Robust LiDAR-Camera Calibration for 3D Mapping in Complex Natural Environments

Giho Lee[1], Seohyeong Jeong[2], and Changwon Kim[1*]

[1]Control and Instrumentation Engineering, Changwon National University, Changwon 51140, Republic of Korea
[2]AI Convergence Engineering, Changwon National University, Changwon 51140, Republic of Korea
**Corresponding Author E-mail: cwkm22@changwon.ac.kr*

ABSTRACT

Accurate extrinsic calibration between LiDAR and cameras is essential for reliable multi-sensor fusion and three-dimensional mapping in environmental monitoring applications. However, calibration in complex natural environments is particularly challenging because vegetation, irregular terrain, weak structural regularity, sparse or uneven LiDAR measurements, and substantial cross-modal appearance differences can reduce the reliability of conventional feature- or target-based calibration methods. This study proposes a robust targetless LiDAR-camera calibration framework for 3D mapping in complex natural environments. The proposed method directly utilizes raw camera images and LiDAR measurements and constructs complementary structural representations from image features and LiDAR range information. A staged optimization framework is employed to progressively estimate the six-degree-of-freedom extrinsic transformation while reducing the influence of unreliable local correspondences. Candidate transformations are evaluated using observability-aware criteria that consider structural alignment, spatial coverage, regional projection residuals, point-distribution support, and cross-metric geometric consistency. Consensus-based refinement and safety constraints are further introduced to prevent unstable parameter updates in scenes containing insufficient or locally ambiguous geometric information. The calibration parameters are estimated without using ground-truth extrinsic information during optimization. Quantitative experiments on the KITTI dataset evaluate rotational and translational calibration accuracy as well as projection consistency across multiple frames, while additional experiments using the HAInich forest dataset distributed through Dryad examine the applicability of the proposed framework to unstructured natural scenes. The proposed approach provides a practical calibration framework for reliable LiDAR-camera fusion and supports robust 3D mapping for environmental sensing and monitoring systems.

Keywords: LiDAR-camera calibration / Extrinsic calibration / 3D mapping / Natural environment / Environmental monitoring / Multi sensor fusion

1. INTRODUCTION

Accurate three-dimensional (3D) perception of natural environments is important for environmental monitoring, forest inventory, ecosystem observation, and autonomous field robotics. Complex outdoor environments such as forests contain irregular terrain, dense vegetation, and highly variable geometric structures, making reliable spatial perception challenging. Representative datasets such as HAInich and RELLIS-3D demonstrate the importance of multi-sensor perception in forest and unstructured off-road environments [1], [2].

Cameras and LiDAR provide complementary information for 3D perception. Cameras capture dense visual and semantic information, whereas LiDAR directly measures 3D geometry and distance. Their fusion is widely used for mapping and robotic perception, as demonstrated by datasets such as KITTI [3]. However, accurate fusion requires precise extrinsic calibration between the LiDAR and camera coordinate systems. Even small rotational or translational errors can cause projection misalignment and degrade subsequent tasks such as 3D mapping, depth association, and object

localization. Traditional LiDAR-camera calibration commonly relies on dedicated targets such as planar checkerboards [4]. Although accurate under controlled conditions, target-based approaches are less practical in large outdoor environments. Targetless methods instead exploit naturally occurring scene information. Pandey et al. [5] proposed calibration based on mutual information, while Yuan et al. [6] utilized natural edge features. EdgeCalib [7] further introduced multi-frame weighted edge features to improve calibration robustness. Nevertheless, targetless calibration remains difficult in natural environments because image structures may originate from texture or illumination without corresponding LiDAR geometry, while LiDAR measurements are sparse and unevenly distributed.

To address these challenges, this study proposes a robust targetless LiDAR-camera calibration framework for 3D mapping in complex natural environments. The proposed method estimates the six-degree-of-freedom extrinsic transformation directly from camera and LiDAR measurements without using ground-truth extrinsic parameters during optimization. A staged optimization process evaluates candidate transformations using complementary structural and geometric criteria, while poorly supported parameter updates are restricted. Information from multiple frames is then combined to obtain a more stable calibration result.

