3D Point Cloud Labeling Tool for Driving Automatically

MingHui Li* and Yanning Zhang†
*Shenzhen Unity-Drive Innovation Technology Co,Ltd. Shenzhen, China
E-mail: liminghui@unity-drive.com Tel: +86-17838921968
†School of Intelligent Engineering, Zhengzhou University of Aeronautics, Zhengzhou, China
E-mail: yanshan@zua.edu.cn Tel: +86-13783583354

***Abstract*—LIDAR（light detection and ranging） and visual perception are the key factors for the success of high level (L4-L5) automatic pilot obstacle avoidance.The combination of deep learning and 3D point cloud undoubtedly lays a solid foundation for the rapid development of automatic driving.At the same time, the demand of a large amount of data urges us to improve and perfect the point cloud marking tools.This article describes a newly developed 3D point cloud annotation tool, it supports PCD and bin formats. Using point cloud tracking P2B algorithm to achieve semi-automatic labeling, and using the reference of the z-axis heading Angle automatic detection function to simplify the complexity of the pull frame, It achieves the conversion of 3D bounding box coordinate information to 2D bounding box coordinates of point clouds and images acquired after the joint calibration of camera and lidar. It simplifies the operation of labeling tools and improves the efficiency of labeling.**

I. INTRODUCTION

The cost of the configuration is a major obstacle to the development of driverless vehicles. When we try to reduce the cost of hardware, the requirements for technology will be higher.The sensing system is the eyes of the autopilot, The camera recognizing traffic lights and traffic signs, The light detection and ranging of lidar are used to avoid obstacles and identifying the surrounding objects and their running speed.There is no doubt that the combination of deep learning[2] and 3D point cloud provides a broader development platform for driving automatically, and deep learning is more stable, robust and accurate. The combination of deep learning and point cloud is also under constant trial and trial, but the demand for data falls like a downpour.

Although the open data setting KITTI[1] has already existed, but the project actual application still requires us to conduct annotation training on our own real road data. Current tagging tools are also in contention. 3d_bat[3] uses linear interpolation to realize point cloud tracking, which is suitable for object labeling with uniform linear motion, the "TUBS"[4] labeling tool has also possibility to detect and remove the ground to better locate road users on the ground "latte"[5] utilize image-based detection algorithms to automatically pre-label a calibrated image, and transfer the labels to the point cloud, simplify the annotation to just one click on the target object, and automatically generate the bounding box for the target. We hope to develop a practical and applicable labeling tool, which could be like two-dimensional image labeling tools Labelme[6] and Labelimg, which are online and lightweight.

Through the exploration and application of several kinds of annotation tool, we developed a web-based 3d point clouds annotation tool, and considering that when we collect the cloud data online, the average PCD file will takes about 2M memory more than bin file lead to more resource consumption, in practice we use bin files more, therefore we develop annotation tool supporting the loading both PCD and bin file. In actual operation, our driverless car will encounter scenes with multiple objects such as parking lots and other relatively static scenes, so we adopt the principle of translational invariance of relatively static objects to achieve the function of paste and copy for multiple objects.

In addition, what our main contributions are: Firstly, we integrate tracking algorithm P2B[7] into the annotation tool, which can specify tracking frame number, category and ID to realize semi-automatic annotation. Secondly, coordinate transformation formula of point cloud and image obtained after joint calibration of camera and radar, which realizing the mapping from 3D bounding box to 2d image. Thirdly, the z-Axis Angle automatic detection function of R3Det[8] is integrated to simplify the complexity of drawing 3D frame.We developed strict standards for point cloud labeling, and the efficiency of the new labeling tools was significantly improved under the testing of the data group staff.

II. REALATED WORK

A. 3D object tracking

At present, two mainstream methods account for half of 3d point cloud tracking. First of all, RGB-D information[9,10,11,12] , taking [9] as an example, it proposes a 3D object tracking algorithm using 3-D LIDAR, RGB camera and INS (GPS/IMU) sensor data. By

