26.

DRONE 4K. *DRONE 4K PONTE ESTAIADA SAO PAULO SP*. 2022. Available at <https://www.youtube.com/watch?v=Cxdx4cCMYZI>. [Accessed 04-02-2025]. Cited on page 48.

DUQUE, L.; SEO, J.; WACKER, J. Bridge deterioration quantification protocol using uav. *Journal of Bridge Engineering*, American Society of Civil Engineers (ASCE), v. 23, n. 10, out. 2018. ISSN 1943-5592. DOI:10.1061/(asce)be.1943-5592.0001289. Cited on page 11.

ELKHRACHY, I. Modeling and visualization of three dimensional objects using low-cost terrestrial photogrammetry. *Int. J. Archit. Heritage: Conserv. Anal. Restor.*, Informa UK Limited, v. 14, n. 10, p. 1456–1467, nov. 2020. Cited on page 50.

FAN, X. et al. *Deep Geometry Post-Processing for Decompressed Point Clouds*. 2022. Cited on page 40.

GEE, A. et al. Engineering a freehand 3d ultrasound system. *Pattern Recognition Letters*, v. 24, n. 4, p. 757–777, 2003. ISSN 0167-8655. DOI:10.1016/S0167-8655(02)00180-0. Cited on page 26.

GOESELE, M.; CURLESS, B.; SEITZ, S. M. Multi-View stereo revisited. In: *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2 (CVPR'06)*. [S.l.]: IEEE, 2006. Cited on page 18.

HADJI-KYRIACOU, A.; ARANDJELOVIĆ, O. Raymarching distance fields with CUDA. *Electronics (Basel)*, MDPI AG, v. 10, n. 22, p. 2730, nov. 2021. Cited on page 26.

HERMAN, G. T. Three-dimensional computer graphic display in medicine: The mipg perspective. In: HÖHNE, K. H. (Ed.). *Pictorial Information Systems in Medicine*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1986. p. 181–210. ISBN 978-3-642-82384-8. Cited on page 25.

HERMAN, G. T.; LIU, H. K. Three-dimensional display of human organs from computed tomograms. *Computer Graphics and Image Processing*, v. 9, n. 1, p. 1–21, 1979. ISSN 0146-664X. DOI:10.1016/0146-664X(79)90079-0. Cited on page 25.

HOHNE, K. H.; BERNSTEIN, R. Shading 3d-images from ct using gray-level gradients. *IEEE Transactions on Medical Imaging*, v. 5, n. 1, p. 45–47, 1986. DOI:10.1109/TMI.1986.4307738. Cited on page 25.

HOSAMO, H. H.; HOSAMO, M. H. Digital twin technology for bridge maintenance using 3d laser scanning: A review. *Advances in Civil Engineering*, v. 2022, n. 1, p. 2194949, 2022. DOI:10.1155/2022/2194949. Cited on page 14.

HU, K. et al. High-fidelity 3d reconstruction of plants using neural radiance fields. *Computers and Electronics in Agriculture*, Elsevier BV, v. 220, p. 108848, maio 2024. ISSN 0168-1699. DOI:10.1016/j.compag.2024.108848. Cited 2 times on pages 8 and 23.

HUANG, H.; TIAN, G.; CHEN, C. Evaluating the point cloud of individual trees generated from images based on neural radiance fields (NeRF) method. *Remote Sens. (Basel)*, MDPI AG, v. 16, n. 6, p. 967, mar. 2024. Cited 2 times on pages 8 and 23.

HUANG, H. et al. Adaptive cloud-to-cloud (ac2c) comparison method for photogrammetric point cloud error estimation considering theoretical error space. *Remote Sensing*, v. 14, n. 17, 2022. ISSN 2072-4292. DOI:10.3390/rs14174289. Cited on page 41.

HUANG, J. et al. FastSplats: optimized splatting on rectilinear grids. In: *Proceedings Visualization 2000. VIS 2000 (Cat. No.00CH37145)*. [S.l.]: IEEE, 2002. Cited on page 26.

HUANG, J. et al. City3d: Large-scale building reconstruction from airborne lidar point clouds. *Remote Sensing*, v. 14, n. 9, 2022. ISSN 2072-4292. DOI:10.3390/rs14092254. Cited on page 15.

