2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

MOE: A Dense LiDAR MOving Event Dataset, Detection Benchmark and LeaderBoard

Zhiming Chen*, Haozhe Fang*, Jiapeng Chen, Michael Yu Wang, Hongyu Yu†

***Abstract*— Detecting moving events produced by moving objects is a crucial task in the realms of autonomous driving and mobile robots. Moving objects have the potential to create ghost artifacts in mapped environments and pose risks to autonomous navigation. LiDAR serves as a vital sensor for autonomous systems due to its ability to provide dense and precise range measurements. However, existing LiDAR datasets often lack sufficient discussion on the motion labeling of moving objects, containing only a limited representation of moving entities within a single scene. Furthermore, the methodologies for Moving Event Detection (MED) on LiDAR sensors have not been comprehensively explored or evaluated. To address these gaps, this study focuses on constructing a diverse LiDAR moving event dataset encompassing multiple scenes with a high density of moving objects. A thorough review of current MED techniques is conducted, followed by the establishment of a performance benchmark based on evaluating these methods using our dataset. Additionally, part sequences of the dataset are utilized to host an online MED competition, aimed at fostering collaboration within the research community and advancing related studies.**

I. INTRODUCTION

During the utilization of autonomous agents such as self-driving cars or mobile robots, one of the key challenges in their deployment is the presence of moving objects like pedestrians crossing suddenly and vehicles moving at high speeds. These moving objects can leave ghost artifacts during the SLAM process [1] [2] [3] [4]. These artifacts can lead to a noisy map, potentially resulting in imprecise localization or impeding the creation of optimal and safe trajectories. Dealing with incidents arising from the sudden appearance of objects is crucial for autonomous navigation. Detecting moving objects or their components is essential for the autonomous system's functionality, a task often referred to as Moving Event Detection (MED) or event detection, typically conducted using Event Cameras. However, LiDAR emerges as a more widely used and crucial sensor for autonomous agents due to its dense and precise depth measurements, offering a more geometric approach to moving event detection [5]. Some experts in the LiDAR perception domain also term this task as Moving Object Removal (MOR) [6] or Moving Object Segmentation (MOS) [7]. While numerous LiDAR datasets are accessible for research purposes, they predominantly concentrate on conventional tasks such as 3D object detection and semantic segmentation in the context of autonomous driving, rather than on the detection of moving events [8] [9] [10]. The existing datasets primarily address semantic aspects or class labels of objects, which do not provide insights into the dynamic status of the detected objects. For instance, prevalent 3D object detection algorithms like [11] are capable of identifying a stationary car but are unable to determine if it is in motion. Only a limited number of datasets incorporate motion labels for moving objects,

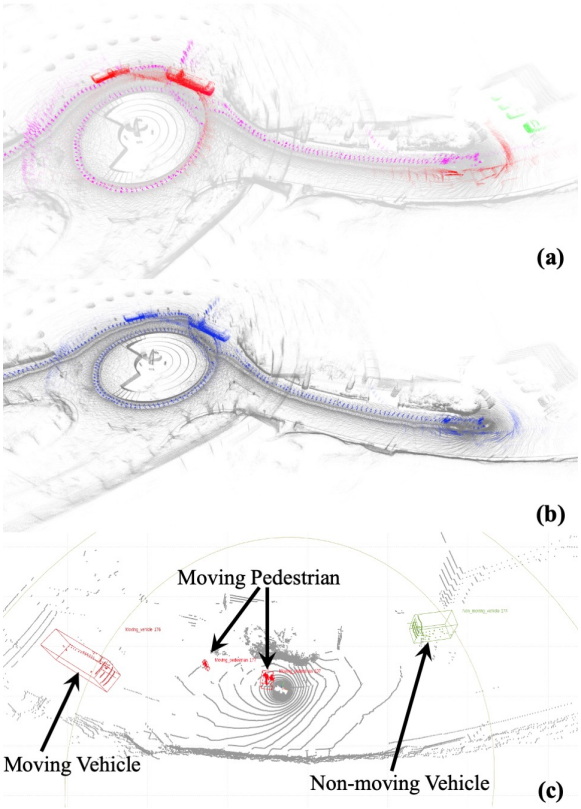

Fig. 1: A typical labeled scene. (a) Different semantic classes are distinguished by various colors. (b) Moving points are indicated by the color blue, while non-moving points are represented in grey. (c) A LiDAR frame in this scene with moving points and non-moving points is labeled with different instances.

* denotes equal contributions, † denotes corresponding author.

This work was supported in part by the Innovation and Technology Commission under Grant ITS/036/21FP of HKSAR and in part by the Project of Hetao ShenzhenHong Kong Science and Technology Innovation Cooperation Zone under Grant HZQB-KCZYB-2020083.

Zhiming Chen, Haozhe Fang, Hongyu Yu are with Robotics Institute, The Hong Kong University of Science and Technology. Hongyu Yu is also with HKUST Shenzhen-Hong Kong Collaborative Innovation Research Institute, Futian, Shenzhen. Emails: zhiming.chen@connect.ust.hk, hfangah@connect.ust.hk, hongyuyu@ust.hk

Michael Yu Wang is with the School of Engineering, Great Bay University, Songshan Lake, Dongguan, Guangdong, China. Email: mywang@gbu.edu.cn. Jiapeng Chen is an individual researcher. Email: jiapeng.chen@kaylordut.com

