INTERNATIONAL JOURNAL OF APPLIED METHODS IN ELECTRONICS AND COMPUTERS

www.ijamec.org

International Open Access

Volume 14
Issue 01
March, 2026

Research Article

Simulation-Based Evaluation of LiDAR–Photogrammetry Fusion via NeRF Reconstruction and ICP Registration in Urban Scenes

Rytis Maskeliūnas [a,*], ***Sarmad Maqsood*** [a], ***Ahmad Qurthobi*** [a], ***Irfan Abbas*** [a], ***Mantas Vaškevičius*** [b], ***Julius Gelšvartas*** [b]

[a] *Centre of Real Time Computer Systems, Faculty of Informatics, Kaunas University of Technology, LT-51386 Kaunas, Lithuania.*
[b] *Matomai UAB, LT-51423 Kaunas, Lithuania*

ARTICLE INFO

Article history:
Received 28 January 2026
Accepted 9 March 2026

Keywords:
NeRF,
LiDAR–Photogrammetry Fusion,
Point-Cloud Registration,
Simulation,
Urban Scenes

ABSTRACT

Reliable integration of LiDAR and photogrammetric point clouds is essential for high-precision 3D mapping, yet systematic evaluations of fusion accuracy under controlled conditions remain limited. This study presents a simulation-based assessment framework for LiDAR–photogrammetry fusion using a neural radiance field (NeRF) representation and iterative closest point (ICP) registration. Experiments were conducted in the CARLA simulator (Town 10), where a drone-mounted multi-sensor platform (RGB, LiDAR, GPS, IMU) generated spatially aligned datasets. Photogrammetric reconstructions were produced using the Nerfstudio–Nerfacto pipeline with varied architectural and sampling configurations to analyze their impact on cross-modal registration. Quantitative evaluation employed Chamfer and cloud-to-cloud (C2C) distances to assess NeRF reconstruction fidelity, and ICP root-mean-square error (RMSE), inlier fitness, and runtime to evaluate registration performance. Results show that tuning NeRF's hidden dimensions and sampling levels yields up to 25% lower ICP RMSE and faster convergence across object categories. The proposed framework enables reproducible benchmarking of LiDAR–photogrammetry fusion and provides a foundation for extending NeRF-based methods to real-world urban mapping scenarios.

This is an open access article under the CC BY-SA 4.0 license.
(https://creativecommons.org/licenses/by-sa/4.0/)

1. Introduction

High-fidelity 3D mapping and reconstruction are critical in robotics, geospatial sensing, autonomous navigation, and smart cities [1], [2]. Light detection and ranging (LiDAR) sensors offer precise depth measurements and geometric accuracy even in low texture environments but typically lack rich color or texture information [3]. Photogrammetry, on the other hand, reconstructs dense, textured surfaces from image collections but often suffers from scale drift, holes in textureless or repetitive-surface regions, and sensitivity to lighting and viewpoint. A promising approach is to fuse LiDAR and photogrammetric point clouds to combine geometric accuracy with dense appearance and surface detail. The idea of hybrid LiDAR-photogrammetry fusion has been explored in prior works [4].

However, achieving robust and accurate fusion remains nontrivial. Disparate sensor modalities lead to differences in point densities, noise characteristics, and coordinate referencing [5]. Photogrammetry reconstructions may drift spatially or warp subtly, particularly in textureless or weakly constrained regions. Likewise, LiDAR scans carry measurement noise and partial occlusions [6], [7], [8]. Standard registration methods, such as iterative closest point (ICP), are widely used but can fail under poor initialization, noise, or low overlap; in particular, ICP is known to be sensitive to local minima and initialization quality [9].

Recently, neural radiance field (NeRF) representations

* Corresponding author. E-mail address: *rytis.maskeliunas@ktu.lt*
DOI: 10.58190/ijamec.2026.162

have emerged as a powerful alternative to traditional photogrammetric reconstruction, enabling continuous, photorealistic scene modeling and implicit geometry encoding. Approaches such as bundle-adjusting (BARF) NeRF [10] jointly optimize camera poses and scene radiance fields, improving geometric consistency, while DReg-NeRF [11] extends this idea toward learned registration between neural fields. Nevertheless, prior literature rarely isolates where the true bottleneck lies, whether in the reconstruction quality, the registration algorithm, or their interaction across sensing modalities. This lack of controlled evaluation motivates our work.