HUANG, T. S. Computer vision: Evolution and promise. In: . [S.l.: s.n.], 1996. Cited on page 25.

IEEE. Ieee recommended practice for three-dimensional (3d) medical modeling. *IEEE Std 3333.2.1-2015*, p. 1–32, 2015. DOI:10.1109/IEEESTD.2015.7063875. Cited on page 25.

IMMEL, D. S.; COHEN, M. F.; GREENBERG, D. P. A radiosity method for non-diffuse environments. *SIGGRAPH Comput. Graph.*, Association for Computing Machinery, New York, NY, USA, v. 20, n. 4, p. 133–142, ago. 1986. ISSN 0097-8930. DOI:10.1145/15886.15901. Cited on page 39.

JANS-SINGH, M. et al. Digital twin of an urban-integrated hydroponic farm. *Data-Centric Engineering*, v. 1, p. e20, 2020. DOI:10.1017/dce.2020.21. Cited on page 13.

JAYASINGHE, S. et al. Innovative digital twin with artificial neural networks for real-time monitoring of structural response: A port structure case study. *Ocean Engineering*, v. 312, p. 119187, 2024. ISSN 0029-8018. DOI:10.1016/j.oceaneng.2024.119187. Cited on page 13.

JO, H. C.; SOHN, H.-G.; LIM, Y. M. A LiDAR point cloud data-based method for evaluating strain on a curved steel plate subjected to lateral pressure. *Sensors (Basel)*, MDPI AG, v. 20, n. 3, p. 721, jan. 2020. Cited on page 7.

JUNG, J. T.; MERKLE, D.; REITERER, A. Automated camera pose generation for high-resolution 3d reconstruction of bridges by unmanned aerial vehicles. *Remote Sensing*, v. 16, n. 8, 2024. ISSN 2072-4292. DOI:10.3390/rs16081393. Cited on page 20.

KAARTINEN, E.; DUNPHY, K.; SADHU, A. LiDAR-based structural health monitoring: Applications in civil infrastructure systems. *Sensors (Basel)*, MDPI AG, v. 22, n. 12, p. 4610, jun. 2022. Cited 2 times on pages 7 and 11.

KAJIYA, J.; HERZEN, B. von. Ray tracing volume densities. *ACM SIGGRAPH Computer Graphics*, v. 18, p. 165–174, 07 1984. DOI:10.1145/964965.808594. Cited on page 28.

KARAMOOZIAN, A. et al. Applications of big data and opportunities for the construction industry. In: *2023 9th International Conference on Big Data and Information Analytics (BigDIA)*. [S.l.]: IEEE, 2023. Cited on page 7.

KAUFMAN, A.; MUELLER, K. Overview of volume rendering. In: *Visualization Handbook*. [S.l.]: Elsevier, 2005. p. 127–174. Cited on page 25.

KAZEMNEJAD, A. et al. *The Impact of Positional Encoding on Length Generalization in Transformers*. 2023. Cited on page 29.

KIM, G.; CHA, Y. 3D pixelwise damage mapping using a deep attention based modified nerfacto. *Autom. Constr.*, Elsevier BV, v. 168, n. 105878, p. 105878, dez. 2024. Cited on page 8.

KIM, J.-Y.; LEE, D.; KIM, G.-H. Measurement of work progress using a 3d laser scanner in a structural framework for sustainable construction management. *Sustainability*, v. 16, n. 3, 2024. ISSN 2071-1050. DOI:10.3390/su16031215. Cited on page 17.

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*. [S.l.], 2006. Cited on page 46.

LEE, J.; KIM, R. E. Noncontact dynamic displacements measurements for structural identification using a multi-channel lidar. *Struct. Contr. Health Monit.*, Hindawi Limited, v. 29, n. 11, nov. 2022. Cited on page 7.

LERCARI, N. Monitoring earthen archaeological heritage using multi-temporal terrestrial laser scanning and surface change detection. *Journal of Cultural Heritage*, v. 39, p. 152–165, 2019. ISSN 1296-2074. DOI:10.1016/j.culher.2019.04.005. Cited on page 17.