analyzing the positioning data of continuous 2D-RGB, 3D point cloud and Ego-Vehicle and the output trajectory tracking object, the velocity of flow was estimated and the position was predicted in a time step under the 3D world coordinate system. Tracing starts with the initial 3D border of the known object. Two parallel mean shift algorithms are used to detect and locate targets in 2d images and 3D point clouds, and then a robust fusion and tracking algorithm based on 2D / 3D Kalman filter is adopted. Secondly, Siamese Tracking[13,14,15,16].Based on studying the potential possibility of using Shape to complete the Tracking of 3D objects in the LIDAR point cloud, they designed a twin tracker that encodes model shapes and candidate shapes into compact potential representations, then the potential representation of this object is then decoded into a model shape, but the tracking algorithm we use is P2B which begins by sampling the seeds from the point cloud of the template and the search area respectively, then the permutation invariant feature is enhanced, and the target clues in the template are embedded into the seeds of the search area and represented by the specific features of the target.Therefore, the seeds of the extended search area are returned to the potential target centers through Hoff votes, seed orientation scores further strengthen these centers. Finally, each center clusters its neighbors to leverage the power of integration to jointly propose 3D targets and validate.

B. Rotating frame detection

There are four common detection methods of rotating frame：

Representative: RRPN[17]: RRPN (peace-oriented Scene Text Detection via Rotation Proposals), based on FasterRCNN, "in place" Rotation of 6 angles (30 degrees) for each box generated according to the original method, that is, for each box of each position, each Scale, and each Ratio, they are derived into 6 different angles. This way of thinking is simple. Generate more dense rotating Anchor to adapt to the task of rotating frame detection.

R3Det[18]: R3Det(the combination of violence and speed) mainly solves the speed problem in the process of RRPN. The location information of the current Refined boundary box is recoded as the corresponding Feature point through Feature interpolation, so as to realize Feature reconstruction and alignment.

RPN[19]: ROITransfomer (Learning RoI Transformer for Detecting Oriented Objects in Aerial Images) = three-stage method (two-stage frame detection + one-stage boundary detection), and the accuracy of three-stage algorithm is improved to a relatively high degree of recurrence, so it is maybe a little hard for people to actually use it. The reasons as below:

(1) use all RoI (RRoI) indicators to transform a region of interest (HRoI) to a region of rotational interest (RRoI).

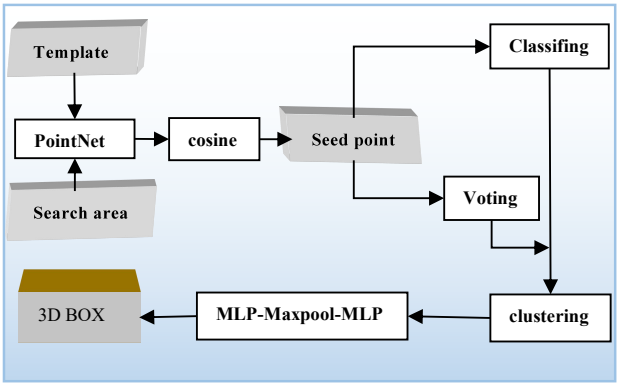

Fig. 1 The strength of traget-specific feature augmentation and 3D target proposal and Verfication is that the point-state tracking model can assist in the better utilization of three-dimensional local geometric information to describe targets within a point cloud. Secondly, an end-to-end approach is adopted to develop 3d target tracking tasks, which has a strong ability to track changes in the 3D appearance of the target.

(2) The module RPS-RoI-align was used to extract the rotation invariant features so as to promote subsequent classification and regression.

(3) RoI Transformer can be embedded in detectors for directed border object detection.

The most skilled three-stage model all you Need is Boundary[20] (All you need is Boundary Toward Arbitrary - Shaped Text Spotting), skills are extremely complex, repetition is extremely difficult, for speed, although use the detectron2, three phases is not fast the central idea and the third are similar, butmore tricky, the Text with the boundary point to represent Arbitrary shape, the method of solving the Arbitrary shape of the Text in natural scene images end-to-end recognition problem, the Oriented Rectangular Box Detector. The Boundary Point Detection Network and the Recognition Network. For the Oriented Rectangular Box Detector, this paper firstly uses RPN network to extract candidate areas.

III. METHOD

A. Tracking

Linear interpolation is only applicable to the Object of uniform velocity in a straight line, the ideal is not very common, with Tracking algorithm will be slightly better than the interpolation, we rely on P2B algorithm for Tracking the basis of the model, it regard a 3d Point Clouds Object Tracking problem as a Object Detection, mainly using the VoteNet [21] to achieve target Detection. Take the target template point cloud (for example, the point cloud of an automobile model needs to be located in the Search area to the location of a car similar to the template point cloud) and the search area point cloud (the searched point cloud) as input, and the 3D box information of the Target template point cloud

in the search area point cloud as output. The network structure is Fig.1, which is divided into 2 steps. The first step calculates the seed point, and the second step calculates the 3D box

