Ninth Hungarian Conference on Computer Graphics and Geometry, Budapest, 2018

Városi objektum felismerés mindösszesen néhány LIDAR szkennelési síkból

Z. Rozsa,[12] and T. Sziranyi[12]

[1] Department of Material Handling and Logistic Systems, Budapest University of Technology and Economics, Budapest, Hungary
[2] Research Institute for Computer Science and Control (MTA SZTAKI), Budapest, Hungary

Abstract

A LIDAR szenzorok tárgy és szabadterület detekciós lehetőségükkel szerves részét képezik a mai intelligens járműveknek és közlekedési rendszereknek. Ebben a tanulmányban egy felismerő algoritmust javasolunk, amely olyan LIDAR-ok esetében is alkalmazható, ahol csak néhány szkennelési sík áll rendelkezésre. Ez a módszer az elérhető 3D információ mellett a felismerendő alakzat időbeni változását is figyelembe veszi. A módszert több tízezer mintán validáltuk publikus adatbázison.

1. 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 their limited capability. 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[†] or Velodyne VLP-16[‡]) or even with only one (e.g., SICK LMS5xx series[§]). Dealing with LIDARs with many planes (e.g. Velodyne HDL-64[¶]), 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] we proposed a soulution for Automated Guided Vehicles, where we made a 3D reconstruction 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, if we consider their change over time as a feature. Increasing number of scan planes increase the recognition probability as well.

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

The contribution of the paper:

- New approach for description of plane curves.
- Object representation as set of time varying plane curves.
- Propose voting scheme in order to increase recognition probability.
- Offer solution to recognition problem caused by far objects in case of LIDARs and also in case of any objects of few plane LIDARs.

2. Related works

The related literature mainly corresponds to recognition of objects realized with LIDAR sensors having one or only few planes. Methods working on 3D LIDARs have the potential for the classification of several object classes because

[†] https://www.sick.com/us/en/detection-and-ranging-solutions/3d-lidar-sensors/ld-mrs/c/g91913
[‡] http://velodynelidar.com/vlp-16.html
[§] https://www.sick.com/us/en/detection-and-ranging-solutions/2d-lidar-sensors/lms5xx/c/g179651
[¶] http://velodynelidar.com/hdl-64e.html

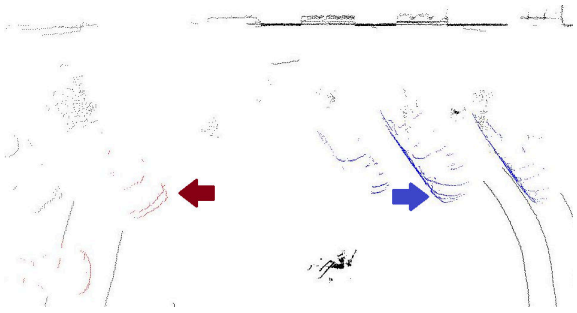


Figure 1: *Velodyne VLP16 sequence. Cars represented only with 2-3 detection plane (cannot be treated as point clouds) are marked with red points. Blue points corresponds to objects which can be treated as point clouds for classification (in sense of extension and continuity of scan planes).*

of having more information in case of separated 2D LIDAR segments. Works like [3] and [4] use 2D or 3D Convulutional networks to classify require point clouds as input. Compared to this, in earlier work [1], we have already proposed solution 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 it is composed from only a few unconnected 2D planar curves, so a combined approach is proposed.

The first applications related to object detection [5] and tracking [6] 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. [7] uses 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 was 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 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. 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, where the data was 3D point cloud. [14] applied motion characteristics to identify humans with baby cart, shopping cart or wheel chairs.

Summarizing, classification methods based on one or few planar scans are most of the time uses tens of geometrical features and Adaboost method to build a strong classifier or neural network ([2], [15]). They rarely use time-variant information ([16]), and also do not use the information provided via multiple plans (only for searching specific body parts), and to the best of our knowledge, these two have never been utilized simultaneously. A few classes are considered for detection. Most of the time these methods are applied for the classification of indoor objects scanned with indoor sensors with limited range (they also mostly depend on range and angular resolution of the sensor). The tests are executed on a few thousands of samples [2]. Compared to these, we list here the main advantages of our method:

- We propose a method extends the classification possibility from few layer LIDARs and also far field classification of 3D LIDARs with utilizing both time-varying shape and multiple plan information.
- Our method designed for outdoor objects and to be invariant of the sensor.
- It is model-free, we do not restrict it by assuming any relation between the sensor planes.
- We evaluate test results from ten thousands of samples.

