Proceeding of the IEEE
International Conference on Robotics and Biomimetics
Dali, China, December 2019

An Efficient 3D Pedestrian Detector with Calibrated RGB Camera and 3D LiDAR

Lei Pang[1,2], Zhiqiang Cao[1,2], Junzhi Yu[1,3], Shuang Liang[1,2], Xuechao Chen[4], and Weimin Zhang[4]
[1]*State Key Lab Management and Control for Complex Systems, Institute of Automation, CAS, Beijing 100190, China*
[2]*University of Chinese Academy of Sciences, Beijing 100049, China*
[3]*Dept. Mech. Eng. Sci., BIC-ESAT, College of Engineering, Peking University, Beijing 100871, China*
[4]*BAICIRS, Beijing 100081, China; Intell. Robot. Inst., Beijing Inst. Technol., Beijing 100081, China*
{*panglei2015, zhiqiang.cao, junzhi.yu, liangshuang2017*}*@ia.ac.cn*, {*chenxuechao, zhwm*}*@bit.edu.cn*

***Abstract*— Pedestrian detection plays an important role in the environmental perception and autonomous navigation for robotics, which provides critical information for the safe operation in complex environments. In this paper, a 3D pedestrian detector with calibrated LiDAR and RGB Camera is proposed, which takes full advantage of the precise range information of 3D LiDAR scanner and semantic information acquired from RGB image. The proposed approach integrates the segmented object clusters of point cloud and 2D bounding boxes generated by a visual object detector. The point cloud is segmented by a three-step segmentation approach, which can segment point cloud with both high precision and high efficiency. By fusion multi-sensor information in the image domain, the proposed approach provides 3D information of pedestrians in both LiDAR and camera coordinate systems. Experiments was conducted to evaluate the performance of the proposed approach.**

***Index Terms*— LiDAR segmentation, visual detector, multi-sensor fusion, 3D pedestrian detection.**

I. Introduction

Pedestrian detection is necessary for robots to work in the human environments, which can further be utilized for collision avoidance, motion prediction, social behavior of robotics, etc. As is well-known, vision-based object detectors have achieved remarkable breakthroughs recently, which can provide multiple 2D bounding boxes of various of categories on a single image [1], [2]. However, precise 3D positions of pedestrians are required to perceive in many scenarios.

Due to the absence of range information of single-frame visual detector, multiple kinds of additional methods are used to implement vision-based 3D object detection. Li *et al.* proposed a 3D object detector GS3D with a single 2D image, which determines the 3D box on the basis of 2D box through a series of operations including 2D bounding box generation, orientation prediction, and 3D box refinement [3]. Approaches based on RGB-D sensors and stereo camera are commonly employed for 3D object detection for the convenience of capturing depth map [4], [5]. The performance of 3D object detectors based on RGB-D sensors and stereo cameras are easily affected by the quality of depth maps, which usually suffers from the variations of light condition and the limitation of measuring distance.

LiDAR scanners can provide long-range and high-precision distance measurements regardless of illumination, which has attracted more and more attention for 3D object detection. The point cloud captured by LiDAR scanners are usually segmented into meaningful clusters, which are then used to implemented object classification with hand-crafted features [6] and CNN-based features [7]. However, the ability of object recognition for LiDAR scanners, especially for those sensors with less laser beams, is restricted by the sparse information comprised of 3D coordinates and intensity. Hence, we draw more attention on object clustering of point cloud.

Ground segmentation is the basis for the further object clustering and other advanced applications. Discarding all points lower than a scalar threshold is a simple way for ground removal [8]. Thrun *et al.* presented a grid-based approach [9], which divides the grid into ground cells and non-ground cells according to the maximum absolute difference between the heights of points inside the cells. RANSAC-based plane-fitting approaches may satisfy more complex situations, which usually generate some seed points and then grows regions from those points according to predefined criteria [10], [11].