The proposed framework is quantitatively evaluated on the KITTI dataset using reference extrinsic parameters [3]. Additional experiments using the HAInich forest dataset distributed through Dryad [1] examine its applicability to complex natural environments. Since independent LiDAR-camera extrinsic ground truth is unavailable for the forest experiments, the evaluation focuses on projection consistency and practical applicability.

2. METHODOLOGY

2.1 Proposed Calibration Method

The proposed framework performs targetless LiDAR-camera extrinsic calibration directly from raw RGB images and raw LiDAR measurements. For each frame, monocular depth is independently estimated from the RGB image using Depth Anything V2 [8], while a range-image representation is generated from the original LiDAR points. Structural information obtained from the two modalities is then used to estimate the six-degree-of-freedom (6-DoF) LiDAR-to-camera transformation. For the initial cross-modal alignment, feature representations from WAFT-Stereo [9] are used to generate correspondence candidates. For a LiDAR point the transformation into the camera coordinate system is defined as

$$P_C = R\,P_L + t \tag{1}$$

where R and t denote the rotation and translation between the sensors. The transformed LiDAR points are projected onto the image using the camera intrinsic parameters.

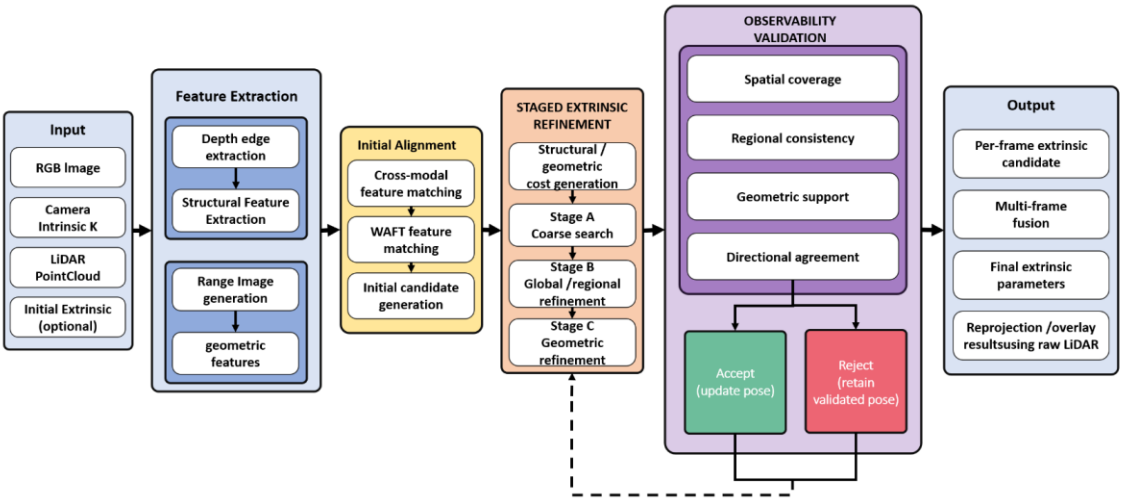


Figure 1. Overall architecture of the proposed **targetless** LiDAR-camera calibration framework

2.2 Staged Calibration and Observability Validation

The calibration is progressively refined through three stages. Stage A estimates an initial frame-wise SE(3) transformation using normalized nearness, multi-scale structural matching, and edge consistency between the camera and LiDAR representations. Stage B refines this estimate by evaluating global and regional consistency over near-, mid-, far-, focus-, and horizontal spatial-bin regions, rejecting updates that improve only local areas while degrading other supported regions. Stage C performs the final cross-metric geometric refinement and retains an update only when complementary edge, projection, spatial-coverage, and geometric criteria remain mutually consistent.

Because Depth Anything V2 provides relative rather than metric depth, its output is treated as a disparity-like quantity in which larger values indicate nearer structure. For both modalities, valid values are robustly normalized using the 2nd and 98th percentiles; the normalized LiDAR depth is inverted so that both camera and LiDAR representations express nearness on a common relative scale. The subsequent comparison therefore uses relative structural consistency rather than direct metric-depth differences.