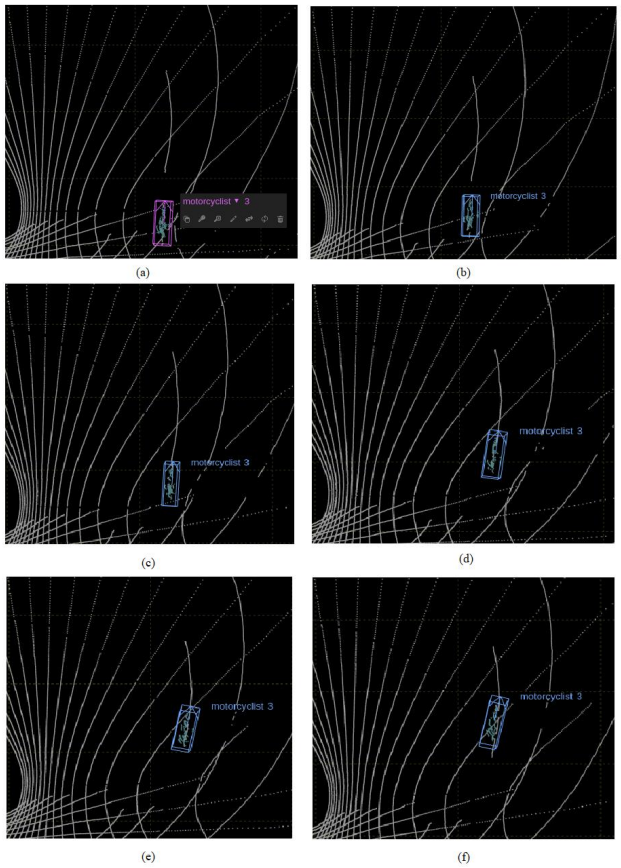

Fig.2 To track the motorcyclist. This picture selects an object that is not in a straight line and moving at a non-uniform speed. We can specify the tracking object and the number of tracking frames. (a) Start, (f) end, with turning and shifting in the middle.

through the seed Point.

The training-specific feature augmentation focuses on preprocessing the template point cloud and the Search Area point cloud, generating a seed point set for the Search Area point cloud to perform voting operations in subsequent modules, and can be considered to be mergin g the template point cloud and the point cloud to be searched, giving rise to a new seed point cloud. 3D Target proposal and Verification takes seed point cloud as input, locates the box of target template point cloud in the Search area point cloud, and selects the highest predicted value of Score as the final result.

We transplanted P2B algorithm into our annotation tool, and the tracking effect is shown in Fig.2. From the perspective of Angle transformation and displacement distance difference, obviously, tracking is better than linear interpolation.

B. Vision and point cloud information fusion

Obstacle detection can use laser radar for object

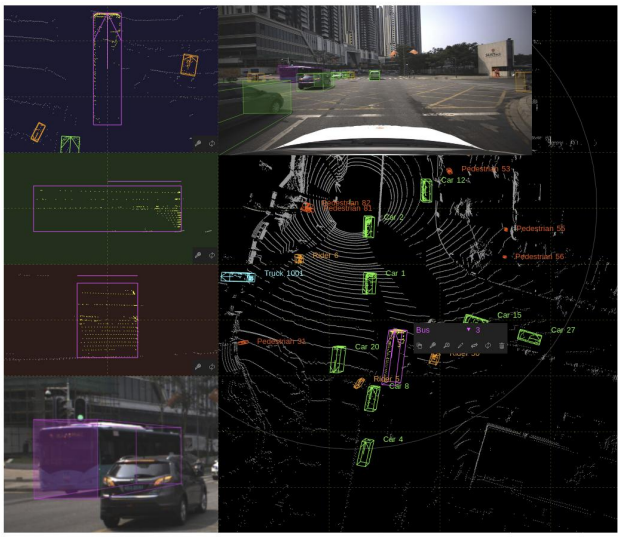

Fig.3 3D point cloud Bounding Box maps to 2D images

clustering, but we are using the velodyne 16 line, the number of lines is still a little sparse, for more distant objects but with sparse line the clustering effect will be not good, so consider using visual when marking point cloud help us distinguish distant objects, when marking the point cloud box, 3 d frame will be projected on the 2d image simultaneously. 3D bounding Boxx coordinate information in the point cloud is projected to the image to obtain 2D bounding box coordinates, and the camera and lidar are jointly calibrated, that is, the spatial conversion relationship between their coordinate systems is obtained. The role of joint calibration is to establish the correspondence between the point cloud and the image Pixel. External parameters of camera and lidar need to be acquired and the points in the three-dimensional coordinate system of point cloud are projected into the three-dimensional coordinate system of camera.

