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

Dynamic Obstacle Tracking Based On High-Definition Map In Urban Scene

Yuxiao Li[1234], Can Wang[124*], Zhilong Su[1234], Shengcai duan[1234], Xinyu Wu[124],

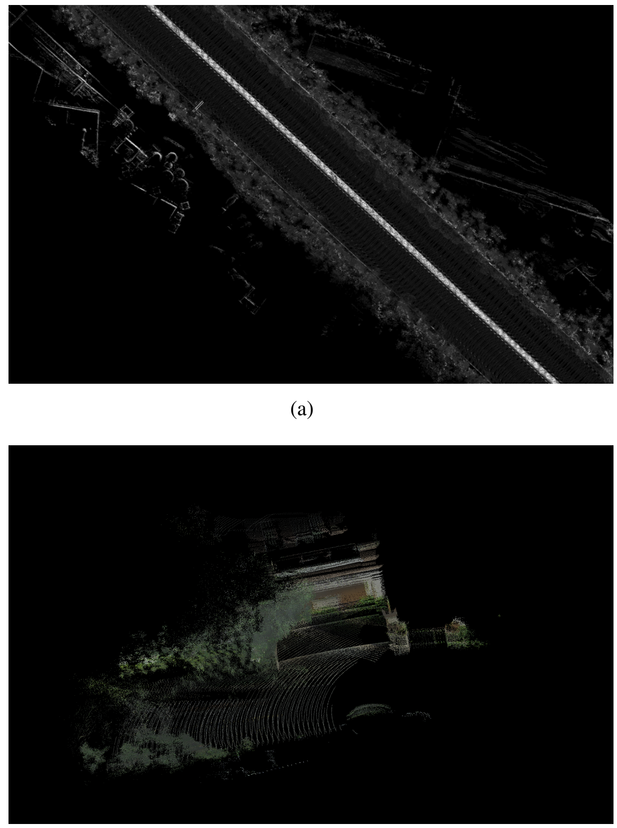
***Abstract*— The application of High-Definition map can realize centimeter-level position in urban scenes, the development of deep learning has made great breakthrough in the point cloud dynamic obstacle recognition.All these technologies make obstacle detection and tracking based on High-Definition map effectively realize in the modern smart city scene. It is different from the previous obstacle detection and tracking methods based purely on vision the using of High-Definition map can provide high-precision positioning and reduce the difficulty of point cloud classification. The using of lidar also solves the problem of up-to-scale in dynamic detection. In this paper, we put forward a new Multi-camera Lidar Point Cloud Map, we complete the map at normal speed on the highway and get a satisfactory result. At the same time, we also find a robust combination of traditional kalman filter, Hungary algorithm and current deep learning to solve dynamic obstacle tracking and detection. The experimental results show that our system can effectively complete the special problem of generating map and target tracking on urban scene.**

***Index Terms*— High-Definition Map, Lidar, Obstacle Tracking**

I. INTRODUCTION

Self-driving taxi services have been tested in local areas[1], unmanned aerial vehicles (UAV) have achieved good intelligent flight, and AR/VR technology is defining new human-computer interaction modes[2]. With the increasing popularity of 5G, autonomous driving technology, smart city, Large-Scale High-Definition Color-Point Map gradually recognized by the world[3].The research combines location and target tracking based on High-Definition Map also faces great challenges.People are also looking for less noise, more accurate information, more robust High-Definition map, Fig.1.

Location is the starting point of unmanned devices work in urban environments, With the development of GNSS, when the signal conditions are excellent, we can achieve the location with centimeter accuracy in the urban environment. However, the blocking of buildings in cities will cause signal loss, complex electromagnetic environment will lead to signal quality decline, and diffuse reflection between buildings will make GNSS signal unreliable. The participation of IMU in GNSS/IMU system makes the location system more robust and accurate[6], but the inherent defects of IMU, accumulated error and angular offset, make the cumulative location information of the IMU has a great error during long-term location. Simultaneous localization and mapping(SLAM) provides a good supplementary location information for location[7]. The GNSS/IMU location data is used when the signal conditions are good, and the effective SLAM location result when the GNSS signal is lost can continuously provide the location service.



(a)

(b)

Fig. 1. Comparison of Multi-camera Lidar Point Cloud Map and High-Definition map.(a)Multi-camera Lidar Point Cloud Map ; (b) High-Definition map.

