2014 IEEE International Conference on Robotics & Automation (ICRA)
Hong Kong Convention and Exhibition Center
May 31 - June 7, 2014. Hong Kong, China

Robust Vehicle Detection using 3D Lidar under Complex Urban Environment

Jian Cheng, Zhiyu Xiang※, Teng Cao and Jilin Liu
Department of Information Science & Electronic Engineering
Zhejiang University, Hangzhou, 310027, China
21131083@zju.edu.cn

***Abstract*—Robust vehicle detection is one of the key task for autonomous vehicle under the complex urban environment. Using 3D Lidar, the difficulty of the task lies in that the appearance of a vehicle in the sparse range data changes greatly with the distance, the angle of view, as well as occlusions. In this paper we present a new algorithm to detect vehicles using the finely segmented 3D object points. In segmentation, RGLOS (Ring Gradient based Local Optimal Segmentation) algorithm is proposed. Instead of processing in the grid map, the point-wise segmentation method keeps the connection between points and is able to extract object points in far distance. Using the local optimal ground height, it produces much more correct object points with less wrong ground points. In feature extraction stage, three novel features: position-related shape, object height along length and reflective intensity histogram are proposed. Finally the kernel based SVM is used to finish the classification task. Experiments are carried out using the real data acquired from urban environment. The results demonstrate the superior performance comparing with previous methods, thanks to the improved segmentation and new features.**

I. INTRODUCTION

With the development of artificial intelligence, sensors and control systems technology, the research of autonomous vehicle has made great progress[1][2]. Autonomous vehicles can replace human driving in the future driver systems, it can be used to help solving urban traffic problems and reducing traffic accidents. In urban complex environments, autonomous vehicle has to interact with other traffic participants, of which the main are variety of vehicles. This calls for the robust vehicle detecting abilities, which can be used to predict the movement of the vehicles and prevent some dangerous in advance.

For the detecting task, the difficulty lies in that the appearance of the object changes with the illumination, the distance to the object, the angle of view, occlusions, etc. The noisy data provided from sensors are also unreliable. 3D Laser radar (Lidar), with the advantage of high-precision and little interference to the illumination, is becoming more and more popular in autonomous vehicles. They are often used to detect obstacles and classify objects. The common processing flows are: data acquisition, segmentation, feature extraction and classification. Segmentation is an important and fundamental step in the algorithm. Some representative algorithms [11] [12] carry out segmentation in a 2D grid map and do object Classification with 3D point clouds. However, grid map based segmentation tends to lose object points in long distance due to the sparse lidar data. Meanwhile, the low resolution the grid map used will also induce too much ground points into the object cluster, leading to unstable classification of the object.

There are mainly two classification methods used in the detection: model based (Model matching) [6] and feature based classification (Machine learning classifier) [4] [5] [7]. In this paper, we propose three novel features to better distinguish vehicles and others, i.e., position-related shape, object height along length and reflective intensity histogram.

This paper presents a robust vehicle detection algorithm using 3D Lidar data. Static vehicles can also be detected since the algorithm is carried out within one frame of data, using no tracking information. Our contributions are two-folds: (1) A RGLOS (Ring Gradient based Local Optimal Segmentation) algorithm is proposed. Instead of processing in the grid map, the point-wise segmentation method keeps the connection between points and is able to extract object points in far distance. Using the local optimal ground height, it produces much more correct object points with less wrong ground points; (2) Three novel features are presented, improving the detecting rate greatly.

This paper is organized as follows: Section II briefly summarize some of previous related work. Section III presents the detailed algorithm for vehicle detection. In Section IV experimental results and some analysis are provided. Section V concludes the paper.

