



# Exploitation of Dense MLS City Maps for 3D Object Detection

Örkény Zováthi<sup>1,2</sup><sup>(✉)</sup>, Balázs Nagy<sup>1,2</sup>, and Csaba Benedek<sup>1,2</sup>

<sup>1</sup> Institute for Computer Science and Control (SZTAKI),  
Kende u. 13-17, 1111 Budapest, Hungary  
{zovathi.orkeny,nagy.balazs,benedek.csaba}@sztaki.hu

<sup>2</sup> Faculty of Information Technology and Bionics,  
3in Research Group, Pázmány Péter Catholic University,  
2500 Esztergom, Hungary

**Abstract.** In this paper we propose a novel method for the exploitation of High Density Localization (HDL) maps obtained by Mobile Laser Scanning in order to increase the performance of *state-of-the-art* real time dynamic object detection (RTDOD) methods utilizing Rotating Multi-Beam (RMB) Lidar measurements. First, we align the onboard measurements to the 3D HDL map with a multimodal point cloud registration algorithm operating in the Hough space. Next we apply a grid based probabilistic step to filter out the object regions on the RMB Lidar data which were *falsely* predicted as dynamic objects by RTDOD, although they are part of the static background scene. On the other hand, to find objects erroneously *missed* by the RTDOD predictions, we implement a Markov Random Field based point level change detection approach between the map and the current onboard measurement frame. Finally, to analyse the changed but previously unclassified segments of the RMB Lidar clouds, we apply a geometric blob separation and a Support Vector Machine based classification to distinguish the different object types. Comparative tests are provided in high traffic road sections of Budapest, Hungary, and we show an improvement of 5,96% in precision, 9,21% in recall and 7,93% in *F*-score metrics against the *state-of-the-art* RTDOD algorithm.

**Keywords:** Lidar · City map · Registration · Change detection · Object detection

## 1 Introduction

Real time dynamic object detection (RTDOD) in 3D sparse point clouds is a key challenge in autonomous driving. During the past few years, several geometric [1] and deep learning [1, 5, 10–12, 14, 15] based approaches appeared in the literature, which operate on raw Rotating Multi-Beam (RMB) Lidar frames and provide output sets of oriented bounding boxes for various object categories such as vehicles, pedestrians or bicycles. As main advantage, these approaches

can simultaneously consider local shape and point density features together with global contextual information for the classification of the different point cloud segments. However, due to the low vertical density of the RMB Lidar sensor data, which quickly decreases as a function of the objects' distance from the sensor, the typical ring pattern of the point clouds, and various occlusion effects in dense urban environment, there are a number of limitations of these approaches. On one hand, *false positive* hits may be detected in point cloud regions containing static scene objects with similar appearance and context parameters to the focused dynamic scene objects. On the other hand, the point cloud blobs of several dynamic objects can be heavily merged or occluded by static street furniture elements, yielding many *false negative* detections.

Mobile Laser Scanning (MLS) technologies may be used to obtain High Density Localization (HDL) maps [6,7,9] of the cities, with providing dense and accurate point clouds from the static environment with homogeneous scanning of the surfaces and a nearly linear reduction of points as a function of the distance. Exploiting low level information from 3D city maps is a quite new research area, with a few related techniques. The HDNET [13] approach uses a prior *road map* with local *ground-height* data as reference, which helps in eliminating false object candidates detected out of the road, or above/under the ground level. However, it does not deal with the confusion of dynamic objects with static entities from the map, and therefore it cannot adjust the missing object rate. To fill this gap, we present a new approach which utilizes dense HDL maps in order to decrease in parallel both the *false negative* and *false positive* hits of RTDOD algorithms.

The key steps of the proposed algorithm are multimodal point cloud registration between the RMB Lidar measurements and the HDL maps, map based object validation, multimodal change extraction and object level change analysis. As a basis of comparison, we have chosen the *PointPillars* [5] state-of-the-art RTDOD method, which can predict object-candidates from multiple classes, together with their 3D oriented bounding boxes and class confidence values.