Point cloud segmentation focuses on clustering non-ground points into multiple clusters. Most real-time 3D segmentation methods use the spatial relation between points in a single frame for clustering without regard to semantic information or relationship between consecutive frames. Euclidean cluster extraction [12] is the commonly used clustering approach for point cloud, which is time-consuming. A coarse-to-fine segmentation approach based on occupied grid cells was presented by Behley *et al.*, where the precision depends on the cell size and under-segmentation always exists in z-axis [13]. Bogoslavskyi *et al.* proposed an efficient segmentation approach based on range image with a single angle condition [8], which suffers from over-segmentation when the scanner

*This work was supported in part by the Beijing Advanced Innovation Center for Intelligent Robots and Systems under Grant 2018IRS21, in part by the National Natural Science Foundation of China under Grants 61633017, 61633020, 61836015, and in part by the Key Research and Development Program of Shandong Province under Grant 2017CXGC0925.

978-1-7281-6321-5/19//$31.00 ©2019 IEEE

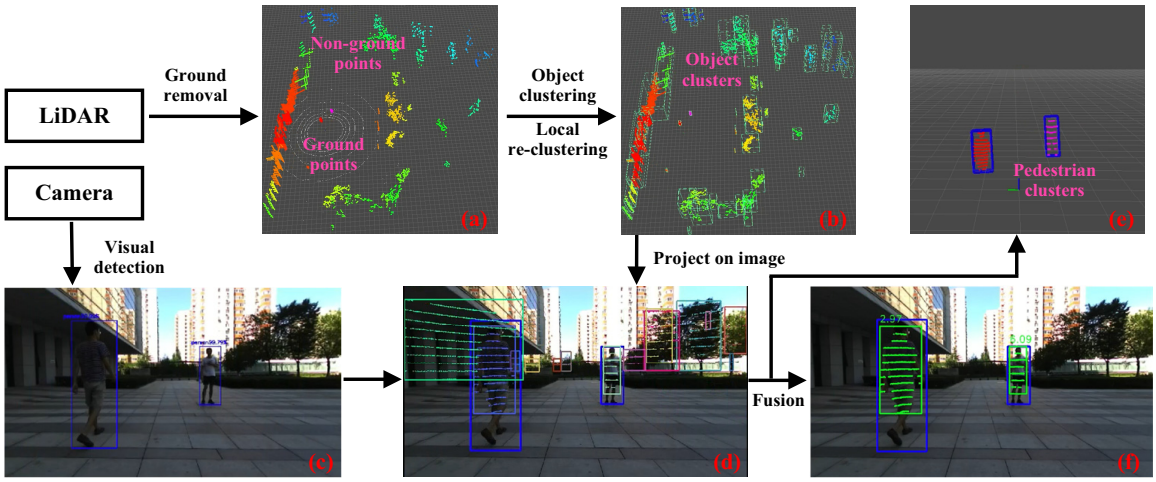

Fig. 1. The schematic diagram of the processes of proposed 3D pedestrian detector. (a) is the result after the process of ground removal, where the raw point cloud from LiDAR scanner are separated into two part, i.e. ground points and non-ground points. (b) demonstrates the object clusters generated by the processes of object clustering and local re-clustering, where each green bounding box denotes an object cluster. Blue bounding boxes in (c) are detection results of the visual detector. (d) presents the visual detection results (blue) and projected points of object clusters with corresponding bounding boxes (random colors). (e) and (f) are detection results obtained by the operation of sensor fusion; (e) presents results of pedestrian clusters in the LiDAR coordinate system; (f) gives bounding boxes of visual detector (blue), projected points with corresponding bounding boxes of pedestrian clusters (green).

locates closely to a wall.

Moreover, multi-sensor approaches are promising to obtain more comparative performance by efficiently integrating advantages of different sensors [14], [15]. In this paper, we focus on 3D pedestrian detection by integrating semantic information provided by visual detector and range information captured by 3D LiDAR scanner. A 3D LiDAR scanner and a RGB camera are employed, where the extrinsic matrix of multiple sensors are calculated by a multi-sensor calibration method. The proposed approach is composed of an agnostic segmentation approach for point cloud, a visual pedestrian detector, and a fusion approach. The point cloud are segmented by three steps including ground removal, object clustering, and local re-clustering. A real-time visual detector is used to compute 2D bounding boxes of pedestrians in image domain. Then the multi-sensor information is integrated in the image domain by considering spatial relationship between 2D bounding boxes generated by visual detector and projected bounding boxes of segmented clusters. The proposed approach outputs 3D information of pedestrians in both LiDAR and camera coordinate systems.