Unlike optimization based only on a single global score, the proposed method evaluates candidate transformations over multiple spatial regions. The regional residual is expressed as

$$\mathrm{J(T)} = \frac{\sum b \in B_v \, \mathrm{w_b r_b}(T)}{\sum b \in B_v \, \mathrm{w_b}}, \quad \mathrm{B} = \{\mathrm{near, mid, far, focus, global}\} \tag{2}$$

where r_b is the residual of region b, and $\mathrm{w_b}$ is its weight, $T = [R|t] \in SE(3)$ is the LiDAR-to-camera transformation defined in Eq. (1), and $B_v \subseteq B$ is the subset of valid regions with sufficient structural support. For the focus region, the image is divided into eight horizontal bins and the two outermost bins are excluded from automatic selection. An interior bin is eligible when it contains at least 30 supported structural points, including at least 20 near-range points (<15 m) and a near-point fraction of at least 0.45, with candidate support considered up to 22 m. The eligible focus bin is selected using a score based on the robust upper-tail edge residual together with structural support, near-point fraction, depth preference, and the fraction of residuals larger than 1 px. A candidate update is accepted only when sufficient structural support and cross-region consistency are maintained; otherwise, the previously validated transformation is retained.

2.3 Multi-Frame Calibration and Fusion

Since a single frame may contain insufficient structural information, calibration is independently performed on multiple frames. For each frame, the observability score q_i is computed only from non-ground-truth Stage-A indicators: multi-scale score, nearness score, edge score, edge residual shift, translation structural cost, near/mid translation cost, translation median residual, number of active translation bins, translation-refinement acceptance, and edge-refinement acceptance. For each indicator, candidate-frame values are converted to normalized rank scores in [0,1], with the rank direction reversed for residual/cost terms for which lower values are better. The ranks are combined by a weighted mean using weights 1.00, 0.80, 1.40, 1.40, 1.20, 1.00, 0.80, 0.40, 0.40, and 0.30, respectively; if the structural edge score is negative, 0.25 is subtracted from q_i as an anti-alignment penalty. The normalized fusion weight for selected frame i is defined as

$$\mathrm{w_i} = \frac{\max(q_i, 0.05)}{\sum j \in S \max(q_j, 0.05)} \tag{3}$$

where q_i denotes the observability score and S is the selected frame set. The fusion weight uses $max(q_i, 0.05)$ and is normalized by the corresponding sum over $j \in S$, preventing a zero denominator for weak scenes. The final translation is obtained by

$$t_f = \sum_{i \in S} w_i t_i \tag{4}$$

while the rotations are combined by quality-weighted geodesic SO(3) fusion. By default, the two highest-ranked frame candidates are fused, reducing the influence of weak or geometrically ambiguous frames. No ground-truth extrinsic parameters are used for frame ranking or fusion.

3. RESULTS AND DISCUSSION

3.1 Experimental Evaluation

The quantitative evaluation was conducted on KITTI Odometry Sequence 00 [3] using 30 trials under each initialization condition. Each trial used five distinct calibration frames, yielding 150 frame-level calibration runs per condition. The five frames were independently calibrated and ranked using the proposed observability score, and the two highest-ranked transformation estimates were selected for multi-frame fusion. Relative Rotation Error (RRE) and Relative Translation Error (RTE) were calculated between the estimated and reference transformations. The Near and Far initializations used fixed, rather than random, SE(3) perturbations: (+1.0°, -0.6°, +0.8°) and (+0.10, -0.06, +0.08) m for Near, and (+5.0°, -3.0°, +4.0°) and (+0.50, -0.30, +0.40) m for Far, corresponding to initial errors of approximately 1.417° / 14.155 cm and 7.143° / 70.696 cm, respectively. The HAInich forest dataset [1] was additionally used for qualitative evaluation in a complex natural environment.

The HAInich forest dataset [1] was additionally used for qualitative evaluation in a complex natural environment containing vegetation, irregular structures, and nonuniform LiDAR measurements.

3.2 Quantitative Calibration Results

Table 1 summarizes the quantitative calibration results on KITTI [3]. The reported values represent the mean performance over 30 independent trials under each Near and Far initialization condition. Rotation and translation errors were measured only after calibration and were not used for optimization, candidate ranking, or multi-frame fusion.