3. The 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.

3.1. Preprocessing

Here, we list known methods that we used in our experiments:

- Registering consecutive frames: Iterative Closest Point (ICP) [17].
- Ground detection: M-estimator SAmple Consensus (MSAC) Plane fitting [18].
- Object detection: Euclidean cluster extraction [19] with distance varying neighborhood radius.
- Change detection: M3C2 distance for determining points with significant change [20].
- Moving object detection: Objects which have many points with significant change (their percentage reaches a given treshold) are considered as moving objects.
- Tracking of these moving objects: based on their location, extension and orientation.
- Objects are separated to plane curves, which are continuously matched in the consecutive frames.

From these steps change detection, moving object detection and tracking are optional, for stationary objects it is not necessary. Note that: In perspective of the LIDAR sensor segments of stationary objects will vary their shape because of the viewpoint change. Illustration of these processing steps can be seen on Fig. 2.

3.2. Descriptor and classification

Here, we assume that objects are represented by plane curves tracked through several frames. In our experiments we used a $f*(n+6)$ matrix as a descriptor of LIDAR segment. Here f is the number of frames we are tracking through the segment and n is the number of Fourier components we use (it determines the minimum number of points which can construct a segment). In the following it will be explained how it is composed.

3.2.1. Fourier descriptor

Instead of extracting geometric features from curves we utilize descriptor which can be used to reconstruct the curve exactly [21]. Fourier descriptor is applicable on closed contours, we construct a closed contour from the segment by adding to it its original points in reverse order [22]. 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.

3.2.2. 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.

3.2.3. Time varying shape

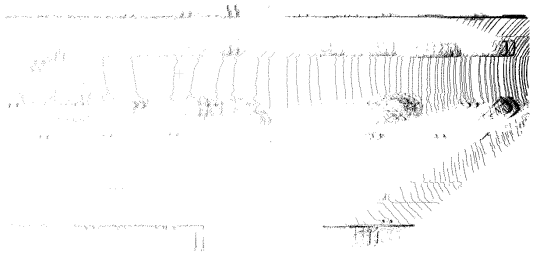
The values of the two previous section are saved through consecutive frames, these will form the rows of our descriptor matrix. The descriptor matrix is illustrated on Fig. 4.

3.2.4. Classification

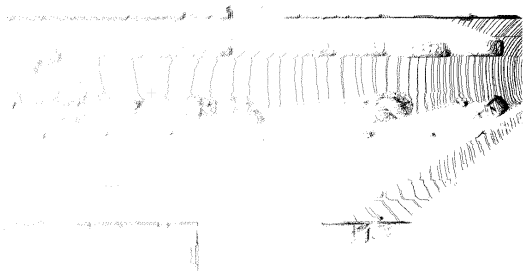
For the classification of the objects we use a Convolutional Neural Network (CNN) [23]. The network architecture we used can be seen in Fig. 3.

3.2.5. Voting

We applied a simple voting scheme to achieve the final decision in the case when an object was build up from multiple planar curves.



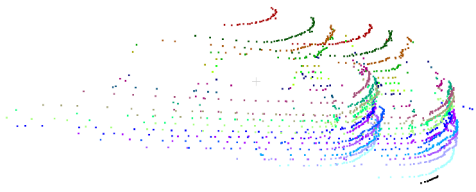
(a) Two consecutive frames without registration



(b) Two consecutive frames registered

(c) Change detection (significant changes marked with red) on one frame without ground

(d) Detected objects



(e) Segments of car are matched on two consecutive frames (Same color indicates the match). Note that: for illustration purposes we chosen an object with several segments, however the method was designed primarily for objects with only a few of those.

Figure 2: *Example of preprocessing steps on KITTI tracking database*

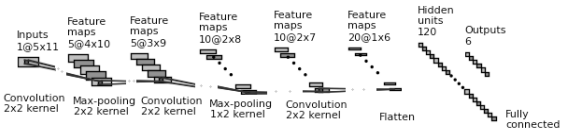
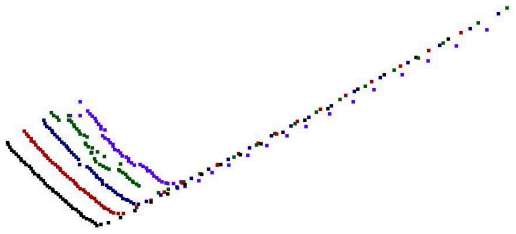


