

The City that Never Settles: Simulation-based LiDAR Dataset for Long-Term Place Recognition Under Extreme Structural Changes

Hyunho Song¹, Dongjae Lee², Seunghun Oh², Minwoo Jung², and Ayoung Kim^{2*}

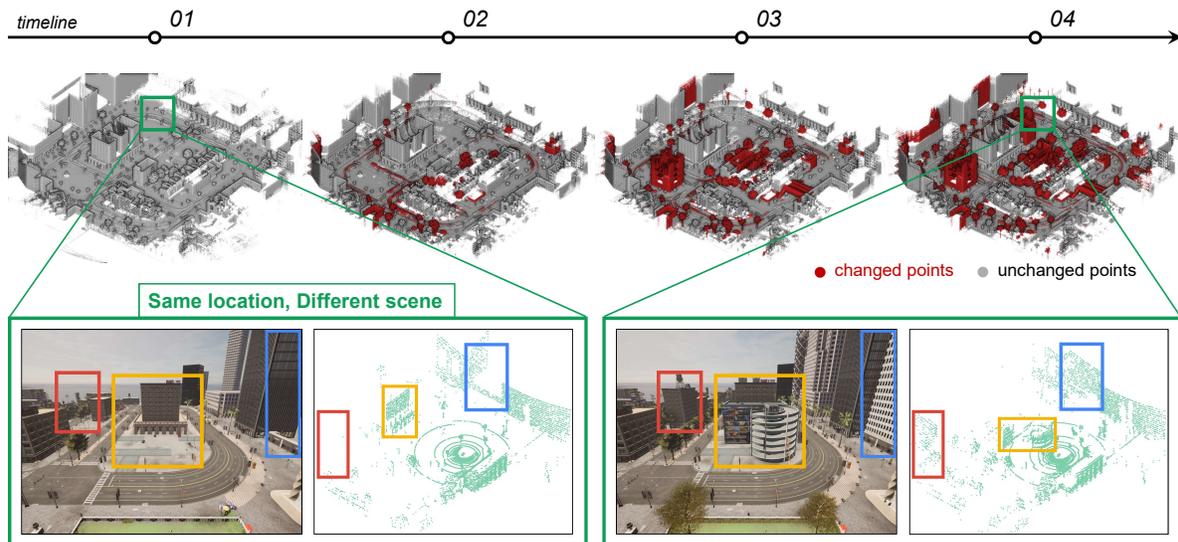


Fig. 1: Structural changes over time in Metropolis. The upper part of the image shows how the city evolves over time through point cloud maps, where gray points represent unchanged areas and red points indicate structural changes. The lower part of the image illustrates two scenes captured at the same location, demonstrating how buildings are constructed over time, resulting in different observations.

Abstract—Large-scale construction and demolition significantly challenge long-term place recognition (PR) by drastically reshaping urban and suburban environments. Existing datasets predominantly reflect limited or indoor-focused changes, failing to adequately represent extensive outdoor transformations. To bridge this gap, we introduce the City that Never Settles (CNS) dataset, a simulation-based dataset created using the CARLA simulator, capturing major structural changes—such as building construction and demolition—across diverse maps and sequences. Additionally, we propose TCR_{sym} , a symmetric version of the original TCR metric, enabling consistent measurement of structural changes irrespective of source-target ordering. Quantitative comparisons demonstrate that CNS encompasses more extensive transformations than current real-world benchmarks. Evaluations of state-of-the-art LiDAR-based PR methods on CNS reveal substantial performance degradation, underscoring the need for robust algorithms capable of handling significant environmental changes. Our dataset is available at https://github.com/Hyunho111/CNS_dataset.

I. INTRODUCTION

Outdoor environments are dynamic and constantly evolving, with construction activities reshaping not only building interiors but also the entire cityscape. In these continuously

changing environments, relying on static maps for localization becomes increasingly difficult. While the Nothing Stands Still (NSS) benchmark [1] addressed extreme structural changes in construction sites, it focused primarily on interior modifications and did not capture large-scale transformations at the urban and suburban scales. Moreover, acquiring real-world data that reflects such massive evolutionary changes over time is impractical due to the long observation periods and logistical constraints.