## 2 Proposed Algorithm

The workflow of the proposed approach is shown in Fig. 1. Initially, we apply a state-of-the-art RTDOD algorithm on the raw RMB Lidar frames - in the paper the *PointPillars* [5] techniques is used for this purpose - which provides us multiple vehicle and pedestrian candidates. To refine the output of RTDOD, first we need to accurately register the input RMB Lidar point cloud to the MLS based High Density Localization (HDL) map, which is achieved by a multimodal point cloud registration algorithm. After the alignment, we apply a probability map based validation step against the HDL map to remove *false positive* RTDOD predictions. Finally, for eliminating the *false negatives*, we subtract the HDL map and the already detected RTDOD objects from the actual Lidar frame, then we extract object candidate blobs in the remaining dynamic regions, and we attempt to identify the *previously undetected* dynamic objects by a Support Vector Machine (SVM) based blob-classifier.

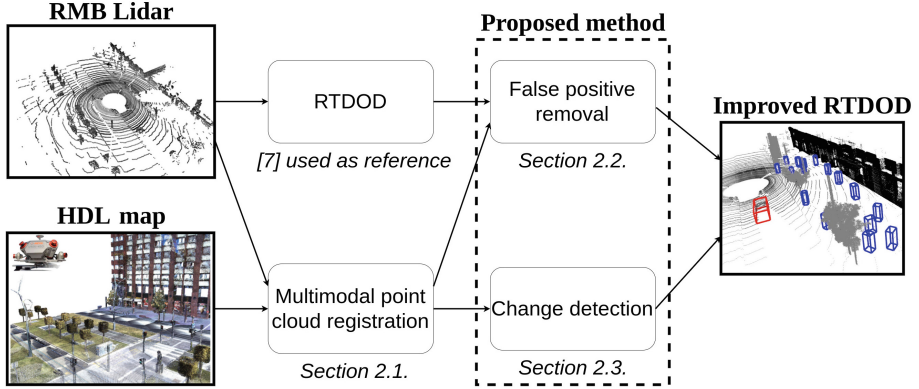


Fig. 1. Workflow of the proposed approach

## 2.1 Multimodal Point Cloud Registration

Let us assume that using internal navigation sensors, the current position of the vehicle is roughly known with a maximal error of 10 meters in the map’s coordinate system. For accurately registering the recorded RMB Lidar frames ( $\mathcal{P}_{\text{RMB}}$ ) to the available HDL map ( $\mathcal{P}_{\text{Map}}$ ), we search for a rigid transform between the two point clouds in the following form:

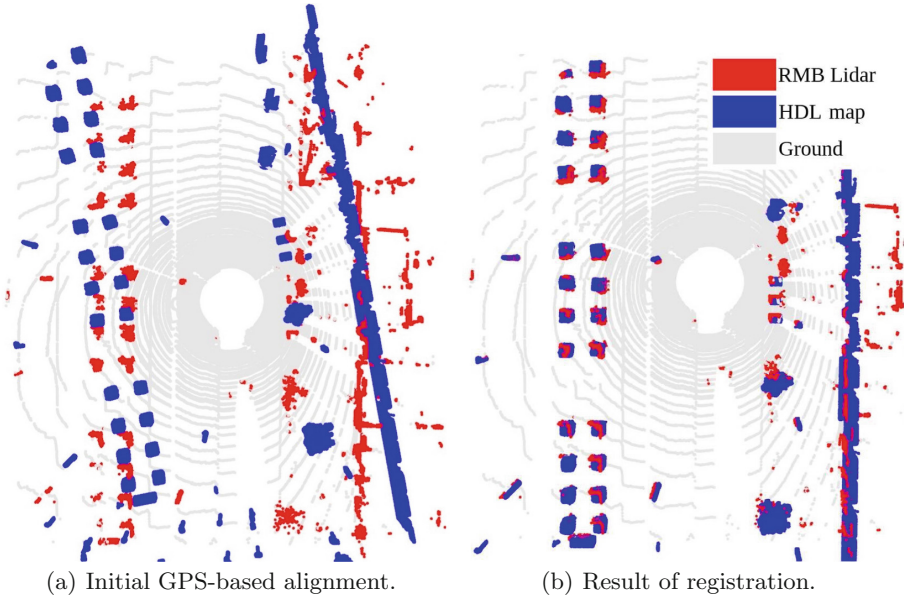
$$\mathbf{T}_{dx,dy,dz,\alpha} = \begin{bmatrix} \cos \alpha & \sin \alpha & 0 & dx \\ -\sin \alpha & \cos \alpha & 0 & dy \\ 0 & 0 & 1 & dz \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

where  $dx, dy, dz$  are the offset parameters and  $\alpha$  is the rotation angle around the vertical axis.