II. RELATED WORK

In recent years, researchers had deep study in vehicle and pedestrian detection using laser radar [2][3][4][5]. Anna Petrovskaya et al. [6] proposed a vehicle detection and tracking algorithm combining geometric and motion model of vehicle in consecutive frames. In their application, they used simple geometric model and the performance of detection mainly depends on the motion information. In other words, they can hardly detect a static vehicle parking along the roadside. M.Himmelsbach et al. [14] strengthened the geometric features of 3D point clouds by using the overall object volume, the reflective intensity of the object and eigenvalues along height. Asma Azim et al. [7]

※ corresponding author and e-mail: xiangzy@zju.edu.cn

employed the contour information, i.e., the length to height ratio and width to height ratio of the object bounding box, to detect vehicles and pedestrian. Those features work well under ideal environments while performs poor under crowd traffic where occlusions from other objects and itself happens frequently. The occlusion changes the characteristics of vehicles' shape in the 3D measurement. Luciano Spinello et al. [9] use three-dimensional laser data to detect the pedestrian, extracting the geometry features of the different layers of the body parts along height and calculate each layer's statistical features. Recent years, some researches proposed to use reflective intensity features [10] [11] such as the distribution of reflect intensity to recognize pedestrians. Using the single line Airborne LiDAR, Wei Yao et al. [15] proposed a framework to extract vehicles and analyze vehicles motion in urban areas. An adaptive mean shift analysis method is presented to facilitate the task of vehicle extraction and SVM is used to classify the objects, Features such as area, vertical position and vertical range are employed in the algorithm.

III. Algorithm Description

Overview of process flow is illustrated in Figure 1, and the detail process will be explained below.

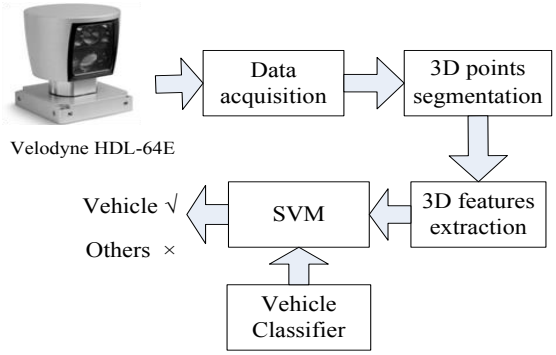


Figure 1. Algorithm flow chart

A. Data acquisition

The Velodyne-HDL-64E is mounted on the top of the autonomous vehicle with unit spins to 600 RPM(10hz) to gather data. The unit inherently delivers a 360-degree horizontal field of view (FOV) and a 26.8 degree vertical FOV. The raw data are received by UDP Ethernet packets. The raw data contains rotational degree, distance and intensity information. Then we make use of the extrinsic calibration and unique calibration file (db.XML) to convert 3D point clouds to the vehicle coordinate system.

B. Segmentation

Data Segmentation is a fundamental step for further accurate vehicle detection. So far, most of the segmentation methods get obstacle points by computing the maximum height difference on each occupancy grid map [8]. In practice, though the occupancy grid map can quickly get the obstacle grids, it has some drawbacks: (1) all of the point within the obstacle grid will be considered as obstacle points, leading to some ground points wrongly classified into objects. (2) In long distance, with the very sparse laser data, some obstacle points tends to be missed due to insufficient laser point in the grid; (3) the low grid resolution tends to merge the neighboring object into one, which leads to false classification results. On the other hand, traditional gradient –based methods are sensitive to object edges and tend to lose the laser points inside the object.

Here we proposed a ring gradient based local optimal segmentation (RGLOS) method to extract and clustering the true object points. The RGLOS segmentation algorithm is composed of four steps: (1) Ring gradient based object point extraction and clustering; (2) Filtering out false positives; (3) Object point retrieving; (4) Clustering.

(1) Ring gradient based object point extraction and clustering.

We calculate lidar adjacent ring gradient to get initial object points. In this step, the mesh structure is easily constructed using the Lidar's scanning mode and each pair of neighboring points along the same ring (scan line) and the next ring are used to compute the ring gradients, as shown in Figure 2. A minimum distance between the points is set to decrease the affection of measurement noise. Those with gradients exceeding a preset threshold are marked as object points. Instead of computing on the grid, the point-wise gradient computing reserves the high resolution that original data has. This will greatly benefit the following clustering process, helping to produce a clear boundary between the neighboring objects. Meanwhile, the connectedness of the mesh guarantees the effective segmentation of object points in the long distance where very sparse data exists in each grid.

