Street object classification via LIDARs with only a single or a few layers

1st Zoltan Rozsa
Hungarian Academy of Sciences
Institute for Computer Science and Control (MTA SZTAKI)
and
Budapest University of Technology and Economics
Faculty of Transportation Engineering and Vehicle Engineering (BME-KJK)
Budapest, Hungary
zoltan.rozsa@logisztika.bme.hu

2nd Tamas Sziranyi
Hungarian Academy of Sciences
Institute for Computer Science and Control (MTA SZTAKI)
and
Budapest University of Technology and Economics
Faculty of Transportation Engineering and Vehicle Engineering (BME-KJK)
Budapest, Hungary
sziranyi.tamas@sztaki.mta.hu

Abstract—**LIDAR sensors are part of the sensor system of several intelligent vehicles and transportation systems providing both object and free-space detection capabilities. In this paper a recognition method is proposed for LIDARs with only a few detection planes. Our method is especially useful in the case when angular resolution of the scan is sufficient, but in the vertical direction the planes are far from each other. The proposed method uses new features including Fourier based descriptor, deep learning classification and exploits additional 3D information if it is available. We tested the method on ten thousands of samples from a large public database. This paper gives an effective solution for a hard problem of LIDAR based recognition problems, namely the far-object detection in case of mobile LIDARs of limited or poor vertical resolution.**

Index Terms—**LIDAR, Intelligent vehicles, recognition**

I. INTRODUCTION

Autonomous driving requires different sensor modalities to work together in order to ensure safe transportation. There are ways of task allocation between sensors which are proved to be efficient, like using depth sensors as LIDARs for free-space or object candidate detection, vision for object recognition. However, relying only on one sensor in case of any task (for example cameras for classification) is just not enough to minimize probability of accidents in any circumstances because of the limited capabilities of the sensors. That is why we have to maximize the efficiency of each sensor modality for each task. We aim to improve the overall classification performance with LIDAR sensors in this paper.

Vehicles are frequently equipped with LIDARs with only a few detection planes (e.g. SICK LD-MRS[1] or Velodyne VLP-16[2]) or even with only one (e.g., SICK LMS5xx series[3]). Dealing with LIDARs with many planes (e.g. Velodyne HDL-64[4]), we will experience that far objects will be represented in only a few planes and they cannot be treated as point clouds (Figure 1). In [1], the authors proposed a solution for the relatively slow Automated Guided Vehicles, where a

3D reconstruction was made by fusing the separated planes. However, in case of autonomous vehicles, their fast movement requires even faster decision. In this paper we propose a solution to this problem by handling all the object candidates as set of plane curves. We will show that these plane curves are suitable for object recognition and increasing number of scan planes increases the recognition probability as well. Instead of point clouds of very poor vertical resolution at far-distances, we use the advantage of good in-plane resolution of few-layer LIDARS, considering the under-sample situation at the vertical direction.

Recent works (e.g. [2], [3]) show good detection performance for a few categories (about 95 % recall for four categories) in case of 2D LIDARs. We aim to enhance these methods and apply to the present problem.
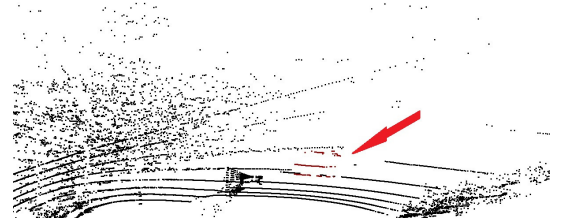

Fig. 1. Velodyne VLP16 sequence. Car represented only with 3 detection plane (and so cannot be treated as point cloud) is marked with red points. The car's distance to the sensor is about 13 m.

Addressing the above problems usual in recognition tasks from LIDAR point cloud we contribute a new methodology listed now:

- New approach for description of plane curves.
- Object representation as set of plane curves with altitude.
- Convolutional Neural Network (CNN) with classification at the output.
- Extension is possible for tracking and/or multiple planar curves.
- Propose voting scheme in order to increase recognition probability.
- Offer solution to recognition cases of limited number of LIDAR planes scanning an object, including far objects cut by only a few scan plane.