The remainder of this paper is structured as follows. The detailed approach is introduced in Section II. We present the verification experiments in Section III. Finally, Section IV concludes this paper with an outline of future work.

II. Proposed Approach

The proposed approach implements 3D pedestrian detection by fully utilizing the information captured from the calibrated 3D LiDAR scanner and RGB camera. The schematic diagram of the proposed approach in the Fig. 1. The input point cloud is firstly handled with the segmentation approach for sparse point cloud to obtain object clusters. Meanwhile, the visual detector is responsible for outputting 2D bounding boxes in the image domain. Then the spatial relationship between 2D bounding boxes from the image and 3D information of segmented clusters is integrated to obtain the 3D detection results in LiDAR coordinate system and camera coordinate system.

A. LiDAR Segmentation

An agnostic segmentation approach for point cloud is proposed to cluster raw point cloud captured by LiDAR scanner into multiple clusters. An input scan of point cloud $\{P\}$ is first classified as ground point set $\{P_g\}$ and non-ground point set $\{P_{ng}\}$ by the process of ground removal. Then non-ground points in $\{P_{ng}\}$ are clustered as multiple clusters $\{C\}$ belonging to different objects with the operations of object clustering and local re-clustering.

1) Ground removal: The ground points are classified with a plane mathematical model through multiple iterative operations [16]. A specified amount of points with lowest height are selected as initial seeds, which can reduce the computing time. The performance of ground fitting can be further improved by splitting ground plane into multiple parts and fitting them separately with separate plane models.

2) Object clustering: Based on the fact that point cloud is comprised of sparse and irregular points with cartesian coordinate (x, y, z), the point cloud need to be rearranged in a structured space, such as grids on the ground plane [13], range image [17], voxels [18], layered-based organization [16], etc. In this paper, the input points cloud is projected on

Algorithm 1: Clustering objects on the range image

Input: Range image I_r, point cloud $\{P_{ng}\}$
Output: Label image I_l, object set $\{C\}$

$Label = 1$;
$I_l \leftarrow 0 \parallel -1$ (ground points are set to -1);
while $I_l(u,v) : I_l$ **do**
  if $I_l(u,v) = 0$ **then**
    $queue.push(I_l(u,v))$;
    $C_{Label} \leftarrow p(u,v)$ and $I_l(r,c) = Label$;
    while *queue is not empty* **do**
      for $I_l(i,j)$ *: neighbors of queue.top()* **do**
        if $I_r(i,j)$ *satisfies condition (1)* **then**
          $I_l(i,j) = Label$;
          $Cluster \leftarrow p(i,j)$;
          $queue.push(I_l(i,j))$;
        end
      end
      $queue.pop()$;
    end
    $\{C\} \leftarrow C_{Label}$;
    Label++;
  end
end

a cylindrical range image I_r, where the Euclidean distance is regarded as the pixel values. A labeling image I_l with the same dimension as I_r is set to store clustering labels. The pixel values in I_l corresponding to the ground points obtained by the process of ground removal are set to -1, and values of other pixels are initialized as 0.