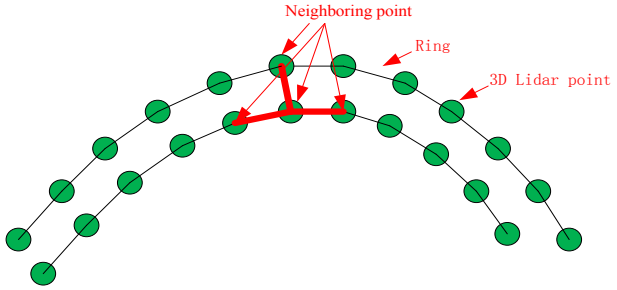


Figure 2. Lidar's scanning Ring and neighboring point to compute the ring gradient.

(2) Filtering out false positives.

The adjacent ring gradient method will produce some false positives, for example, some ground points with relatively large local gradient and small height. It is necessary to filter out them. Objects are regarded as false positive if the maximum height difference within the cluster is below a threshold.

(3) Object point retrieving.

After step (1) and (2), the segmented objects are highly believable, leading to be "seed points" on the corresponding objects. The ring gradient method is good to extract vertical surfaces of the object while insensitive to the parallel surface to the ground. With the "seed points" at hand, the local ground height can be easily obtained by averaging the neighboring points around the boundary of the "seed points". Unlike a global ground height threshold, this local ground height are local optimal and can be safely used as a reliable threshold to retrieve the missing points inside the object. The

retrieving process is done by iteratively searching the grids within and around the "object grid" and adding points whose height is above the local ground height threshold.

(4) Clustering.

Due to the sparsity and noise of the laser data, the initial clustered objects tends be separated into several small parts. It is necessary to merge them back to obtain a complete description of the objects. We use a simple and efficient distance-based region labeling algorithm [11] to finish this task. After clustering, the cluster labels as well as their bounding box can be obtained.

Figure 3 shows the segmentation result from different method mentioned above. As expected, the result from RGLOS method preserves the object points in long distance, while get rid of some wrong ground points appeared in the bottom part of (b). Meanwhile, it retrieves the most object point comparing with (b) and (c). To make it more clear, the segmented data of the car with blue rectangles in Figure 3 are enlarged and shown in Figure 4. Obviously the results from (d) are the best, with most points on the object and least points on the ground. The better segmentation results will benefit the following classification step greatly.

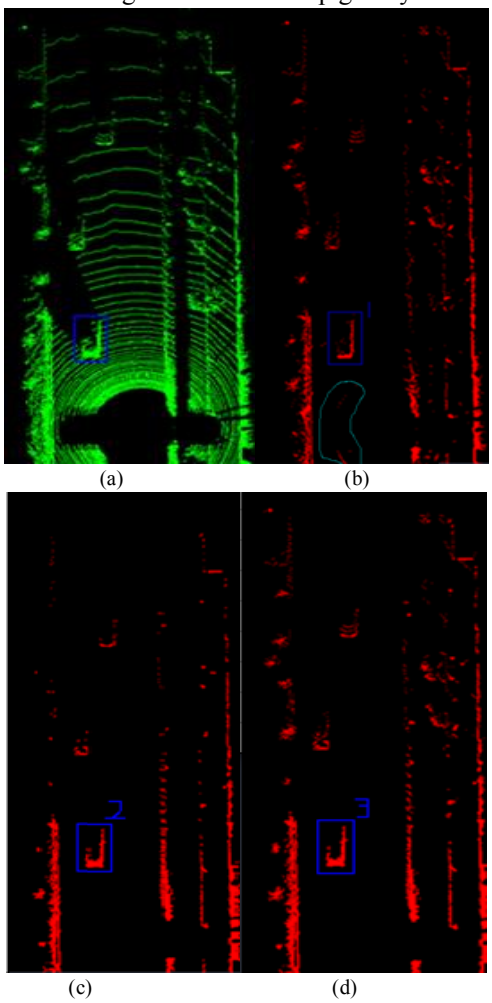


Figure 3. Raw lidar data and segmentation results. (a) 64 HDL-Raw data, (b) Results from pure ring gradient method; (c) Results from grid method and (d) Results from our RGLOS method.