Given a 3d point P = (x, y, z), the 3d points obtained by laser radar scanning are mapped, and the mapping method is as follows:

$$\begin{aligned} \theta &= \arctan 2(y,x) \\ \phi &= \arcsin\left(z/\sqrt{x^2+y^2+z^2}\right) \\ r &= \lfloor \theta/\Delta\theta \rfloor \\ c &= \lfloor \phi/\Delta\phi \rfloor \end{aligned} \tag{1}$$

Then, the mapping point corresponding to any point P(x,y,z) in 3D space is:

$$(r', c', z')$$

Among them, r', c' represents 2D coordinates of the mapping point, and the mapping formula is shown in

(2). z' represents the value of 2 channels, and the calculation method of the value of the two channels is:

$$z'[0] = \sqrt{x^2 + y^2} \qquad (2)$$
$$z'[1] = z$$

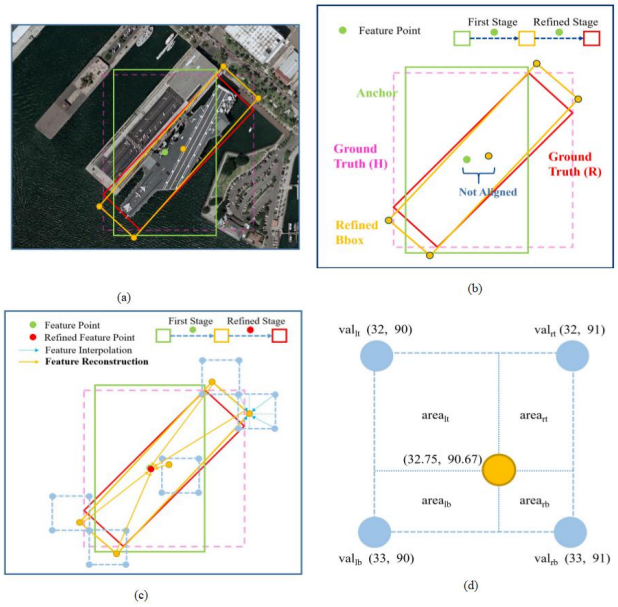

Fig. 4 (a)original image. (b)The matrix is refined without considering the feature misalignment caused by the bounding box position,.(c)Through the reconstruction of feature map, the aligned feature can be used to refine the box. (d)feature interplotation.

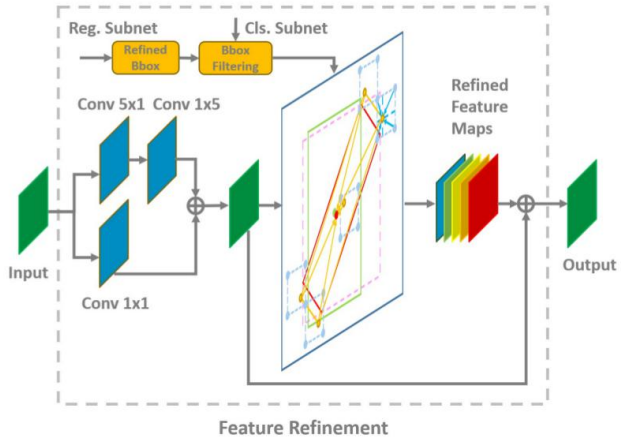

Fig. 5 The feature subdivision module mainly includes fine boundary box filtering, feature interpolation and feature graph reconstruction.

If a point on a 2D plane has no corresponding 3D point, it is filled with (0,0).The fusion effect of vision and point cloud is shown in Fig.3.

C. *Z-axis direction Angle automatic detection*

On a point cloud map, vehicles and people walk in different directions. We import an automatic direction Angle detection auxiliary pull frame to help determine yaw Angle. The current conventional horizontal target detector has limitations for many practical applications. A feature refining module is proposed by R3Det to obtain more accurate features to improve the performance of rotating target detection. First, RetinaNet was used to construct a single-stage target detection framework. Seconed, RefineDet thought was used to refine the detection results of the FirstStage. Third, FRM module is imported to realize the operation similar to ROIPooling in the one-stage model.

The Fig.4 shows the box refining process without feature alignment, resulting in feature inaccuracies, which are detrimental to large width-to-height ratios and small Numbers of samples.