LEVOY, M. Display of surfaces from volume data. *IEEE Computer Graphics and Applications*, v. 8, n. 3, p. 29–37, 1988. DOI:10.1109/38.511. Cited on page 26.

LI, S. et al. Fusion-3D: Integrated acceleration for instant 3D reconstruction and real-time rendering. In: *2024 57th IEEE/ACM International Symposium on Microarchitecture (MICRO)*. [S.l.]: IEEE, 2024. p. 78–91. Cited 2 times on pages 8 and 22.

LIU, L. et al. *On the Variance of the Adaptive Learning Rate and Beyond*. 2021. Cited on page 37.

LIU, M. et al. Review of digital twin about concepts, technologies, and industrial applications. *J. Manuf. Syst.*, Elsevier BV, v. 58, p. 346–361, jan. 2021. Cited on page 7.

LIU, Y. et al. Point2building: Reconstructing buildings from airborne lidar point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 215, p. 351–368, 2024. ISSN 0924-2716. DOI:10.1016/j.isprsjprs.2024.07.012. Cited on page 15.

LOMBARDI, S. et al. Neural volumes: learning dynamic renderable volumes from images. *ACM Trans. Graph.*, Association for Computing Machinery, New York, NY, USA, v. 38, n. 4, jul. 2019. ISSN 0730-0301. DOI:10.1145/3306346.3323020. Cited on page 27.

LOPES, I.; LALONDE, J.-F.; CHARETTE, R. de. Material transforms from disentangled NeRF representations. 2024. Cited 2 times on pages 8 and 22.

LORENSEN, W. E.; CLINE, H. E. Marching cubes: A high resolution 3d surface construction algorithm. *SIGGRAPH Comput. Graph.*, Association for Computing Machinery, New York, NY, USA, v. 21, n. 4, p. 163–169, ago. 1987. ISSN 0097-8930. DOI:10.1145/37402.37422. Cited on page 25.

LOWE, D. G. Object recognition from local scale-invariant features. In: *Proceedings of the Seventh IEEE International Conference on Computer Vision*. [S.l.]: IEEE, 1999. Cited on page 17.

MARZA, P. et al. AutoNeRF: Training implicit scene representations with autonomous agents. 2023. Cited 2 times on pages 8 and 23.

MAX, N. Optical models for direct volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, v. 1, n. 2, p. 99–108, 1995. DOI:10.1109/2945.468400. Cited on page 29.

MCREYNOLDS, T.; BLYTHE, D. Chapter 20 - scientific visualization. In: MCREYNOLDS, T.; BLYTHE, D. (Ed.). *Advanced Graphics Programming Using OpenGL*. San Francisco: Morgan Kaufmann, 2005, (The Morgan Kaufmann Series in Computer Graphics). p. 531–570. ISBN 978-1-55860-659-3. DOI:10.1016/B978-155860659-3.50022-6. Cited on page 26.

MILDENHALL, B. et al. Local light field fusion: practical view synthesis with prescriptive sampling guidelines. *ACM Trans. Graph.*, Association for Computing Machinery, New York, NY, USA, v. 38, n. 4, jul. 2019. ISSN 0730-0301. DOI:10.1145/3306346.3322980. Cited on page 36.

MILDENHALL, B. et al. NeRF: Representing scenes as neural radiance fields for view synthesis. 2020. Cited 5 times on pages 14, 20, 28, 30, and 31.

MIRZAEI, K. et al. Automatic compliance inspection and monitoring of building structural members using multi-temporal point clouds. *Journal of Building Engineering*, v. 72, p. 106570, 2023. ISSN 2352-7102. DOI:10.1016/j.jobe.2023.106570. Cited on page 7.

MOHAMMADI, M. et al. Integration of tls-derived bridge information modeling (brim) with a decision support system (dss) for digital twinning and asset management of bridge infrastructures. *Computers in Industry*, v. 147, p. 103881, 2023. ISSN 0166-3615. DOI:10.1016/j.compind.2023.103881. Cited on page 13.