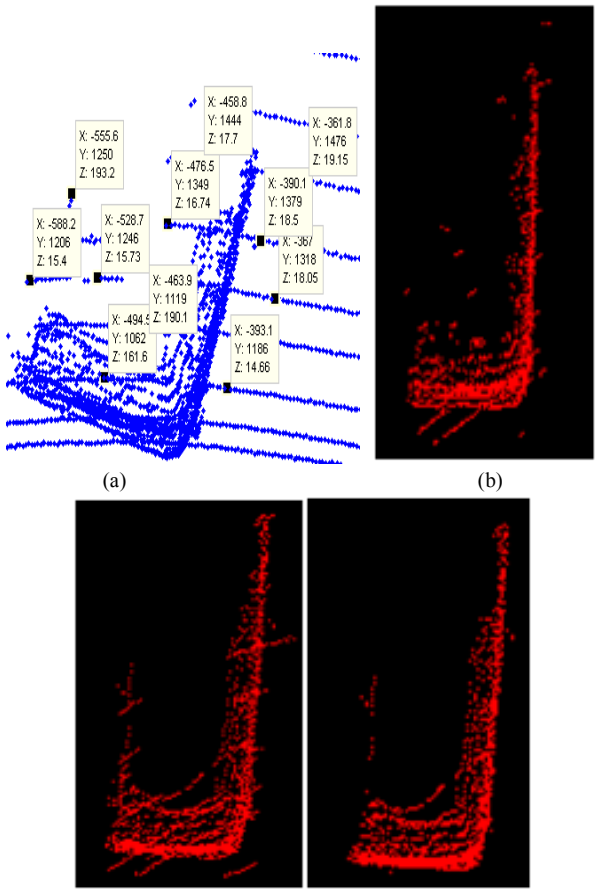


Figure 4. Zoom in view of the car's data in Figure 3. (a) is the real scanline Marked with 3D coordinates, (b), (c) and (d) corresponds blue rectangles marked 1,2 and 3 in Figure 3

C. *Feature extraction*

Before this step, we have obtained some vehicle candidates represented by bounding boxes. Each box's data contains object's 3D points and intensity information, as well as the bounding box's parameters (i.e., center, width and length). In order to improve the detection performance, we propose three novel features below.

(1) Position-related shape

As we all know, the different orientation and angle of view will result in different appearance characteristic of objects. As shown in Figure 5, for the same vehicle, different angle of view corresponds to drastic different point distribution. Same phenomenon happens to the change of objects' orientation. Therefore we need to construct a position-related shape feature to well compensate the variance of the position (angle of view and orientation). In this feature, both shape and position information is included. Shape features are the width to length ratio and width to height ratio of the bounding box. The position information: (a) the distance to object; (b) angle of view and (c) orientation characteristics. As shown in Figure 5, θ represents the angle of view of the object and β the object's orientation. θ is defined in the range of [0,360], and β is in [0,180], both with the 0 from the x-axis.

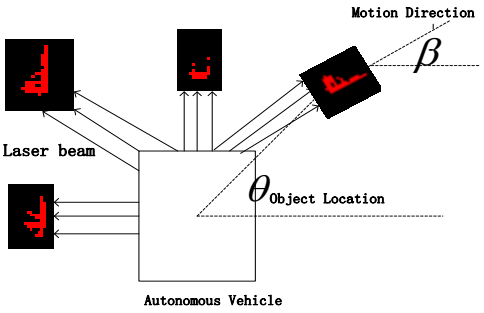


Figure 5. Describe object position-related features.