The comparison includes DVLC (Direct Visual-LiDAR Calibration) [10], MIcalib (Mutual-Information-based LiDAR-Camera Calibration) [5], MFcalib (Single-shot and Automatic Extrinsic Calibration for LiDAR and Camera in Targetless Environments Based on Multi-Feature Edge) [11], and the proposed method. DVLC is based on the targetless direct visual LiDAR-camera calibration

toolbox proposed by Koide et al. [10], while MIcalib denotes the mutual-information-based targetless calibration method of Pandey et al. [5]. MFcalib is a single-shot targetless calibration method based on multi-feature edge information [11].

Table 1. Comparison of extrinsic calibration accuracy under Near and Far initialization conditions corresponding to small and large initial calibration errors

Condition	Method	RRE (°) ↓	RTE (cm) ↓	ΔRRE (%) ↑	ΔRTE (%) ↑
Near	DVLC	0.173313	4.673196	87.771	66.986
	MIcalib	0.365054	14.732779	74.243	-4.082
	MFcalib	1.103777	8.094556	22.121	42.815
	Proposed	0.596871	9.343854	57.887	33.989
Far	DVLC	4.989942	71.724404	30.146	-1.454
	MIcalib	3.448984	64.304865	51.718	9.041
	MFcalib	8.964018	106.909513	-25.487	-51.224
	Proposed	0.211227	13.180926	97.043	81.356

RRE and RTE denote the rotational and translational errors with respect to the reference extrinsic parameters, respectively. ΔRRE and ΔRTE represent the percentage reduction in rotational and translational errors relative to the initial calibration, where higher values indicate greater improvement and negative values indicate error deterioration. Lower RRE and RTE values indicate better calibration accuracy. Near and Far denote calibration conditions with relatively small and large initial extrinsic perturbations, respectively; the initial errors were 1.417° / 14.155 cm for Near and 7.143° / 70.696 cm for Far.

Table 2. Comparison of projection and geometric alignment performance under Near and Far initialization conditions corresponding to small and large initial calibration errors.

Condition	Method	2D Reproj. (px) ↓	Edge Align. (px) ↓	3D Align. (cm) ↓	Inlier Ratio @5 px (%) ↑
Near	DVLC	1.5245	12.6753	3.3190	39.13
	MIcalib	4.7949	12.6249	10.132	39.80
	MFcalib	12.9182	12.3719	30.0940	40.27
	Proposed	3.0171	12.7148	7.3716	39.39
Far	DVLC	45.9394	11.9747	111.8476	42.71
	MIcalib	41.9975	14.8166	74.9869	34.56
	MFcalib	102.1134	12.5830	246.7062	44.10
	Proposed	3.8620	12.8168	11.9571	39.36

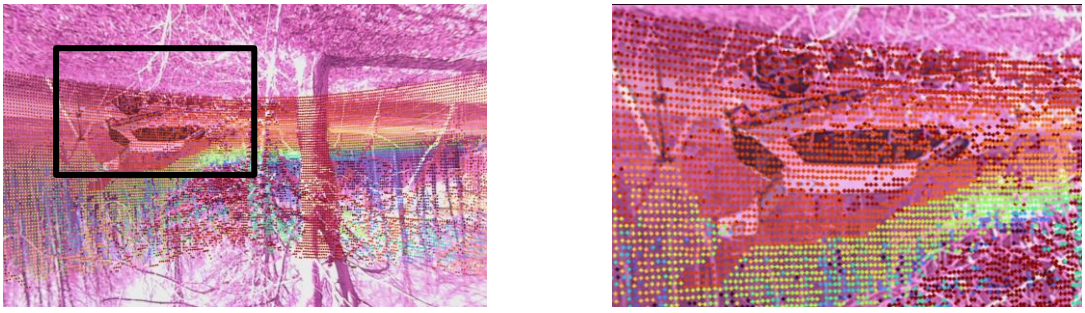
The values in Table 2 are also reported as the mean performance over the same 30 independent trials. The 2D reprojection error is computed on raw LiDAR points that are visible under the KITTI reference extrinsic and remain in front of the estimated camera; it is the Euclidean pixel displacement between their projections obtained with the estimated and reference extrinsics. The 3D alignment error is the Euclidean distance, in the camera coordinate system, between the same GT-visible raw LiDAR points transformed by the estimated and reference SE(3) transformations. Edge Align. is the mean distance between projected LiDAR structural features and image edges, while Inlier Ratio @5 px is the percentage of structural projections lying within 5 pixels of an image edge.