All unlabeled pixels in I_l will be labeled by traversing with a strategy of breadth-first search (BFS). According to the range information stored in I_r, the neighbor pixels of an unlabeled pixel $I_l(u,v)$, i.e. left, right, lower, and top pixels, will be determined whether they belong to the same object with $I_l(u,v)$. Since there is large difference between vertical resolution and horizontal resolution for the LiDAR scanner, different thresholds T_h and T_v are set for better segmentation of objects with various shapes.

$$\begin{cases} |I_r(u,v) - I_r(u+i, v+j)| < T_h, i = \pm 1, j = 0 \\ |I_r(u,v) - I_r(u+i, v+j)| < T_v, i = 0, j = \pm 1 \end{cases} \quad (1)$$

Meanwhile, each non-ground point $p(u,v) \in \{P_{ng}\}$ corresponding to the pixel $I_l(u,v)$ will be added into the cluster $C_{I_l(u,v)}$, and a cluster set $\{C\}$ containing N^C objects will be computed after the clustering process. The detailed object clustering approach is presented in Algorithm 1. The speed of the clustering algorithm depends on the number of points, where each point in the range image is visited at maximum twice.

3) Local re-clustering: Limited by the inherent property of LiDAR scanners that the laser ray is easily absorbed by the objects with dark color, which influences the performance of point cloud segmentation. As shown in Fig. 2, the points on the man are segmented into multiple parts due to the missing points located in the black T-shirt. Hence, a local re-clustering process on obtained cluster set $\{C\}$ is employed to handle the problem of over segmentation. Firstly, the centroids of $\{C\} = \{C_i | i = 1, ..., N_C\}$ are calculated, which is labeled as $\{\Theta\} = \{\Theta_i | i = 1, ..., N_C\}$. The horizontal angles $\{\varphi\}$ and distances $\{d\}$ of $\{\Theta\}$ w.r.t. the forward direction of the LiDAR scanner are computed, respectively. Then clusters are sorted by the horizontal angles $\{\varphi\}$. N_l clusters jointed with the cluster $C_i \in \{C\}$ are considered to be re-clustered as a new cluster. Grouping a cluster C_j into an angular adjacent cluster C_i should satisfy the following re-clustering condition.

$$|\varphi_i - \varphi_j| < T_\varphi \bigcap |d_i - d_j| < T_d \quad (2)$$

where T_φ and T_d are given thresholds. A new cluster set $\{C^R\}$ containing N^R object clusters are obtained by the process of object re-clustering.

Fig. 2. The example of over-segmented scenario.

B. Visual Pedestrian Detector

The regression-based detectors that simultaneously predict multiple bounding boxes and corresponding class probabilities achieve a better tradeoff between speed and accuracy. In this paper, YOLOv3 [2] is applied as the visual pedestrian detector, which is used to output the 2D bounding boxes of all pedestrians.

C. Calibration Between 3D LiDAR and Camera

To integrate multi-sensor information into one system, intrinsic calibration of RGB camera and extrinsic calibration between RGB camera and 3D LiDAR are necessary. The intrinsic matrix M_C of the RGB camera is computed by the ROS *camera_calibration* tool. An joint calibration approach with encoded ArUco markers and rectangular board [19] is utilized to obtain the extrinsic matrix $M_{LC} = [R_{LC}|t_{LC}]$ between LiDAR and camera.

D. Multi-Sensor Information Fusion

For the time t, 2D bounding boxes $B^V = \{B_j^V | j = 1, 2, ..., N^V\}$ on the RGB image I and object clusters $\{C^R\} = \{C_i^R | i = 1, ..., N^R\}$ are computed by the visual detector and agnostic segmentation approach of point cloud,

respectively. Semantic information and spatial information will be integrated in the image domain. Each point $p_j^L = (x_j^L, y_j^L, z_j^L)$, $p_j^L \in \{P_{ng}\}$ is transformed to the camera coordinate and then projected in the image I to get its corresponding pixel coordinate $p_j^I = (u_j, v_j)$.

$$\begin{bmatrix} u_j \\ v_j \\ 1 \end{bmatrix} = M_C \begin{bmatrix} R_{LC} & t_{LC} \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} x_j^L \\ y_j^L \\ z_j^L \\ 1 \end{bmatrix} \tag{3}$$