(2) Object height along length

The geometric description of bounding boxes in x-y plane is not enough to discriminate vehicle and others. We consider the vehicle's height variation along the length direction. The vehicle contour can be described that we make the bounding box into ten parts along length direction and record each part's average height into one feature vector. See Figure 6.

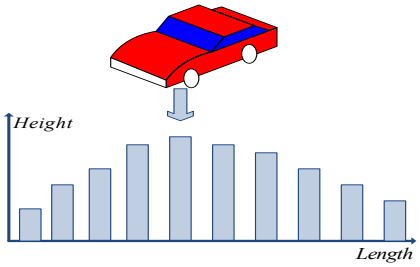


Figure 6. Object height along length

(3) Reflective intensity histogram

The reflective intensity of object has some relationship with the object material [17]. Intensity variation can be seen as a stable feature invariant to object shape. The laser reflective intensity values are in the range of [0,255]. Figure 7 shows an example distribution of reflective intensity for a vehicle and a parterre. It can be seen from the curve that the vehicle's reflection are mainly focus on the lower part of the intensity values, while the parterre in a higher part. The characteristic intensity distribution for the vehicle may caused by the absorption and mirror reflection of the vehicle body. In the reflective intensity histogram, we use 25 bins each of which has an interval of 10 to model the intensity distribution.

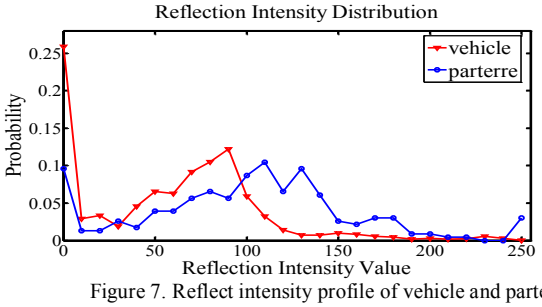


Figure 7. Reflect intensity profile of vehicle and parterre

Besides the proposed three features, other five features are also used as the input to the SVM. Table I lists all the features used to describe characteristics of the vehicle obstacle in our detecting system. The final full feature vector has a dimension of 59.

In Table I, f5 is the inertia tensor matrix used to capture the overall distribution of all points, using six independent elements [4], as shown in (1).

$$M=\begin{bmatrix} \sum_{i=1}^{n}(y_i^2+x_i^2) & -\sum_{i=1}^{n}x_iy_i & -\sum_{i=1}^{n}x_iz_i \\ -\sum_{i=1}^{n}x_iy_i & \sum_{i=1}^{n}x^2+z^2 & -\sum_{i=1}^{n}y_iz_i \\ -\sum_{i=1}^{n}x_iz_i & -\sum_{i=1}^{n}y_iz_i & \sum_{i=1}^{n}(x^2{}_i+y_i^2) \end{bmatrix}. \quad (1)$$

f6 is covariance matrix with six independent elements as features [4], as shown in (2)(3).

$$\mathrm{cov}(x,y)=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{n-1}. \quad (2)$$

$$C=\begin{pmatrix} \mathrm{cov}(x,x) & \mathrm{cov}(x,y) & \mathrm{cov}(x,z) \\ \mathrm{cov}(y,x) & \mathrm{cov}(y,y) & \mathrm{cov}(y,z) \\ \mathrm{cov}(z,x) & \mathrm{cov}(z,y) & \mathrm{cov}(z,z) \end{pmatrix}. \quad (3)$$

f7 [13][14] describes the eigen values of the covariance matrix. The eigenvalues are sorted according to value d1> d2> d3, and using equation (4) to normalize to the region of [0, 1]. The feature values of L1, L2, L3 are defined as (4):

$$L_1=d_1,\ L_2=d_1-d_2,\ L_3=d_2-d_3. \quad (4)$$

Table I features for vehicle detection

No.	Description	Dim
f1	Position-related shape	5
f2	Object height along length	10
f3	Reflective intensity histogram	25
f4	The number of 3D points	1
f5	The normalized moment of inertia tensor	6
f6	3D covariance matrix of bounding box	6
f7	Eigenvalues of 3D of covariance matrix	3
f8	Max intense, mean intense, covariance of intense	3