In this work, we address that gap by providing a controlled simulation-based evaluation of LiDAR–photogrammetry fusion. Our contributions are:

A reproducible CARLA-based testbed with a drone-mounted sensor suite (RGB, LiDAR, GPS, IMU) in urban scenes, enabling tight control over trajectories and noise.

An ablation study of NeRF-based photogrammetric reconstruction (varying hidden dimension, sampling levels, etc.) to quantify how reconstruction quality affects alignment.

Quantitative fusion benchmarks, reporting Chamfer / cloud-to-cloud (C2C) distances for reconstruction fidelity, and ICP RMSE, inlier fitness, and runtime for registration, including a noise-sensitivity analysis.

Our results show that properly tuned NeRF reconstructions can reduce ICP registration error by up to ~25% compared to naive or default configurations, yield more stable convergence across object types (e.g. poles, road signs, benches), and maintain viable runtimes, underscoring the importance of reconstruction design in effective LiDAR–photogrammetry fusion.

2. Related works

2.1. Photogrammetry and NeRF for 3D Reconstruction

Traditional photogrammetry frameworks reconstruct dense surface models from overlapping 2-D images using structure-from-motion and multi-view-stereo algorithms [12]. Despite high visual fidelity, these methods often suffer from scale drift, poor depth recovery in textureless regions, and lighting sensitivity. Recent advances in neural implicit representations, particularly the NeRF [13], overcome many of these limitations by encoding scene geometry and appearance within a continuous volumetric function. Follow-up variants such as BARF NeRF [10] and DReg-NeRF [11] introduced joint camera-pose optimization and learned inter-NeRF registration, respectively, further improving geometric consistency. While these methods excel at reconstructing compact scenes, few studies have quantitatively examined their alignment behavior when fused with external LiDAR data under controlled noise conditions.

2.2. LiDAR–Image Fusion and Registration

Integrating LiDAR with optical data has long been explored to combine the complementary strengths of accurate geometry and rich texture [14], [15]. Hybrid fusion strategies range from colorizing LiDAR point clouds [5] to joint optimization of photometric and geometric constraints [6]. Rigid registration techniques such as ICP [9] remain a standard baseline, with numerous variants addressing outlier rejection, multi-scale matching, and probabilistic correspondences [10]. Nevertheless, cross-modal registration between LiDAR and photogrammetric or NeRF-derived point clouds remains challenging because of differing point densities and incomplete overlap [16], [17], [18]. Previous studies have focused primarily on real-world data, where measurement noise and uncontrolled motion complicate ground-truth evaluation [19], [20],[21].

2.3. Simulation for 3D Perception

Physics-based simulation offers a practical solution for repeatable benchmarking. Platforms such as CARLA [19] and Blender [22] can synthesize photorealistic scenes with configurable sensors (LiDAR, RGB, GPS, IMU) and controllable noise models. These environments allow rigorous testing of perception and mapping algorithms before deployment to the field. In this context, our work leverages CARLA Town 10 to emulate urban scenes and Blender for high-quality image rendering, enabling reproducible LiDAR–photogrammetry fusion experiments and quantitative comparison of reconstruction and registration accuracy.

3. Simulation Testbed and Dataset

Experiments were carried out in CARLA Town 10, a realistic urban environment containing multi-story buildings, road intersections, traffic signs, light poles, benches, bus stops, and playground structures. This setting provides diverse geometric and textural features for evaluating LiDAR–photogrammetry fusion under controlled conditions. A drone-based sensor platform was simulated to acquire aerial data, ensuring wide-area coverage and varied viewpoints. The sensor suite included synchronized RGB, LiDAR, GPS, and IMU modules, mounted on the virtual drone to emulate real-world survey operations.

High-resolution RGB images were rendered using Blender to generate visually consistent inputs for photogrammetry. These images were reconstructed using Nerfstudio (Nerfacto), producing NeRF-based point clouds that represent the photogrammetric perspective of the environment. The resulting NeRF reconstructions were aligned with simulated LiDAR scans through the ICP algorithm, enabling geometric fusion and performance comparison in a shared coordinate frame.