Note that only the part of projected points within the image boundary will be reserved. Then rectangular bounding box set $\{B^L\} = \{B_i^L | i = 1, ..., N^L\}$ for the projected clusters are generated by surrounding all projected points belonging to each cluster, where N^L is the number of obtained bounding boxes. A demonstration of the projected bounding boxes is shown in Fig. 1 (d). The intersection-over-union (IOU) area ratios of B_j^V and $\{B^L\}$ are computed to find the matched bounding boxes with B_j^V, which is labeled as $\{B_j^M\}$.

$$\frac{B_i^L \cap B_j^d}{B_i^L \cup B_j^d} \geq T_{iou}, i = 1...N^L, j = 1...N^V \tag{4}$$

where T_{iou} is the given threshold. Since the objects detected by the visual detector are usually in the foreground of the image, the one in $\{B_j^M\}$ with minimum distance will be selected as the matched bounding box with B_j^V. In this way, the semantic labels are attached to the cluster set $\{C^R\}$, i.e. the clusters belonging to the pedestrians will be distinguished from other clusters. Besides, Θ_i of cluster belonging to i-th pedestrian should be transformed to the camera coordinate system, where the range of the transformed point will be regarded as the distance of the i-th pedestrian. Hence, both of 2D bounding boxes and corresponding range information are obtained in the camera coordinate system.

III. Experiments and Discussion

A. Setup and System Implementation

The experimental setup is shown in Fig. 3. The Velodyne VLP-16 is chosen as the 3D LiDAR scanner, which is mounted at a height of 1.15 m from the floor. The vertical vertical field-of-view (FOV) of VLP-16 is $30°(\pm 15°)$ with the resolution of $2°$, and its horizontal angular resolution is set to $0.2°$. The left camera of ZED camera is applied as the RGB camera with the resolution of 1280×720, where its horizontal FOV and vertical FOV are $85°$ and $54°$, respectively.

B. Performance of Point Cloud Segmentation

In this section, experiments are conducted in the outdoor environments to evaluate the performance of the proposed segmentation approach for point cloud. Besides, the existing method (AEC) [20] is compared with our proposed approach, which removes the ground points by a predefined scalar threshold and adopts an adaptive Euclidean clustering algorithm to cluster objects. Fig. 4 illustrates several results in representative scenarios of our approach and AEC. Experiment 1 was conducted in the scenario as shown in the Fig. 4 (a), and our result is shown in Fig. 4 (c). Fig. 4 (b) is the segmentation result of AEC, where points with heights below 1.1 m are regarded as ground plane (height of LiDAR scanner is 1.15 m). Focusing on the pedestrian A wearing black T-shirt, A is segmented into multiple parts by AEC, but our proposed approach can correctly segment the point cloud, which can be seen in Fig. 4 (b) and (c), respectively. Besides, our approach can accurately segment the uneven ground plane.

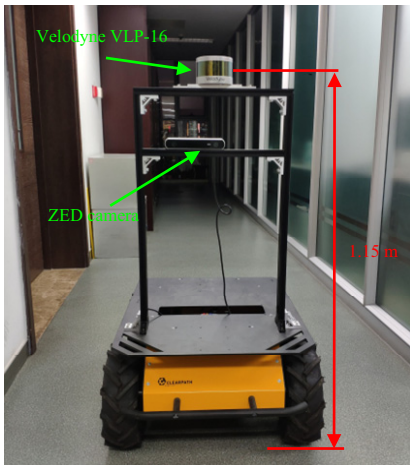

Fig. 3. The experimental setup for the proposed 3D pedestrian detector.

The scenario of Experiment 2 is shown as Fig. 4 (d). The limited minimum clustering height of AEC are set to -0.9 m for better comparison of clustering performance without the influence of poor ground segmentation of AEC [20]. The results of AEC and our approach are given in Fig. 4 (e) and (f), respectively. Three target pedestrians A, B, and C (see in Fig. 4 (d)) located in different distance are segmented well by our approach, but target A and C are over-segmented by AEC. Besides, as demonstrated in Fig. 4 (f), our proposed approach can better segment the low steps on the both sides than AEC.