[1] Guangdong Provincial Key Lab of Robotics and Intelligent System, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China. {yx.li, can.wang, zl.su, sc.duan, xy.wu}@siat.ac.cn

[2]CAS Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institutes of Advanced Technology, Shenzhen 518055, China.

[3]Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, Beijing 100049, China.

[4]Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong.

*Corresponding Author

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

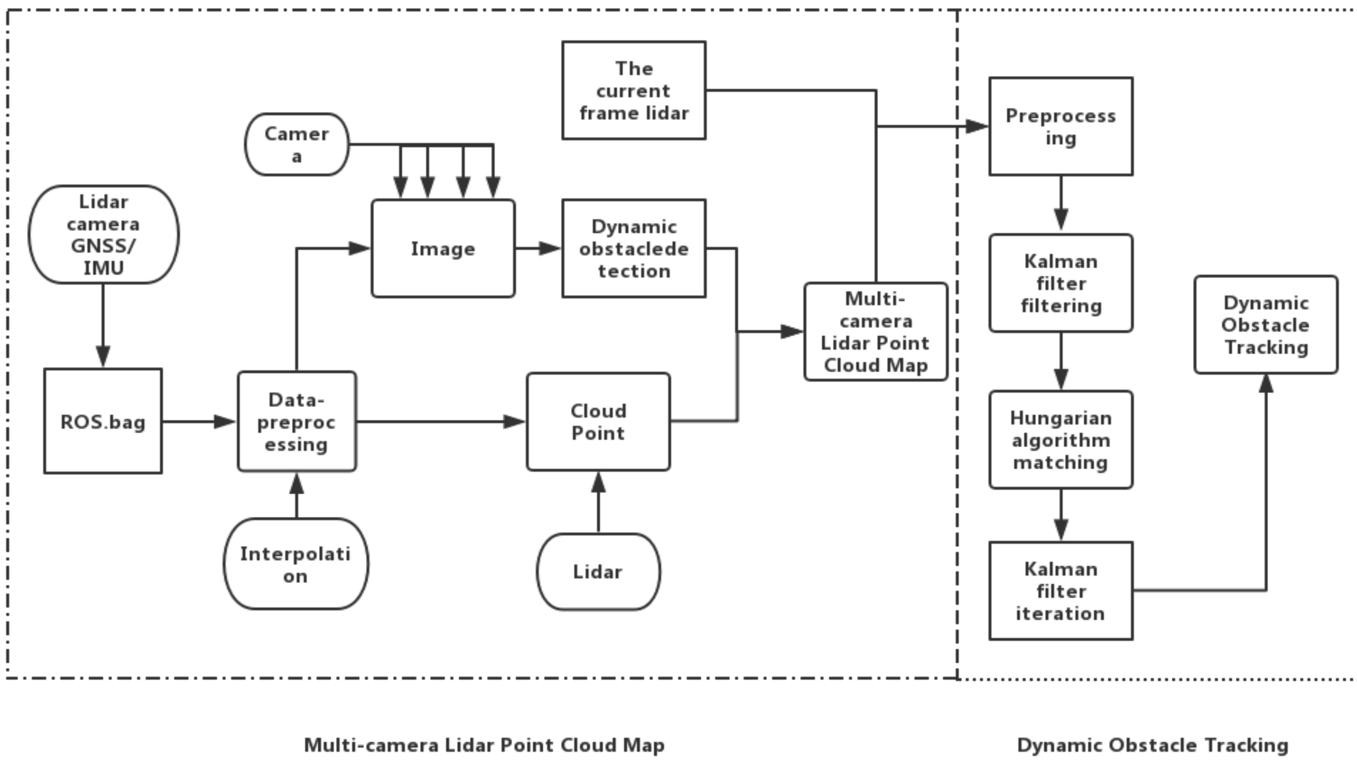


Fig. 2. Dynamic Obstacle Tracking Based On High-Definition Map System.

Orb-slam[8] is a 3d simultaneous localization and mapping construction algorithm based on ORB features, orb-slam is based on PTAM architecture, which adds map initialization and closed-loop detection functions, optimizes key frame selection and map construction methods, and achieves good results in processing speed, tracking effect and map accuracy.However, it should be noticed that map is sparse,it is not robust enough to rely on the map for relocation and the accuracy can not arrive the requirements.