To estimate the optimal transform, we apply a robust blob level voting technique in the Hough space based [8]. First, we remove the road points by a locally adaptive ground filter, and extract object-like connected blobs – called abstract objects – by region-growing in both the actual measurements and the HDL map’s point cloud (see Fig. 2). Let us denote the two obtained blob sets by  $\mathcal{O}_{\text{RMB}}$  and  $\mathcal{O}_{\text{Map}}$ , respectively. Since we can assume that the HDL map is free of dynamic objects [7], we exclude from the  $\mathcal{O}_{\text{RMB}}$  set the blobs which overlap with the initial object candidate regions provided by the RTDOD. Thereafter, based on [8] we extract 8 keypoints in each abstract object candidate of  $\mathcal{O}_{\text{RMB}} \cup \mathcal{O}_{\text{Map}}$ . For finding the optimal parameter quartet  $(dx, dy, dz, \alpha)$  we iterate through all possible keypoint pairs in  $\mathcal{O}_{\text{RMB}} \times \mathcal{O}_{\text{Map}}$ , and aggregate their votes in the Hough space.

## 2.2 False Positive Removal by Map Based Validation

False objects of the RTDOD algorithms often overlap with static obstacles of the background scene, thus they can be identified through analyzing their location



**Fig. 2.** Registration of an onboard measurement to the HDL map

in the registered HDL map. We have proposed a 2D probabilistic approach to manage this problem (see Fig. 3). First, taking a top view analysis, we project both the RMB Lidar and the registered HDL map point clouds to a discrete grid on the ground plane, with a resolution of 10 cm. Thereafter, we assign to each  $(i, j)$  cell two competing potentials describing the foreground ( $P_{\text{fg}}(i, j)$ ) and background likelihoods ( $P_{\text{bg}}(i, j)$ ). Foreground values are determined by the RTDOD output: for each cell covered by an object candidate, we take  $P_{\text{fg}}(i, j) \in [0, 1]$  as the prediction score (i.e. confidence value) of the RTDOD network regarding the given object. The remaining cells receive  $P_{\text{fg}}(i, j) = 0$ . On the other hand, the background likelihoods are calculated from the projected MLS point cloud. If cell  $(i, j)$  is occluded by a static obstacle in the HDL map, we set  $P_{\text{bg}}(i, j) = 1$ , while for cells near to the boundaries of static objects we use a distance-based Gaussian attenuation in the  $P_{\text{bg}}$  until 1 meter in any directions (with variance parameter  $\sigma = 10$ ). For the remaining cells, we set  $P_{\text{bg}}(i, j) = 0$ .

Using the constructed likelihood maps, we remove all RTDOD object candidates, which cover any cell  $(i, j)$ , where  $P_{\text{bg}}(i, j) \geq P_{\text{fg}}(i, j)$ . Note that the adopted Gaussian soft boundary also ensures robustness of the approach against small registration errors.

### 2.3 Search for Missing Objects via Change Detection

Decreasing the number of missing dynamic objects of RTDOD is highly challenging, since we cannot exploit here the HDL map's object-information in a