To overcome these challenges, we propose the CNS dataset, which simulates large-scale structural changes in both urban and suburban environments using the CARLA simulator [2]. As shown in Fig. 1, we simulate large-scale transformations such as construction and demolition across different cityscapes, providing a new benchmark for testing long-term place recognition (PR). Additionally, we introduce the TCR_{sym} metric, a modification of the TCR from the NSS benchmark [1], to quantitatively compare the rate of change between real-world and our dataset. This enables an objective comparison of the extent of structural variations, highlighting the larger-scale transformations captured by our approach.

Our key contributions are as follows:

- 1) We provide a simulation-based dataset which systematically introduces massive construction and demolition across multiple maps and sequences, using CARLA to emulate real-world transformations at scale.
- 2) We propose a new metric, TCR_{sym} , ensuring symmetric measurements of structural changes. Comparing real-

[†]This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. RS-2023-00241758)

¹H. Song is with the Dept. of Future Automotive Mobility, SNU, Seoul, S. Korea hun1021405@snu.ac.kr

²D.Lee, S.Oh, M. Jung and A. Kim are with the Dept. of Mechanical Engineering, SNU, Seoul, S. Korea [pur22, alvin0808, moonshot, ayoungk][@snu.ac.kr](mailto:ayoungk@snu.ac.kr)

world datasets with ours demonstrates far larger transformations, reinforcing the dataset’s value for long-term PR.

- 3) We evaluate state-of-the-art LiDAR-based PR methods on our dataset and reveal substantial performance drops under these extreme changes, highlighting the need for more robust, adaptable algorithms.

II. RELATED WORKS

Numerous datasets containing temporal and structural variations have been developed to support long-term LiDAR-based PR. The NCLT dataset [3], which spanned 27 sessions over a year, captured long-term dynamics including major construction projects. Similarly, the Oxford Robotcar dataset [4] extended over a year and a half, reflecting structural changes caused by city roadworks, while the Boreas dataset [5] primarily tracked seasonal variations but also included some construction-related changes. The Mulran [6] and the HeLiPR dataset [7], collected at the same location with a four-year gap between sessions, revealed notable structural shifts. Although these real-world datasets provide valuable insights, the modifications they depict remain relatively minor, with permanent structures dominating the scene.

Meanwhile, the NSS benchmark [1] does capture drastic spatiotemporal changes in construction sites. However, it was originally focused on point registration and largely confined to indoor spaces, leaving city-scale outdoor transformations unaddressed. As a result, the existing datasets—both real-world and NSS—offer only limited challenges for PR algorithms when faced with massive structural changes.

The CNS dataset is specifically designed to address the challenges posed by extreme structural changes at the urban and suburban scale. By leveraging the CARLA simulator [2], we simulate large-scale transformations, including construction and demolition processes. This provides a more rigorous evaluation setting for LiDAR-based PR methods, enabling the development of robust algorithms capable of handling massive environmental transformations.

III. THE CNS DATASET

In this section, we present the **City that Never Settles (CNS)** dataset, which aims to facilitate the investigation of long-term PR in highly dynamic environments. By leveraging the CARLA simulator, we are able to emulate significant structural changes—such as construction and demolition—across multiple sequences of the same map. The data was acquired using an NVIDIA GeForce RTX 4070 GPU and a 13th Gen Intel Core i9-13900HX x 32 CPU.

A. System Configuration

We configure the following virtual sensors in CARLA: a LiDAR and an IMU. Notably, we do not include a GNSS sensor because CARLA provides a built-in odometry message that already serves as a global ground-truth for vehicle poses. The simulation runs at 20Hz, with all sensor measurements synchronized at each frame.

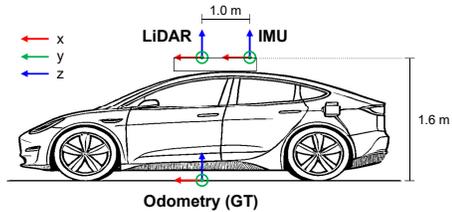


Fig. 2: System Configuration

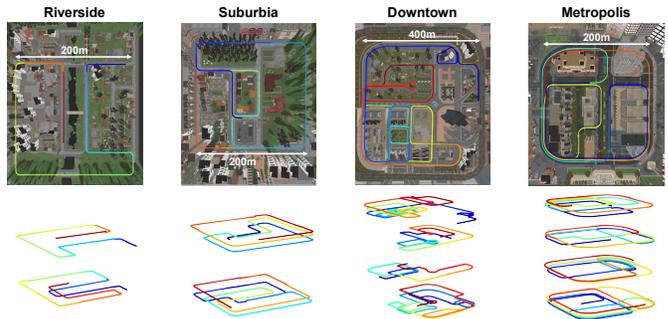


Fig. 3: The upper displays aerial views with trajectories overlaid for Riverside 02, Suburbia 02, Downtown 04, and Metropolis 04. The below shows all the routes on each map, with red being the starting point and blue being the ending point.