The quantitative results show that the proposed method is particularly effective under the Far initialization condition. Under the Near condition, DVLC and MIcalib achieve lower rotational errors than the proposed method, while DVLC and MFcalib achieve lower translational errors. This reflects a limitation of the conservative validation strategy: it is designed to prevent unstable updates under large initial deviations and may provide smaller gains when the initial estimate is already close to the reference transformation. Under the Far condition, however, the proposed method achieves

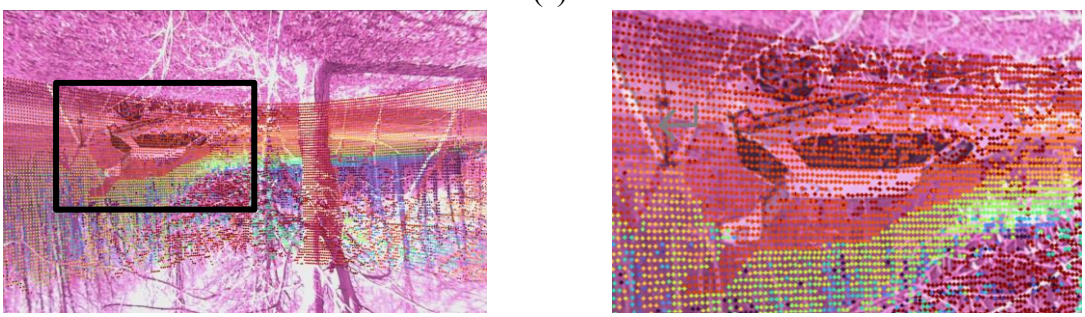
substantially lower RRE, RTE, 2D reprojection error, and 3D alignment error, demonstrating the benefit of the staged optimization and observability-aware validation strategy for large initial errors.

3.3 Qualitative Projection Results

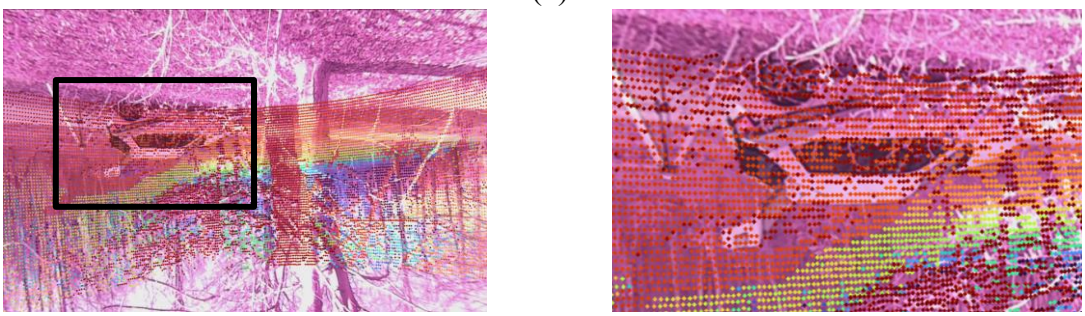
The qualitative performance was further evaluated using the HAInich forest dataset [1]. For all methods, the same CAD-based extrinsic parameters were used as the initial calibration input. Figure 2 compares the initial projection with the results of DVLC, MIcalib, MFcalib, and the proposed method. Since independent ground-truth extrinsic parameters are unavailable, the comparison focuses on visual consistency between projected LiDAR points and scene structures. The boxed regions highlight the front outline of the vehicle and nearby structural boundaries for detailed comparison.