---

[1]https://www.sick.com/us/en/detection-and-ranging-solutions/3d-lidar-sensors/ld-mrs/c/g91913

[2]http://velodynelidar.com/vlp-16.html

[3]https://www.sick.com/us/en/detection-and-ranging-solutions/2d-lidar-sensors/lms5xx/c/g179651

[4]http://velodynelidar.com/hdl-64e.html

II. Related works

The related literature mainly corresponds to recognition of objects realized with LIDAR sensors having one or only a few planes. Methods working on 3D LIDARs have the potential for the classification of several object classes, because these methods have more information than in the case of dealing with separated 2D LIDAR segments. Works like [4] and [5] use 2D or 3D Convolutional networks for classification, but they require point clouds as input. Compared to this, in work of [1] a solution was proposed for the problem, where 2.5D point clouds are not available, only partial (but connected) object data. However, objects in the far plane cannot be handled even with this type of methods, because they are scanned by only a few unconnected 2D planar curves, so here a combined approach is proposed.

The first applications related to object detection [6] and tracking [7] with laser range finders have been already introduced in the early 2000s. The primary goal of these early approaches were to find and track people; more than one object class was not considered. Today, it is still an actual topic in robotics and autonomous driving. Now, the development of sensors and computer vision algorithms offer the possibility to consider more than one class to recognize even in this planar contour data. [2] used the width of an obstacle and the measured intensity. The authors were capable of differentiating four categories with good accuracy based on euclidean distance. Later, adding one more feature to the descriptor (range variance) they were able to increase their classification accuracy [8]. Another approach was presented in [9] where the detected blobs were converted to a 5x5 binary image and SVM was used to classify the objects as vehicles or pedestrians. [10] propose a distant-invariant feature for segmentation and detection of people without walking aids, people with walkers, people in wheelchairs and people with crutches.

There are further works, gathering information from multiple planes, either by using more than one planar LIDARs or utilizing multi-planar ones. The authors of [11] detect different body parts at different heights by using more than 10 features acquired from the scans and AdaBoost algorithm to train a strong classifier and based on that and their model they predict people's shape. A similar approach is presented in [12] but they use multiple laser range-finder instead of a multi-layered one and in [13] as well. [14] applied motion characteristics to identify humans with baby cart, shopping cart or wheel chairs.

Summarizing, classification methods dealing with on one or a few planar scans, most of the cases use tens of geometrical features and Adaboost or neural network methods to build a strong classifier ([3], [15]). They do not use the information provided via multiple planes (only for searching specific body parts). A few classes are considered for detection. Most of the time these methods are applied for the classification of objects of industrial halls scanned with indoor sensors with limited range (they also mostly depend on range and angular resolution of the sensor). These tests have been executed on a few thousands of samples [3]. Compared to these, we list

(a) Original frame

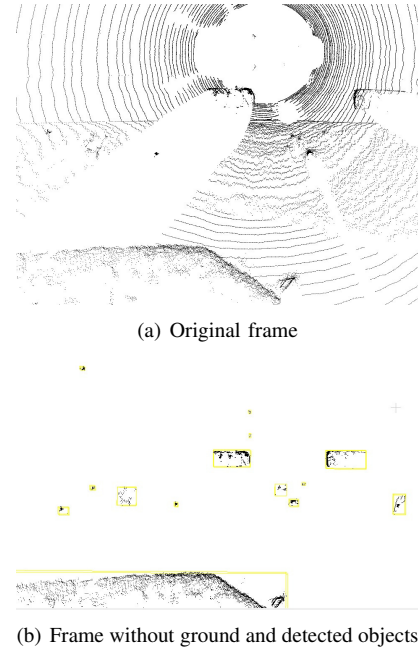(b) Frame without ground and detected objects

Fig. 2. Example of preprocessing steps on KITTI tracking database

here the main advantages of our method:

- We propose a method for classification of data acquired by LIDARs with a few layers and far field data of 3D LIDARs with utilizing the multi-plane information.
- Our method is designed for outdoor object classification, and it is suitable for several classes.
- We validated our method on ten thousands of samples.

III. Our proposed method

In the following we will explain our method in details. First, preprocessing steps will be described then the classification procedure which is the contribution of the paper. We will assume a few-layer LIDAR in the following.