NASR, M. A.-S.; ALRAHMAWY, M. F.; TOLBA, A. Multi-scale structural similarity index for motion detection. *Journal of King Saud University - Computer and Information Sciences*, v. 29, n. 3, p. 399–409, 2017. ISSN 1319-1578. DOI:10.1016/j.jksuci.2016.02.004. Cited on page 33.

NGUYEN, C. H. P.; CHOI, Y. Triangular mesh and boundary representation combined approach for 3D CAD lightweight representation for collaborative product development. *J. Comput. Inf. Sci. Eng.*, ASME International, v. 19, n. 1, p. 011009, mar. 2019. Cited on page 14.

NIEMEYER, M. et al. Differentiable volumetric rendering: Learning implicit 3D representations without 3D supervision. 2019. Cited 2 times on pages 14 and 20.

OMAR, T.; NEHDI, M. L. Remote sensing of concrete bridge decks using unmanned aerial vehicle infrared thermography. *Automation in Construction*, v. 83, p. 360–371, 2017. ISSN 0926-5805. DOI:10.1016/j.autcon.2017.06.024. Cited on page 12.

PAL, A. et al. Activity-level construction progress monitoring through semantic segmentation of 3d-informed orthographic images. *Automation in Construction*, v. 157, p. 105157, 2024. ISSN 0926-5805. DOI:10.1016/j.autcon.2023.105157. Cited on page 24.

PALAMADAI, J.; YU, W. *VistaFlow: Photorealistic Volumetric Reconstruction with Dynamic Resolution Management via Q-Learning*. 2025. Cited on page 25.

PAN, X. et al. Vision-based real-time structural vibration measurement through deep-learning-based detection and tracking methods. *Engineering Structures*, v. 281, p. 115676, 2023. ISSN 0141-0296. DOI:10.1016/j.engstruct.2023.115676. Cited on page 12.

PARK, J. J. et al. *DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation*. 2019. Cited on page 38.

Pedro Manuel Calas; MURIETA, E. Sainz de (Ed.). *Multi-span large bridges*. London, England: CRC Press, 2015. Cited on page 46.

PEPINO, L.; RIERA, P.; FERRER, L. Study of positional encoding approaches for audio spectrogram transformers. In: *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. [S.l.]: IEEE, 2022. p. 3713–3717. DOI:10.1109/icassp43922.2022.9747742. Cited on page 30.

PHUA, A.; DAVIES, C.; DELANEY, G. A digital twin hierarchy for metal additive manufacturing. *Computers in Industry*, v. 140, p. 103667, 2022. ISSN 0166-3615. DOI:10.1016/j.compind.2022.103667. Cited on page 13.

POTENZA, F. et al. A robotics and computer-aided procedure for defect evaluation in bridge inspection. *J. Civ. Struct. Health Monit.*, Springer Science and Business Media LLC, v. 10, n. 3, p. 471–484, jul. 2020. Cited on page 12.

RABBY, A. S. A.; ZHANG, C. *BeyondPixels: A Comprehensive Review of the Evolution of Neural Radiance Fields*. 2024. Cited on page 33.

REMONDINO, F. et al. A critical analysis of NeRF-based 3D reconstruction. *Remote Sens. (Basel)*, MDPI AG, v. 15, n. 14, p. 3585, jul. 2023. Cited 2 times on pages 8 and 21.

REUTER, P. et al. Surface reconstruction with enriched reproducing kernel particle approximation. In: *Proceedings Eurographics/IEEE VGTC Symposium Point-Based Graphics, 2005.* [S.l.: s.n.], 2005. p. 79–87. DOI:10.1109/PBG.2005.194068. Cited on page 40.

ROLIN, R. et al. From point cloud data to structural analysis through a geometrical hBIM-oriented model. *J. Comput. Cult. Herit.*, Association for Computing Machinery (ACM), v. 12, n. 2, p. 1–26, jun. 2019. Cited on page 14.

SCHONBERGER, J. L.; FRAHM, J.-M. Structure-from-Motion revisited. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.]: IEEE, 2016. Cited 3 times on pages 17, 20, and 37.