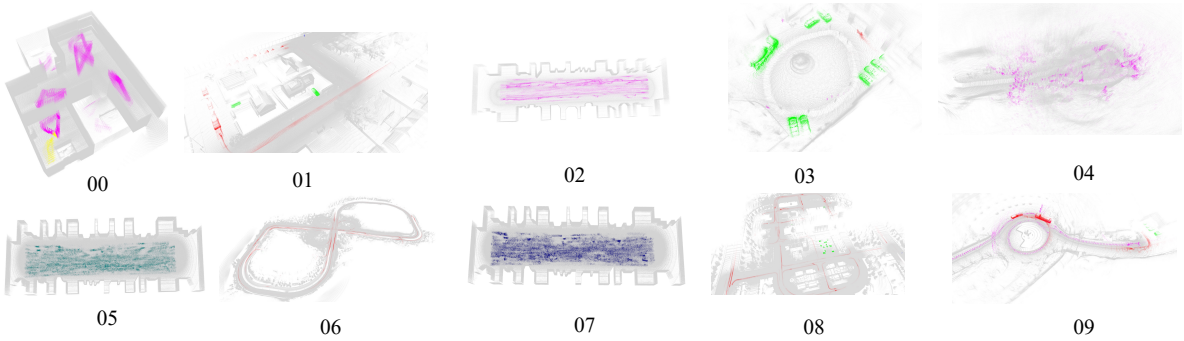

Fig. 2: This overview presents the 10 sequences included in our MOE dataset. Non-moving points are depicted in grey, while points from various movable classes are represented by non-grey colors. Sequence 00 features a simulated indoor apartment environment obtained from a mobile robot in the Webots simulator. Sequences 01, 06, and 08 showcase outdoor city scenes captured by an autonomous driving car in the Carla simulator. Sequences 02, 05, and 07 were recorded in the Gazebo simulator, illustrating a crowded pedestrian setting with varying numbers of walking pedestrians: 50, 100, and 150 respectively. Finally, Sequences 03, 04, and 09 were gathered using a legged robot at different sites on the HKUST campus.

such as [12]. However, these datasets exhibit limited scene diversity and lack high object density, thereby not adequately challenging algorithmic evaluation. While some benchmarks have been introduced for LiDAR moving event detection or moving object segmentation [7] [13], they assess only a fraction of the current methods using relatively simplistic datasets. In this study, we introduce a more rigorous LiDAR moving event dataset and establish a diversified benchmark to assess the latest advancements in Moving Event Detection (MED) approaches.

To summarize, our contributions are outlined as follows:

- We propose a diverse-scene LiDAR dataset with a high density of moving objects focusing on moving event detection.
- We conducted a comprehensive literature review on the latest methodologies, encompassing both non-learning-based and learning-based approaches for detecting events involving motion.
- We assessed existing detection methods for events involving motion and established a benchmark for such methods.
- We are organizing a competition utilizing our dataset and have created a leaderboard for participants aiming to encourage advancements in future research on event detection involving motion.

For more information about the MOE Dataset or the competition and leaderboard, please visit our project page https://sites.google.com/view/moe-dataset.

II. RELATED WORK

A. LiDAR Datasets

LiDAR serves as a fundamental sensor in the realm of autonomous driving and mobile robotics. While certain LiDAR datasets like S3DIS [14], Paris-lille-3D [15], and Semantic3D [16] primarily focus on static environments, prominent LiDAR datasets within the fields of robotics and computer vision, such as KITTI [8], nuScenes [9], and Waymo Open Dataset [10], are tailored towards facilitating 3D object detection or semantic segmentation for autonomous driving applications. These datasets offer bounding box labels for object detection or point-wise labels for semantic information, catering to multi-modal sensors and extensive scenarios. These annotations empower researchers to extract object category information effortlessly. Nevertheless, while object semantics can indicate whether an object is movable, they do not provide insights into the immediate motion characteristics of an object, such as whether it is moving at a particular time or not. Conventional object detection methods [11] [17] [18] and semantic segmentation algorithms [19] [20] [21] also fall short in determining whether a stationary vehicle parked by the roadside is in motion or not, despite being able to identify its category through 3D object detection or semantic segmentation.

Some datasets primarily focus on single-class moving objects such as pedestrians rather than vehicles. For instance, the L-CAS People Dataset [22] offers LiDAR labels for individual persons and groups of people when the robot is in motion or stationary. However, this dataset is limited to a single fixed indoor setting and lacks diverse outdoor scenarios. The DOALS dataset [23] gathers LiDAR data of pedestrians from various indoor environments. Nonetheless, the movement labels are generated through an automated approximation process, which may not be as precise as manual labeling. On the other hand, the UofTPed Dataset [24] ensures accurate positioning data by equipping pedestrians with GPS devices. Nevertheless, the sequences in this dataset exhibit relatively low pedestrian density and only feature one pedestrian per sequence. Furthermore, these pedestrian datasets do not include explicit movement labels, focusing instead on object class labels.

SemanticKITTI [12] is a dataset derived from the well-known KITTI [8], where the authors manually annotate point-wise semantics and motion supervision signals. The dataset distinguishes between parked cars and moving cars. However, most sequences in SemanticKITTI [12] have sparse moving objects, typically one or two vehicles. On the contrary, SemanticPoss [25] captures a sequence in a bustling city block, allowing for more abundant moving objects com-

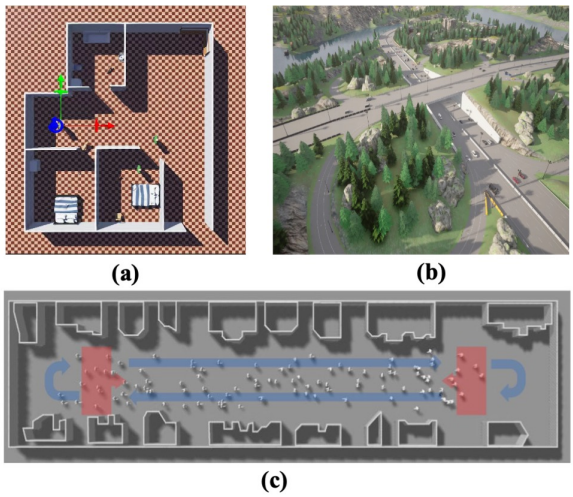

Fig. 3: Simulation environments. (a) A simulated apartment indoor environment in the Webots simulator. (b) A typical city scene in the Carla simulator. (c) A high-density crowd scene in the Gazebo scene.