LiDAR. We employ a 32-channel LiDAR sensor, modeled after the *Ouster OS1* design, mounted on the roof of a simulated vehicle. The vertical field of view is set to cover $\pm 22.5^\circ$, with a maximum range of 120m. Because the sensor data is generated within a simulation, each full scan is inherently *deskewed* (i.e., free of motion-induced distortions).

IMU. We employ an IMU that provides *ideal* ground-truth inertial data—namely, zero bias and noise.

B. Dataset Generation

To create environments that challenge long-term PR, we introduce substantial structural modifications in CARLA. Starting with base maps Riverside, Suburbia, Downtown, and Metropolis, we manually delete key buildings, trees, and landmarks using the Unreal Engine Editor across multiple sequences. As illustrated in Fig. 3, each modified sequence is then traversed by a scripted autonomous driving agent following partially overlapping, yet distinct routes, capturing structural changes from various viewpoints and ensuring meaningful PR experiments.

C. Sequence Description

Our dataset comprises 4 maps and 12 sequences in total: Downtown and Metropolis each include 4 sequences (01–04), while Riverside and Suburbia each contain 2 sequences (01–02). Each sequence reflects a distinct stage of structural changes, where lower-numbered sequences represent maps with fewer structures, and higher-numbered sequences (e.g., 04 or 02) represent the original, fully built environment. Thus, moving from lower to higher numbers simulates construction, and from higher to lower numbers simulates demolition as shown in Fig. 1.

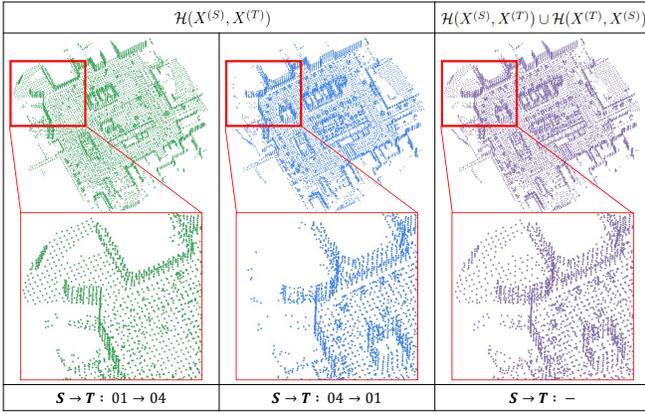


Fig. 4: Visualization of the set \mathcal{H} for Metropolis. The left two columns illustrate how \mathcal{H} changes under the source–target ordering, causing variations in TCR values. To ensure consistency regardless of the ordering, we introduce the union of both sets, as shown in the right column, which remains invariant to the ordering.

TABLE I: Temporal Change Ratio comparison

Map	Riverside	Suburbia	Downtown	Metropolis	KAIST*
Sequences	01 ↔ 02	01 ↔ 02	01 ↔ 04	01 ↔ 04	01 ↔ 06
TCR _{sym}	0.0986	0.1840	0.1120	0.2920	0.1819

IV. TEMPORAL CHANGE RATIO

We propose the TCR_{sym} (symmetric Temporal Change Ratio) metric to quantify large-scale structural changes between sequences. Unlike the original TCR from the NSS benchmark [1], which is sensitive to the source–target ordering, TCR_{sym} ensures consistent measurements of structural changes irrespective of the ordering. As shown in Fig. 4, the original TCR can vary depending on the source–target direction, while the symmetric version eliminates this dependency. The TCR_{sym} is defined as:

$$\text{TCR}_{\text{sym}} = 1 - \frac{|\mathcal{O}(X^{(S)}, X^{(T)}; \tau) \cup \mathcal{O}(X^{(T)}, X^{(S)}; \tau)|}{|\mathcal{H}(X^{(S)}, X^{(T)}) \cup \mathcal{H}(X^{(T)}, X^{(S)})|}, \quad (1)$$

where the unchanged point set \mathcal{O} and the convex hull set \mathcal{H} are defined as follows:

$$\mathcal{O}(X^{(S)}, X^{(T)}; \tau) = \left\{ x \in X^{(S)} \mid \text{NN}(x, X^{(T)}) \leq \tau \right\}, \quad (2)$$

$$\begin{aligned} \mathcal{H}(X^{(S)}, X^{(T)}) &= \left\{ x \in X^{(S)} \mid \text{ConvexHull}(X^{(T)}) \right. \\ &= \left. \text{ConvexHull}(X^{(T)} \cup \{x\}) \right\}. \quad (3) \end{aligned}$$

A higher TCR_{sym} indicates more structural changes, such as significant building construction or demolition, whereas a lower TCR_{sym} reflects smaller-scale modifications. To focus primarily on structural rather than transient changes (e.g., vehicles, pedestrians), we voxelize the point clouds at a resolution of 5m and set the distance threshold $\tau = 4.5\text{m}$. This ensures TCR_{sym} predominantly captures large-scale structural changes rather than transient objects.

Table I summarizes the TCR_{sym} values observed between sequences in our dataset alongside a real-world comparison. The KAIST entry represents the structural changes occurring over more than four years (from MulRan’s KAIST 01 to

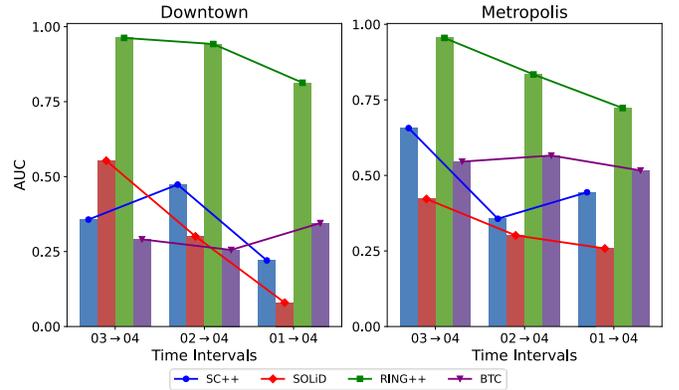


Fig. 5: AUC for each baseline as time intervals increase

HeLiPR’s KAIST 06 sequences), aligned using the LT-SLAM module from LT-mapper [8]. Notably, our dataset exhibits TCR_{sym} values that meet or even surpass the changes observed in the KAIST sequences. This highlights the capability of our dataset to effectively emulate extensive long-term urban transformations, offering a more rigorous benchmark for long-term PR research.

V. BENCHMARK RESULTS WITH OUR DATASET

In this section, we evaluate representative LiDAR-based PR methods—Scan Context [9], SOLID [10], RING++ [11], and BTC [12]—on our CNS dataset. We compare these baselines across varying degrees of structural changes, using standard metrics such as the Precision–Recall curve, Area Under the Curve (AUC), Recall at N ($R@N$) and the maximum F1 score along with TCR_{sym}.

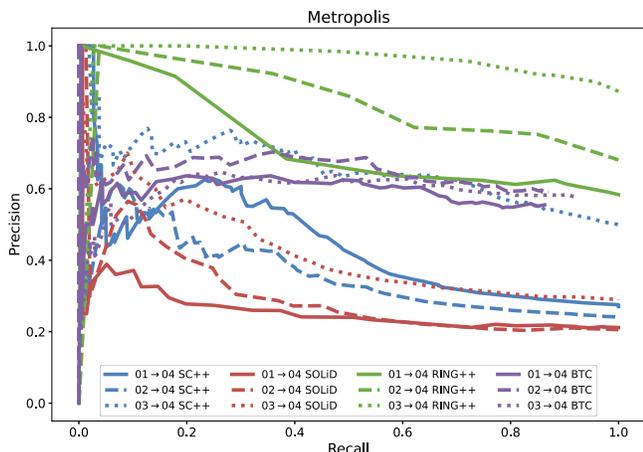
For the PR task, we preprocess our dataset as follows. Query sequences are sampled at intervals of 10 meters along their trajectories, while database sequences are sampled at 5 meters intervals. Given a query scan, candidate matches are retrieved from the database based on their spatial similarity. A candidate is considered a *true positive* if its Euclidean distance to the query location is within 7.5 meters. Additionally, to ensure consistency across methods, each raw LiDAR point cloud is uniformly limited to a range of $[-100\text{m}, 100\text{m}]$. For evaluating BTC specifically, we merge 10 sequential scans into one submap as described in the original paper.