A. Preprocessing

The input of the pipeline is a full scan of a LIDAR sensor, which we call frame in the following. By segmenting the ground we can detect objects clusters. Here, we are listing known methods that we used in our experiments:

- Ground detection: M-estimator SAmple Consensus (MSAC) Plane fitting [16]. MSAC uses the loss function:

$$Loss(e) = \begin{cases} e^2 & |e| < T \\ T^2 & \text{otherwise} \end{cases} \quad (1)$$

  where e is the error and T is the threshold for inliers.
- Object detection: Euclidean cluster extraction [17] with distance varying neighborhood radius.

Illustration of these processing steps can be seen on Fig. 2. After we found object clusters, if an object is represented on more than one ring, we segment it to plane curves in order to separately evaluate it.

Fig. 3. Example of description of a car from 5 segments (Purple: Curve 1, Green: Curve 2, Blue: Curve 3, Red: Curve 4, Black: Curve 5)

B. Descriptor and classification

Here, we assume that objects are represented by plane curves. In our experiments we used a fx$(n + 6)$ matrix as a descriptor of LIDAR segments. Here f is the number of curves representing an object and n is the number of Fourier descriptor components we use (n is also the minimum number of points which can construct a segment). In the following it will be explained how it is composed.

1) Fourier descriptor: Instead of extracting geometric features from curves, we found that utilizing a descriptor which can be used to reconstruct the curve exactly [18] gives better classification results. Fourier descriptor is applicable on closed contours, we construct a closed contour from the segment by adding to the original 2D cloud its points in reverse order [19]. By subtracting the mean from the 2D point cloud and by using the absolute value of the Fourier transformed contour we get a translation and rotation invariant representation of the plane curve. This representation also shows robustness against varying point density.

2) Statistical measures: Other than shape properties of the plane curve are stored in a simple form. The mean and standard deviation values of altitude, distance to the sensor and intensity values are also part of our descriptor.

3) Multiple plane: We use the f geometrically nearest curve of the same object, these will form the rows of our descriptor matrix. In our experiments we used $f = 5$ and $n = 5$, if the object has $f < 5$ curves, we used the original curve more than once in order to always get 5x(5 + 6) descriptor dimension. As tested, this simple but useful replication solved the lack of enough samples at the input. The descriptor matrix is illustrated on Fig. 3 and Table I.

4) Classification: For the classification of the objects we use a Convolutional Neural Network [20]. The network architecture we used can be seen in Fig. 4. We use this classifier and this structure because we found it to be superior to other classifiers we tried in case of our descriptor (e.g. multilayer fully-connected network, Nearest Neighbor Classification or Support Vector Machine [21]) and it also has the advantage of grouped curve classification. This model is prepared for using 5 segments from an object as input.

5) Voting: We applied voting scheme in cases when an object was built up from more than 5 planar curves. Each of the segments is evaluated separately, but the final decision is made at the object level.

$$C = argmax_i N_i \qquad (2)$$

TABLE I
THE (TRANSPOSE MATRIX OF THE) DESCRIPTOR OF THE 2D POINT CLOUD
SET ABOVE (FDx INDICATES THE xTH FOURIER COMPONENT, z IS THE
ALTITUDE, r IS THE DISTANCE TO THE ORIGIN AND I MEANS INTENSITY)

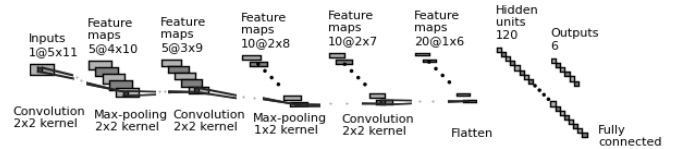|	Curve 1	Curve 2	Curve 3	Curve 4	Curve 5
FD1	0.2477	0.2774	0.3139	0.2839	0.3363
FD2	0.0774	0.0642	0.0418	0.0555	0.0504
FD3	0.0312	0.0649	0.0268	0.0253	0.0394
FD4	0.0203	0.0380	0.0128	0.0081	0.0390
FD5	0.0120	0.0361	0.0171	0.0299	0.0178
mean(z)	-0.1315	-0.6662	-0.8657	-1.1011	-1.3704
std(z)	0.0021	0.0023	0.0023	0.0037	0.0041
mean(r)	42.8106	41.5523	41.0485	40.7877	40.8404
std(r)	0.2516	0.1127	0.0857	0.1172	0.1063
mean(I)	0.0	0.0	0.0	0.1808	0.0793
std(I)	0.0	0.0	0.0	0.2080	0.1044