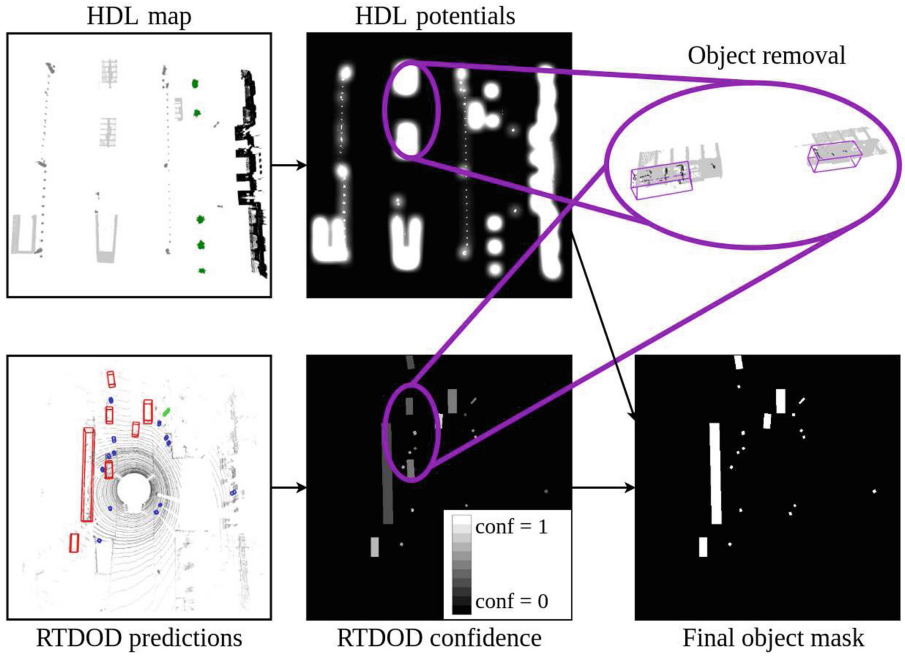


Fig. 3. False positive removal



(a) Range image representation of the RMB Lidar measurement.

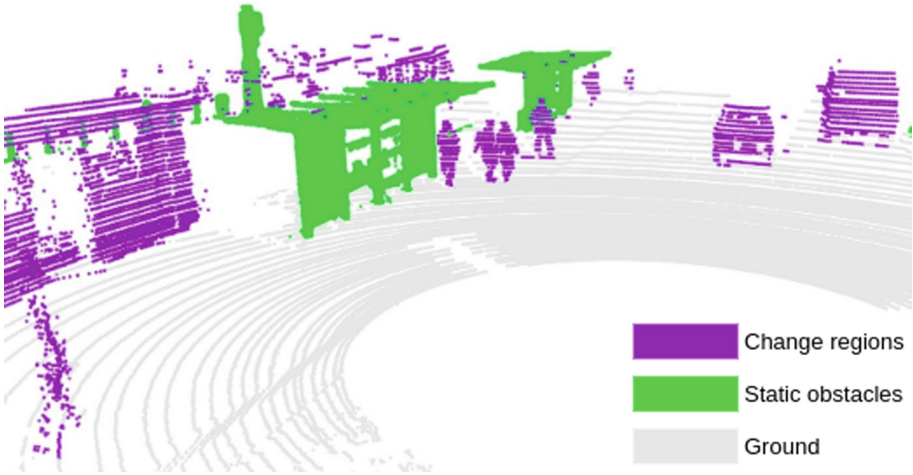


(b) Range image obtained from the HDL Map with ray tracing



(c) Markov Random Field based change mask between (a) and (b)

Fig. 4. Change detection in the range image domain



**Fig. 5.** Changes backprojected to the RMB Lidar frame

straightforward way, and especially for deep neural network (DNN) based detectors, the sources of mis-detections are often hard to explain intuitively. The key idea of our approach is that we separate at point level the dynamic and static regions of the input RMB Lidar cloud  $\mathcal{P}_{\text{RMB}}$ , and search for further possible objects of interest in the dynamic segments of  $\mathcal{P}_{\text{RMB}}$ . In this way, we expect improvement for two reasons: *first* we re-locate our previously undetected objects into a different context, where a second-round detection can be successful. *Second*, if a dynamic object's point cloud is heavily merged with a static obstacle's blob, the elimination of background points can highlight the object's real shape, which step can highly facilitate the classification.

We separate the dynamic and static regions of  $\mathcal{P}_{\text{RMB}}$  through multimodal background subtraction, where the registered HDL map  $\mathcal{P}_{\text{Map}}^*$  provides the background point cloud. Following our earlier approach [2], we transform the point clouds  $\mathcal{P}_{\text{RMB}}$  and  $\mathcal{P}_{\text{Map}}^*$  into range images by ray tracing, and apply a Markov Random Field based binary change segmentation in the 2D image domain (Fig. 4). Finally, we backproject the obtained change labels from the range images to the 3D point cloud (Fig. 5).

The next step is object separation within the change regions of the  $\mathcal{P}_{\text{RMB}}$  cloud, which is performed by our efficient two-level grid based clustering method introduced in [1]. This process is implemented using a 2D cell map: at super cell level, a region growing algorithm is executed, where empty cells act as stopping criterion. This step may merge several nearby entities into the same object. Therefore, we apply a refinement step at the sub-cell level: a super-level object is divided into different parts, if we find a separator line composed of low density sub-cells at the fine resolution.

The above process provides as output a set of blobs, where some of them might represent further objects of interest ignored by RTDOD, while the other

ones belong to a general street clutter class, which is currently out of our focus. For the final decision, we trained a Support Vector Machine with a Radial Basis Function (RBF) [4] kernel, which classifies the blobs based on the set of features listed in Table 1. After classification, the blobs labeled as vehicles or pedestrians are added to the object list of the detector.