The dataset provides synchronized RGB–LiDAR

sequences, corresponding 6-DoF ground-truth poses, and NeRF reconstructions suitable for both training and validation tasks. These assets collectively allow reproducible analysis of reconstruction fidelity and registration accuracy. Fig. 1 shows representative Town 10 views used for capture, Fig. 2 shows the visualization with the CARLA simulator (point cloud), and Table I summarizes the simulated sensor suite and dataset organization.



Figure 1. Representative views of CARLA Town 10 used for data capture.

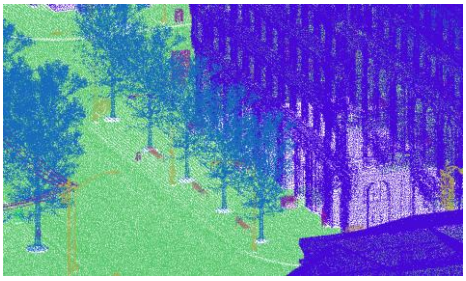


Figure 2. Town 10 point-cloud view used for geometric evaluation.

Table 1. Summary of simulated sensors and dataset composition

Sensor	Modality	Description	Output
RGB Camera	Visual	High-quality images rendered in Blender	Input for NeRF reconstruction
LiDAR	Depth	Simulated CARLA 3D geometry scans	Ground truth for registration
GPS	Position	Provides reference coordinates	Pose alignment
IMU	Motion	Orientation and acceleration	Temporal synchronization

4. Methodology

4.1. Photogrammetry via NeRF

Photogrammetric reconstruction was performed using Nerfstudio (Nerfacto), trained on synthetic RGB images rendered in Blender under controlled lighting. A subset of frames was selected from each drone trajectory to ensure diverse viewpoints and minimal motion blur. Camera poses were derived directly from CARLA's simulated GPS/IMU data and refined through bundle adjustment during NeRF optimization. The complete processing workflow is illustrated in Fig. 3.

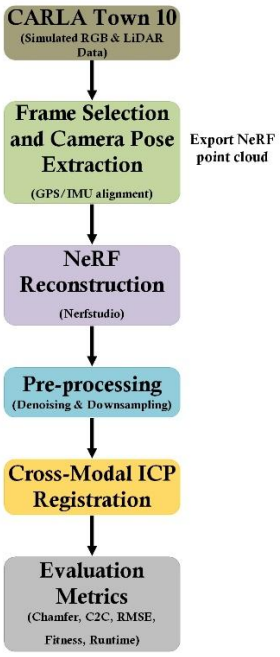


Figure 3. End-to-end LiDAR–Photogrammetry fusion framework.

The NeRF network used the standard volumetric rendering equation as:

$$C(r) = \int_{t_n}^{t_f} T(t)\ \sigma\big(r(t)\big)\, c\big(r(t)\big)\, dt, \tag{1}$$

where $C(\mathbf{r})$ is the predicted pixel color along a camera ray $\boldsymbol{r}$, σ denotes volumetric density, and $\boldsymbol{c}$ represents emitted radiance, and $T(t) = \exp\left(-\int_{t_n}^{t} \sigma\big(r(s)\big)\ ds\right)$ is the accumulated transmittance.

Key configuration parameters included hidden layer size (*hidden_dim* = 256), hash-grid depth (*num_levels* = 32), and adaptive ray sampling. Models were trained for 20,000 iterations using the Adam optimizer ($\mathrm{lr} = 1 \times 10^{-3}$), with exponential decay after 10,000 steps. Denoising and voxel down-sampling were applied to exported point clouds (voxel size = 0.02 m) to suppress background noise before LiDAR alignment. A qualitative example is shown in Fig. 4.

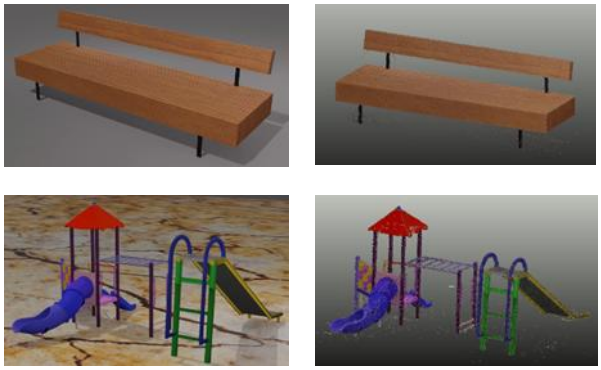


Figure 4. Qualitative NeRF reconstructions (Nerfacto): bench and playground after architecture tuning.