Table II provides the AUC, $R@N$ and the F1 score for each method on different inter-session pairs of our dataset. Bold and underlined values indicate the best and second-best performance, respectively. As observed in the Table II and Fig 5, PR performance decreases as temporal change increases. This indicates that, as the structural changes in the environment become more significant, PR algorithms generally face increasing difficulty in maintaining high accuracy.

However, BTC exhibits relatively consistent performance regardless of the increase in TCR_{sym}. This can be attributed to its effective local descriptor construction from the submaps and the robust verification logic applied during candidate matching. The recall values of BTC increase steadily, while maintaining relatively high precision, as seen in the Fig 6. Despite this, the absolute performance of BTC remains lower compared to other methods, which suggests limitations in its

TABLE II: Quantitative results for long-term PR

Map	DB \rightarrow query	TCR _{sym}	Method											
			Scan Context++ [9]			SOLiD [10]			RING++ [11]			BTC [12]		
			AUC	R@1	F1 score	AUC	R@1	F1 score	AUC	R@1	F1 score	AUC	R@1	F1 score
Riverside	01 \rightarrow 02	0.0986	0.1423	0.1271	0.2222	0.0703	0.0678	0.1119	0.8856	0.7664	0.8678	<u>0.4126</u>	<u>0.4538</u>	<u>0.6178</u>
	02 \rightarrow 01		0.1268	0.1695	0.2032	0.1199	0.1356	0.1827	0.8291	0.9626	0.8653	<u>0.2702</u>	<u>0.4091</u>	<u>0.5243</u>
Suburbia	01 \rightarrow 02	0.1840	0.4480	0.3732	0.5146	0.2066	0.1901	0.3094	0.8974	0.7192	0.8367	<u>0.6359</u>	<u>0.6414</u>	<u>0.7706</u>
	02 \rightarrow 01		0.4699	0.3285	0.4839	0.1442	0.1168	0.2000	0.8872	0.6667	0.8000	<u>0.5479</u>	<u>0.5521</u>	<u>0.6928</u>
Downtown	03 \rightarrow 04	0.0678	0.3567	0.4560	0.5192	<u>0.5530</u>	<u>0.5520</u>	<u>0.5308</u>	0.9626	0.9816	0.9907	0.2907	0.3828	0.4923
	02 \rightarrow 04	0.1120	<u>0.4734</u>	<u>0.5674</u>	<u>0.5613</u>	0.3007	0.3333	0.3800	0.9422	0.8945	0.9443	0.2554	0.3147	0.4562
	01 \rightarrow 04	0.1674	0.2204	0.2303	0.2833	0.0796	0.1292	0.1402	0.8131	0.7038	0.8262	<u>0.3450</u>	<u>0.4532</u>	<u>0.5986</u>
Metropolis	03 \rightarrow 04	0.2528	<u>0.6566</u>	0.5185	0.6693	0.4220	0.2963	0.4495	0.9550	0.8721	0.9317	0.5456	<u>0.5605</u>	<u>0.7092</u>
	02 \rightarrow 04	0.2552	<u>0.3566</u>	0.2680	0.4039	0.3015	0.2288	0.3431	0.8341	0.6807	0.8100	<u>0.5658</u>	<u>0.5877</u>	<u>0.7037</u>
	01 \rightarrow 04	0.2920	0.4442	0.2993	0.4623	0.2582	0.2449	0.3512	0.7228	0.5837	0.7371	<u>0.5164</u>	<u>0.5362</u>	<u>0.6756</u>


Fig. 6: Precision–Recall curves of each baseline for Metropolis

ability to handle significant transformations effectively.

In contrast, RING++ stands out with superior performance compared to the other methods. This can be explained by its use of Bird’s Eye View (BEV) images, which incorporate contour information from roads and surrounding distances. This strategy appears to be highly effective for long-term PR tasks, as it allows the method to capture broader contextual information. However, RING++ shows a clear performance drop as TCR_{sym} increases, revealing a limitation in handling extreme environmental changes.