**Table 1.** Feature vector used for SVM classification

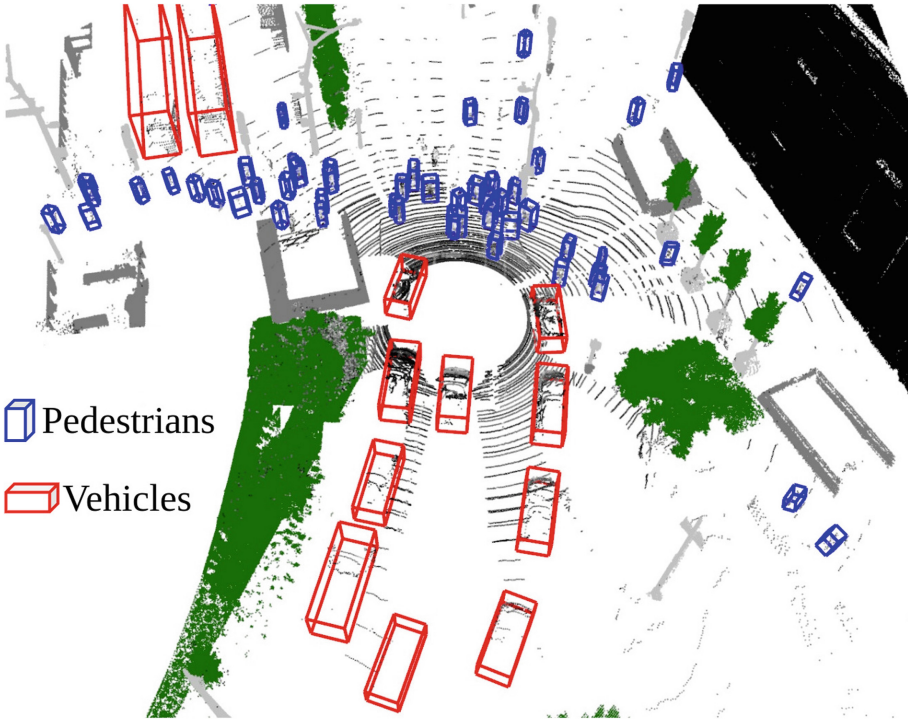
No.	Description	Dim.
$f_1$	Number of points included in the object	1
$f_2$	The minimum distance to the object center	3
$f_3$	3D covariance matrix of the object points	6
$f_4$	Principial component of the object	3
$f_5$	3D bounding box sizes (height, width, depth)	3

### 3 Evaluation

We have evaluated the proposed technique on real dynamic point cloud sequences recorded by a car mounted Velodyne HDL 64-E Lidar scanner, on roads with heavy traffic in Budapest, Hungary. The High Density Localization (HDL) map was prepared in our laboratory from high resolution point clouds of a Riegl VMX-450 MLS system, provided by the city’s road management company (Budapest Közút Zrt). During HDL map generation, the raw MLS data also undergo a semantic segmentation step [7] for ghost object removal and road detection. As state-of-the-art RTDOD method, we used the *PointPillars* technique [5] which was trained on a mixed dataset composed of the KITTI [3] benchmark, and additional annotated samples from our Budapest data [1].

Qualitative results are shown in Fig. 6 and 7. Figure 6 displays a large scene where the proposed model provides us a comprehensive scene interpretation, although several vehicles (both cars and trams) and pedestrians are jointly present. Figure 7 demonstrates the improvements of using our map-based approach versus the pure RTDOD technique. At the top (Fig. 7(a)(b)), we find two false vehicle predictions which are successfully removed based on the background cloud, while at the bottom (Fig. 7(c)(d)), we illustrate that even in a crowded scene with multiple pedestrians we can find new previously undetected people with our change detection based approach.