D. SVM classifier

Since the range of values between the different features vary greatly that will result the imbalance of weights, it is necessary to be normalized before training. Normalization can make the classifier converge faster and improve the classification performance. In scale of normalization for each component is shown in (5) [16]:

$$Scale(x)=\begin{Bmatrix} -1, & x<=\min \\ -1+\frac{2*(x-\min)}{\max-\min}, & \min<x<\max \\ 1 & else \end{Bmatrix}. \quad (5)$$

Taking into account that the radial basis function can handle with nonlinear problems, and radial basis function range is (01]. It will not generate an infinite value as a linear function. The RBF kernel has less parameter than the polynomial kernel. So we choose the RBF kernel in our system.

$$K(x,y)=e^{-\gamma\|x-y\|^2}. \quad (6)$$

Taking RBF as the kernel function, there are two main parameters to be decided: penalty factor and RBF function

parameter γ .We use grid optimization (grid-search) and cross-validation method in libsvm toolbox [16] to select the optimal parameters.

IV. Experimental Results

To evaluate the performance of our proposed method, we manually drove our autonomous vehicle in the urban environment and collected raw lidar data, as shown in Figure 8. Then we divided the data into two parts for training and testing respectively. The training samples are set up by hand-labeling the training set. Some of the positive and negative samples in the training set are shown in Figure 9. The positive samples are of different size and orientation, while negative samples are from flower beds, trees, etc. The size of the training and testing set are shown in Table II.

Figure 8. Traffic scene of real urban environment.

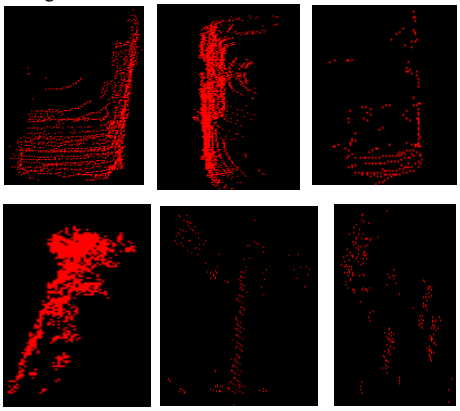


Figure 9. Typical hand-label sample: first row are vehicle positive samples and second row are negative samples.

Table II. Classification sample set

Data set	Total	Positive samples	Negative Samples
Training set	1250	420	830
Test set	1123	358	765

Detecting result corresponding to Figure 8 is shown in Figure 12. As expected, three vehicles are successfully detected among tens of candidates.

The ROC curve of the detection results are shown in Figure 10. For comparison, the detection result from the method in [14] is also presented. [14] used 60 dimension of features while ours use 59. Superior performance is obtained thanks to the fine segmentation results from the RGLOS method and three effective new features used for classification.

To further analysis the detection performance of each single feature in Table I, each of the features is used alone to construct a SVM and the ROC curves are shown in Figure 11. It can be seen from Figure 11 that all of the three features we proposed have better performance than the others. Among them, the reflective intensity histogram is the best, followed by position-related shape and object height along length. In particular, the position-related shape, which has only 5 dimensions, has much better performance than the covariance matrix eigenvalues and the inertia tensor. Meanwhile, further comparison between with and without 3 dimensional position information in feature f1 verifies the importance of the posture.

Another real time vehicle detection test is carried out under urban environment. With 50 consecutive frames where there are 82 real vehicles, 72 were correctly detected, resulting in a detection rate of 87.8%. Most of the false positives are in the distant range ahead of the Lidar, with too few scanning points to produce good detection performance.