(a)



(b)



(c)

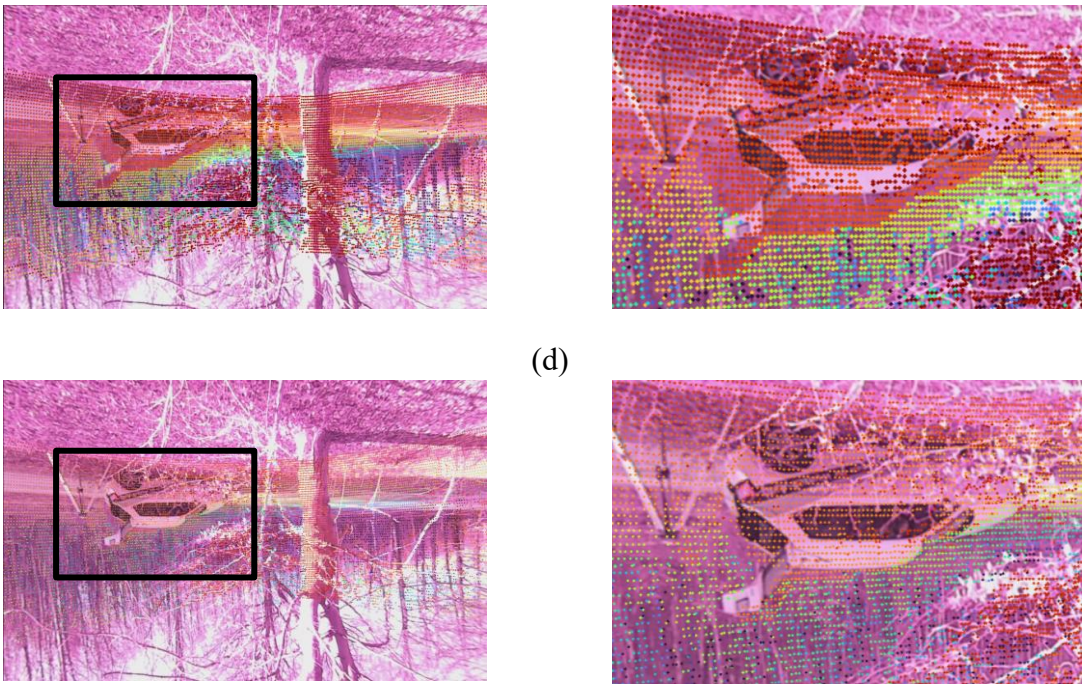


Figure 2. Qualitative comparison of LiDAR-camera calibration results on the HAInich forest dataset using the same CAD-based extrinsic initialization: (a) CAD-based initial projection, (b) DVLC, (c) MIcalib, (d) MFcalib, and (e) Proposed. Black boxes mark the regions of interest in the full views on the left; the corresponding enlarged views are shown on the right.

The qualitative comparison on the HAInich forest dataset further shows clear projection differences among the calibration methods. In the enlarged boxed regions, the alignment of projected LiDAR points with the front outline of the vehicle and nearby structural boundaries provides an intuitive indication of local projection consistency. Because reference extrinsic parameters are unavailable for the forest data, these results indicate practical projection consistency rather than absolute calibration accuracy.

4. CONCLUSIONS

Based on the quantitative and qualitative evaluations conducted in this study, the following conclusions were drawn:

The proposed targetless LiDAR-camera calibration method successfully estimated the 6-DoF extrinsic parameters without using ground-truth extrinsic information during optimization. Under the Far initialization condition, the proposed method achieved an RRE of 0.2112° and an RTE of 13.1809 cm, corresponding to error reductions of 97.043% and 81.356%, respectively.

The proposed method showed particularly strong robustness when the initial extrinsic deviation was large. Under the Far condition, it achieved a 2D reprojection error of 3.8620 px and a 3D alignment error of 11.9571 cm, the lowest values among the compared methods for these two metrics; it was not the best in Edge Alignment or Inlier Ratio. Under the Near condition, DVLC and MIcalib achieved lower rotational errors, while DVLC and MFcalib achieved lower translational errors than the proposed method. These results indicate a limitation in fine calibration accuracy when the initial estimate is already close to the reference transformation, which may reflect the conservative validation strategy.

Qualitative experiments on the HAInich forest dataset using the same CAD-based initialization demonstrated the applicability of the proposed method to complex natural environments containing dense vegetation and irregular geometric structures. Since independent extrinsic ground truth was unavailable for the forest data, these results indicate practical projection consistency rather than absolute

calibration accuracy. Overall, the proposed method provides a robust calibration approach for LiDAR-camera fusion and 3D mapping in complex environmental sensing applications.