Fig. 4. Network architecture: all convolutional layers are followed by ReLUs and the fully-connected layer is followed by a softmaxlayer not illustrated in the scheme.

where C is the final decision about the object class, i is the class number and N_i is the number of vote for the ith class, which we get by counting the all the segments classified as member of the ith class from all the n segments of the object.

$$N_i = \sum_{j=1}^{n} [S_j = i] \qquad (3)$$

IV. TEST RESULTS

For the comparable test purposes we used the well-known KITTI database including Velodyne 64 data, for which we randomly selected out the vertically far under-sampled planes, which results infrequent random few-plane sampling. We have also tested other devices, like Velodyne VLP-16 and Qunaergy M8 [5] in real-word conditions with similar results, however there were not enough annotated data to show relevant comparison here. So, we conducted our quantitative, proof of concept tests in the training set of the KITTI tracking database [22]. In this set labeled objects are annotated through different number of frames in 21 sequences. It allows us to investigate our classification algorithm independently from the quality of the preprocessing. In these tests we gathered all the not occluded and not truncated objects from 8 categories (car, van,

[5] https://quanergy.com/m8/

truck, pedestrian, person sitting, cyclist, tram, misc) having at least 1 segment with minimum 5 points. These objects were cut out based on their annotated 3D bounding box and then we divided them into segments by the scanner planes. This resulted us 197,256 samples, which we divided into training (70 %), validation (15 %) and test (15 %) sets randomly, however there were completely new sequences in the tests as well. From the original KITTI categories of car and van and also pedestrian and person sitting are combined, because they are 'neighboring' categories. The categories are the following: 1: Car and Van, 2: Truck, 3: Pedestrian and Person Sitting, 4: Cyclists, 5: Tram, 6: Misc.

First, we tested our method on single planar segments without using information from the neighboring curves with one $(n + 6)$ data vector at the input. The result is visible in Table III. We implemented this in order to be able to compare our method to the state of the art one applied on 2D LIDAR databases. The method proposed in [2] and [8] was tested in our database (Table II). In the test of Table II a nearest neighbor classification was made based on Euclidean distance to the train database built from width, range variance and intensity data as the authors of [8] proposed. Comparing results of Tables II and III it can be seen that our method is superior in almost every aspect. However, our method has been developed for multiple curves, so if we use information of them and voting scheme we get significant improvements. Confusion matrices for these cases are shown in Tables IV and VI. Table IV uses 5 planar segments of an object as CNN input, Table VI uses only 1 planar segment as CNN input, but voting is applied on object level on the output. In the second way all the segments of an object can be considered for the decision. Average F-measure is indicated as $\overline{F}$, F-measure weighted by sample number of each category denoted by $\overline{F_w}$:

$$\overline{F_w} = \sum F_i \cdot \frac{n_i}{N} \qquad (4)$$

Where F_i is the F-measure and n_i is the number of samples of ith category, and N is the cardinality of all the samples.

TABLE II
CONFUSION MATRIX FOR SINGLE PLANAR CURVES BY METHOD PROPOSED IN [2], [8]. *(1: Car and Van, 2: Truck, 3: Pedestrian and Person Sitting, 4: Cyclists, 5: Tram, 6: Misc)*

	1	2	3	4	5	6	Precision (%)
1	**10024**	460	1136	289	16	267	**82.2**
2	466	**464**	281	31	12	43	**35.7**
3	1146	314	**11170**	924	2	112	**81.7**
4	325	32	919	**541**	0	40	**29.1**
5	18	12	2	0	**15**	1	**31.3**
6	255	46	103	40	1	**81**	**15.4**
Recall (%)	81.9	34.9	82.1	29.6	32.6	14.9	$\overline{F}$:**0.4595** $\overline{F_w}$:**0.753**

TABLE III
CONFUSION MATRIX FOR SINGLE PLANAR CURVES BY OUR PROPOSED METHOD. *(1: Car and Van, 2: Truck, 3: Pedestrian and Person Sitting, 4: Cyclists, 5: Tram, 6: Misc)*