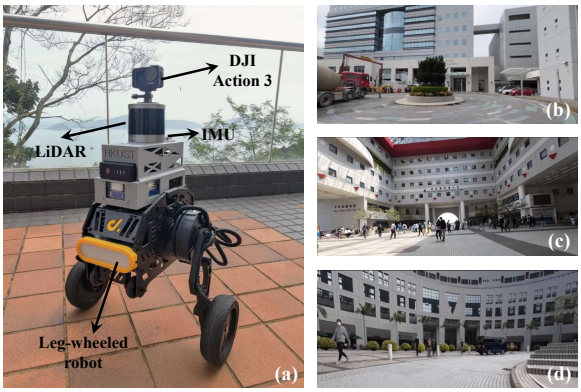

Fig. 4: (a) The legged-wheeled robot used for collecting data in HKUST campus. (b) A small outdoor scene near the CYT building on the HKUST campus. (b) A half-indoor scene in the Jockey Club playground on the HKUST campus. (c) An outdoor scene around the Redbird playground on the HKUST campus.

pared to SemanticKITTI [12]. Nevertheless, SemanticPoss is restricted to a single structured outdoor setting and lacks motion labels for moving objects. Building on the ideas of SemanticKITTI [12] and SemanticPoss [25], we gather various LiDAR sequences encompassing indoor and outdoor environments with a substantial number of moving objects such as pedestrians and cars. We meticulously annotate hierarchical motion labels for each point to create a more varied and densely populated dataset of moving events. A study closely related to ours is presented by Q. Zhang et al. [13]. However, their work heavily depends on existing datasets like [12], with the addition of a modest self-collected dataset. Furthermore, their benchmark evaluation includes only three non-learning-based methods, which may not fully capture the current research landscape.

B. Detection Methods

Moving Event Detection (MED), sometimes called Moving Object Removal (MOR) or Moving Object Segmentation (MOS) depending on different goals after detection, is a research hot spot in the field of LiDAR sensing. The traditional MED methods can be divided into three classes: map-based, visibility-based, and segmentation-based approaches. Besides,learning-based methods are also appealing to researchers' interest recently.

Numerous map-based techniques employ ray-casting, as evidenced in Octomap [26], Dynablox [27], and the work by Pagad et al. [28]. These methodologies analyze the intersections and non-intersections of laser scans within the 3D voxel space to determine the probability of space occupancy or to navigate the voxel occupancy grid, akin to the Peopleremover algorithm [29]. Nevertheless, the computational demands of ray-casting or ray-tracing can be prohibitive in three-dimensional space. Other techniques concerning the partitioning of map space fall into this category as well. In contrast to the division of map space into an occupancy grid or voxel grid based on Cartesian coordinates, certain methods segment the map space into a polar coordinate frame. These methods utilize specific descriptors to distinguish dynamic points on moving objects; however, descriptors like those in ERASOR [30] are superficial and lack robustness in highly dynamic environments. A recent advancement, ERASOR2 [31], has been introduced with enhancements focusing on instance segmentation. Nonetheless, its fundamental concept, derived from ERASOR [30], reliant on pseudo occupancy predicated on a significant variance in free space percentage within the LiDAR sweep space for most collected scans utilized in mapping, may encounter similar limitations as ERASOR [30] in dynamic environments characterized by a high density of moving objects. Certain methodologies, such as the one presented by Falque et al. [32], opt to leverage point cloud normals instead of geometric descriptors to discern moving objects.

Another category of MED techniques is referred to as the visibility-based approach, which entails evaluating the visibility contrast between a query point and a base point within a limited field of view (FOV). The underlying principle is that if the query point is obscured by the base point, then the nearer base point should be considered dynamic. Typically, a preconstructed noisy map is frequently utilized as the primary reference for the base point in visibility assessments. Nevertheless, these methodologies [33] [34] may encounter challenges in scenarios with a substantial incident angle or when obstructed by a significant obstacle, commonly known as *visibility issues*. Removert [6] aims to mitigate these challenges by initially removing dynamic points aggressively, followed by an iterative process to rectify misclassified static points. On the other hand, M-Detector [5] employs a series of tests to validate the motion of each point based on visibility evaluations between consecutive scans. Despite its effectiveness, M-Detector still requires parameter adjustments to accommodate new LiDAR sequences.

Segmentation-based techniques, as discussed in [35] and [36], rely on plane fitting. These methods utilize a ground plane model to distinguish between moving objects and

the static ground plane. However, in scenarios with a high number of dynamic objects present on the ground, these approaches may face challenges. A recent development, MapCleaner [37], introduces a terrain segmentation approach. Nevertheless, the researchers did not address strategies for handling situations where dynamic points are incorrectly classified as ground points. Additionally, works such as DORF [38] attempt to combine multiple approaches to achieve a better balance between detecting moving objects and preserving static objects. But specific tuning for hyperparameters is still needed in DORF.