VINS-Mono[9] is a Visual and Inertial fusion SLAM algorithm. Through Sliding Windows optimization, the tight coupling between vision and IMU is realized as VIO. In addition, DBoW is added as Loop Closure, which is a complete SLAM system.Front end of the algorithm by Harris corner detection and LK optical flow tracking implementation. To optimize the back-end Visual Inertial fusion algorithm of vision and innovative marketing information, they use the pre-integration to realize the Visual and IMU tight coupling. However, the drift and accumulated error of IMU signal will make the system unreliable after a long period of operation.

In the latest multi-sensor slam research, the combination of camera and lidar is an interesting study. Limo-slam[10] is to filter feature points through lidar distance information and estimate motion using BA algorithm based on robust key frame. Semantic tagging is used for outlier culling and weighting of vegetation landmarks. This method has been proved to be reliable and time-robust in experiments, but the extraction of visual key frames can still produce drift without texture.

The advent of High-Definition maps provides a powerful aid to our positioning tasks. Through High-Definition map urban information can transmit to UAV and AGVs, information interaction between cities and robots. Neither 2D-Grid-Map nor 3D-Map pay more attention to the semantic information in the map[11]. The above maps can only be used to enable UAV or AGV repositioning or path planning,but they can not provide detailed road condition information, whereas electronic maps are generated from satellite images and then located by GPS, which can achieve meter accuracy.High-Definition maps require centimeter-level accuracy, which is not enough for satellites and GPS. High-Definition[12] map is one of the core technologies of unmanned driving. Accurate map is crucial to the positioning, navigation, control and safety of unmanned vehicles. The High-Definition map should feed back road information to the vehicle such as the state of the signal lamp in front of the road and judge the state of the lane line in front of the road. The High-Definition map contains a lot of driving assistance information, including the geometric structure of the road surface, the position of the marked line and the point cloud model of the surrounding road environment. An accurate high-definition point cloud map can help the vehicle to better complete the perception and location in the process of driving, and help robots complete obstacle recognition, target tracking in the task of Dynamic Obstacle Tracking Based On High-Definition Map In Urban Scene.

With the help of the High-Definition map NDT[13] can provide high-precision UAV urban environment position, normal distribution transform (NDT) algorithm is a registration algorithm, it is applied to the 3D point of statistical model, using standard optimization techniques to determine the optimal matching between two point cloud.In the process of registration, it does not use the characteristics of the corresponding points and matching calculation, so it cost less time than other methods. Thanks to the high-precision semi-dense point cloud information collected by the lidar sensor, the positioning is robust, fast and accurate. Visual object tracking is an important research direction in computer vision, which has a wide range of applications, such as: video monitoring, human-computer interaction, unmanned driving and so on. Recently, the object tracking method based on deep learning has achieved satisfactory results and made a breakthrough in the technology of object tracking.

The tracking task of visual target is to predict the size and position of the target in the subsequent frame given the size and position of the target in the initial frame of a video sequence[14]. This basic task flow can be divided into the following framework:Motion Model,Feature Extractor,Observation Model,Model Update,Ensemble Method.

Although a good solution has been proposed by visual target tracking, the problems of Occlusion, Deformation, Background Clutter and Scale Variation have not been solved.

Through the high-precision map module, we are able to adopt the ukf-based system to conduct target tracking with the use of lidar point cloud information and NDT location information, which effectively solved these problems[15].

II. SYSTEM OVERVIEW

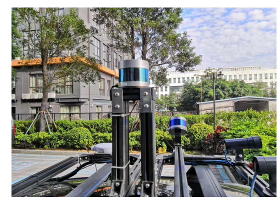
The construction of Large-Scale High-Definition Map is the foundation of modern smart city[4]. Large-Scale High-Definition Map can provide centimeter-level positioning function and accurate urban functional building location information, which plays an important role in real-time positioning, path planning and target tracking.A High-Definition point cloud Map can provide positioning services and effective road information at the same time. By eliminating map point cloud, we can obtain the initial suspected dynamic obstacle queue, which provides great convenience for dynamic obstacle detection and tracking.

The appearance of Lidar sensor makes it possible for us to construct Large-Scale High-Definition map. Different from the sparse point clouds obtained by traditional radar, the point clouds obtained by Lidar sensor are dense and highly accurate. Through the post-processing of the data, we can obtain the Large-Scale High-Definition map and location tracking point cloud clusters.The acquisition of these basic data is the starting point of this paper.