	1	2	3	4	5	6	Precision (%)
1	**11876**	337	232	168	26	330	**91.6**
2	175	**974**	0	3	14	58	**79.6**
3	113	4	**12395**	776	0	74	**92.8**
4	47	10	970	**874**	0	48	**44.8**
5	0	0	0	0	**6**	0	**100.0**
6	23	3	14	4	0	**34**	**43.6**
Recall (%)	97.1	73.3	91.2	47.9	13.0	6.3	$\overline{F}$:**0.572** $\overline{F_w}$:**0.878**

TABLE IV
CONFUSION MATRIX FOR PLANAR CURVES BY OUR PROPOSED METHOD, USING MAXIMUM 5 SEGMENTS OF AN OBJECT AS DESCRIPTOR (CNN INPUT). *(1: Car and Van, 2: Truck, 3: Pedestrian and Person Sitting, 4: Cyclists, 5: Tram, 6: Misc)*

	1	2	3	4	5	6	Precision (%)
1	**11957**	286	67	90	13	299	**94.1**
2	205	**1031**	0	0	20	93	**76.4**
3	24	1	**13182**	914	0	105	**92.7**
4	41	10	361	**822**	0	28	**65.1**
5	0	0	0	0	**13**	0	**100.0**
6	7	0	1	0	0	**19**	**70.4**
Recall (%)	97.7	77.6	96.9	45.0	28.3	3.5	$\overline{F}$:**0.619** $\overline{F_w}$:**0.902**

TABLE V
CONFUSION MATRIX FOR PLANAR CURVES (OF FAR OBJECTS) BY OUR PROPOSED METHOD, USING MAXIMUM 5 SEGMENTS OF AN OBJECT AS DESCRIPTOR (CNN INPUT). *(1: Car and Van, 2: Truck, 3: Pedestrian and Person Sitting, 4: Cyclists, 5: Tram, 6: Misc)*

	1	2	3	4	5	6	Precision (%)
1	**1119**	1	0	7	5	12	**97.8**
2	31	**2**	0	0	2	0	**5.7**
3	3	0	**310**	23	0	1	**92.0**
4	0	0	4	**10**	0	2	**62.5**
5	0	0	0	0	**0**	0	**0.0**
6	3	0	0	0	0	**6**	**66.7**
Recall (%)	96.8	66.7	98.7	25.0	0.0	28.6	$\overline{F}$:**0.465** $\overline{F_w}$:**0.939**

	1	2	3	4	5	6	Precision (%)
1	**12170**	62	8	61	15	373	**96.0**
2	62	**1266**	0	0	18	42	**91.2**
3	1	0	**13444**	755	0	79	**94.2**
4	0	0	157	**1009**	0	46	**83.3**
5	0	0	0	0	**13**	0	**100**
6	1	0	2	0	0	**4**	**57.1**
Recall (%)	**99.5**	**95.3**	**98.8**	**55.3**	**28.3**	**0.8**	$\overline{F}$:0.660 $\overline{F_w}$:0.932

The confusion matrix in Table III shows that even one 2D contour can produce good initial results with our method, and both using multiple curve information Table IV and simple voting scheme VI is effective to increase the accuracy of the classification even more. Detailed results divided by categories:

- The results of car and pedestrian categories are convincing both in terms of precision and recall.
- The performance in case of truck category is acceptable, the main source of confusion is that they are frequently categorized as Car or Van, which can be reasonable.
- There is a similar situation in case of cyclists, which are frequently categorized as Pedestrian or Person Sitting. The performance measurements in case of this category are not satisfying in case of 2D contours, but it has to be noted there were much less samples in this case. If we merge this category to the other human related one (pedestrian and person sitting), we get cc. 99 % precision and recall for this category and about 0.96 F-measure weighted by sample numbers of each category.
- The results on tram class are hardly sufficient, the recall of the category is increasing by using multiple curves of the same object and voting. However, it is not representative because of the very small number of samples.
- Finally, in case of misc category our proposed method did not performed well at all, because of the variety of the objects hard to identify in 2D contours and distinguish from vehicles (e.g. trailer, caravan).

In Table V a separate evaluation is presented for far objects. Here an object is considered far if it builds up from maximum five scan planes. In this case the average distance of center of gravity from the sensor is about 38.5 m. The table shows that the increasing distance does not influence the method. Note that: some categories are not present in the far field in this database or just with very few samples, results about these cases are not representative.
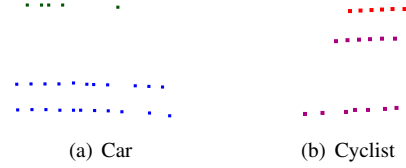

(a) Car   (b) Cyclist

Fig. 5. Far object examples of the KITTI database: the colors of the points correspond to the output category of the algorithm (Red - Pedestrian, Purple - Cyclist, Blue - Car, Green - Truck).