Moreover, driven by the rapid advancement of deep learning, conventional dynamic entities such as vehicles and pedestrians are now detectable through sophisticated deep neural network architectures in the realm of 3D object detection [11] [17] [18] , semantic segmentation [19] [20] [21] [39], instance segmentation [40] [41], and panoramic segmentation [42]. Despite the provision of semantic labels by deep learning techniques, distinguishing whether an object is in motion remains a challenge. Current learning-based methodologies fall short in adequately addressing the need for detecting moving events. Noteworthy scholars, X. Chen and C. Stachniss, have made significant contributions to the integration of deep learning for object detection based on motion cues. Their research group has introduced a range of learning-based approaches [7] [43] [44] [45] [46]. LIDAR-MOS [7] employs a sequence of residual range images as an intermediary representation, coupled with a Convolutional Neural Network (CNN), to enhance detection speed compared to the LiDAR frame rate. In contrast, 4DMOS [43] utilizes a set of downsampled LiDAR scans within a sparse 4D CNN to jointly extract spatial and temporal features for predicting moving objects. Similarly, akin to LIDAR-MOS [7], RVMOS [47] proposed by J. Kim et al. integrates sequential range images with customized feature extraction networks. While prior studies have combined spatial-temporal information, MotionSeg3D [44] separates spatial and temporal features into distinct branches within the network, integrated using a sparse 3D CNN. AutoMOS [45] amalgamates traditional and learning-based methods in a two-stage process. Initially, it employs ERASOR [30] for automatic detection and tracking of moving objects, followed by training a neural network in the second stage using the detected moving objects as supervisory signals. InsMOS [48] extracts motion attributes, spatio-temporal features, and instance particulars from varied modules, merging them to achieve moving object segmentation. Lastly, MapMOS [46] upholds a volumetric belief map, integrating new predictions through a voxel-wise binary Bayes filter to rectify earlier erroneous estimations, thereby boosting robustness.

Bird's Eye View (BEV) is another research trend in LiDAR perception. MotionBEV [49] converts 3D LiDAR scans into a 2D polar BEV representation to improve computational efficiency, and segments moving objects with appearance and motion features in the BEV domain.

III. Dataset Construction

A. Data Collection

We gathered multiple sequences from diverse indoor and outdoor environments. Since LiDAR data exclusively provides geometric information that can be readily simulated through ray-tracing, and obtaining labels is more straightforward compared to manual labeling, we utilized renowned robotics simulators such as Webots [50], Gazebo [51], and Carla [52] to augment our dataset. Illustrated in Fig. 3, we replicated an apartment scenario featuring a robot navigating amidst randomly moving individuals within the residence. To emulate a scenario with high pedestrian density, we captured data from a Gazebo setting with 50-150 pedestrians occupying a 70m$\times$10m rectangular space, where each frame captured by the LiDAR sensor mounted on the robot detected numerous pedestrians in motion. Moreover, we simulate a large amounts of moving vehicles and pedestrians in Carla to get several urban traffic sequences. Furthermore, as depicted in Fig. 4, we also acquired data from various outdoor urban settings in addition to the aforementioned indoor scenarios. A rough overview of our dataset sequences can be seen in Fig. 2. In these simulations, the accurate ground truth poses and motion labels were easily obtained through the simulator's API. Whereas, for the data collected in real-world settings, the poses were estimated using a reliable SLAM method [3] and the motion labels were assigned through manual annotation, as shown in Fig. 1.

B. Data Annotation

To efficiently represent the useful information about moving objects, we propose a new hierarchical motion label system for our dataset. As shown in Fig. 5, the hierarchical motion label system consisted of 3 layers. The first layer indicates if the object is moveable or not (1 for movable, 0 for non-movable). Then, the second layer shows whether the object's status is moving or not (1 for moving, 0 for non-moving). Last, we will add the semantic class labels for movable objects (-1 for non-movable objects).

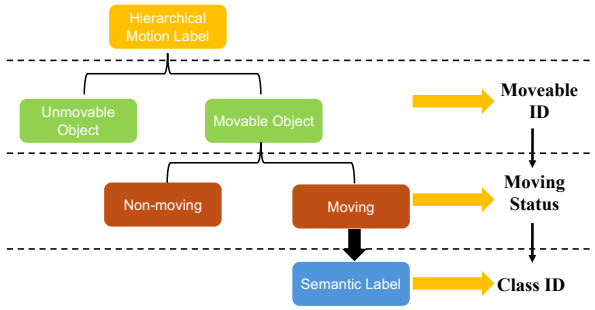

Fig. 5: Hierarchical structure of our proposed motion label, which consists of three layers. The first layer indicates whether the object is movable like a car or not. The second layer represents the moving status of the object - moving or non-moving. The third layer illustrates the semantic class id for the moving object.

C. Dataset Analysis

This dataset primarily focuses on densely populated environments with dynamic movements of pedestrians and vehicles. A comparative analysis is conducted with well-known datasets such as S3DIS [14], Paris-lille-3D [15], Semantic3D [16], KITTI [8], SemanticPOSS [25], and SemanticKITTI [12], as delineated in Table I. Notably, the majority of these foundational datasets do not explicitly include motion labels, making it challenging to discern whether a vehicle is in motion or non-moving. The dataset most akin to our proposed MOE Dataset is SemanticKITTI [12], which pioneers the use of explicit motion labels to denote the movement status of objects. Our MOE Dataset, featuring a hierarchical motion labeling system, exhibits a significantly higher average of moving points per second compared to its predecessor [12].

IV. Experiments and Benchmark

A. Evaluation Metrics

For quantifying the MED performance, we use the commonly applied Jaccard Index or intersection-over-union (IoU) metric over detected moving points, which is given by:

$$IoU = \frac{TP}{TP + FP + FN}$$

where TP, FP, and FN represent the counts of true positive, false positive, and false negative predictions related to the moving class. In this context, positive points denote moving points, whereas negative points refer to non-moving points.

