Scalable Digital Twins for Complex Architecture: Preserving Neural Geometry via Deterministic Alignment

Nirab Shrestha[1], Vasit Sagan[1,2] *, Derek Tesser[1], Ashutosh Pawar[1], Hadi Ali Akbarpour[2]

[1] Department of Earth, Environmental and Geospatial Science, Saint Louis University, St. Louis, USA
[2] Department of Computer Science, Saint Louis University, St. Louis, USA

* vasit.sagan@slu.edu

Keywords: neural radiance fields, point cloud registration, Structure-from-Motion, photogrammetry

Introduction

Urban digital twinning and architectural preservation require highly scalable, metrically accurate 3D spatial data. Structure-from-Motion (SfM) photogrammetry is widely used for image-based 3D reconstruction because of its established geometric framework and compatibility with geospatial workflows. However, SfM and subsequent multi-view stereo reconstruction depend on successful image correspondence and can become incomplete around sharp architectural edges, repetitive or homogeneous surfaces, occluded regions, and structures observed against the sky.

Neural Radiance Fields (NeRF) provide an alternative by representing scenes as continuous radiance and density fields learned from posed imagery (Mildenhall et al., 2020). Recent studies show complementary behaviour between neural and conventional photogrammetric reconstruction. Large-scale aerial assessments indicate that NeRF can improve geometric completeness and the representation of fine structural detail, whereas conventional photogrammetry often retains an advantage in metric accuracy (Xu et al., 2023). Civil-infrastructure experiments have also evaluated NeRF and photogrammetry under progressively reduced image availability, demonstrating that their relative performance depends on data availability, geometric requirements, processing demands, and intended application (Atencio et al., 2026). These findings motivate the use of neural reconstruction for scalable and acquisition-constrained digital-twin workflows, but geospatial integration still requires recovery of physical scale and geodetic position.

A fundamental limitation of NeRF-based reconstruction is that the resulting geometry is generally expressed in an arbitrary local coordinate system without inherent physical scale, geodetic position, or conventional map orientation (Remondino et al., 2023). This limits direct quantitative comparison with established geospatial datasets such as photogrammetric products and airborne LiDAR and complicates the integration of neural geometry into spatially explicit digital twins. A robust registration strategy is therefore required to transform the arbitrary neural coordinate system into a metrically meaningful geodetic reference frame while preserving the internal geometry produced by the neural reconstruction.

To bridge this gap, this study investigates a deterministic alignment pipeline for scaling and anchoring UAV-derived neural geometry within a geodetic coordinate system using airborne LiDAR as the fixed spatial reference. The proposed workflow uses spatially distributed structural correspondences and a RANSAC-Umeyama similarity transformation to recover the unknown global scale, rotation, and translation of the NeRF reconstruction. The method is evaluated using correspondence residuals, leave-one-out cross-validation, and comparison with rigid Iterative Closest Point (ICP) refinement. In addition, LiDAR-referenced bidirectional point-cloud distances are used to evaluate the positional accuracy and geometric completeness of Nerfacto and conventional SfM reconstruction across different architectural height zones.

Specifically, this research addresses the following question: To what extent can a correspondence-constrained similarity alignment accurately transform arbitrary-scale neural geometry into a geodetic coordinate system while preserving its structural integrity, and how does its geodetic performance compare with conventional ICP-based refinement?

Method

The study focused on the clock tower on the Saint Louis University north campus, a complex architectural structure containing vertical masonry walls, clock-face features, sharp roof edges, an upper apex and spire, and regions affected by occlusion. RGB imagery was acquired using a Skydio X10 UAV. Traditional photogrammetric reconstruction was performed using Pix4Dmapper, while the neural reconstruction was generated using the Nerfacto implementation in Nerfstudio with camera poses estimated through COLMAP. A dense point cloud was subsequently extracted from the trained neural representation. Airborne LiDAR acquired using a RIEGL VUX-120 and referenced to WGS 84 / UTM Zone 15N (EPSG:32615) served as the fixed geodetic reference for registration and geometric assessment.