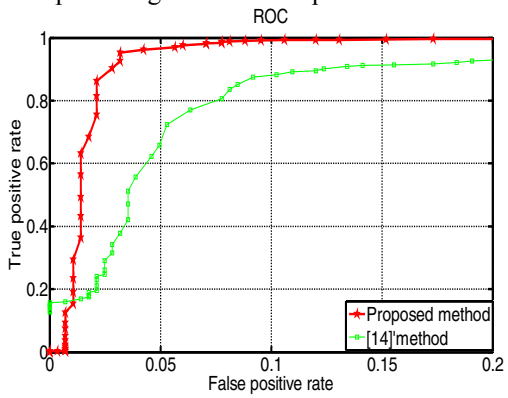


Figure10. Vehicle detection performance for combined feature

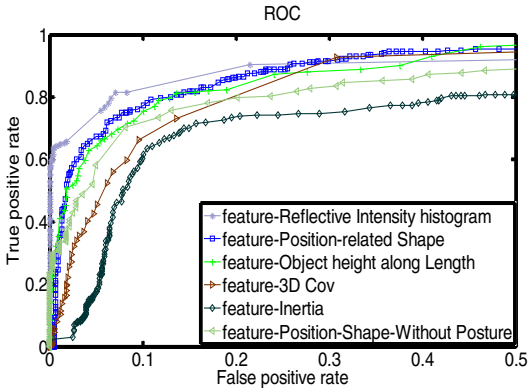
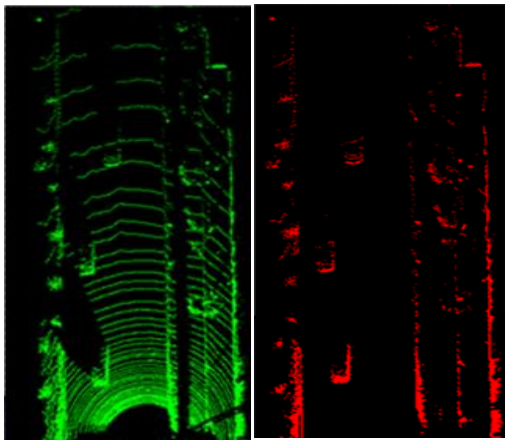
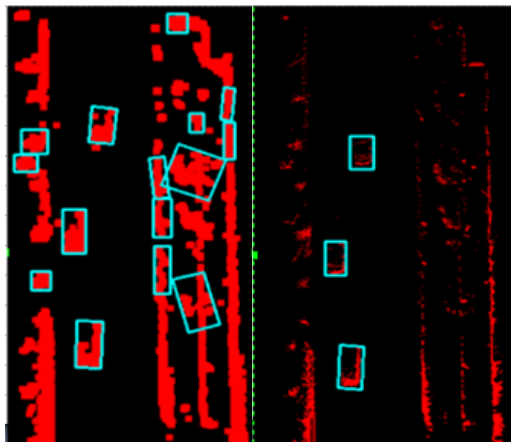


Figure 11. Vehicle detection performance for each feature



(a) (b)



(c) (d)

Figure 12. A typical detection scene and the results. (a) the raw data; (b) segmentation results from RGLOS; (c) candidate clustering results; (d) final detection results.

V. CONCLUSION

In this paper, we present an approach for reliable detection of vehicles under real complex unban environment. We proposed a RGLOS method to well segment the object points, leading to a better presentation of the candidate vehicles. Then three novel features are proposed to robustly classify the true vehicles. The better segmentation results and more effective features contribute greatly to the better detection performance. The experimental results demonstrate our success.

However, vehicles detection remains a challenge under crowded traffic environment. The heavy occlusion in the scene brings great difficulty to segmentation. In future research we will continue to improve the segmentation method and make our effort to find better features for traffic participants.

ACKNOWLEDGMENTS

The authors thank the National Science Foundation of China (NSFC) under Grant No. 61071219.