B. Results and Discussion

We assess eight state-of-the-art (SOTA) techniques using the 00, 01, and 02 sequences from our newly introduced MOE Dataset. Within these cutting-edge approaches, three belong to the offline category, namely Removert [6], ERASOR [30], and Octomap [26]; three are non-learning online methods, including Dynablox [27], Dynamic Object Detection (DOD) [32], and M-detector [5]; and two represent distinct learning-based methodologies, namely InsMOS [48] and MotionBEV [35].

Since certain methodologies such as ERASOR [30] generate a voxelized result map, we uniformly reduce the resolution of all result maps and the ground truth point cloud map to facilitate rapid and fair evaluation. Specifically, we downsample the point cloud maps for sequences 00 and 02 to a resolution of 0.2 meters, while for the larger scene in sequence 01, we use a resolution of 0.5 meters. In the case of non-learning methods, we adjusted their parameters for each sequence. Conversely, for learning-based approaches like InsMOS [48] and MotionBEV [35], we validate their generalization capacity by employing models trained on the extensive SemanticKITTI datasets [12] to mitigate overfitting issues that may arise from training and testing on the same dataset. The evaluation in Table II reveals that DOD achieves the highest average IoU compared to other advanced methods. Nonetheless, the obtained average IoU of 0.508 is considered relatively low for moving event detection. Despite learning-based methods like InsMOS [48] exhibiting slightly inferior performance to the traditional non-learning method - DOD [32], we anticipate their potential to address the moving event detection challenge effectively in the future, attributing this to the robust fitting capability of deep neural networks and the rapid advancements within the associated research community.

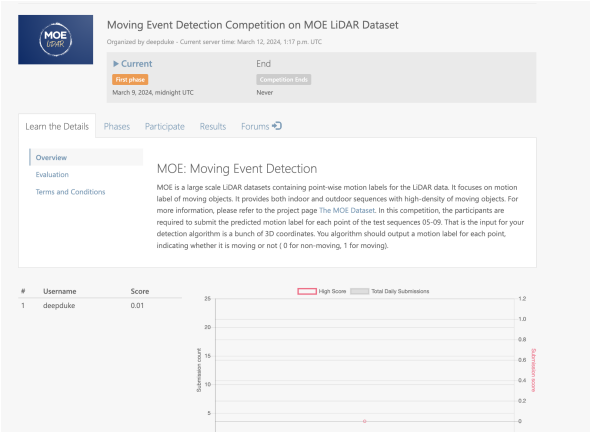

Fig. 6: A preview for the interface of our competition and leaderboard.

C. Leaderboard and Competition

We believe our proposed MOE Dataset is more challenging than previous datasets for LiDAR moving event detection. In order to facilitate the development of related research fields, we use part sequences of the MOE Dataset to host a competition for the moving event detection task with LiDAR, as shown in Fig. 6. For more details, please visit our project page.

V. Conclusions

In this work, we introduce a comprehensive LiDAR dataset called MOE, specifically designed for multi-scene, and dense moving event detection. Additionally, we conduct an extensive review of the latest state-of-the-art (SOTA) techniques used in detecting moving events using LiDAR sensors. Furthermore, we assess the performance of these advanced algorithms on our MOE Dataset to establish a performance benchmark. Moreover, we leverage segments of the MOE Dataset to organize an online competition aimed at advancing research in the field of moving event detection. Moving forward, our focus will be on developing a learning-based approach to enhance the generalized detection of moving events with LiDAR technology.

Acknowledgment

We would like to thank Mr. Chuqiao Zhao for his kind help in Webots simulation. We also want to thank Mr. Hongbo Zhu for his kind help in testing some algorithms. We also want thank Dr. Tingxiang Fan and Mr. Bowen Shen for their kind help in collecting data in Gazbeo simulation.

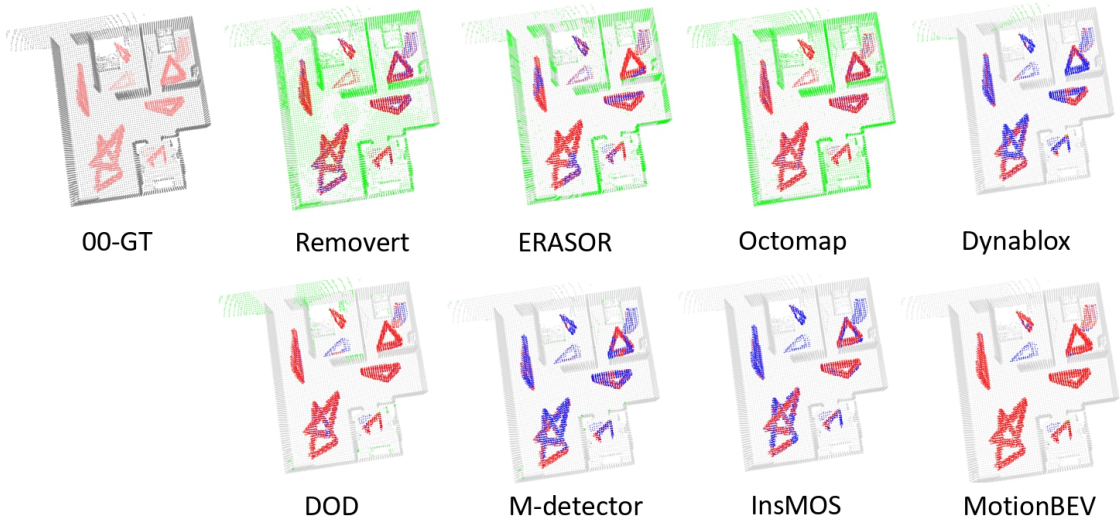

Fig. 7: Quality evaluation on the sequence 00 of our proposed MOE Dataset. In the first subfigure (00-GT), the coral color is the moving point while the grey is the non-moving point. In the remaining subfigures, the red is the True Positive (TP) point, the green is the False Positive (FP) point, the blue is the False Negative (FN) point, and the grey is the True Negative (TN) point.