A distributed set of six structural correspondences spanning the upper, middle, and lower portions of the tower was manually identified between the local NeRF point cloud and the UTM-georeferenced LiDAR reference. A fixed-seed RANSAC procedure with 10,000 iterations and a 0.50 m inlier threshold was first applied to identify a consensus inlier set and reduce the influence of potentially inconsistent correspondences (Fischler & Bolles, 1981). The accepted correspondences were subsequently processed using the closed-form Umeyama estimator (Umeyama, 1991) to a global similarity transformation as shown in equation 1:

$$X_{UTM} = s\, RX_{NeRF} + t \tag{1}$$

where s is the global scale factor, R is the three-dimensional rotation matrix, and t is the translation vector. The transformation recovered the scale, orientation, and position required to transform the NeRF reconstruction from its arbitrary local coordinate system into EPSG:32615. A single global Sim(3) transformation was then applied to the complete neural point cloud, preserving its internal geometry without non-rigid modification.

Registration performance was evaluated using three complementary approaches. First, correspondence residuals were calculated for the complete similarity solution. Second, Leave-One-Out Cross-Validation (LOOCV) was performed by sequentially withholding one structural correspondence, estimating the similarity transformation from the remaining five points, and predicting the location of the excluded point. This provided a held-out assessment of transformation stability and sensitivity to individual correspondence. Third, rigid Iterative Closest Point (ICP) refinement was evaluated as an alternative nearest-neighbor registration approach to determine whether improved aggregate surface correspondence also improved agreement with the structural geodetic controls.

Georeferencing and reconstruction-quality assessment were treated as separate evaluation problems. The correspondence-constrained RANSAC-Umeyama transformation was retained as the primary NeRF georeferencing solution. For the subsequent photogrammetry-versus-NeRF geometric comparison, equivalent locally centered rigid ICP procedures were applied to both reconstructions to minimize residual global pose differences while preserving their internal geometry and scale. The evaluation region was defined from the LiDAR reference so that areas absent from either reconstruction remained represented in the completeness assessment. All point clouds were voxel-downsampled using a common 0.10 m voxel size to reduce point-density bias prior to comparison. Bidirectional nearest-neighbor Cloud-to-Cloud (C2C) distances were then calculated, with reconstruction-to-LiDAR distances representing positional accuracy and LiDAR-to-reconstruction distances representing geometric completeness. Finally, the LiDAR height extent was stratified into Lower (0–33%), Middle (33–66%), and Upper (66–100%) zones to quantify height-dependent reconstruction performance across different architectural regions.

Results

The correspondence-constrained RANSAC-Umeyama transformation successfully placed the NeRF reconstruction into the LiDAR geodetic reference frame. All six structural correspondences were retained as inliers, yielding a global scale factor of 3.260 and a correspondence fitting RMSE of 0.137 m. Leave-One-Out Cross-Validation produced a 3D RMSE of 0.205 m, while omission of individual correspondences changed the recovered scale by less than 0.27%, indicating a stable similarity solution. Subsequent rigid ICP refinement increased the structural-control RMSE from 0.137 m to 0.211 m; therefore, the RANSAC-Umeyama solution was retained as the primary georeferencing transformation.

LiDAR-referenced comparison revealed complementary performance between the two reconstruction approaches. Pix4D photogrammetry achieved greater reconstruction-to-LiDAR positional accuracy, whereas NeRF provided greater LiDAR-to-reconstruction geometric completeness. The largest difference occurred in the Upper zone, which includes the roof, apex, and spire, where completeness RMSE decreased from 0.175 m for Pix4D to 0.077 m for NeRF. Within a 0.10 m distance threshold, NeRF represented 90.7% of the LiDAR reference geometry, compared with 78.5% for Pix4D. These results indicate that neural reconstruction provides greater geometric continuity in complex architectural regions, while conventional SfM retains superior point-level positional accuracy, as shown in Figure 1.