R3Det proposes to recode the position information of the current refined bounding box (orange rectangle) into corresponding feature points (red points), and then achieve feature alignment by rebuilding the entire feature map.In order to accurately obtain the location feature information of the refined boundary box, the bilinear interpolation method is used, and the formula is as follow:

$$val = val_{lt} \cdot area_{rb} + val_{rt} \cdot area_{lb} + val_{rb} \cdot area_{lt} + val_{lb} \cdot area_{rt} \qquad (3)$$

Feature refining module, as shown in Fig.5.New features are obtained by superimposing the feature map with double convolution. In the refining stage, only the bounding box with the highest score for each feature point is retained to improve the speed and ensure that only one refined bounding box is corresponding to each feature point. For each feature point of the feature graph, the corresponding feature vector on the feature graph is obtained according to the 5 coordinates of the refined bounding box (one center point, four corner points). More accurate eigenvectors can be obtained by bilinear interpolation. Next, 5 eigenvectors are added to replace the previous eigenvectors. After the feature points are traversed, the entire feature map is reconstructed. Finally, the reconstructed feature map is added to the original feature map to complete the whole process.

In the absence of an automatic direction detection module, the starting direction of the mouse is the direction of the car. As shown in Fig.6, auto labeling is taken as an example to demonstrate the function of automatic direction angle detection.

IV. PERFORMANCE SHOW

A. *performance show*

We demonstrate the function of the annotation tool in practical application of data. Our annotation file is output

in JSON format, with variables such as upper-left coordinates (x,y,z), length, width, height, z-Zaix direction angles, labels, and the name of the dot cloud file.

In this section we show the basic function of a Point size +/-,Point brightness +/-, toggle background, toggle box, toggle obj color, toggle id, toggle the category, auto play point cloud, quick adjustment direction Angle and long wide high.

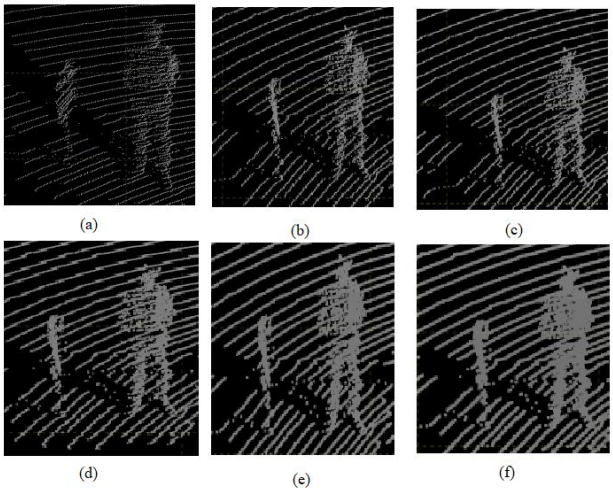

Fig. 6 (a) original figure. (b)~(f) size of points gradually increases.

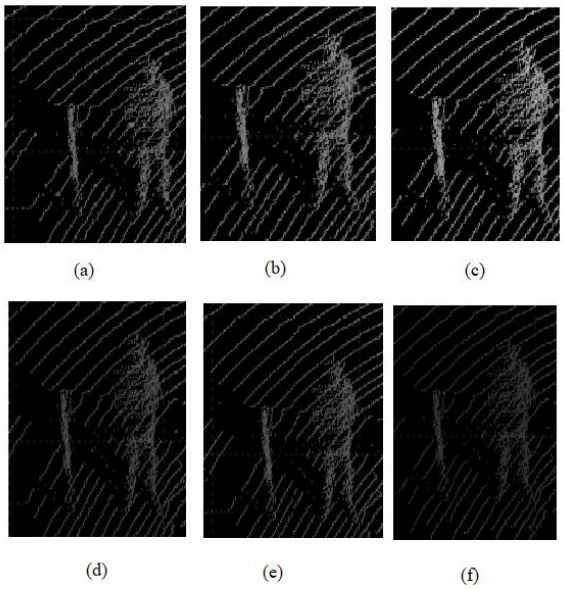

Fig. 7 (a) original picture. (b) and (c) points gradually become brighter. (d)~(f) points gradually become darker

B. mark specification

The 3D Bounding Box we mark is aimed at making ground truth. Our indexers strictly follow the labeling requirements of