Table 2 lists the NeRF configurations and corresponding training settings.

Table 2. NeRF configurations compared

Config ID	Hidden Dim	Num Levels	Comment
A	128	16	Lower-capacity baseline
B	256	32	Best trade-off (reported)
C	512	32	Higher compute; marginal gain

4.2. Cross-Modal Registration (ICP)

To align NeRF-derived and LiDAR point clouds, we employed the ICP algorithm. Each point cloud was pre-processed by statistical outlier removal and normal estimation. Coarse initialization was obtained using centroids derived from GPS pose priors. ICP minimizes the squared distance between corresponding points:

$$E(R,t) = \frac{1}{N}\sum_{i=1}^{N} ||\boldsymbol{p}_i - (R\boldsymbol{q}_i + t)||^2, \quad (2)$$

where $\boldsymbol{p}_i$ and $\boldsymbol{q}_i$ denote LiDAR and NeRF points, respectively, and (R, t) represent the rotation and translation to be estimated.

A voxel resolution of 0.05 m was used for down-sampling. Maximum correspondence distance was set to 0.1 m, with 10 iterations per run. Convergence was declared when the change in mean-squared error dropped below 10^{-6} m² or pose-update norms fell below 10^{-5}.

4.3. Evaluation Metrics

Reconstruction Quality:

Accuracy between NeRF-reconstructed and LiDAR ground-truth point clouds was measured using the Chamfer distance as:

$$d_{CD}(P,Q) = \frac{1}{|P|}\sum_{p\in P} min_{q\in Q} \|\mathbf{p}-\mathbf{q}\|_2^2 + \frac{1}{|Q|}\sum_{q\in Q} min_{p\in P}\|\mathbf{q}-\mathbf{p}\|_2^2. \quad (3)$$

and cloud-to-cloud (C2C) distances computed via nearest-neighbor search.

Registration Accuracy:

After ICP, alignment performance was reported using RMSE, fitness score (inlier ratio within correspondence threshold), and runtime per iteration, along with convergence curves.

Runtime and Convergence:

Training and inference times were recorded for each NeRF configuration, while ICP convergence curves were used to analyze stability across object categories (e.g., poles, benches, playgrounds).

5. Experiments

5.1. Ablation on NeRF Capacity and Sampling

We first analyzed how the NeRF network's representational capacity affects reconstruction and subsequent LiDAR alignment. As described in Section IV-A, the parameters *hidden_dim* ∈ {32,64,128,256,512} and *num_levels* ∈ {32,64,128,256,512} were varied. Increasing *hidden_dim* improved structural detail recovery but incurred diminishing returns beyond 256 units. Fig. 5 illustrates this trend, showing that configuration B (*hidden_dim* = 256, *num_levels* = 32) achieved the lowest Chamfer and C2C distances. This setup also yielded faster and more stable ICP convergence, thus serving as the baseline for subsequent experiments.

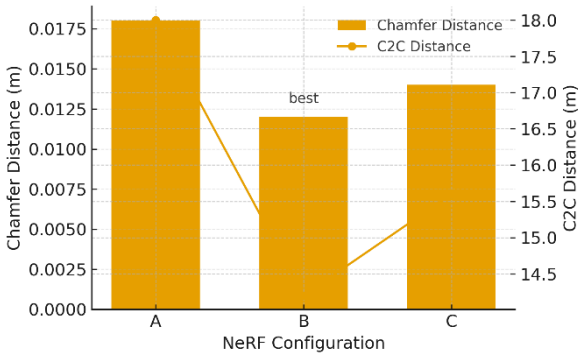


Figure 5. Reconstruction quality vs. NeRF configuration (Chamfer & C2C).