C. Performance of Pedestrian Detection

Experiments are implemented to prove the effectiveness and accuracy of the proposed 3D pedestrian detector. Experiment 3 and Experiment 4 are conducted to testify the performance when pedestrians located closely in the horizontal direction. Results of Experiment 3 are given in Fig. 5 (a) and (b), which reveals that the proposed approach correctly detected target people standing side by side without influence of under segmentation. Experiment 4 is conducted to test the performance in the scenario when two people stand closed in the horizontal direction with different distances. As shown in

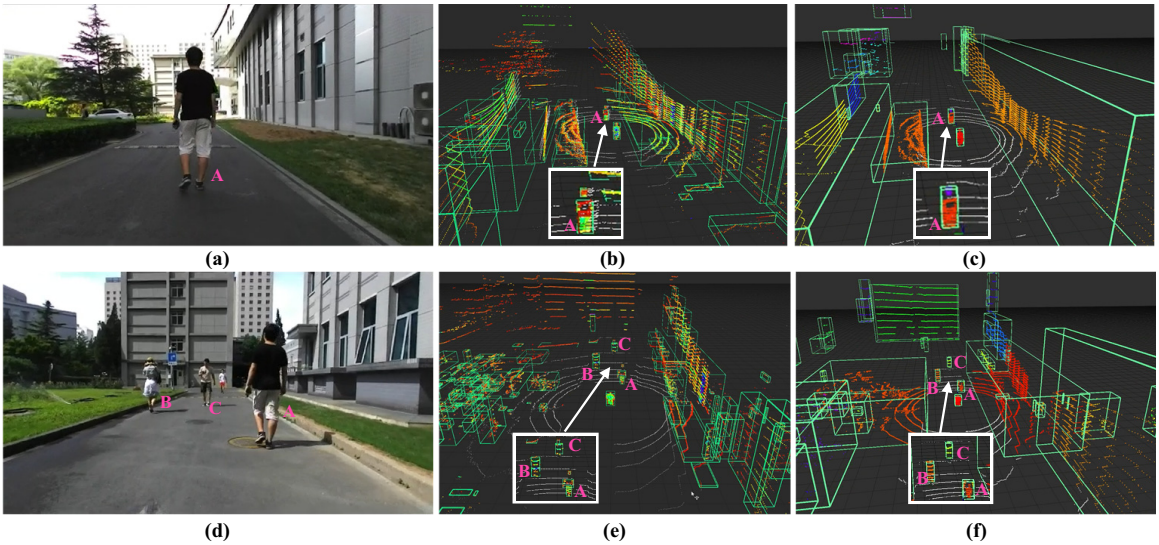

Fig. 4. Comparison of segmentation of our proposed approach with AEC [20]. In (b), (c), (e), and (f), colored points denote non-ground points, white points represent ground points, green 3D bounding boxes denotes segmented object clusters, and white boxes are the local enlarged images of corresponding parts. (a)–(c) illustrate the experimental scenario, result of AEC, and our result for Experiment 1, respectively. Experimental scenario, result of AEC, and our result of Experiment 2 are respectively given in (d)–(f).

TABLE I

MEASURED DATA OF EXPERIMENT 6.

Ground truth (m)	2.0	2.5	3.0	3.5	4.0	4.5	5.0	6.0	7.0	8.0	9.0	10.0
P_1	2.02	2.62	3.07	3.5	4.01	4.5	5.05	5.98	7.01	8.03	9.01	10.01
P_2	1,99	2.53	3.02	3.54	3.98	4.5	4.96	5.93	6.97	7.96	8.99	9.94
P_3	1.97	2.49	2.98	3.48	4.01	4.5	4.99	5.96	7.02	7.93	8.97	9.94
MAD	0.020	0.053	0.036	0.02	0.013	0.000	0.033	0.043	0.020	0.046	0.017	0.043

Fig. 5 (c) and (d), measuring results provided by the proposed detector are closed to the distances measured by the tapeline. Besides, Experiment 5 is implemented to test the performance of the proposed approach for remote targets. As shown in Fig. 5 (e) and (f), a person standing 20 m away from the LiDAR are detected correctly. Note that detection range of the proposed approach depends on the mounting height of LiDAR and calibration precision between sensors.