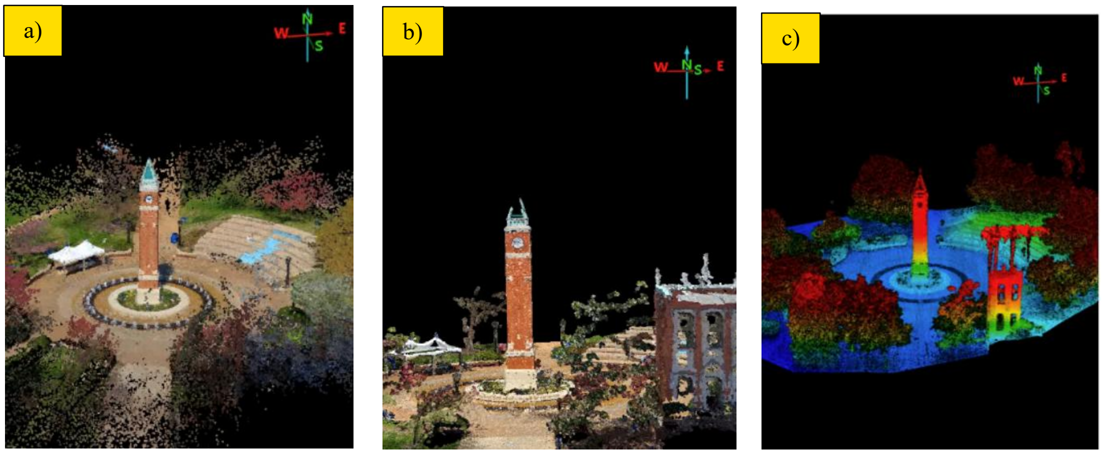

Figure 1: Comparison of 3D reconstructions and spatial analysis of the clocktower. (a) Complete continuous volumetric reconstruction utilizing Neural Radiance Fields (NeRF), demonstrating structural integrity (b) Traditional Structure-from-Motion (SfM) photogrammetric point clouds, exhibiting topological truncation and geometric degradation (c) High-density airborne LiDAR point cloud, utilized as the definitive geodetic baseline for structural validation

Discussion and Conclusion

This study demonstrates that neural georeferencing and reconstruction quality should be evaluated as separate components of a geospatial digital-twin workflow. The correspondence-constrained RANSAC-Umeyama transformation provided a stable means of recovering the unknown scale, orientation, and position of the NeRF reconstruction, while subsequent ICP refinement did not improve agreement with the structural geodetic controls.

The LiDAR comparison further demonstrates that SfM and NeRF provide complementary geometric characteristics. Conventional photogrammetry remains advantageous when local metric accuracy is the primary objective, whereas NeRF can provide greater structural completeness in architecturally complex regions where feature-based reconstruction becomes incomplete. The proposed alignment and validation framework therefore provides a reproducible pathway for integrating arbitrary-scale neural geometry into spatially explicit digital twins while preserving transparent control over geodetic accuracy and geometric completeness.

Acknowledgements

The authors gratefully acknowledge the Remote Sensing Lab at Saint Louis University for providing the essential research infrastructure, equipment, and support for this study.

References

Atencio, E., Duarte, F., Lozano-Galant, F., Porras, R., & Xia, Y. (2026). AI-Based 3D Modeling Strategies for Civil Infrastructure: Quantitative Assessment of NeRF and Photogrammetry. *Sensors*, *26*(3). https://doi.org/10.3390/s26030852

Foley, J. D., Fischler, M. A., & Bolles, R. C. (1981). *Graphics and Image Processing Random Sample Consensus: A Paradigm for Model Fitting with Apphcatlons to Image Analysis and Automated Cartography*.

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2020). NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. *CoRR*, *abs/2003.08934*. https://arxiv.org/abs/2003.08934

Remondino, F., Karami, A., Yan, Z., Mazzacca, G., Rigon, S., & Qin, R. (2023). A Critical Analysis of NeRF-Based 3D Reconstruction. *Remote Sensing*, *15*(14). https://doi.org/10.3390/rs15143585

Umeyama, S. (1991). Least-Squares Estimation of Transformation Parameters Between Two Point Patterns. *IEEE Trans. Pattern Anal. Mach. Intell.*, *13*(4), 376–380. https://doi.org/10.1109/34.88573

Xu, N., Qin, R., Huang, D., & Remondino, F. (2023). Multi-tiling neural radiance field (NeRF)—geometric assessment on large-scale aerial datasets. *The Photogrammetric Record*, *39*, 718–740. https://api.semanticscholar.org/CorpusID:263334081