SELBESOGLU, M. O.; BAKIRMAN, T.; GOKBAYRAK, O. Deformation measurement using terrestrial laser scanner for cultural heritage. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W1, p. 89–93, 2016. DOI:10.5194/isprs-archives-XLII-2-W1-89-2016. Cited on page 16.

SHAN, J. *Topographic laser ranging and scanning*. Second edition. | Boca Raton : Taylor & Francis, CRC Press, 2018.: CRC Press, 2018. Cited 3 times on pages 0, 14, and 15.

SHAN, J. et al. Democratizing photogrammetry: an accuracy perspective. *Geo Spat. Inf. Sci.*, Informa UK Limited, v. 26, n. 2, p. 175–188, abr. 2023. Cited on page 50.

SHEN, W. et al. Distilled feature fields enable few-shot language-guided manipulation. jul. 2023. Cited 2 times on pages 8 and 22.

SHIM, C.; KANG, H.; DANG, N. Digital twin models for maintenance of cable-supported bridges. *International Conference on Smart Infrastructure and Construction 2019 (ICSIC)*, p. 737–742, 2019. DOI:10.1680/icsic.64669.737. Cited on page 13.

SHRESTHA, A.; DANG, J.; WANG, X. Development of a smart-device-based vibration-measurement system: Effectiveness examination and application cases to existing structure. *Structural Control and Health Monitoring*, v. 25, n. 3, p. e2120, 2018. DOI:10.1002/stc.2120. Cited on page 12.

SNELLENBURG, J. J. et al. Forward ray tracing for image projection prediction and surface reconstruction in the evaluation of corneal topography systems. *Opt. Express*, Optica Publishing Group, v. 18, n. 18, p. 19324–19338, Aug 2010. DOI:10.1364/OE.18.019324. Cited on page 20.

SONY, S.; LAVENTURE, S.; SADHU, A. A literature review of next-generation smart sensing technology in structural health monitoring. *Structural Control and Health Monitoring*, v. 26, n. 3, p. e2321, 2019. DOI:10.1002/stc.2321. Cited on page 11.

SRAMEK, M. 20 years of volume rendering. In: *Proceedings of the 22nd Spring Conference on Computer Graphics*. New York, NY, USA: Association for Computing Machinery, 2006. (SCCG '06), p. 7–16. ISBN 9781450328296. DOI:10.1145/2602161.2602162. Cited on page 26.

SUN, Z. et al. Approach towards the development of digital twin for structural health monitoring of civil infrastructure: A comprehensive review. *Sensors*, v. 25, n. 1, 2025. ISSN 1424-8220. DOI:10.3390/s25010059. Cited on page 12.

SWEENEY, J.; MUELLER, K. Shear-warp deluxe: The shear-warp algorithm revisited. In: *VisSym*. [S.l.: s.n.], 2002. Cited on page 26.

ŚWIERCZYŃSKA, E. J.; KURDEK, D.; JANKOWSKA, I. Accuracy of the application of mobile technologies for measurements made in headings of the kłodawa salt mine. *Rep. Geod. Geoinformatics*, Walter de Gruyter GmbH, v. 117, n. 1, p. 55–68, jun. 2024. Cited on page 48.

TANCIK, M. et al. *Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains*. 2020. Cited on page 29.

TANCIK, M. et al. Nerfstudio: A modular framework for neural radiance field development. In: *ACM SIGGRAPH 2023 Conference Proceedings*. [S.l.: s.n.], 2023. (SIGGRAPH '23). Cited on page 38.

TATARCHENKO, M.; DOSOVITSKIY, A.; BROX, T. Octree generating networks: Efficient convolutional architectures for high-resolution 3d outputs. *2017 IEEE International Conference on Computer Vision (ICCV)*, p. 2107–2115, 2017. Cited on page 27.

TENG, J. et al. Review on the research and applications of tls in ground surface and constructions deformation monitoring. *Sensors*, v. 22, n. 23, 2022. ISSN 1424-8220. DOI:10.3390/s22239179. Cited on page 16.

TERVEN, J. et al. *Loss Functions and Metrics in Deep Learning*. 2024. Cited on page 32.

The Editors of Encyclopaedia Britannica. *Minaret*. [S.l.]: Encyclopedia Britannica, 2023. Accessed: 2025-02-24. Cited on page 16.