Figure 3: *Network architecture: all convolutional layers are followed by ReLUs and the fully-connected layer is followed by a softmaxlayer not illustrated in the scheme.*



(a) Segments of car (Purple: Frame n, Green: Frame n-1, Blue: Frame n-2, Red: Frame n-3, Black: Frame n-4)

	Frame n	Frame n-1	Frame n-2	Frame n-3	Frame n-4
FD1	0.5882	0.5885	0.6228	0.6396	0.5616
FD2	0.3791	0.3778	0.4107	0.4078	0.3325
FD3	0.2693	0.2662	0.2831	0.2712	0.2376
FD4	0.1794	0.1763	0.1755	0.1639	0.1649
FD5	0.0996	0.0953	0.0911	0.0944	0.1111
mean(z)	0.6807	0.6946	0.7112	0.7247	0.6374
std(z)	0.0611	0.0606	0.0656	0.0648	0.0484
mean(r)	12.8307	13.0329	13.2751	13.4702	13.5801
std(r)	0.8878	0.8818	0.9543	0.9434	0.7830
mean(I)	0.4256	0.4243	0.4178	0.5896	0.4195
std(I)	0.1069	0.1147	0.1169	0.3084	0.2418

(b) 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 origo and I means intensity)

Figure 4: *Example of description of a vehicle segment from 5 consecutive frames*

4. Test results

We conducted our proof of concept tests in the training set of the KITTI tracking database [24]; here it is guaranteed that we can have information of an object through at least several frames. In this set labeled objects are annotated through different number of frames in 21 sequences. It allowed 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 category (car, van, truck, pedestrian, person sitting, cyclist, tram, misc) which can be tracked through at least 5 frames and contains at least 1 segment in each of these frames with minimum 5 points. These objects were cutted out based on their annotated 3D bounding box and than 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. The categories of car and van and also pedestrian and person sitting are combined, because they are 'neighboring' categories. The accuracy on the train set is 89.63 %, 89.73 % on the validation and 89.69 % on the test set. Confusion matrices for all the samples are shown in Tables 2, 3, 4 and 5.

Table 1: *Confusion matrix by method proposed in [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	**73682**	2298	2274	986	162	1705	**0.909**
2	2315	**5757**	36	39	59	320	**0.675**
3	2471	39	**79742**	8473	0	679	**0.872**
4	944	30	8328	**2827**	1	215	**0.229**
5	164	50	0	2	**65**	8	**0.225**
6	1691	321	618	211	7	**737**	**0.206**
Recall	**0.907**	**0.678**	**0.876**	**0.226**	**0.221**	**0.201**	

Table 2: *Confusion matrix from a single planar curve with the 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	**79969**	1851	544	781	66	1664	**0.942**
2	394	**6429**	1	2	13	89	**0.928**
3	300	21	**87382**	4564	0	766	**0.939**
4	270	15	2915	**7111**	0	84	**0.684**
5	13	49	0	0	**215**	0	**0.776**
6	321	130	156	80	0	**1061**	**0.607**
Recall	**0.984**	**0.757**	**0.960**	**0.567**	**0.731**	**0.290**	

The confusion matrix in Table 2 shows that even one 2D contour can produce good initial results and Tables 3-5 show that the simple voting scheme we use is effective to increase accuracy of the classification. Detailed results divided by categories:

- The results of car and pedestrian categories are promising 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.

Table 3: *Confusion matrix resulted by voting planar curves of objects in one frame with the 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	**81075**	401	24	183	28	1521	**0.974**
2	144	**8045**	0	0	8	0	**0.982**
3	2	0	**90371**	4018	0	857	**0.945**
4	8	0	563	**8337**	0	36	**0.932**
5	7	49	0	0	**258**	0	**0.822**
6	31	0	40	0	0	**1250**	**0.946**
Recall	**0.998**	**0.947**	**0.993**	**0.665**	**0.878**	**0.341**	

Table 4: *Confusion matrix resulted by voting planar curves of objects in five frames with the proposed method. (1: Car, Van, 2: Truck, 3: Pedestrian and Person Sitting, 4: Cyclists, 5: Tram, 6: Misc)*