Experiment 6 is conducted to evaluate the precision for measuring range of the proposed 3D pedestrian detector. As shown in the Fig. 6, the target pedestrian locates in multiple positions (P_1, P_2, P_3) with different angles w.r.t. the sensors. Detailed measured data is given in Table I, where the values of mean absolute deviations (MAD) are computed.

$$MAD = \frac{1}{3}\sum |x_i - x_{ref}|, i = \{1, 2, 3\} \tag{5}$$

where x_i is the result measured in $P_i, i = \{1, 2, 3\}$, and x_{ref} is the ground truth measured by tapeline. This experiment proves that the measurement accuracy is favorable for robotic applications.

IV. CONCLUSION AND FUTURE WORK

In this paper, we have proposed a 3D pedestrian detector with calibrated camera and LiDAR, which is comprised of an agnostic segmentation approach for point cloud, a visual detector, and a fusion method in the image domain. Experiments validate the effectiveness of the presented approach. By adding the detecting classes of visual detector, the proposed 3D detector can be extended to detect other categories of objects. A drawback of the proposed detector is the limited effective detection perspective, which is restricted by the FOV of camera. In the future, we will focus on 3D pedestrian detection in broader perspective by employing multiple cameras or panorama camera.

REFERENCES

[1] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. 28th Int. Conf. Neural Inf. Process. Syst.*, 2015, pp. 91–99.

[2] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," arXiv:1804.02767, 2018.

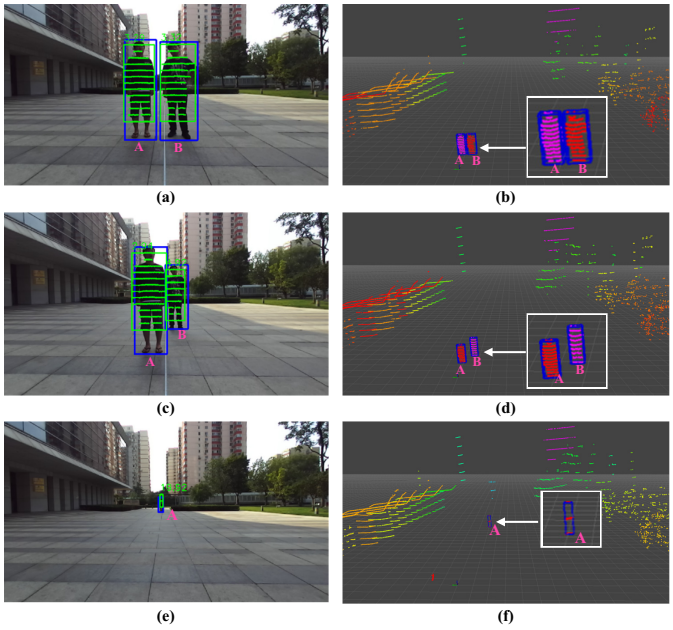

Fig. 5. Results of Experiment 3, Experiment 4, and Experiment 5. (a), (c), and (e) are the 2D pedestrian detection results and distance information of Experiment 3, Experiment 4, and Experiment 5, respectively, where blue bounding boxes denote detection results of visual detector, green bounding boxes and points indicate matched bounding box sets and surrounded points for pedestrians. Blue 3D bounding boxes in (b), (d), and (f) represent 3D pedestrian detection results in LiDAR coordinate system.

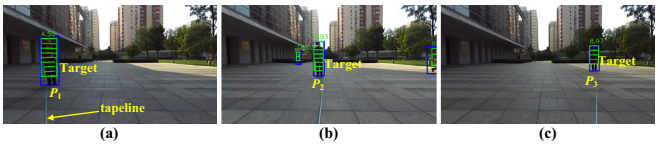

Fig. 6. Sample scenarios of Experiment 6.