TORZONI, M. et al. A digital twin framework for civil engineering structures. *Computer Methods in Applied Mechanics and Engineering*, v. 418, p. 116584, 2024. ISSN 0045-7825. DOI:10.1016/j.cma.2023.116584. Cited on page 13.

TZACHOR, A. et al. Potential and limitations of digital twins to achieve the sustainable development goals. *Nat. Sustain.*, Springer Science and Business Media LLC, v. 5, n. 10, p. 822–829, jul. 2022. Cited on page 13.

UDUPA, J. Display of 3d information in discrete 3d scenes produced by computerized tomography. *Proceedings of the IEEE*, v. 71, n. 3, p. 420–431, 1983. DOI:10.1109/PROC.1983.12599. Cited on page 25.

USIBD. *Guide for USIBD Document C220TM: Level of Accuracy (LOA) Specification for Building Documentation*. [S.l.], 2016. Document C120TM, Version 2.0. Cited on page 51.

VASWANI, A. et al. Attention is all you need. 2023. Cited 3 times on pages 28, 29, and 30.

VIEN, B. S. et al. Strain monitoring strategy of deformed membrane cover using unmanned aerial vehicle-assisted 3d photogrammetry. *Remote Sensing*, v. 12, n. 17, 2020. ISSN 2072-4292. DOI:10.3390/rs12172738. Cited on page 19.

WANG, J. et al. Integrating bim and lidar for real-time construction quality control. *Journal of Intelligent & Robotic Systems*, v. 79, n. 3, p. 417–432, Aug 2015. ISSN 1573-0409. DOI:10.1007/s10846-014-0116-8. Cited on page 17.

WANG, Q.; KIM, M.-K. Applications of 3D point cloud data in the construction industry: A fifteen-year review from 2004 to 2018. *Adv. Eng. Inform.*, Elsevier BV, v. 39, p. 306–319, jan. 2019. Cited on page 7.

WANG, Q. et al. A comprehensive survey of loss functions in machine learning. *Annals of Data Science*, Springer Science and Business Media LLC, v. 9, n. 2, p. 187–212, abr. 2020. ISSN 2198-5812. DOI:10.1007/s40745-020-00253-5. Cited on page 32.

WANG, Z. et al. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, v. 13, n. 4, p. 600–612, 2004. DOI:10.1109/TIP.2003.819861. Cited on page 33.

WEILER, M. et al. Level-of-detail volume rendering via 3d textures. In: *Proceedings of the 2000 IEEE Symposium on Volume Visualization*. New York, NY, USA: Association for Computing Machinery, 2000. (VVS '00), p. 7–13. ISBN 1581133081. DOI:10.1145/353888.353889. Cited on page 26.

WEIß, S. K. *Deep Learning for Volume Visualization*. 159 p. Tese (Doutorado) — Technische Universität München, 2022. Cited on page 27.

WESTERMANN, R.; ERTL, T. Efficiently using graphics hardware in volume rendering applications. In: *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques*. New York, NY, USA: Association for Computing Machinery, 1998. (SIGGRAPH '98), p. 169–177. ISBN 0897919998. DOI:10.1145/280814.280860. Cited on page 26.

WRIGHT, L.; DAVIDSON, S. How to tell the difference between a model and a digital twin. *Adv. Model. Simul. Eng. Sci.*, Springer Science and Business Media LLC, v. 7, n. 1, dez. 2020. Cited on page 7.

WU, J. et al. *Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling*. 2017. Cited on page 27.

WU, Z. et al. 3D ShapeNets: A deep representation for volumetric shapes. 2014. Cited on page 14.

XING, S. et al. High-efficient computer-generated integral imaging based on the backward ray-tracing technique and optical reconstruction. *Opt. Express*, Optica Publishing Group, v. 25, n. 1, p. 330–338, Jan 2017. DOI:10.1364/OE.25.000330. Cited on page 20.

XIONG, M.; WANG, H. Digital twin applications in aviation industry: A review. *Int. J. Adv. Manuf. Technol.*, Springer Science and Business Media LLC, v. 121, n. 9-10, p. 5677–5692, ago. 2022. Cited on page 7.