After obtaining the dynamic obstacle queue, UKF can effectively and robustly estimate the motion state information of the current point cloud cluster. Through the location distance,direction distance, point number distance,histogram distance information, the Hungarian algorithm will match the point cloud by comparing the information characteristics of cloud clusters at different points between the before and after frames. Finally, the matched points are iterated according to the prediction at the previous moment and the observation at this time.

Based on the above research, we divided the whole Dynamic Obstacle Tracking Based On High-Definition Map into three parts: High-Definition map construction in Urban Scene, unmanned device achieves accurate positioningreal-time scene task target tracking,Fig.2.We will discuss all in detail in the following sections of this article

(a)



(b)

(c)

Fig. 3. Demonstration of experimental equipment.(a) Camera. (b) Lidar (c) Test unmanned vehicle with GNSS.

III. HIGH-DEFINITION MAP AND DYNAMIC OBSTACLE TRACKING

The realization of Dynamic Obstacle Tracking Based On High-Definition Map In Urban Scene relies on the collaborative work of the two modules of positioning and tracking. The positioning module is realized through the NDT[13] matching after the generation of high-precision map. In this section, we will introduce the generation of of Multi-camera Lidar Point Cloud Map in detail, as well as the algorithm flow of the UKF system using lidar point cloud information of dynamic obstacle tracking.

A. Multi-Camera Lidar Point Cloud Map

High-Definition map generated depends on accurate data acquisition and processing, integration of the GNSS/IMU information, lidar point cloud information, IMU information, different from the traditional High-Definition map we also incorporate the camera information. At the same time in order to deal with the scenario of effective information acquisition and generation, we also used the surround camera, implements the panorama of the pixel fusion. Multi-camera lidar point cloud map information generated by the following parts:

Data-preprocessing: The camera information, GNSS information, lidar point cloud information and IMU information collected by the sensor need to be synchronized with the timestamp. As the information with different collection frequency camera is 30hz, GNSS/IMU information is as high as 100hz, while the lidar information is only 10hz. In this case, we adopted interpolation synchronization method to synchronize the timestamp Fig.2.

After data synchronization, point cloud needs to be densed according to time stamp. In this case, we adopted the previous method of dislocation projection, which realized the density by projecting the point cloud near each base point cloud together.

Point-cloud Projection: Before the point cloud projection, we need to deal with the moving obstacles in the point cloud. We adopt a deep learning model YOLO2[16] to realize this function Fig.3. The use of YOLO2 makes this complex work simple and fast. During point cloud reprojection, we need to filter the marked obstacles to ensure that there will be no obstacle points. At the same time, in order to adapt to the high-speed acquisition scene, we use the surround camera, which enables us to conduct multiple cameras during the re-projection. When the color coincidence point is encountered, we need to make a weight judgment.

The color reprojection of the processed point cloud can be realized through the parameters between camera and lidar.

B. Dynamic Obstacle Tracking

When dealing with obstacle tracking through the visual Method, we usually implement it through the five steps of Motion Model, Feature Extractor, Observation Model, Model Update and Ensemble Method. There are many problems such as Scale Variation, Occlusion and Background Clutter. Due to the combination of lidar sensor and high-precision map, we use the method based on kalman filter for iteration of the sate and use the Hungarian Match algorithm to match point cloud cluster.

Pre-processing: In the pre-processing stage, we subtracted the local map from the point cloud in each frame, then converted the center, center of gravity, point cloud and polygonal convex hull of the cluster point cloud from lidar coordinate system to local ENU coordinate system, and identified obstacles to create a tracking list. Kalman filter is used to Predict the position and speed of the center of gravity of the tracking object in the tracking list at the current moment.The properties of the observation equation are part of the properties of the equation of state and we can assume that State Equation:

$$X_t = A_{t,t-1}X_{t-1} + W_t \tag{1}$$

$W_t \rightarrow N(0, Q)$.

Observation Equation:

$$Z_t = C_t X_t + V_t \tag{2}$$

$V_t \rightarrow N(0, R)$.

Kalman filtering has two stages, Predict and Update.