	1	2	3	4	5	6	Precision
1	**81058**	357	0	52	0	1457	**0.978**
2	205	**8138**	0	0	0	0	**0.975**
3	0	0	**90675**	2895	0	867	**0.960**
4	0	0	314	**9591**	0	17	**0.967**
5	0	0	0	0	**294**	0	**1.000**
6	4	0	9	0	0	**1323**	**0.990**
Recall	**0.997**	**0.958**	**0.997**	**0.765**	**1.000**	**0.361**	

- 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 only one 2D contour, but it has to be noted there were much less samples in this case, and also that by voting these performances can be significantly increased.
- The results on tram class are sufficient, however it should be noted that is not representative because of the very small number of samples.
- Finally, in case of misc category our proposed method did not performed well because of the variety of the objects hard to identify in 2D contours. Although as Table 5 shows, if such an object could be tracked in sufficient number of frames, it is likely that we can differentiate it from the other categories.

Fig. 5 shows examples of categorized plane curves.

Table 5: *Confusion matrix resulted by voting planar curves of objects in all the available frames with the 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	**81164**	288	0	0	0	1048	**0.984**
2	103	**8207**	0	0	0	0	**0.988**
3	0	0	**90998**	1957	0	841	**0.970**
4	0	0	0	**10581**	0	0	**1.000**
5	0	0	0	0	**294**	0	**1.000**
6	0	0	0	0	0	**1775**	**1.000**
Recall	**0.999**	**0.966**	**1.000**	**0.844**	**1.000**	**0.484**	

Table 6: *Confusion matrix resulted by voting from planar curves of far objects in one frame with the proposed method. (1: Car, Van and Truck, 2: Pedestrian and Person Sitting, 3: Cyclists, 4: Tram, 5: Misc)*

	1	2	3	4	5	Precision
1	**5185**	14	11	8	31	**0.988**
2	2	**842**	25	0	0	**0.969**
3	0	25	**150**	0	56	**0.649**
4	2	0	0	**27**	0	**0.931**
5	12	23	7	0	**26**	**0.382**
Recall	**0.997**	**0.931**	**0.777**	**0.771**	**0.230**	

The results are promising considering that pedestrian detection robust against about 30 % occlusion [25] on 2D images, and in a similar dataset [24] best detection results using both vision and LIDAR data [26] is about 90 %. We tested the method proposed in [8] used for pedestrian, cyclist and car detection in 2D LIDAR scans in our database. In this test we made a nearest neighbor classification based on euclidean distance to the train database built from width, range variance and intensity data. The results can be seen in Table 1 which can be compared to the results of our method Table 2. It can be seen that our method is superior in every aspect. Table 6 shows a separate evaluation with objects represented with maximum 4 scanning plane (the mean distance between the LIDAR and their center of mass is about 41 m). Truck category was added to Car and Van category in this case, because they show much similarity in the far field in point of view of the LIDAR.

(a) Car (b) Pedestrian

Figure 5: *Examples of the KITTI database: the colors indicates the output category of the algorithm (Red - Pedestrian, Purple - Cyclist, Blue - Car, Green - Misc). Note that: for illustration purposes we choosed above object with several segments, however the method was designed primarily for objects with only a few of those.*

5. Conclusion

In the paper we proposed a 2D recognition method exploiting time varying information, using additional 3D information if available. This method is designed to solve the recognition problem of far objects from LIDAR clouds or the general recognition problem for 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. To the best of our knowledge there is no similar method in the literature. However, we compared it to a method used for object detection in 2D LIDAR clouds, and our one is proved to be superior. We suggest to use it as extension to 3D recognition methods on environment they cannot process. In the future we would like to combine such a method with our one and also would like to investigate the influence of curve representation and the number of tracked frames.

Acknowledgment

This work was supported by the Hungarian Scientific Research Fund (No. OTKA/NKFIH 120499)