[3] B. Li, W. Ouyang, L. Sheng, X. Zeng, and X. Wang, "GS3D: An efficient 3D object detection framework for autonomous driving," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Long Beach, CA, USA, Jun. 2019, pp. 1019–1028.

[4] L. Spinello and K. O. Arras, "People detection in RGB-D data," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, San Francisco, CA, USA, Sep. 2011, pp. 3838–3843.

[10] S. Thrun *et al.*, "Processing dense stereo data using elevation maps: Road surface, traffic isle, and obstacle detection," *IEEE Trans. Veh. Technol.*, vol. 59, no. 3, pp. 1172–1182, 2010.

[5] P. Li, X. Chen, and S. Shen, "Stereo r-cnn based 3d object detection for autonomous driving," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Long Beach, CA, USA, Jun. 2019, pp. 7644–7652.

[6] K. Kidono, T. Miyasaka, A. Watanabe, T. Naito, and J. Miura, "Pedestrian recognition using high-definition LIDAR," in *Proc. IEEE Intell. Veh. symp.*, Baden-Baden, Germany, Jun. 2011, pp. 405–410.

[7] Y. Zhou and O. Tuzel, "VoxelNet: End-to-end learning for point cloud based 3D object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 4490–4499.

[8] I. Bogoslavskyi and C. Stachniss, "Fast range image-based segmentation of sparse 3D laser scans for online operation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, Daejeon, South Korea, Oct. 2016, pp. 163–169.

[9] S. Thrun *et al.*, "Stanley: The robot that won the DARPA Grand Challenge," *J. Field Robot.*, vol. 23, no. 9, pp. 661–692, 2006.

[11] A. Ošep, A. Hermans, F. Engelmann, D. Klostermann, M. Mathias, and B. Leibe, "Multi-scale object candidates for generic object tracking in street scenes," in *Proc. IEEE Int. Conf. Robot. Autom.*, Stockholm, Sweden, May 2016, pp. 3180–3187.

[12] R. B. Rusu, "Semantic 3d object maps for everyday manipulation in human living environments," Ph.D. dissertation, Technischen Univ. München, Munich, Germany, 2009.

[13] J. Behley, V. Steinhage, and A. B. Cremers, "Laser-based segment classification using a mixture of bag-of-words," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, Tokyo, Japan, Nov. 2013, pp. 4195–4200.

[14] X. Chen, H. Ma, J. Wan, B. Li, and T. Xia, "Multi-view 3D object detection network for autonomous driving," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Honolulu, HI, USA, Jul. 2017, pp. 6526–6534.

[15] Z. Wang, W. Zhan, and M. Tomizuka, "Fusing bird's eye view LIDAR point cloud and front view camera image for 3D object detection," in *Proc. IEEE Intell. Veh. symp.*, Singapore, Singapore, Jun. 2018, pp. 1–6.

[16] D. Zermas, I. Izzat, and N. Papanikolopoulos, "Fast segmentation of 3D point clouds: A paradigm on LiDAR data for autonomous vehicle applications," in *Proc. IEEE Int. Conf. Robot. Autom.*, Changshu, China, May 2017, pp. 5067–5073.

[17] F. Moosmann, O. Pink, and C. Stiller, "Segmentation of 3D lidar data in non-flat urban environments using a local convexity criterion," in *Proc. IEEE Intell. Veh. symp.*, Xi'an, China, Jun. 2009, pp. 215–220.

[18] B. Douillard et al., "On the segmentation of 3D LIDAR point clouds," in *Proc. IEEE Int. Conf. Robot. Autom.*, Shanghai, China, May 2011, pp. 2798–2805.

[19] A. Dhall, K. Chelani, V. Radhakrishnan, and K. M. Krishna, "LiDAR-camera calibration using 3D-3D point correspondences," arXiv:1705.09785, 2017.

[20] Z. Yan, T. Duckett, and N. Bellotto, "Online learning for human classification in 3D LiDAR-based tracking," in *Proc. IEEE Int. Conf. Robot. Autom.*,Vancouver, BC, Canada, Sep. 2017, pp. 864–871.