Predict phase: Using the optimal estimation on the time $t-1$, X_{t-1}forecast state of the current moment :

$$X_{t,t-1} = A_{t,t-1}X_{t-1} \tag{3}$$

$X_{t,t-1}$is just to estimate the states of the time. Using the optimal covariance matrix P_{t-1} on $t-1$ to predict the current moment covariance matrix:

$$P_{t,t-1} = A_{t,t-1}P_{t-1}{A_{t,t-1}}^T + Q \tag{4}$$

$P_{t,t-1}$is not the optimal covariance on t.

Update phase:Using $X_{t,t-1}$ estimate:

$$X_t = X_{t,t-1} + H_t[Z_t - C_t X_{t,t-1}] \tag{5}$$

optimal state on t, where

$$H_t = P_{t,t-1}{C_t}^T[C_t P_{t,t-1}{C_t}^T + R]^{-1} \tag{6}$$

Using$P_{t,t-1}$ estimate

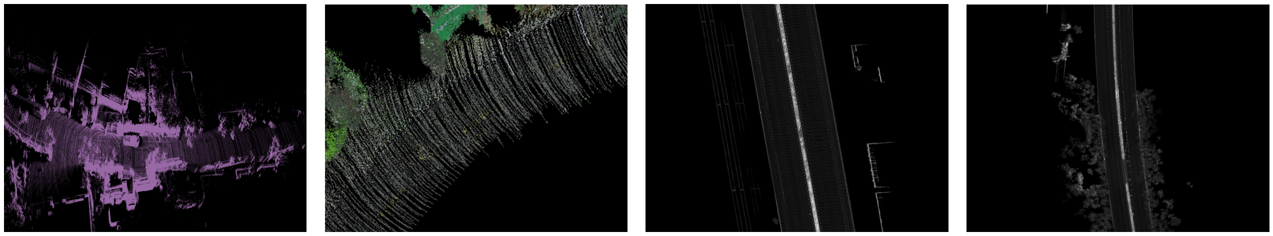
$$P_t = [I - H_t C_t]P_{t,t-1} \tag{7}$$

optimal covariance matrix on t.

Tracking: Tracking is realized through the Hungarian algorithm, which matches by calculating the score of the coordinate distance of the center of gravity position, the score of the object direction, the score of the size of the calibration box, the score of the number of point clouds and the score of the appearance features. We can assume that the barycenter coordinates of Object are (x1,y1,z1), the direction is (dx1,dy1,dz1), the size of bbox is (l1,w1,h1), and shape featrue is a 30-dimensional vector sf1, including the original point cloud number n1. TrackedObject's barycenter coordinates are (x2,y2,z2), direction is (dx2,dy2,dz2), bbox size is (l2,w2,h2), shape featrue is a 30-dimensional vector sf2, containing the original point cloud number n2.

location distance:

$$location - distance = \sqrt{(x1-x2)^2 + (y1-y2)^2} \tag{8}$$



(a) (b) (c) (d)

Fig. 4. A comparison of High-Definition map(a) The original local map. (b) Color point cloud map. (c)(d) multi-camera lidar point cloud map

direction distance:

$$cos\theta = ab/(|a||b|) \tag{9}$$

point number distance:

$$point - num - distance = |n1 - n2|/max(n1, n2) \tag{10}$$

histogram distance:

$$histogram - distance = \sum_{m=0}^{30} |sf1[m] - sf2[m]| \tag{11}$$

Hungarian Match algorithm is used to match each subgraph. Finally, the matched point clusters are iterated by Kalman filter based on the prediction at the previous moment and the observation at this time.

IV. EXPERIMENT AND RESULT

In this section, we show the process of point cloud map generation and important experimental results. We also make a set of demonstration of our final target tracking results, and the experiment proves that our algorithm can run robuly.

A. Experimental Setup

The original data collection of the high-precision map is completed in a set of vehicle-mounted equipment,Fig.3:Lidar is the HESAI PANDAR 40 Scanning Method Mechanical RotationRange 0.3 m 200 m (at 10 percent reflectivity) Range Accuracy 5 cm (0.3 m 0.5 m), 2 cm (0.5 m 200 m) Frame Rate 10 Hz20 Hz.The navigation is provided by NovAtel PwrPak7D, its multi-frequency dual antenna input allows to utilize NovAtel CORRECT with RTK and ALIGN functionality. PwrPak7D has a Global Navigation Satellite System (GNSS). The PC comes with Intel Core i7-8850H CPU @ 2.60GHz 12, and with a GeForce MX130 GPU. Point-cloud project and the dynamic obstacle detection are all running on this device.