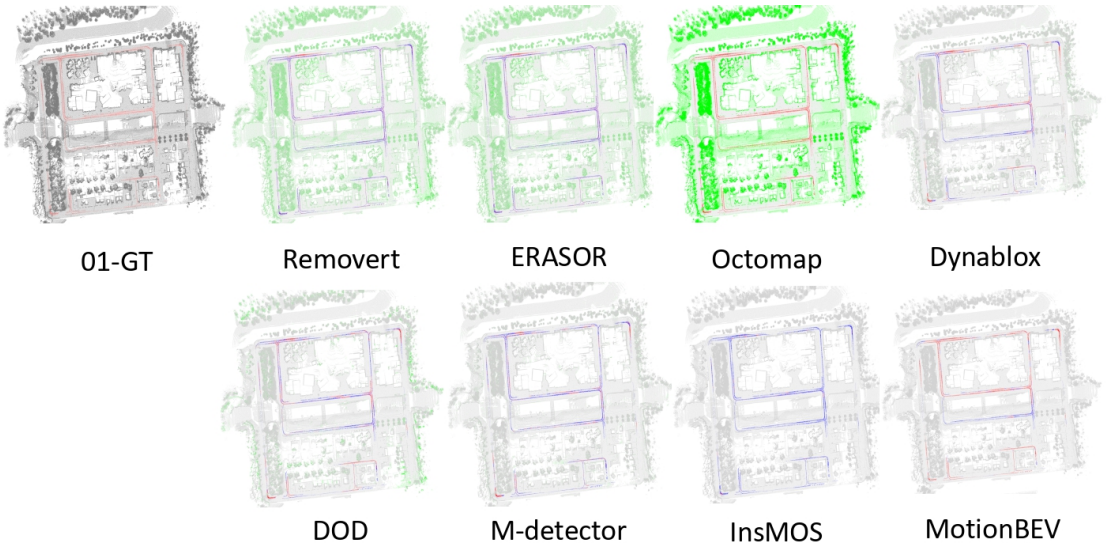

Fig. 8: Quality evaluation on the sequence 01 of our proposed MOE Dataset. In the first subfigure (01-GT), the coral color is the moving point while the grey is the non-moving point. In the remaining subfigures, the red is the True Positive (TP) point, the green is the False Positive (FP) point, the blue is the False Negative (FN) point, and the grey is the True Negative (TN) point.

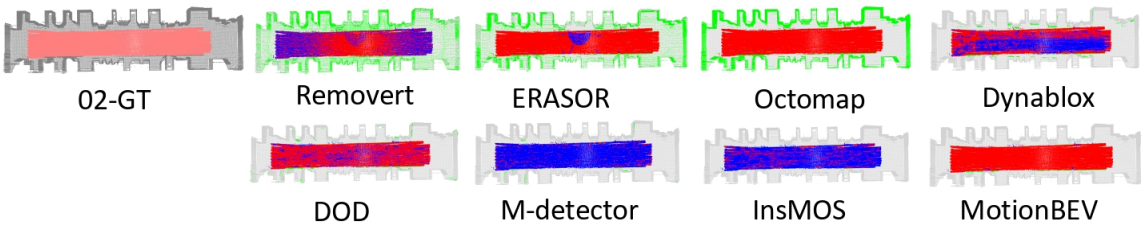

Fig. 9: Quality evaluation on the sequence 02 of our proposed MOE Dataset. In the first subfigure (02-GT), the coral color is the moving point while the grey is the non-moving point. In the remaining subfigures, the red is the True Positive (TP) point, the green is the False Positive (FP) point, the blue is the False Negative (FN) point, and the grey is the True Negative (TN) point.

Dataset \ Metric	Frames	Points	Type	Annotation	Scene	Average Moving Points Per Frame
S3DIS [14]	5	215M	static point clouds	point-wise	Indoor	No motion label
Paris-lille-3D [15]	3	143M	static point clouds	point-wise	Outdoor	No motion label
Semantic3D [16]	30	4009M	static point clouds	point-wise	Outdoor	No motion label
KITTI [8]	14999	1799M	sequential point clouds	bounding box	Outdoor	No motion label
SemanticPOSS [25]	2988	216M	sequential point clouds	point-wise	Outdoor	No motion label
SemanticKITTI [12]	23201	2848M	sequential point clouds	point-wise	Outdoor	308
MOE	23406	1177M	synthetic/sequential point clouds	point-wise	Diverse	2068

TABLE I: Comparision with current LiDAR datasets

Seq# \ Method	Removert[6]	ERASOR[30]	Octomap[13]	Dynablox[27]	DOD[32]	M-Detector[5]	MotionBEV[49]	InsMOS[48]
00	0.297	0.378	0.328	0.320	**0.786**	0.305	0.002	0.495
01	0.028	0.028	0.031	0.195	0.142	0.174	0.055	**0.282**
02	0.421	0.627	**0.652**	0.492	0.595	0.044	0.069	0.379
Average IoU[%]	0.249	0.344	0.337	0.336	**0.508**	0.174	0.042	0.385

TABLE II: Evaluation on some SOTA methods with IoU metric

REFERENCES

[1] J. Zhang and S. Singh, "Loam: Lidar odometry and mapping in real-time," in *Robotics: Science and Systems*, 2014.

