Validating AI-Based Depth Estimation for Road Scene Reconstruction Using Dashcam Images and Low-Cost LiDAR

Bilguunmaa Myagmardulam[1]*, Kazuyoshi Takahashi[2]

[1]*Toyama Prefectural University, 5180 Kurokawa, Imizu, Toyama, Japan.*
Bilguunmaa@pu-toyama.ac.jp
[2]*Nagaoka University of Technology, 1603-1 Kamitomiokamachi, Nagaoka, Niigata, Japan.*
ktakaha@vos.nagaokaut.ac.jp

Abstract: Monocular depth estimation powered by deep learning has the potential to significantly reduce the cost and complexity of 3D scene reconstruction in mobile applications. This study evaluates the performance of the VGGT model using a low-cost Mobile Mapping System (MMS) consisting of a consumer dashcam, a GNSS/IMU unit, and a compact LiDAR sensor. Dashcam images were undistorted with MATLAB-derived calibration parameters before depth estimation. The predicted depth maps were transformed into sparse 3D point clouds and compared with LiDAR-derived reference data in CloudCompare. Structure-from-Motion (SfM) outputs were also used for cross-validation. Results from snow-free urban road segments indicate that the VGGT model can reconstruct consistent roadway geometry and capture features such as lane boundaries and crosswalk markings. While errors occurred in regions containing moving vehicles, shadows, or reflective surfaces, the findings confirm the feasibility of applying AI-based depth estimation to low-cost road monitoring. Future work will extend validation to snow-covered conditions to assess winter applicability.

Keywords: Monocular Depth Estimation, VGGT, Low-Cost MMS LiDAR, Dashcam, Road Scene Reconstruction

1. Introduction

Accurate 3D reconstruction of road environments is crucial for infrastructure management, safety monitoring, and winter road maintenance. Traditional MMS platforms equipped with high-end LiDAR and GNSS/INS units are effective but often prohibitively expensive for continuous patrol and monitoring. Recent advances in deep learning have enabled monocular depth estimation from a single camera, with models such as MiDaS (Ranftl et al., 2022), ZoeDepth, Depth-Anything-V2, and VGGT (Wang et al., 2025) demonstrating strong performance across varied environments (2). Despite these advances, few studies have validated such models using **low-cost MMS data** in real-world road conditions. Moreover, the potential of integrating dashcam imagery with affordable LiDAR for ground-truth validation has not been fully explored. While earlier work validated MMS point cloud accuracy for snow monitoring (Kobayashi et al., 2025), the integration of AI-based monocular depth estimation with such systems has not been systematically explored.

The primary objective of this work is to assess the feasibility and accuracy of AI-based monocular depth estimation for road-scene reconstruction and to establish protocols for validating depth outputs against LiDAR ground truth.

2. Methodology

A low-cost Mobile Mapping System (MMS) was assembled using a consumer-grade dashcam, a GNSS/IMU unit, and a compact 3D LiDAR sensor (Fig 1). The dashcam continuously recorded video frames of urban road environments, while GNSS/IMU measurements provided positioning and trajectory information. The LiDAR data served as a reference dataset for validation.

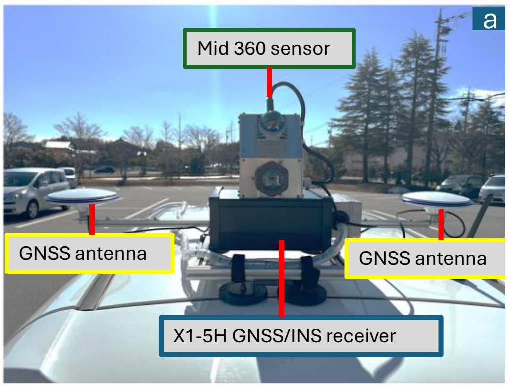

Figure 1: Low-cost MMS.

Lens distortion parameters were estimated in MATLAB and applied to correct raw dashcam images prior to inference. The corrected frames were processed with the VGGT model to generate monocular depth maps, which were then converted into sparse 3D point clouds. To evaluate accuracy, the reconstructed point clouds were aligned with LiDAR point clouds using CloudCompare. SfM reconstructions derived from VGGT outputs were additionally used for cross-validation. The study area consisted of snow-free road segments in Niigata Prefecture, Japan. Data were collected under varied illumination and traffic conditions. Feature extraction experiments focused on roadway elements, including crosswalks, to assess the system's ability to capture critical infrastructure features.