XU, C.; SUN, G.; LIANG, R. A survey of volume visualization techniques for feature enhancement. *Vis. Inform.*, Elsevier BV, v. 5, n. 3, p. 70–81, set. 2021. Cited on page 25.

XU, H. et al. Integration of terrestrial laser scanning and nurbs modeling for the deformation monitoring of an earth-rock dam. *Sensors*, v. 19, n. 1, 2019. ISSN 1424-8220. DOI:10.3390/s19010022. Cited on page 16.

XU, J. et al. Developing a digital twin model for monitoring building structural health by combining a building information model and a real-scene 3D model. *Measurement (Lond.)*, Elsevier BV, v. 217, n. 112955, p. 112955, ago. 2023. Cited on page 12.

XU, N. et al. Multi-tiling neural radiance field (nerf)—geometric assessment on large-scale aerial datasets. *The Photogrammetric Record*, Wiley, v. 39, n. 188, p. 718–740, abr. 2024. ISSN 1477-9730. DOI:10.1111/phor.12498. Cited 2 times on pages 8 and 24.

XU, X.; YANG, H. Intelligent crack extraction and analysis for tunnel structures with terrestrial laser scanning measurement. *Advances in Mechanical Engineering*, v. 11, n. 9, p. 1687814019872650, 2019. DOI:10.1177/1687814019872650. Cited on page 16.

YANG, L. et al. Fast registration algorithm for laser point cloud based on 3d-sift features. *Sensors*, v. 25, n. 3, 2025. ISSN 1424-8220. DOI:10.3390/s25030628. Cited on page 41.

YE, C. et al. A digital twin of bridges for structural health monitoring. In: *Structural Health Monitoring 2019*. Lancaster, PA: DEStech Publications, Inc., 2019. Cited on page 13.

YE, Y. et al. Digital twin for the structural health management of reusable spacecraft: A case study. *Engineering Fracture Mechanics*, v. 234, p. 107076, 2020. ISSN 0013-7944. DOI:10.1016/j.engfracmech.2020.107076. Cited on page 13.

YU, A. et al. *PlenOctrees for Real-time Rendering of Neural Radiance Fields*. 2021. DOI:2103.14024. Cited on page 20.

ZHANG, Q.; EAGLESON, R.; PETERS, T. M. Volume visualization: a technical overview with a focus on medical applications. *J. Digit. Imaging*, Springer Science and Business Media LLC, v. 24, n. 4, p. 640–664, ago. 2011. Cited 2 times on pages 25 and 26.

ZHANG, R. et al. The unreasonable effectiveness of deep features as a perceptual metric. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2018. p. 586–595. DOI:10.1109/CVPR.2018.00068. Cited 5 times on pages 0, 33, 34, 36, and 69.

ZHANG, T. et al. Spatiotemporally enhanced photometric loss for self-supervised monocular depth estimation. In: *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. [S.l.]: IEEE, 2022. Cited on page 32.

ZHAO, H. et al. Loss functions for image restoration with neural networks. *IEEE Transactions on Computational Imaging*, v. 3, n. 1, p. 47–57, 2017. DOI:10.1109/TCI.2016.2644865. Cited on page 32.

ZHAO, Q. et al. Optimization algorithm for point cloud quality enhancement based on statistical filtering. *Journal of Sensors*, v. 2021, n. 1, p. 7325600, 2021. DOI:10.1155/2021/7325600. Cited on page 40.

ZHAO, X. et al. Experimental research on quick structural health monitoring technique for bridges using smartphone. *Advances in Materials Science and Engineering*, v. 2016, n. 1, p. 1871230, 2016. DOI:10.1155/2016/1871230. Cited on page 12.

ZHOU, Y. et al. High-precision monitoring method for bridge deformation measurement and error analysis based on terrestrial laser scanning. *Remote Sens. (Basel)*, MDPI AG, v. 16, n. 13, p. 2263, jun. 2024. Cited on page 48.

ZYGUN, D. Implicit neural representations for photorealistic 3d scene reconstruction using signed distance fields. 10 2024. Cited on page 14.