References

1. Z. Rozsa and T. Sziranyi, "Obstacle prediction for automated guided vehicles based on point clouds measured by tilted LIDAR sensor," *IEEE Transactions on Intelligent Trasnportation Systems*, 2018, in press. 1, 2
2. L. Kurnianggoro and K. H. Jo, "Object classification for LIDAR data using encoded features," in *2017 10th International Conference on Human System Interactions (HSI)*, July 2017, pp. 49–53. 1, 2
3. A. Borcs, B. Nagy, and C. Benedek, "Instant object detection in Lidar point clouds," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 7, pp. 992–996, July 2017. 2
4. D. Maturana and S. Scherer, "VoxNet: A 3D Convolutional Neural Network for Real-Time Object Recognition," in *IROS*, 2015. 2
5. K. O. Arras, O. M. Mozos, and W. Burgard, "Using boosted features for the detection of people in 2D range data," in *Proceedings 2007 IEEE International Conference on Robotics and Automation*, April 2007, pp. 3402–3407. 2
6. A. Fod, A. Howard, and M. A. J. Mataric, "A laser-based people tracker," in *Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No.02CH37292)*, vol. 3, 2002, pp. 3024–3029. 2
7. M. Lee, S. Hur, and Y. Park, "Obstacle classification method based on 2d lidar database," *Pattern Recognition Letters*, vol. 8, no. 8, pp. 1442–1446, 2014. 2
8. ——, "An obstacle classification method using multi-feature comparison based on 2D lidar database," in *2015 12th International Conference on Information Technology - New Generations*, April 2015, pp. 674–679. 2, 4, 5
9. F. Galip, M. H. Sharif, M. Caputcu, and S. Uyaver, "Recognition of objects from laser scanned data points using SVM," in *2016 First International Conference on Multimedia and Image Processing (ICMIP)*, June 2016, pp. 28–35. 2
10. C. Weinrich, T. Wengefeld, M. Volkhardt, A. Scheidig, and H.-M. Gross, *Generic Distance-Invariant Features for Detecting People with Walking Aid in 2D Laser Range Data*. Cham: Springer International Publishing, 2016, pp. 735–747. 2
11. A. Carballo, A. Ohya, and S. Yuta, "People detection using range and intensity data from multi-layered laser range finders," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct 2010, pp. 5849–5854. 2
12. O. M. Mozos, R. Kurazume, and T. Hasegawa, "Multi-part people detection using 2D range data," *International Journal of Social Robotics*, vol. 2, no. 1, pp. 31–40, Mar 2010. 2
13. L. Spinello, K. O. Arras, R. Triebel, and R. Siegwart, "A layered approach to people detection in 3D range data," in *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence*, ser. AAAI'10. AAAI Press, 2010, pp. 1625–1630. 2
14. Z. Yücel, T. Ikeda, T. Miyashita, and N. Hagita, "Identification of mobile entities based on trajectory and shape information," in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, 2011, pp. 3589–3594. 2
15. L. Beyer, A. Hermans, and B. Leibe, "Drow: Real-time

deep learning-based wheelchair detection in 2-D range data," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 585–592, April 2017. 2

16. B. Qin, Z. J. Chong, S. H. Soh, T. Bandyopadhyay, M. H. Ang, E. Frazzoli, and D. Rus, *A Spatial-Temporal Approach for Moving Object Recognition with 2D LIDAR*. Cham: Springer International Publishing, 2016, pp. 807–820. 2

17. P. Besl and N. D. McKay, "A method for registration of 3-D shapes," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 14, no. 2, pp. 239–256, Feb 1992. 2

18. P. Torr and A. Zisserman, "Mlesac: A new robust estimator with application to estimating image geometry," *Computer Vision and Image Understanding*, vol. 78, no. 1, pp. 138 – 156, 2000. 2

19. R. B. Rusu, "Semantic 3D object maps for everyday manipulation in human living environments," Ph.D. dissertation, Computer Science department, Technische Universitaet Muenchen, Germany, October 2009. 2

20. D. Lague, N. Brodu, and J. Leroux, "Accurate 3D comparison of complex topography with terrestrial laser scanner: Application to the rangitikei canyon (n-z)," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 82, no. Supplement C, pp. 10 – 26, 2013. 2

21. J. Cooley, P. Lewis, and P. Welch, "The finite Fourier transform," *IEEE Transactions on Audio and Electroacoustics*, vol. 17, no. 2, pp. 77–85, Jun 1969. 3

22. A. Licsar and T. Sziranyi, "User-adaptive hand gesture recognition system with interactive training," *Image and Vision Computing*, vol. 23, no. 12, pp. 1102 – 1114, 2005. 3

23. J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, and T. Chen, "Recent advances in convolutional neural networks," *Pattern Recognition*, 2017. 3

24. A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the KITTI vision benchmark suite," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012. 4, 5

25. D. Varga and T. Sziranyi, "Robust real-time pedestrian detection in surveillance videos," *Journal of Ambient Intelligence and Humanized Computing*, vol. 8, no. 1, pp. 79–85, Feb 2017. 5

26. X. Chen, H. Ma, J. Wan, B. Li, and T. Xia, "Multi-view 3D object detection network for autonomous driving," in *CVPR*, 2017. 5