B. High-Definition map

In Fig.4,(a) is local map with simple point cloud stitching, (b) is Color point cloud map of our previous work, and (c) (d) are multi-camera map. it can be clearly seen that compared with the local map, Color point cloud map is relatively clear and visible, but there are still a lot of dislocation and re-projection errors in the point cloud, while the multi-camera lidar point cloud map in (c) and (d) shows a straight lane line, clear ground and few noise points.

C. Tracking

Fig. 5 shows the tracking results of moving obstacles, the green dot in the graph is in the process of the scanning laser radar is considered part of the ground and fixed buildings, the alternatives of track targets to the queue in the box, after posture iterative and Hungary algorithm of kalman filtering, the match for the next moment the direction and speed of prediction. It is easy to distinguish pedestrians from vehicles in the speed relationship.

V. CONCLUSION AND FUTURE WORK

In this paper, we propose a new approach to produce multi-camera high definition point cloud map,which to be used as the local map in the urban positioning and tracking system. The multi-camera high definition point cloud map has few noise points and extremely high accuracy, which can meet the requirements of rapid acquisition and high-precision positioning in the urban scene. At the same time, we propose a traditional target tracking method based on High-Definition map, which is proved to be effective and robust by experiments. In the next step, we will combine UAV and 5G technology to realize the formation pursuit and communication of heterogeneous robots in urban environment.

ACKNOWLEDGMENT

The work described in this paper is partially supported by the National Key Research and Development Program of China (2017YFB1302303) and the NSFC-Shenzhen Robotics Research Center Project (U1613219).

The authors would also like to thank all subjects who participate in experiments, and the members of SIAT exoskeleton team from the Center for Intelligent and Biomimetic Systems.

REFERENCES

[1] Seif, Heiko G., and Xiaolong Hu. "Autonomous driving in the iCityHD maps as a key challenge of the automotive industry." Engineering 2.2 (2016): 159-162.

[2] Todeschini E. Augmented-reality signature capture: U.S. Patent 9,251,411[P]. 2016-2-2.