For quantitative evaluation, we have selected five heavy traffic road sections recorded in the city center of Budapest, near Deák square (Fig. 6), Múzeum boulevard, Fővám square, Károly boulevard and Kálvin square, respectively. From each location, the evaluation dataset contains 50 different frames; and in average 5 vehicles and 16 pedestrians are present in a single time frame. The numerical performance results compared to the original *PointPillars* [5] output



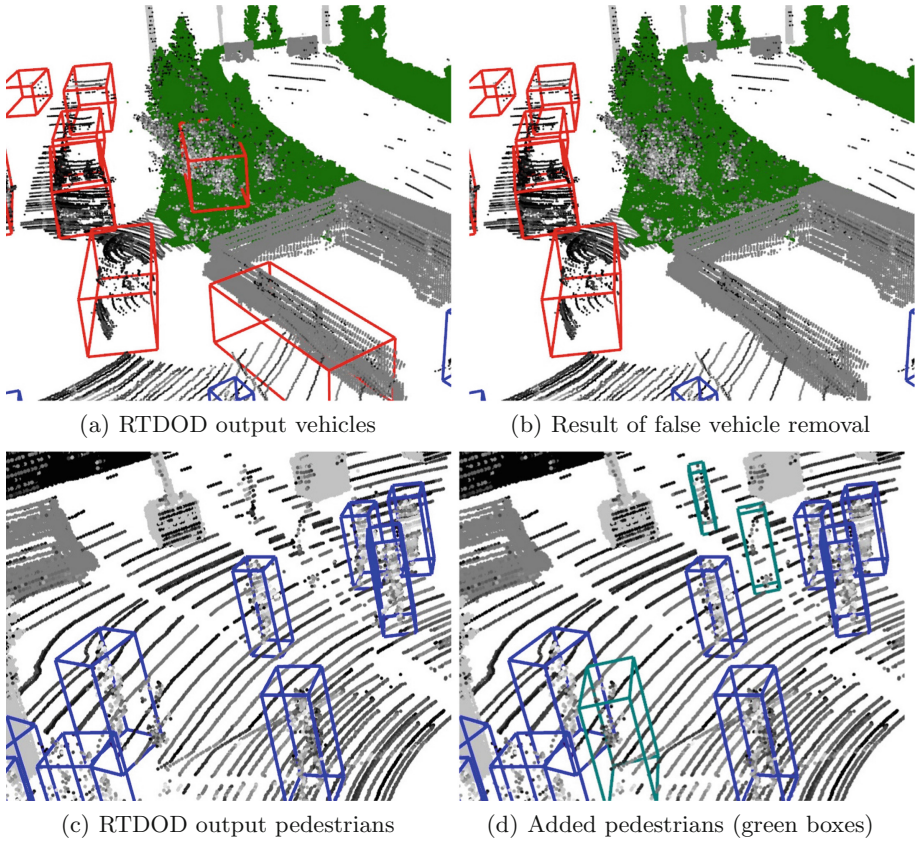
**Fig. 6.** Result of object detection by Deák square, Budapest

**Table 2.** Quantitative results versus *PointPillars* [5]

Method	Class	Precision	Recall	F-score
Only RTDOD [5]	Pedestrian	<b>95,62%</b>	67,42%	79,08%
	Vehicle	75,19%	88,19%	81,11%
	All	88,75%	72,22%	79,64%
RTDOD with the proposed method	Pedestrian	94,60%	<b>84,43%</b>	<b>87,52%</b>
	Vehicle	<b>95,02%</b>	88,19%	<b>91,38%</b>
	All	<b>94,71%</b>	<b>81,43%</b>	<b>87,57%</b>

are shown in Table 2. With the proposed false-positive removal step (Sect. 2.2), we obtained a 19, 83% *precision* improvement for the vehicle class, by eliminating many false vehicle-like regions of the RMB Lidar measurements. As result of the change detection based blob classification step (Sect. 2.3), we could significantly improve the *recall* rate of pedestrians with 17.01%. In general for both classes, we observed an overall improvement of 7, 93% in *F-score*, versus relying purely on the state-of-the-art RTDOD approach.





**Fig. 7.** Qualitative demonstration of improvements by using the proposed model (right) versus RTDOD (left) (Color figure online)

The algorithms were tested on a desktop computer with CPU implementation, where the average computation time was 100 ms per frame for the registration and 80 ms per frame for the change detection step, respectively.

## 4 Conclusion and Future Work

We introduced a new method to exploit High Definition Localization (HDL) maps for performance improvement of state-of-the-art Lidar based dynamic object detection algorithms. We have shown that the proposed approach can efficiently balance the precision and recall values with significant overall improvement for both vehicles and pedestrians. Most of the remaining detection errors were related to pedestrians too close to each other, and vehicles far away where the point cloud is much sparser, which problems may be reduced by adopting object tracking in the future.