[2] T. Shan and B. Englot, "Lego-loam: Lightweight and ground-optimized lidar odometry and mapping on variable terrain," *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4758–4765, 2018.

[3] T. Shan, B. Englot, D. Meyers, W. Wang, C. Ratti, and D. Rus, "Lio-sam: Tightly-coupled lidar inertial odometry via smoothing and mapping," *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5135–5142, 2020.

[4] C. Bai, T. Xiao, Y. Chen, H. Wang, F. Zhang, and X. Gao, "Faster-lio: Lightweight tightly coupled lidar-inertial odometry using parallel sparse incremental voxels," *IEEE Robotics and Automation Letters*, vol. 7, pp. 4861–4868, 2022.

[5] H. Wu, Y. Li, W. Xu, F. Kong, and F. Zhang, "Moving event detection from lidar point streams," *Nature Communications*, vol. 15, 2024.

[6] G. Kim and A. Kim, "Remove, then revert: Static point cloud map construction using multiresolution range images," *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 10 758–10 765, 2020.

[7] X. Chen, S. Li, B. Mersch, L. Wiesmann, J. Gall, J. Behley, and C. Stachniss, "Moving object segmentation in 3d lidar data: A learning-based approach exploiting sequential data," *IEEE Robotics and Automation Letters*, vol. 6, pp. 6529–6536, 2021.

[8] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *The International Journal of Robotics Research*, vol. 32, pp. 1231 – 1237, 2013.

[9] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuscenes: A multimodal dataset for autonomous driving," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11 618–11 628, 2019.

[10] P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine, V. Vasudevan, W. Han, J. Ngiam, H. Zhao, A. Timofeev, S. M. Ettinger, M. Krivokon, A. Gao, A. Joshi, Y. Zhang, J. Shlens, Z. Chen, and D. Anguelov, "Scalability in perception for autonomous driving: Waymo open dataset," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2443–2451, 2019.

[11] W. Zheng, W. Tang, L. Jiang, and C.-W. Fu, "Se-ssd: Self-ensembling single-stage object detector from point cloud," *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14 489–14 498, 2021.

[12] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, and J. Gall, "Semantickitti: A dataset for semantic scene understanding of lidar sequences," *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9296–9306, 2019.

[13] Q. Zhang, D. Duberg, R. Geng, M. Jia, L. Wang, and P. Jensfelt, "A dynamic points removal benchmark in point cloud maps," *ArXiv*, vol. abs/2307.07260, 2023.

[14] I. Armeni, S. Sax, A. Zamir, and S. Savarese, "Joint 2d-3d-semantic data for indoor scene understanding," *ArXiv*, vol. abs/1702.01105, 2017.

[15] X. Roynard, J.-E. Deschaud, and F. Goulette, "Paris-lille-3d: A large and high-quality ground-truth urban point cloud dataset for automatic segmentation and classification," *The International Journal of Robotics Research*, vol. 37, pp. 545 – 557, 2017.

[16] T. Hackel, N. Savinov, L. Ladicky, J. D. Wegner, K. Schindler, and M. Pollefeys, "Semantic3d.net: A new large-scale point cloud classification benchmark," *ArXiv*, vol. abs/1704.03847, 2017.

[17] Y. Zhang, Q. Zhang, Z. Zhu, J. Hou, and Y. Yuan, "Glenet: Boosting 3d object detectors with generative label uncertainty estimation," *International Journal of Computer Vision*, vol. 131, pp. 3332–3352, 2022.

[18] Q. Xu, Y. Zhou, W. Wang, C. Qi, and D. Anguelov, "Spg: Unsupervised domain adaptation for 3d object detection via semantic point generation," *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 15 426–15 436, 2021.

[19] X. Wu, L. Jiang, P.-S. Wang, Z. Liu, X. Liu, Y. Qiao, W. Ouyang, T. He, and H. Zhao, "Point transformer v3: Simpler, faster, stronger," *ArXiv*, vol. abs/2312.10035, 2023.

[20] X. Lai, Y. Chen, F. Lu, J. Liu, and J. Jia, "Spherical transformer for lidar-based 3d recognition," *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17 545–17 555, 2023.

[21] H. Tang, Z. Liu, S. Zhao, Y. Lin, J. Lin, H. Wang, and S. Han, "Searching efficient 3d architectures with sparse point-voxel convolution," in *European Conference on Computer Vision*, 2020.

[22] Z. Yan, T. Duckett, and N. Bellotto, "Online learning for human classification in 3d lidar-based tracking," in *In Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Vancouver, Canada, September 2017.

[23] P. Pfreundschuh, H. F. C. Hendrikx, V. Reijgwart, R. Dubé, R. Y. Siegwart, and A. Cramariuc, "Dynamic object aware lidar slam based on automatic generation of training data," *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 11 641–11 647, 2021.

[24] K. Burnett, S. Samavi, S. L. Waslander, T. D. Barfoot, and A. P. Schoellig, "autotrack: A lightweight object detection and tracking system for the sae autodrive challenge," *2019 16th Conference on Computer and Robot Vision (CRV)*, pp. 209–216, 2019.

[25] Y. Pan, B. Gao, J. Mei, S. Geng, C. Li, and H. Zhao, "Semanticposs: A point cloud dataset with large quantity of dynamic instances," *2020 IEEE Intelligent Vehicles Symposium (IV)*, pp. 687–693, 2020.