We present a comparison (Table VII) with state of the art 3D recognition method as well. The test dataset is presented in [4], it contains segmented objects. Intensity data is not provided, so it was left out from our descriptor. There are four object categories in this urban data, namely: vehicle, street furniture, pedestrian and facade. Results show that our method perform better in case of almost every measure. Vehicle category is an exception, however, authors of [4] execute a contextual refinement for this class.

Categories	Precision (%)		Recall (%)		F-rate	
	[4]	proposed	[4]	proposed	[4]	proposed
Vehicle	**98**	96	**99**	94	**0.99**	0.95
Street Furniture	92	**94**	97	**100**	0.94	**0.97**
Pedestrian	78	**97**	78	**100**	0.78	**0.98**
Facade	**93**	90	77	**97**	0.84	**0.94**
Average	90	**94**	87	**98**	0.89	**0.96**

Fig. 5 shows examples of categorized plane curves. The results are promising considering that pedestrian detection robust against about 30 % occlusion [23] on 2D images, and in a similar dataset [22] best detection results using both vision and LIDAR data [24] is about 82 % for cars and less for pedestrians and cyclists. In Fig. 6 an illustration of the executed tests are visible, respectively from Tables III, IV and VI. In Fig. 6(a) one can observe that, human segments of cyclists objects are frequently categorized as pedestrian (and also some cases car segments categorized as truck). In Fig. 6(b) single mis-categorized curves are not present; different decision clusters can be seen in one object by evaluating 5 neighboring curve simultaneously. Finally, in Fig. 6(c) decisions are made on object level by voting of separately evaluated curves of an object; here most of the cyclists are predicted correctly, however some of them predicted as pedestrian.

V. CONCLUSIONS

In the paper we proposed a novel 2D recognition method using additional 3D information if it is available. This method is designed to solve the recognition problem of far objects

(a) Separately evaluated single planar curves in CNN

(b) Maximum 5 scans evaluated simultaneously in CNN

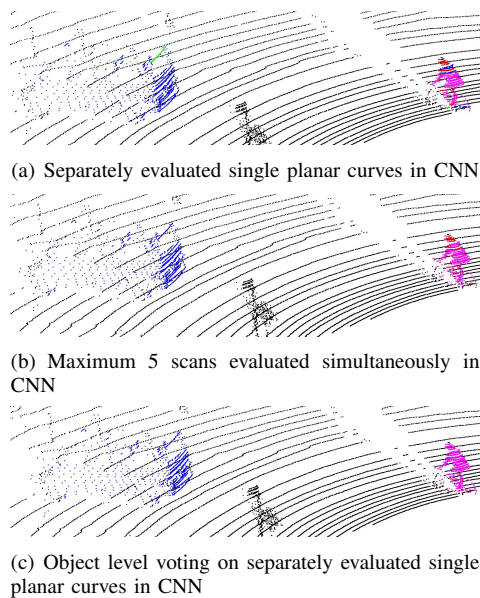(c) Object level voting on separately evaluated single planar curves in CNN

Fig. 6. Examples on KITTI database (Colormap: Blue - Car and Van, Green - Truck, Red - Pedestrian and Person sitting, Purple - Cyclist, Cyan - Tram, Yellow - Misc).

from LIDAR clouds or the general recognition problem for a few layer LIDARs. We demonstrated that our method is capable of categorizing noisy 2D clouds on a large public database. We proposed a method with the advantages of being model-free and also designed for outdoor objects by being invariant of the sensor we use. We compared it to a method used for object detection in 2D LIDAR clouds, our method is proved to be superior. In case of 5 categories 0.96 F-measure is reachable. We compared our method to 3D recognition methods as well. Our proposed method using CNN deep learning makes possible the grouped valuation of multiple planar curves (on the local or temporal - during tracking - neighboring planes). We suggest to use it as extension to 3D recognition methods on environment they cannot process. In the future we would like to combine our method with tracking to increase the recognition performance, evaluate the method on different databases, implement more sophisticated decision and execute remote scanning (far object) tests.