3. Results/Findings

The VGGT model was applied to dashcam imagery collected in urban road environments and compared against LiDAR-derived depth maps. The reconstructed road surfaces showed good structural consistency with LiDAR point clouds, especially for flat road planes and large static objects. Cross-validation using Structure-from-Motion (SfM) outputs confirmed that the estimated geometry was stable and reproducible. Qualitative analysis indicated that roadway features such as lane boundaries and crosswalk markings were captured in the depth maps, demonstrating the potential for detecting critical infrastructure elements. However, depth errors were more pronounced around moving vehicles, reflective surfaces, and in regions with strong shadows, where alignment with LiDAR was less reliable. Overall, the findings suggest that monocular depth estimation can generate meaningful 3D reconstructions of road scenes from consumer-grade dashcam video, supporting the feasibility of low-cost mobile mapping for road monitoring. Figure 2 illustrates an example of crosswalk feature extraction from dashcam and LiDAR data.

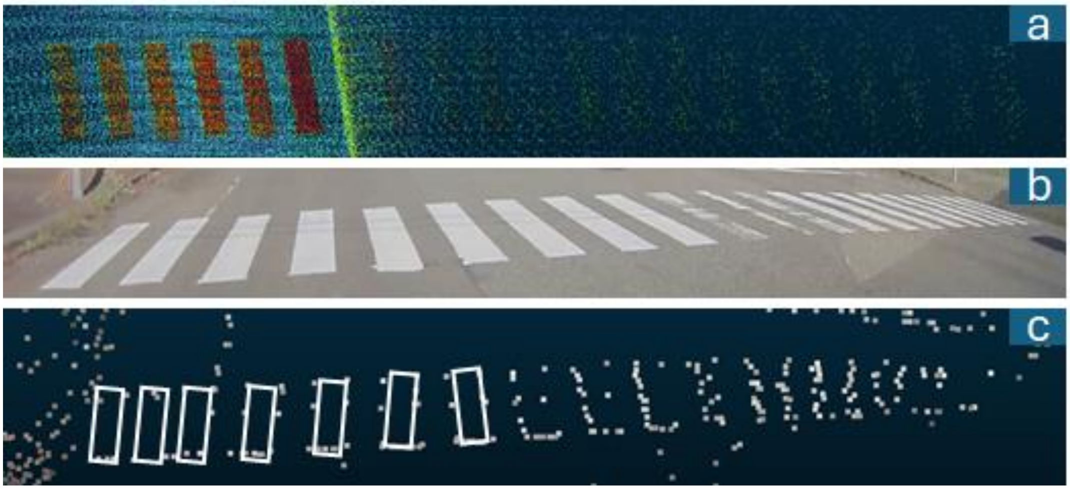

Figure 2: Crosswalk feature extraction using low-cost MMS. (a) LiDAR point cloud showing high-density road surface points, (b) dashcam reference image, and (c) reconstructed sparse features with bounding boxes around crosswalk markings.

4. Conclusion

This study demonstrates that monocular depth estimation using VGGT, when validated against a low-cost MMS with LiDAR, can generate meaningful 3D reconstructions of urban road environments. Although the results are sparser and less accurate than LiDAR measurements, essential roadway elements such as crosswalks and lane boundaries were preserved. These findings support the potential application of AI-based monocular depth estimation in affordable road monitoring systems. Future work will extend this validation to snow-covered roads and incorporate additional deep learning models to evaluate robustness under winter conditions.

Acknowledgment

This study was financially supported by a Research Grant from the Remote Sensing Technology Center of Japan (RESTEC) in 2024.

References

Kobayashi, M., Nakamura, T., Myagmardulam, B. and Takahashi, K., 2025. Fundamental study on road snow depth measurement using a low-cost MMS. *Doboku Gakkai Ronbunshuu (Journal of JSCE)*, 81(22), Paper No. 24-22003.

Ranftl, R., Bochkovskiy, A. and Koltun, V., 2022. Vision transformers for dense prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12179–12188.

Wang, J., Chen, M., Karaev, N., Vedaldi, A., Rupprecht, C. and Novotny, D., 2025. VGG-T: Visual Geometry Grounded Transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5294–5306.