[26] K. M. Wurm, A. Hornung, M. Bennewitz, C. Stachniss, and W. Bur-

gard, "Octomap : A probabilistic , flexible , and compact 3 d map representation for robotic systems," 2010.
[27] L. M. Schmid, O. Andersson, A. Sulser, P. Pfreundschuh, and R. Y. Siegwart, "Dynablox: Real-time detection of diverse dynamic objects in complex environments," *IEEE Robotics and Automation Letters*, vol. 8, pp. 6259–6266, 2023.
[28] S. Pagad, D. Agarwal, S. Narayanan, K. Rangan, H. Kim, and V. Yalla, "Robust method for removing dynamic objects from point clouds," *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 10 765–10 771, 2020.
[29] J. Schauer and A. Nüchter, "The peopleremover—removing dynamic objects from 3-d point cloud data by traversing a voxel occupancy grid," *IEEE Robotics and Automation Letters*, vol. 3, pp. 1679–1686, 2018.
[30] H. Lim, S. Hwang, and H. Myung, "Erasor: Egocentric ratio of pseudo occupancy-based dynamic object removal for static 3d point cloud map building," *IEEE Robotics and Automation Letters*, vol. 6, pp. 2272–2279, 2021.
[31] H. Lim, L. Nunes, B. Mersch, X. Chen, J. Behley, H. Myung, and C. Stachniss, "Erasor2: Instance-aware robust 3d mapping of the static world in dynamic scenes," *Robotics: Science and Systems XIX*, 2023.
[32] R. Falque, C. L. Gentil, and F. Sukkar, "Dynamic object detection in range data using spatiotemporal normals," *ArXiv*, vol. abs/2310.13273, 2023.
[33] D. J. Yoon, T. Y. Tang, and T. D. Barfoot, "Mapless online detection of dynamic objects in 3d lidar," *2019 16th Conference on Computer and Robot Vision (CRV)*, pp. 113–120, 2018.
[34] F. Pomerleau, P. Krüsi, F. Colas, P. T. Furgale, and R. Y. Siegwart, "Long-term 3d map maintenance in dynamic environments," *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3712–3719, 2014.
[35] A. Dewan, T. Caselitz, G. D. Tipaldi, and W. Burgard, "Motion-based detection and tracking in 3d lidar scans," *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4508–4513, 2016.
[36] M. Arora, L. Wiesmann, X. Chen, and C. Stachniss, "Static map generation from 3d lidar point clouds exploiting ground segmentation," *Robotics Auton. Syst.*, vol. 159, p. 104287, 2022.
[37] H. Fu, H. Xue, and G. Xie, "Mapcleaner: Efficiently removing moving objects from point cloud maps in autonomous driving scenarios," *Remote. Sens.*, vol. 14, p. 4496, 2022.
[38] Z. Chen, K. Zhang, H. Chen, M. Y. Wang, W. Zhang, and H. Yu, "Dorf: A dynamic object removal framework for robust static lidar mapping in urban environments," *IEEE Robotics and Automation Letters*, vol. 8, pp. 7922–7929, 2023.
[39] C. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," in *Neural Information Processing Systems*, 2017.
[40] H. Zhang, Y. Wang, J. Cai, H.-M. Hsu, H. Ji, and J.-N. Hwang, "Lifts: Lidar and monocular image fusion for multi-object tracking and segmentation," 2020.
[41] A. Milioto, I. Vizzo, J. Behley, and C. Stachniss, "Rangenet ++: Fast and accurate lidar semantic segmentation," *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4213–4220, 2019.
[42] Z. Zhou, Y. Zhang, and H. Foroosh, "Panoptic-polarnet: Proposal-free lidar point cloud panoptic segmentation," *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13 189–13 198, 2021.
[43] B. Mersch, X. Chen, I. Vizzo, L. Nunes, J. Behley, and C. Stachniss, "Receding moving object segmentation in 3d lidar data using sparse 4d convolutions," *IEEE Robotics and Automation Letters*, vol. 7, pp. 7503–7510, 2022.
[44] J. Sun, Y. Dai, X. Zhang, J. Xu, R. Ai, W. Gu, and X. Chen, "Efficient spatial-temporal information fusion for lidar-based 3d moving object segmentation," *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 11 456–11 463, 2022.
[45] X. Chen, B. Mersch, L. Nunes, R. Marcuzzi, I. Vizzo, J. Behley, and C. Stachniss, "Automatic labeling to generate training data for online lidar-based moving object segmentation," *IEEE Robotics and Automation Letters*, vol. 7, pp. 6107–6114, 2022.
[46] B. Mersch, T. Guadagnino, X. Chen, I. Vizzo, J. Behley, and C. Stachniss, "Building volumetric beliefs for dynamic environments exploiting map-based moving object segmentation," *IEEE Robotics and Automation Letters*, vol. 8, pp. 5180–5187, 2023.
[47] J. Kim, J. Woo, and S. Im, "Rvmos: Range-view moving object segmentation leveraged by semantic and motion features," *IEEE Robotics and Automation Letters*, vol. PP, pp. 1–8, 2022.
[48] N. Wang, C. Shi, R. Guo, H. Lu, Z. Zheng, and X. Chen, "Insmos: Instance-aware moving object segmentation in lidar data," *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 7598–7605, 2023.
[49] B. Zhou, J. Xie, Y. Pan, J. Wu, and C. Lu, "Motionbev: Attention-aware online lidar moving object segmentation with bird's eye view based appearance and motion features," *IEEE Robotics and Automation Letters*, vol. 8, pp. 8074–8081, 2023.
[50] O. Michel, "Cyberbotics ltd. webots™: Professional mobile robot simulation," *International Journal of Advanced Robotic Systems*, vol. 1, 2004.
[51] N. P. Koenig and A. Howard, "Design and use paradigms for gazebo, an open-source multi-robot simulator," *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No.04CH37566)*, vol. 3, pp. 2149–2154 vol.3, 2004.
[52] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," in *Proceedings of the 1st Annual Conference on Robot Learning*, 2017, pp. 1–16.