Acknowledgement

This research was financially supported by the Korea Ministry of Climate, Energy and Environment as Waste to Energy-Recycling Human Resource Development Project; the ANCHOR program through the ANCHOR Center, Gyeongsangnam-do, funded by the Ministry of Education(MOE) and the Gyeongsangnam-do Provincial Government, Republic of Korea.(2026-ANCHOR-16-002); and by the Technology Innovation Program (RS-2026-25546292, Development of experimental and measurement technologies to estimate ice loads and derive design parameters for vessels navigating Arctic shipping routes) funded by the Ministry of Trade, Industry and Resources(MOTIR, Korea)

References

[1] Milz, S., Wäldchen, J., Abouee, A., Ravichandran, A. A., Schall, P., Hagen, C., Borer, J., Lewandowski, B., Wittich, H.-C., & Mäder, P. (2023). The HAInich: A multidisciplinary vision data-set for a better understanding of the forest ecosystem. Scientific Data, 10, 168. https://doi.org/10.1038/s41597-023-02010-8

[2] Jiang, P., Osteen, P. R., Wigness, M., & Saripalli, S. (2021). RELLIS-3D dataset: Data, benchmarks and analysis. In 2021 IEEE International Conference on Robotics and Automation (ICRA) (pp. 1110-1116). IEEE. https://doi.org/10.1109/ICRA 48506.2021.9561251

[3] Geiger, A., Lenz, P., Stiller, C., & Urtasun, R. (2013). Vision meets robotics: The KITTI dataset. The International Journal of Robotics Research, 32(11), 1231-1237. https://doi.org/10.1177/0278364913491297

[4] Zhang, Q., & Pless, R. (2004). Extrinsic calibration of a camera and laser range finder (improves camera calibration). In 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (Vol. 3, pp. 2301-2306). IEEE. https://doi.org/10.1109/IROS.2004.1389752

[5] Pandey, G., McBride, J. R., Savarese, S., & Eustice, R. M. (2012). Automatic targetless extrinsic calibration of a 3D LiDAR and camera by maximizing mutual information. Proceedings of the AAAI Conference on Artificial Intelligence, 26(1), 2053-2059. https://doi.org/10.1609/AAAI.V26I1.8379

[6] Yuan, C., Liu, X., Hong, X., & Zhang, F. (2021). Pixel-level extrinsic self calibration of high resolution LiDAR and camera in targetless environments. IEEE Robotics and Automation Letters, 6(4), 7517-7524. https://doi.org/10.1109/LRA.2021.30 98923

[7] Li, X., Duan, Y., Wang, B., Ren, H., You, G., Sheng, Y., Ji, J., & Zhang, Y. (2024). EdgeCalib: Multi-frame weighted edge features for automatic targetless LiDAR-camera calibration. IEEE Robotics and Automation Letters, 9(11), 10073-10080. https://doi.org/10.1109/LRA.2024.3455895

[8] Yang, L., Kang, B., Huang, Z., Zhao, Z., Xu, X., Feng, J., & Zhao, H. (2024). Depth Anything V2. Advances in Neural Information Processing Systems, 37, 21875-21911. https://doi.org/10.52202/079017-0688

[9] Wang, Y., & Deng, J. (2026). WAFT-Stereo: Warping-Alone Field Transforms for Stereo Matching. arXiv preprint arXiv:2603.24836

[10] Koide, K., Oishi, S., Yokozuka, M., & Banno, A. (2023). General, single-shot, target-less, and automatic LiDAR-camera extrinsic calibration toolbox. In 2023 IEEE International Conference on Robotics and Automation (ICRA) (pp. 11301-11307). IEEE. https://doi.org/10.1109/ICRA48891.2023.10160691

[11] Ye, T., Xu, W., Zheng, C., & Cui, Y. (2024). MFCalib: Single-shot and automatic extrinsic calibration for LiDAR and camera in targetless environments based on multi-feature edge. In 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 864-871). IEEE. https://doi.org/10.1109/IROS58592.2024.10802395