5.2. Object-wise and Scene-wise Registration

For cross-modal registration, five representative urban objects, trash can, pole, playground, road sign, and bench, were evaluated. The NeRF point clouds reconstructed with the optimal configuration (256/32) were aligned with the corresponding LiDAR scans using ICP (10 iterations, voxel = 0.05 m). Table III summarizes the object-wise results, while Fig. 6 presents the ICP convergence curves for "easy" (road sign, pole) and "hard" (playground, bench) objects.

Table 3. Object-wise ICP Results

Object	Initial RMSE (m)	Final RMSE (m)	Best Iter	Fitness (%)	Runtime (s)
Trash can	0.098	0.096	3	97.5	0.92
Pole	0.069	0.069	4	98.2	0.95
Playground	0.200	0.072	7	94.6	1.12
Road sign	0.035	0.034	6	99.1	0.87
Bench	0.150	0.133	4	96.3	0.98

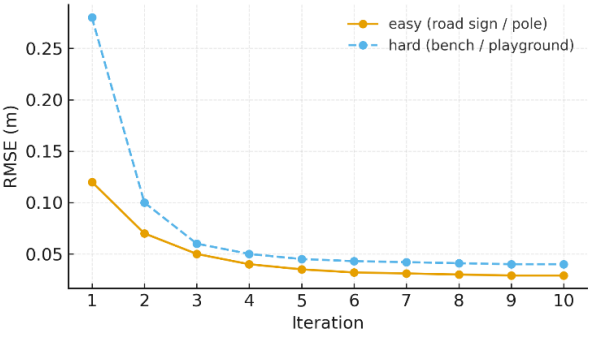


Figure 6. ICP convergence curves (RMSE vs. iteration) for "easy" vs. "hard" objects.

As shown in Fig. 6, structured objects exhibit rapid error

reduction and smooth convergence, whereas complex geometries converge more slowly. Fig. 7(a–e) provides visual overlays of the initial NeRF reconstruction and the final LiDAR–NeRF alignment, highlighting the improved correspondence achieved after ICP registration.

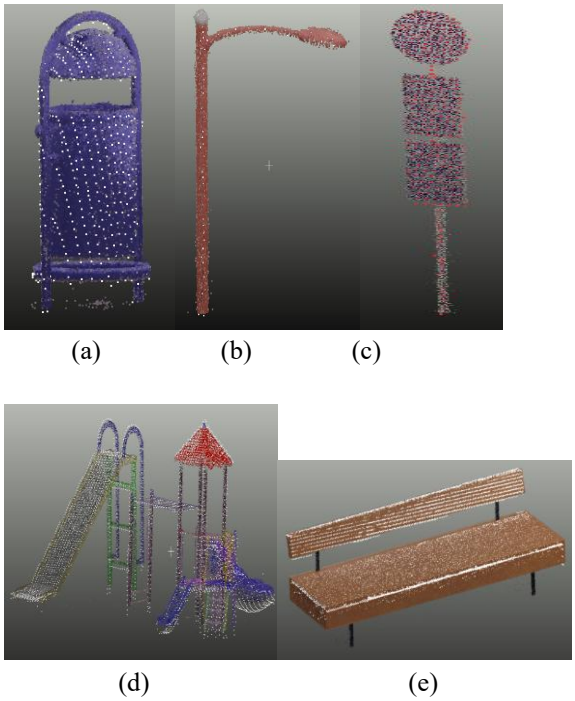


Figure 7. (a–e) Per-object LiDAR–NeRF overlays (trash can, pole, playground, road sign, bench).

5.3. Noise Sensitivity

To evaluate robustness under sensor perturbations, simulated Gaussian noise was applied to LiDAR range and GPS/IMU pose data at three levels (low, medium, high). Table IV summarizes the averaged ICP performance under each condition. At medium noise, RMSE increased by roughly 11%, while high noise caused a 27% degradation and a 2-iteration delay in convergence. Despite this, denoised and voxel-downsampled NeRF inputs maintained > 90% fitness, demonstrating that moderate pre-processing reduces most noise-induced drift.