**Acknowledgements.** This work was supported by the National Research, Development and Innovation Fund under grant number K-120233, by the European Union and the Hungarian Government from the projects *Thematic Fundamental Research Collaborations Grounding Innovation in Informatics and Infocommunications* under grant number EFOP-3.6.2-16-2017-00013 (Örkény Zováthi and Balázs Nagy) and *Intensification of the activities of HU-MATHS-IN - Hungarian Service Network of Mathematics for Industry and Innovation* under grant number EFOP-3.6.2-16-2017-00015, and by the Michelberger Master Award of the Hungarian Academy of Engineering (Csaba Benedek).

## References

1. Börcs, A., Nagy, B., Benedek, C.: Instant object detection in Lidar point clouds. *IEEE Geosci. Remote Sens. Lett.* **14**(7), 992–996 (2017)
2. Gálai, B., Benedek, C.: Change detection in urban streets by a real time Lidar scanner and MLS reference data. In: Karray, F., Campilho, A., Cheriet, F. (eds.) *ICIAR 2017*. LNCS, vol. 10317, pp. 210–220. Springer, Cham (2017). [https://doi.org/10.1007/978-3-319-59876-5\\_24](https://doi.org/10.1007/978-3-319-59876-5_24)
3. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? The KITTI vision benchmark suite. In: *Conference on Computer Vision and Pattern Recognition (CVPR)* (2012)
4. Kidono, K., Miyasaka, T., Watanabe, A., Naito, T., Miura, J.: Pedestrian recognition using high-definition LIDAR. *J. Robot. Soc. Japan* **29**, 405–410 (2011). <https://doi.org/10.1109/IVS.2011.5940433>
5. Lang, A., Vora, S., Caesar, H., Zhou, L., Yang, J., Beijbom, O.: PointPillars: fast encoders for object detection from point clouds. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2019)
6. Matthaei, R., Bagschik, G., Maurer, M.: Map-relative localization in lane-level maps for ADAS and autonomous driving. In: *IEEE Intelligent Vehicles Symposium Proceedings*, Dearborn, MI, USA, pp. 49–55, June 2014. <https://doi.org/10.1109/IVS.2014.6856428>
7. Nagy, B., Benedek, C.: 3D CNN-based semantic labeling approach for mobile laser scanning data. *IEEE Sens. J.* **19**(21), 10034–10045 (2019)
8. Nagy, B., Benedek, C.: Real-time point cloud alignment for vehicle localization in a high resolution 3D map. In: Leal-Taixé, L., Roth, S. (eds.) *ECCV 2018*. LNCS, vol. 11129, pp. 226–239. Springer, Cham (2019). [https://doi.org/10.1007/978-3-030-11009-3\\_13](https://doi.org/10.1007/978-3-030-11009-3_13)
9. Seif, H.G., Hu, X.: Autonomous driving in the iCity-HD maps as a key challenge of the automotive industry. *Engineering* **2**(2), 159–162 (2016). <https://doi.org/10.1016/J.ENG.2016.02.010>
10. Shin, K., Kwon, Y., Tomizuka, M.: . RoarNet: a robust 3D object detection based on region approximation refinement. In: *IEEE Intelligent Vehicles Symposium*, pp. 2510–2515 (2018)
11. Simon, M., Milz, S., Amende, K., Groß, H.M.: Complex-YOLO: real-time 3D object detection on point clouds. *arXiv abs/1803.06199* (2018)

12. Yan, Y., Mao, Y., Li, B.: SECOND: sparsely embedded convolutional detection. *Sensors* **18**, 3337 (2018). <https://doi.org/10.3390/s18103337>
13. Yang, B., Liang, M., Urtasun, R.: HDNET: exploiting HD maps for 3D object detection. In: Conference on Robot Learning. Proceedings of Machine Learning Research, 29–31 Oct 2018, vol. 87, pp. 146–155. <http://proceedings.mlr.press/v87/yang18b.html>
14. Yang, B., Luo, W., Urtasun, R.: PIXOR: real-time 3D object detection from point clouds. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7652–7660 (2018). <https://doi.org/10.1109/CVPR.2018.00798>
15. Zhou, Y., Tuzel, O.: VoxelNet: end-to-end learning for point cloud based 3D object detection. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4490–4499 (2018). <https://doi.org/10.1109/CVPR.2018.00472>