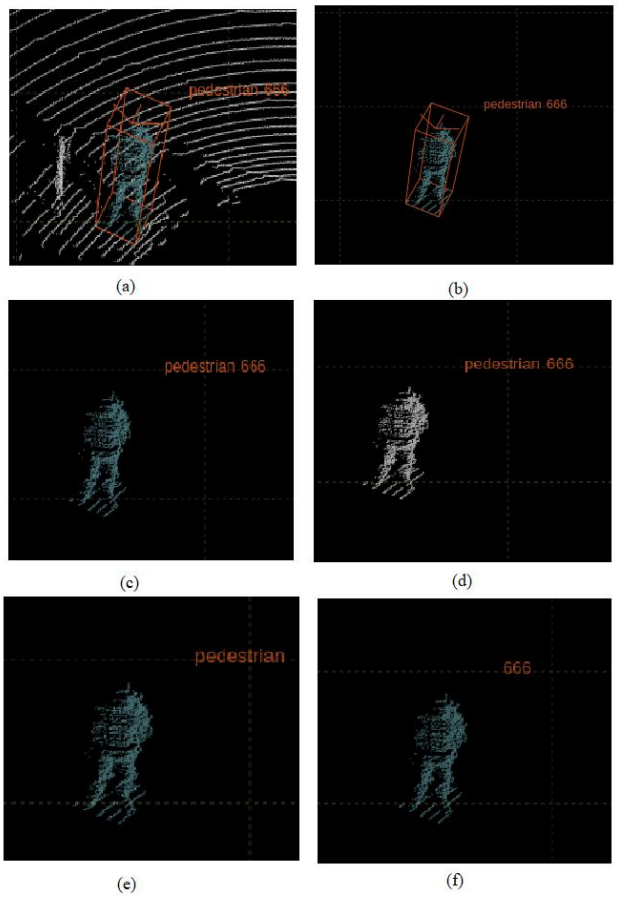

Fig. 8 (a) Original image. (b) Toggle Background. (c) Toggle Box. (d) Toggle obj color. (e) Toggle ID. (f) Toggle the Category

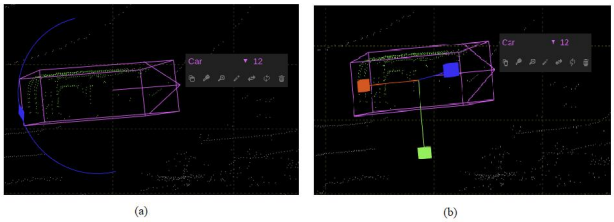

Fig. 9 (a) adjustment direction Angle. (b) adjustment longth width highth

Table.1 and Table.2, and we also analyze the efficiency improvement brought by our labeling tools with this standard.

Table.1 represents the average size of commonly used models in China, which may vary slightly due to different manufacturing processes of different manufacturers, but can be used as the average size of point cloud marking. Table.2 is the standardization requirements of labeling personnel to label vehicles when using labeling tools.

Absolute accuracy of labeling as groundtruth is the guarantee of deep learning.

Table.1 Reference dimensions for each category

	length	width	height
car	4.0-4.9	1.6-1.9	1.3-1.6
SUV/MPV	4.6-5.1	1.6-1.9	1.7-2.1
pickup	5.3	1.85	1.8
truck	12.-18	2.4	2.7
van	6-10	2.4	1.8-2.7
Small bus	6	2.1	2.7
big bus	10-18	2.6	3.2-4.2
pedestrians	0.5	0.5	1.5-1.9
bicycle	1.7	0.7	1
motorcycle	2	7.5	1.2

Table.2 Labeling requirements

1. Objects with less than 10 points need not be marked
2. Mark the target object into the box completely, with no more than 3 leakage points;
3. Objects outside the radius of 70m are not marked
4. There is no obvious Angle deviation in the callout box
5. Do not frame into the ground
6. Make sure you mark the correct driving direction
7. Bounding box should be kept close to object point cloud contour, and bounding box should not be too large.
8. Annotate object classification without error
9. Mark only the entity of the object, and try not to mark the shadow. The residual image and the object entity are judged by 2D and 3D images. When it is impossible to clearly distinguish the residual image and the entity, the residual image is framed
10. Close to the point on the side of the radar vehicle, need to fit. The fit between frame and point, need to have certain space. No missing mark problem