Table 4. ICP Results under Sensor Noise

Noise Level	RMSE (m)	Fitness (%)	Avg. Iters	Observation
Low	0.071	97.8	5	Stable convergence
Medium	0.079	94.5	6	Minor drift
High	0.090	91.2	7	Slower convergence

6. Discussion and Limitations

The experiments revealed that network capacity and pre-processing quality are the most influential factors for improving LiDAR–NeRF alignment. Increasing the hidden layer dimension to 256 and adjusting the hash-grid depth yielded significantly lower Chamfer and C2C distances. Normalization and voxel down-sampling before ICP further enhanced convergence, reducing RMSE in complex objects such as playground structures by over 60%. These findings confirm that reconstruction quality directly governs registration stability.

Nevertheless, several failure cases emerged. NeRF reconstructions degraded for textureless or reflective surfaces, such as metal signs, and repetitive geometries like poles often led to local minima during ICP. Fine details, including thin structures, occasionally appeared blurred due to limited sampling density, indicating a need for adaptive ray sampling and improved surface regularization.

A clear speed–accuracy trade-off was observed: higher-capacity NeRFs improved alignment but increased computation time. While suitable for controlled or offline analyses, real-world applications may require lightweight NeRF variants or partial fine-tuning for on-site mapping. Additionally, the current pipeline was evaluated only in a simulated CARLA environment; real-world data would introduce challenges such as illumination variation, occlusion, and sensor drift, requiring domain adaptation or hybrid calibration strategies.

7. Conclusions

This work presented a controlled simulation framework for evaluating LiDAR–photogrammetry fusion using neural radiance fields (NeRF) and iterative closest point (ICP) registration. Through systematic ablations on NeRF capacity and pre-processing, we demonstrated that moderate network scaling and voxel down-sampling significantly improve alignment accuracy and convergence speed. The proposed CARLA-based setup enables reproducible benchmarking of reconstruction and registration quality under varying noise levels and object complexities. Future work will extend this framework to small-scale real-world pilots, emphasizing robust global initialization across large scenes and the integration of uncertainty-aware ICP to handle real sensor variability and dynamic lighting conditions.

Declaration of Ethical Standards

This study is based entirely on simulation-generated data and did not involve human participants, human tissue, animals, or patient-specific records. therefore, ethical approval and informed consent were not required for this work.

CRediT Authorship Contribution Statement

Rytis Maskeliūnas: Conceptualization, Methodology, Supervision, Project administration, Funding, Writing – review & editing. Sarmad Maqsood: Methodology,

Validation, Formal analysis, Investigation, Writing – original draft, Visualization. Ahmad Qurthobi: Data curation, Visualization. Irfan Abbas: Formal analysis. Mantas Vaškevičius: Software, Resources, Data curation, Writing – review & editing. Julius Gelšvartas: Conceptualization, Resources, Validation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

FUNDING/Acknowledgment

This research project no.: 02-019-K-0044 was funded by the European Union Funds for the period 2021-2027 under the Measure No. 05-001-01-05-07 "Establishing a coherent system for the promotion of innovative activities" under the activity "Stimulating the supply of innovations"' under the action "Investing in activities for the development of new high value added products and enabling researchers to participate in R&D activities of enterprises, promotion of intellectual property, early pilot production of new products being developed and preparation for the market" (region of Central and Western Lithuania).

Data Availability Statement

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Author Note

This study was previously presented as a proceedings paper at the *3rd International Conference on Intelligent Systems and New Applications (ICISNA'25)*, held in Antalya, Türkiye, on December 12–14, 2025.

References

[1] T. Ren and H. Jebelli, "Efficient 3D robotic mapping and navigation method in complex construction environments," *Computer-Aided Civil and Infrastructure Engineering*, vol. 40, no. 12, pp. 1580-1605, 2025.

[2] R. Maskeliūnas, S. Maqsood, "Lightweight Attention-Based Framework for Semantic Segmentation and Compression of 3D LiDAR Data", SETSCI Conference Proceedings, vol. 22, pp. 7-10, 2025.

[3] P. Chen, X. Zhao, L. Zeng, L. Liu, S. Liu, L. Sun, Z. Li, H. Chen, G. Liu, Z. Qiao, Y. Qu, D. Xu, L. Li, and L. Li, "A Review of Research on SLAM Technology Based on the Fusion of LiDAR and Vision," *Sensors*, vol. 25, no. 5, 1447, 2025.