In summary, while BTC and RING++ show resilience to structural changes, they still exhibit significant performance degradation as TCR_{sym} rises. This emphasizes the need for further improvements in long-term PR algorithms, particularly in adapting to large-scale transformations.

VI. CONCLUSION

The CNS dataset pushes the boundaries of long-term PR by incorporating large-scale structural transformations in both urban and suburban settings, all within a reproducible simulation environment. Our evaluation of multiple LiDAR-based PR methods shows that, although certain algorithms maintain resilience under moderate changes, they experience pronounced performance drops as TCR_{sym} increases. This outcome underscores the need for novel descriptors and strategies that can accommodate massive redevelopment scenarios. We hope that by providing reproducible evaluations

and TCR_{sym}-based analysis, our CNS dataset can catalyze further advances in long-term PR for real-world construction and demolition tasks.

REFERENCES

- [1] T. Sun, Y. Hao, S. Huang, S. Savarese, K. Schindler, M. Pollefeys, and I. Armeni, “Nothing stands still: A spatiotemporal benchmark on 3d point cloud registration under large geometric and temporal change,” 2025. [Online]. Available: <https://arxiv.org/abs/2311.09346>
- [2] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, “Carla: An open urban driving simulator,” in *Conference on robot learning*. PMLR, 2017, pp. 1–16.
- [3] N. Carlevaris-Bianco, A. K. Ushani, and R. M. Eustice, “University of michigan north campus long-term vision and lidar dataset,” *Intl. J. of Robot. Research*, vol. 35, no. 9, pp. 1023–1035, 2016.
- [4] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, “1 year, 1000 km: The oxford robotcar dataset,” *Intl. J. of Robot. Research*, vol. 36, no. 1, pp. 3–15, 2017.
- [5] K. Burnett, D. J. Yoon, Y. Wu, A. Z. Li, H. Zhang, S. Lu, J. Qian, W.-K. Tseng, A. Lambert, K. Y. Leung *et al.*, “Boreas: A multi-season autonomous driving dataset,” *Intl. J. of Robot. Research*, vol. 42, no. 1-2, pp. 33–42, 2023.
- [6] G. Kim, Y. S. Park, Y. Cho, J. Jeong, and A. Kim, “Mulran: Multimodal range dataset for urban place recognition,” in *Proc. IEEE Intl. Conf. on Robot. and Automat.* IEEE, 2020, pp. 6246–6253.
- [7] M. Jung, W. Yang, D. Lee, H. Gil, G. Kim, and A. Kim, “Helipr: Heterogeneous lidar dataset for inter-lidar place recognition under spatiotemporal variations,” *Intl. J. of Robot. Research*, vol. 43, no. 12, pp. 1867–1883, 2024.
- [8] G. Kim and A. Kim, “Lt-mapper: A modular framework for lidar-based lifelong mapping,” in *Proc. IEEE Intl. Conf. on Robot. and Automat.*, 2022, pp. 7995–8002.
- [9] G. Kim, S. Choi, and A. Kim, “Scan context++: Structural place recognition robust to rotation and lateral variations in urban environments,” *IEEE Trans. Robot. and Automat.*, vol. 38, no. 3, pp. 1856–1874, 2021.
- [10] H. Kim, J. Choi, T. Sim, G. Kim, and Y. Cho, “Narrowing your fov with solid: Spatially organized and lightweight global descriptor for fov-constrained lidar place recognition,” *IEEE Robot. and Automat. Lett.*, 2024.
- [11] X. Xu, S. Lu, J. Wu, H. Lu, Q. Zhu, Y. Liao, R. Xiong, and Y. Wang, “Ring++: Roto-translation invariant gram for global localization on a sparse scan map,” *IEEE Trans. Robot. and Automat.*, vol. 39, no. 6, pp. 4616–4635, 2023.
- [12] C. Yuan, J. Lin, Z. Liu, H. Wei, X. Hong, and F. Zhang, “Btc: A binary and triangle combined descriptor for 3-d place recognition,” *IEEE Trans. Robot. and Automat.*, vol. 40, pp. 1580–1599, 2024.