REFERENCES

[1] Rozsa, Z., Sziranyi, T.: Obstacle prediction for automated guided vehicles based on point clouds measured by a tilted lidar sensor. IEEE Transactions on Intelligent Transportation Systems **19** (2018) 2708–2720

[2] Lee, M., Hur, S., Park, Y.: Obstacle classification method based on 2D lidar database. Pattern Recognition Letters **8** (2014) 1442–1446

[3] Kurnianggoro, L., Jo, K.H.: Object classification for LIDAR data using encoded features. In: 2017 10th International Conference on Human System Interactions (HSI). (2017) 49–53

[4] Borcs, A., Nagy, B., Benedek, C.: Instant object detection in Lidar point clouds. IEEE Geoscience and Remote Sensing Letters **14** (2017) 992–996

[5] Maturana, D., Scherer, S.: VoxNet: A 3D Convolutional Neural Network for Real-Time Object Recognition. In: IROS. (2015)

[6] Arras, K.O., Mozos, O.M., Burgard, W.: Using boosted features for the detection of people in 2D range data. In: Proceedings 2007 IEEE International Conference on Robotics and Automation. (2007) 3402–3407

[7] Fod, A., Howard, A., Mataric, M.A.J.: A laser-based people tracker. In: Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No.02CH37292). Volume 3. (2002) 3024–3029

[8] Lee, M., Hur, S., Park, Y.: An obstacle classification method using multi-feature comparison based on 2D lidar database. In: 2015 12th International Conference on Information Technology - New Generations. (2015) 674–679

[9] Galip, F., Sharif, M.H., Caputcu, M., Uyaver, S.: Recognition of objects from laser scanned data points using SVM. In: 2016 First International Conference on Multimedia and Image Processing (ICMIP). (2016) 28–35

[10] Weinrich, C., Wengefeld, T., Volkhardt, M., Scheidig, A., Gross, H.M. In: Generic Distance-Invariant Features for Detecting People with Walking Aid in 2D Laser Range Data. Springer International Publishing, Cham (2016) 735–747

[11] Carballo, A., Ohya, A., Yuta, S.: People detection using range and intensity data from multi-layered laser range finders. In: 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems. (2010) 5849–5854

[12] Mozos, O.M., Kurazume, R., Hasegawa, T.: Multi-part people detection using 2D range data. International Journal of Social Robotics **2** (2010) 31–40

[13] Spinello, L., Arras, K.O., Triebel, R., Siegwart, R.: A layered approach to people detection in 3D range data. In: Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence. AAAI'10, AAAI Press (2010) 1625–1630

[14] Yücel, Z., Ikeda, T., Miyashita, T., Hagita, N.: Identification of mobile entities based on trajectory and shape information. In: Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on. (2011) 3589–3594

[15] Beyer, L., Hermans, A., Leibe, B.: Drow: Real-time deep learning-based wheelchair detection in 2-D range data. IEEE Robotics and Automation Letters **2** (2017) 585–592

[16] Torr, P., Zisserman, A.: Mlesac: A new robust estimator with application to estimating image geometry. Computer Vision and Image Understanding **78** (2000) 138 – 156

[17] Rusu, R.B.: Semantic 3D Object Maps for Everyday Manipulation in Human Living Environments. PhD thesis, Computer Science department, Technische Universitaet Muenchen, Germany (2009)

[18] Cooley, J., Lewis, P., Welch, P.: The finite Fourier transform. IEEE Transactions on Audio and Electroacoustics **17** (1969) 77–85

[19] Licsar, A., Sziranyi, T.: User-adaptive hand gesture recognition system with interactive training. Image and Vision Computing **23** (2005) 1102 – 1114

[20] Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., Chen, T.: Recent advances in convolutional neural networks. Pattern Recognition (2017)

[21] Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. Second edn. Springer New York Inc., New York, NY, USA (2008)

[22] 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)

[23] Varga, D., Sziranyi, T.: Robust real-time pedestrian detection in surveillance videos. Journal of Ambient Intelligence and Humanized Computing **8** (2017) 79–85

[24] Chen, X., Ma, H., Wan, J., Li, B., Xia, T.: Multi-view 3D object detection network for autonomous driving. In: CVPR. (2017)