C. comparative analysis

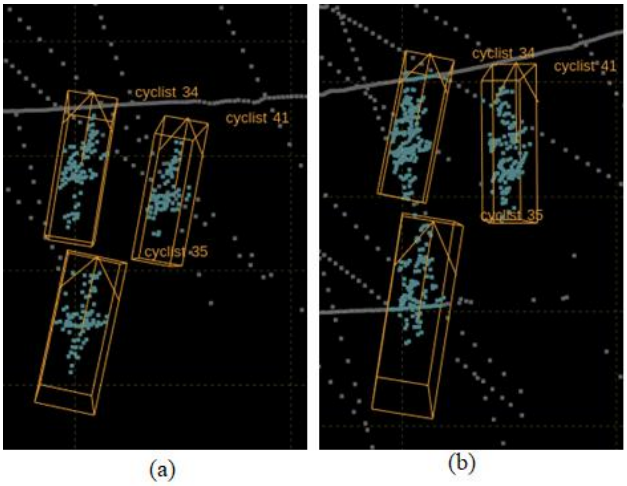

Fig.10 Linear interpolation diagram. (a) is the point cloud of the previous frame of linear interpolation, (b) is the point cloud of the next frame of linear interpolation, and the 3D frame of the point cloud of the 2 frames will be tilted, and the Angle will be different, which will affect the judgment of groundtruth.

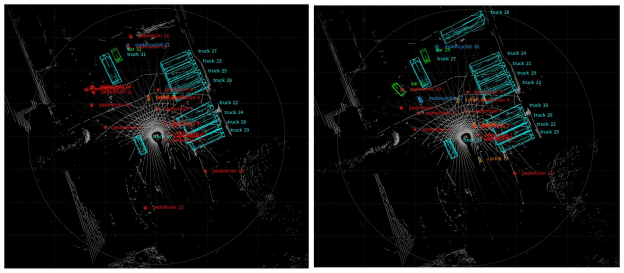

Fig.11 Voting detection algorithm. (a) is the point cloud of the previous frame detected by Voting algorithm, (b) is the point cloud of the last frame detected by Voting algorithm. (A) and (b) point clouds are not significantly different from each other in the detection results, but the IDS of the same object are almost different.

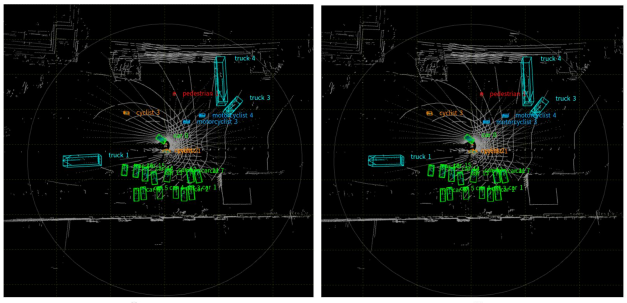

Fig. 12 Tracking. (a) The point cloud of the previous frame tracked by P2B algorithm; (b) The point cloud of the next frame tracked by P2B algorithm; (a) and (b) Point clouds can completely keep the id of the same object before and after frames from the tracking results, and the fitting of 3D frame and object can almost meet the requirements.

Table.3 The average time of each annotation is spent

labelling	average time
Pure manual marking	Ten days
Based on the detection and then id adjustment	Eight days
Based on interpolation	Eight days
Based on tracking	Six days

The whole process of our project is divided into four stages: manual annotation, linear interpolation, ID adjustment based on detection and tracking annotation proposed in this paper.

The idea of using linear interpolation in Fig.10 comes from [3]. We refer to linear interpolation in the annotation tool, but the result is not satisfactory. Because linear interpolation is only applicable to objects moving in a straight line with uniform speed, few objects move in accordance with this logic in the actual road conditions, and interpolation cannot take into account the adjustment of Angle and the fitting of 3D frame. Considering that each object has a different trajectory, only one object can be interpolated linearly at a time.

In the detection algorithm in Fig.11, [22] is used. We can see that the detection accuracy is very high, but the ID of the same object is different in different frames. We can do point cloud detection if we only use tag, but it does not

meet the needs of point cloud tracking. You need to manually adjust the ID to ensure the same ID between the frames, which will cost a lot of manpower and material resources, but it is slightly better than linear interpolation.

Tracking annotation is used in Fig.12. [7] is a tracking algorithm based on voting. Based on the fitting degree of tag ID and box to objects, P2B algorithm is more suitable for integrating into point cloud labeling tool.

Table.3 is a record of the time spent on the annotation point cloud during the development of the annotation tool. With the improvement of the annotation tool integration algorithm, the time spent on the annotation is also reduced and the efficiency is significantly improved.