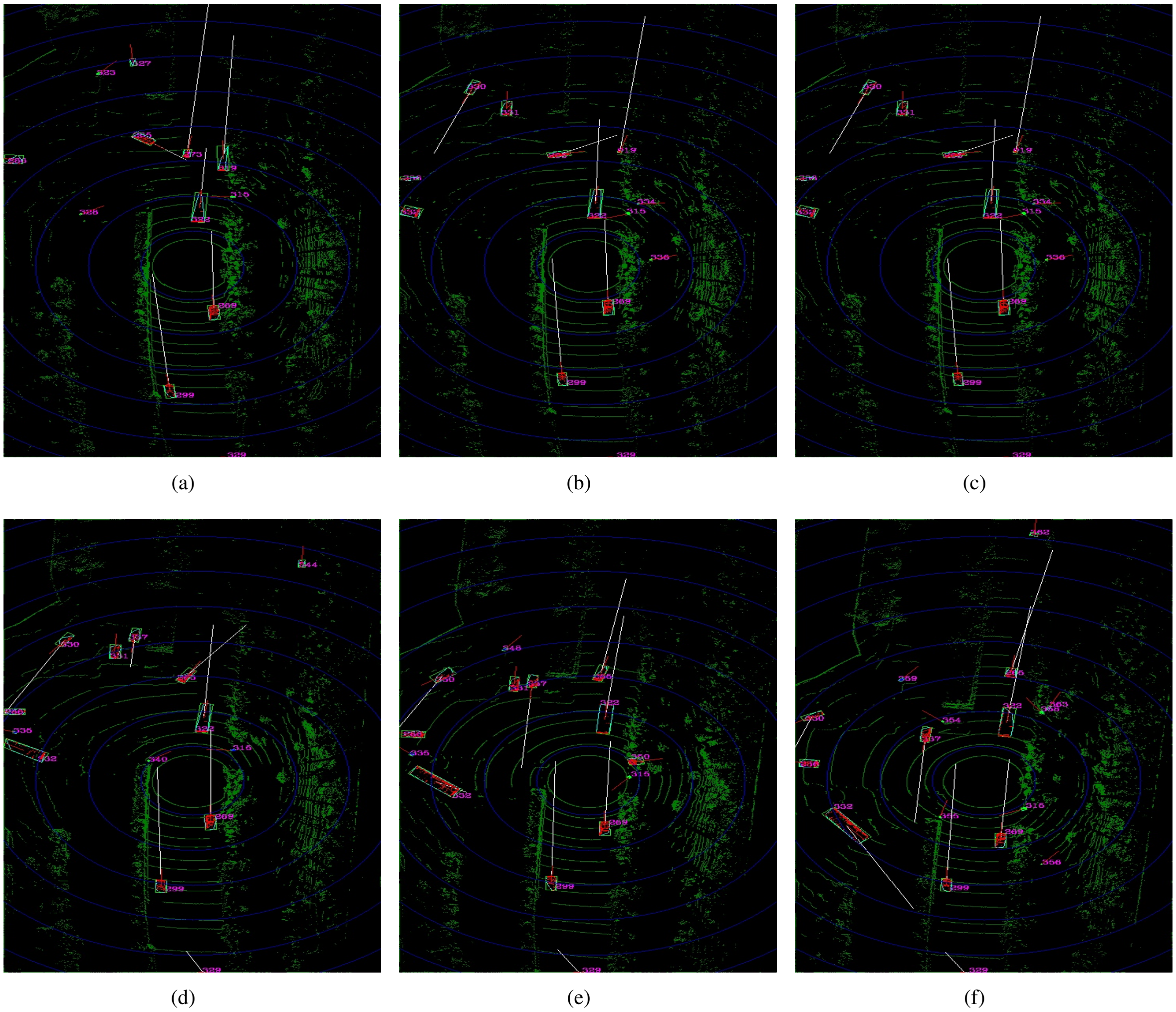


Fig. 5. Tracking results,the green dot in the graph is considered part of the ground and fixed buildings, the alternatives of track targets to the queue in the box

[3] Nagy B, Benedek C. Real-time point cloud alignment for vehicle localization in a high resolution 3D map[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 0-0.

[4] Meijer A, Bolvar M P R. Governing the smart city: a review of the literature on smart urban governance[J]. international review of administrative sciences, 2016, 82(2): 392-408.

[5] Menouar H, Guvenc I, Akkaya K, et al. UAV-enabled intelligent transportation systems for the smart city: Applications and challenges[J]. IEEE Communications Magazine, 2017, 55(3): 22-28.

[6] Montenbruck O, Steigenberger P, Prange L, et al. The Multi-GNSS Experiment (MGEX) of the International GNSS Service (IGS)achievements, prospects and challenges[J]. Advances in space research, 2017, 59(7): 1671-1697.

[7] Barfoot T D. State Estimation for Robotics[M]. Cambridge University Press, 2017.

[8] Mur-Artal, Raul, Jose Maria Martinez Montiel, and Juan D. Tardos. ”ORB-SLAM: a versatile and accurate monocular SLAM system.” IEEE Transactions on Robotics 31.5 (2015): 1147-1163.

[9] Qin, Tong, Peiliang Li, and Shaojie Shen. ”Vins-mono: A robust and versatile monocular visual-inertial state estimator.” IEEE Transactions on Robotics 34.4 (2018): 1004-1020.

[10] Graeter J, Wilczynski A, Lauer M. Limo: Lidar-monocular visual odometry[C]//2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2018: 7872-7879.

[11] Beeson, Patrick, Joseph Modayil, and Benjamin Kuipers. ”Factoring the mapping problem: Mobile robot map-building in the hybrid spatial semantic hierarchy.” The International Journal of Robotics Research 29.4 (2010): 428-459.

[12] Zhang R, Chen C, Di Z, et al. Visual odometry and pairwise alignment for high definition map creation: U.S. Patent Application 10/309,777[P]. 2019-6-4.

[13] Zaganidis A, Magnusson M, Duckett T, et al. Semantic-assisted 3D normal distributions transform for scan registration in environments with limited structure[C]//2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2017: 4064-4069.

[14] Bertinetto L, Valmadre J, Henriques J F, et al. Fully-convolutional siamese networks for object tracking[C]//European conference on computer vision. Springer, Cham, 2016: 850-865.

[15] Wan, Guowei, et al. ”Robust and precise vehicle localization based on multi-sensor fusion in diverse city scenes.” 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2018.

[16] Redmon, Joseph, and Ali Farhadi. ”YOLO9000: better, faster, stronger.” arXiv preprint (2017).