[4] R. Maskeliūnas, S. Maqsood, M. Vaškevičius, and J. Gelšvartas, "Fusing LiDAR and photogrammetry for accurate 3D data: A hybrid approach," *Remote sensing*, vol. 17, no. 3, pp. 1-27, 2025.

[5] Y. Ye, J. Shan, L. Bruzzone, and L. Shen, "Robust registration of multimodal remote sensing images based on structural similarity," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 5, pp. 2941-2958, 2017.

[6] K. Istenič, N. Gracias, A. Arnaubec, J. Escartín, and R. Garcia, "Scale accuracy evaluation of image-based 3D reconstruction strategies using laser photogrammetry," *Remote Sensing*, vol. 11, no. 18, 2093, 2019.

[7] R. Maskeliūnas, and S. Maqsood, "Hybrid attention-based PTv3-SE model for efficient point cloud segmentation," *Remote Sensing Applications: Society and Environment*, vol. 41, 101891, 2026.

[8] M. Gassilloud, B. Koch, and A. Göritz, "Occlusion mapping reveals the impact of flight and sensing parameters on vertical forest structure exploration with cost-effective UAV based laser scanning," *International Journal of Applied Earth Observation and Geoinformation*, vol. 139, 104493, 2025.

[9] J. Yang, H. Li, D. Campbell, and Y. Jia, "Go-ICP: A globally optimal solution to 3D ICP point-set registration," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 11, pp. 2241-2254, 2015.

[10] C. H. Lin, W. C. Ma, A. Torralba, and S. Lucey, "Barf: Bundle-adjusting neural radiance fields," In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 5741-5751, 2021.

[11] Y. Chen and G. H. Lee, "DReg-NeRF: Deep Registration for Neural Radiance Fields," in *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023.

[12] A. T. D. S. Ferreira, C. H. Grohmann, M. C. H. Ribeiro, M. S. T. Santos, R. C. de Oliveira, and E. Siegle, "Beach surface model construction: A strategy approach with structure from motion-multi-view stereo," *MethodsX*, vol. 12, 102694, 2024.

[13] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," *Communications of the ACM*, vol. 65, no. 1, pp. 99-106, 2021.

[14] Y. Li and X. Xiao, "Deep Learning-Based Fusion of Optical, Radar, and LiDAR Data for Advancing Land Monitoring," *Sensors*, vol. 25, no. 16, 4991, 2025.

[15] S. Xu, Q. Xue, Z. Chen, S. Fei, and H. Gao, "Complementary information-guided interactive fusion network for HSI and LiDAR data joint classification," *Expert Systems with Applications*, vol. 298, 129549, 2026.

[16] N. Xu, R. Qin, and S. Song, "Point cloud registration for LiDAR and photogrammetric data: A critical synthesis and performance analysis on classic and deep learning algorithms," *ISPRS Open J. Photogramm. Remote Sens.*, vol. 8, p. 100032, 2023.

[17] K. Ma, F. Yan, S. Li, G. Huang, X. Jia, F. Wang, and L. Chen, "Low-Overlap Registration of Multi-Source LiDAR Point Clouds in Urban Scenes Through Dual-Stage Feature Pruning and Progressive Hierarchical Methods," *Remote Sensing*, vol. 17, no. 17, 2938, 2025.

[18] J. Wang and H. Xu, "Cross-modal deep learning framework for 3D reconstruction and information integration of Zhejiang wood carving heritage," *Scientific Reports*, vol. 16, 465, 2025.

[19] M. Ramezani, Y. Wang, M. Camurri, D. Wisth, M. Mattamala, and M. Fallon, "The Newer College Dataset: Handheld LiDAR, Inertial and Vision with Ground Truth," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Las Vegas, NV, USA, Oct. 2020, pp. 4353–4360.

[20] R. Maskeliūnas, S. Maqsood, M. Vaškevičius, and J. Gelšvartas, "Hybrid deep learning and geometric algorithms for individual object detection in urban LiDAR point clouds," *International Journal of Remote Sensing*, vol. 46, no. 23, pp. 9118–9156, 2025.

[21] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," In *Conference on robot learning,* pp. 1-16, 2017.

[22] T. Roosendaal and Community, *Blender 3.0 User Manual*, Blender Foundation, 2021.