V. CONCLUSIONS

In this paper, we propose a new 3D point cloud marking tool, which mainly integrates three existing technologies : P2B point cloud tracking, visual and point cloud information fusion, and direction Angle automatic detection. The addition of these three technologies improves the efficiency of labeling and reduce the difficulty of labeling. In addition, there are still some problems with our point cloud tagging tool because point clouds are too sparse to track accurately. In the future, we will consider adding point cloud completion function to gradually improve the function.

ACKNOWLEDGMENT

The authors would like to thank for the Key Projects of the Joint Fund of the National Natural Science Foundation of China under Grants No.U1833203 and the Key Research Project of the Henan Provincial Higher Education under Grants No.20A510011.

REFERENCES

[1] Geiger A , Lenz P , Stiller C , et al. Vision meets robotics: The KITTI dataset[J]. International Journal of Robotics Research, 2013, 32(11):1231-1237.

[2] Bojarski M , Del Testa D , Dworakowski D , et al. End to End Learning for Self-Driving Cars[J]. 2016.

[3] Zimmer W , Rangesh A , Trivedi M . 3D BAT: A Semi-Automatic, Web-based 3D Annotation Toolbox for Full-Surround, Multi-Modal Data Streams[C]// 2019 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2019.

[4] Plachetka C , Rieken J , Maurer M . The TUBS Road User Dataset: A New LiDAR Dataset and its Application to CNN-based Road User Classification for Automated Vehicles[C]// 2018 IEEE International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2018.

[5] Wang B , Wu V , Wu B , et al. LATTE: Accelerating LiDAR Point Cloud Annotation via Sensor Fusion, One-Click Annotation, and Tracking[J]. 2019.

[6] Russell B C , Torralba A , Murphy K P , et al. LabelMe: A Database and Web-Based Tool for Image Annotation[J]. International Journal of Computer Vision, 2008, 77(1-3).

[7] H. Qi, C. Feng, Z. Cao, F. Zhao and Y. Xiao, "P2B: Point-to-Box Network for 3D Object Tracking in Point Clouds," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020, pp. 6328-6337, doi: 10.1109/CVPR42600.2020.00636.

[8] Yang X , Liu Q , Yan J , et al. R3Det: Refined Single-Stage Detector with Feature Refinement for Rotating Object[J]. 2019.

[9] Alireza Asvadi, Pedro Girao, Paulo Peixoto, et al. 3D object tracking using RGB and LIDAR data[C]// IEEE International Conference on Intelligent Transportation Systems. IEEE, 2016.

[10] Adel Bibi, Tinahzu Zhang, and Bernard Ghanem. 3d partbased sparse tracker with automatic synchronization and registration. In Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016

[11] Kart U , Lukezic A , Kristan M , et al. Object Tracking by Reconstruction With View-Specific Discriminative Correlation Filters[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2019.

[12] Pieropan, Alessandro, Bergstrom, Niklas, Ishikawa, Masatoshi,et al. Robust 3D tracking of unknown objects[C]// IEEE International Conference on Robotics & Automation. IEEE, 2015.

[13] Giancola S , Zarzar J , Ghanem B . Leveraging Shape Completion for 3D Siamese Tracking[J]. 2019.

[14] Dong X , Shen J . Triplet Loss in Siamese Network for Object Tracking: 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part XIII[M]// Computer Vision – ECCV 2018. 2018.

[15] Anfeng He* †, Chong Luo‡, Xinmei Tian†, et al. A Twofold Siamese Network for Real-Time Object Tracking[J]. 2018.

[16] Li B , Yan J , Wu W , et al. High Performance Visual Tracking with Siamese Region Proposal Network[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2018.

[17] Ma J , Shao W , Ye H , et al. Arbitrary-Oriented Scene Text Detection via Rotation Proposals[J]. ieee transactions on multimedia, 2017, PP(99):1-1.

[18] Yang X , Liu Q , Yan J , et al. R3Det: Refined Single-Stage Detector with Feature Refinement for Rotating Object[J]. 2019.

[19] Ding J , Xue N , Long Y , et al. Learning RoI Transformer for Detecting Oriented Objects in Aerial Images[J]. 2018.

[20] Wang H , Lu P , Zhang H , et al. All You Need Is Boundary: Toward Arbitrary-Shaped Text Spotting[J]. 2019.

[21] Ding Z , Han X , Niethammer M . VoteNet: A Deep Learning Label Fusion Method for Multi-Atlas Segmentation[J]. 2019.

[22] Qi, Charles R , et al. "Deep Hough Voting for 3D Object Detection in Point Clouds." (2019).