REFERENCES

[1] T.Luettel, M. Himmelsbach, and H.-J. Wuensche. “Autonomous ground vehicles: Concepts and a path to the future“. Proceedings of the IEEE, 100(Special Centennial Issue):1831-1839, 13 2012.

[2] Michael Montemerlo, Jan Becker, Suhrid Bhat, Hendrik Dahlkamp and Dmitri Dolgov, Scott Ettinger and Dirk Haehnel, “junior: The Stanford Entry in the Urban Challenge,”Journal of field robotics,pages 569C597,2008.

[3] Huijing Zhao, Quanshi Zhang, Masaki Chiba, RyosukeShibasaki, Jinshi Cui, Hongbin Zha, “Moving Object Classification using Horizontal Laser Scan Data,” 2009 IEEE International Conference on Robotics and Automation, pages 2424-2430, 2009.

[4] LuisE.Navarro-Serment,Christoph Mertz, andMartial Hebert, “Pedestrian Detection and Tracking Using Three-dimensional LADAR Data”, 2011 The International Journal of Robotics, pages 1516-1527,

[5] Nicolai Wojke, Marcel H¨aselich,Active Vision Group, AGAS Robotics,” Moving Vehicle Detection and Tracking in Unstructured Environments”, 2012 IEEE International Conference on Robotics and Automation, pages 3082-3087

[6] Anna Petrovskaya and Sebastian Thrun ,”Model based vehicle detection and tracking for autonomous urban driving”， Auton Robot (2009) 26: 123–139

[7] Asma Azim and Olivier Aycard,” Detection, Classification and Tracking of Moving Objects in a 3D Environment”, 2012 Intelligent Vehicles Symposium , Alcalá de Henares, Spain, June 3-7, pages 802-807, 2012

[8]S. Thrun, "Learning occupancy grid maps with forward sensor models," Auton. Robots, vol. 15, no. 2, pp. 111-127, 2003.

[9]Luciano Spinello, Kai O. Arras Rudolph Triebel Roland Siegwart, “A Layered Approach to People Detection in 3D Range Data,” Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI-10), pages 1625-1630, 2010.

[10]Kiyosumi Kidono, Takeo Miyasaka, Akihiro Watanabe,TakashiNaito and Jun Miura, “Pedestrian Recognition Using High-definition LIDAR“, 2011 IEEE Intelligent Vehicles Symposium (IV) Baden-Baden, Germany, June 5-9, 2011,pages 405-410, 2011.

[11]M. Himmelsbach and Felix v. Hundelshausen and H.-J. Wuensche,”Fast Segmentation of 3D Point Clouds for Ground Vehicles”, 2010 IEEE Intelligent Vehicles Symposium University of California, San Diego, CA, USA,June 21-24, 2010

[12]D. Morris, R. Hoffman, and S. McLean. Ladar-based vehicle detection and tracking in cluttered environments. In Proceedings of the 26th Army Science Conference, 2008.

[13] Jean-Francois Lalonde, Nicolas Vandapel,*Daniel F. Huber, and Martial Hebert, ”Natural Terrain Classification Using Three-Dimensional Ladar Data for Ground Robot Mobility”,Journal of Field Robotics 23(10), 839C861 (2006)

[14]M.Himmelsbach,T.Luettel,H.-J.Wuensche, ”Real-time Object Classification in 3D Point Clouds Using Point FeatureHistograms”The2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 994-1000, 2009.

[15]Yao W, Hinz S, Stilla U. Extraction and motion estimation of vehicles in single-pass airborne LiDAR data towards urban traffic analysis[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2011, 66(3): 260-271.

[16] Chih Chung Chang and Chih-Jen Lin , ”LIBSVM: A Library for Support Vector Machines”, Department of Computer Science National Taiwan University, Taipei, Taiwan , 2001.

[17]D.Williams, Methods of Experimental Physics, Academic Press, vol. 13, 1976.