# Multiple Vehicle-like Target Tracking Based on the Velodyne LiDAR ⋆

**Liang Zhang** * **Qingquan Li** ** **Ming Li** *** **Qingzhou Mao** ***
**Andreas Nüchter** ****

* *State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, Hubei 430079 China (e-mail: zhang.liangtrc@gmail.com)*
** *Shenzhen Key Laboratory of Spatial-temporal Smart Sensing and Services, Shenzhen University, Shenzhen, Guangdong 518052 China*
*** *Engineering Research Center for Spatio-Temporal Data Smart Acquisition and Application, Ministry of Education of China, Wuhan University, Wuhan, Hubei 430079 China*
**** *Automation and Machine Vision Group, Jacobs University, Bremen, 28759 Germany*

**Abstract:** This paper proposes a novel multiple vehicle-like target tracking method based on a Velodyne HDL64E light detection and ranging (LiDAR) system. The proposed method combines multiple hypothesis tracking (MHT) algorithm with dynamic point cloud registration (DPCR), which is able to solve the multiple vehicle-like target tracking in highly dynamic urban environments without any auxiliary information from GPS or IMU. Specifically, to track targets consistently, the DPCR is developed to calculate accurately the pose of the ego-vehicle for the transformation of raw measurements taken in the moving coordinate systems into a static absolute coordinate system; while in turn, MHT helps to improve the performance of DPCR by discriminating and removing the dynamic points from the scene. Furthermore, the proposed MHT method is also able to solve the occlusion problem existing in the point cloud. Experiments on sets of urban environments prove that the presented method is effective and robust, even in highly dynamic environments.

*Keywords:* Multi-Target Tracking; Velodyne Lidar; Multiple Hypothesis Tracking; Dynamic Point Cloud Registration

## 1. INTRODUCTION

Effective environment perception is critical for intelligent vehicle to drive safely since it enables the vehicle to interpret the scene in the surroundings and subsequently take appropriate manoeuvres. In order to navigate successfully in urban environments, complex manoeuvres are required, for example merging into or out of traffic, following or passing the vehicle in front, and crossing an intersection simultaneously with vehicles from other directions. It is difficult or even impossible to handle these situations without perception of the motion of other vehicles. Therefore it is a fundamental task for intelligent vehicles to detect and track moving vehicles on the road. Many video based tracking methods have been developed in recent years (Xu et al. (2011); Benfold and Reid (2011); Rudakova (2010); Andriyenko and Schindler (2011)), however, these methods are apt to be affected by environmental condition such as weather, illumination. Comparing with video camera, the

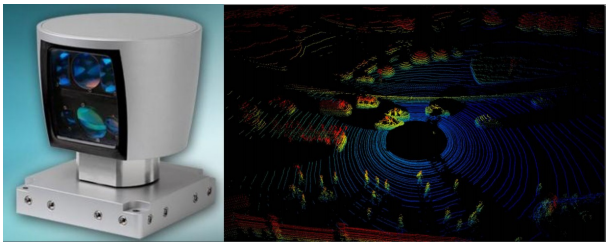

Fig. 1. Velodyne HDL-64E and its Point Cloud

3D LiDAR has many advantages: first, it can work during the day and night in any weather conditions; second, the scale of its measurements is uniform despite of distance; third, its data are easier to segment. In our research, a Velodyne HDL-64E LiDAR is employed which captures over 1.3 million points per second from the surrounding environment (Fig. 1).

Among classical tracking algorithms, the multiple hypothesis tracking (MHT, Koller and Ulmke (2005); Kuo-Chu et al. (1994); Ryoo and Aggarwal (2008); Lau et al. (2010)) has been proven an excellent approach for multiple target tracking. Unlike the Joint Probabilistic Data Association (JPDA, Lennart et al. (2011)) or Global Nearest Neighbourhood (GNN), MHT retains all possible data asso-

⋆ This work is supported by the National Natural Science Foundation of China (Grant No. 2011AA110403), the German Academic Exchange Service (PPP VR China 54368573) and open foundation of the Key Laboratory of Precise Engineering and Industry Surveying, National Administration of Surveying, Mapping and Geoinformation (Grant No.PF2011-14).

ciation hypotheses until there is enough information to resolve the assignment ambiguities occurred in the past (Thomaidis et al. (2010)). MHT is considered as one of the optimal methods for multiple target tracking especially when there exist many ambiguities during data association. Therefore, MHT is employed in this paper to handle the challenging problem of tracking multiple vehicles in urban environments with dense traffic.

To track objects consistently, the coordinates of raw measurements from different frames in the moving coordinate system need to be transformed into a static global coordinate system. Usually, some external information from auxiliary positioning and orientation sensors is used to obtain the transformation parameters. Anna and Thrun (2008), adopted GPS to estimate the ego-motion, while Thomaidis et al. (2010) integrated an inertial navigation system (INS) with vehicle odometer to deal with this problem. However, GPS is sometimes not accurate enough due to signal inaccessibility or multi-pass effects especially in urban environments where tall buildings are present. On the other hand, high-accurate INS systems are extremely expensive and their accuracy decreases with time. Wang et al. (2003) has derived the Bayesian formula of the simultaneous localization and mapping (SLAM) with detection and tracking of moving objects (DATMO) problem just based on the SICK single rangefinder and odometer, however, their method had to include a prior digital map which is might not available everywhere. With the impressive progress in the field of point clouds matching technology (Nüchter et al. (2004); Li et al. (2012); Frank and Christoph (2011)) in recent years, it is possible to accomplish the tracking task only using a 3D laser scanner without relying on GPS, INS or digital map. To achieve this goal, we incorporate dynamic point cloud registration (DPCR) technology and MHT in a single framework. In this framework the DPCR provides accurate ego-motion information for tracking while in turn, MHT will help to improve the performance of DPCR by discriminating and removing the dynamic points from the scene before applying the iterative closes points (ICP)(Besl and McKay (1992)) matching method.

The rest of this paper is organized as follows: Firstly, we give a brief overview of MHT in Sec. 2, then in Sec. 3, we begin with an introduction of DPCR and subsequently elaborates on the integration of MHT with DPCR; Sec. 4 demonstrates the experimental results in real-world urban scenarios with dense traffic flow. Finally, conclusions and remarks are given in Sec. 5.

## 2. OBJECTS SEGMENTATION

Segmenting vehicles from 3D point cloud data is the basic step of tracking. However, instead of detecting real vehicles, we just detect the vehicle-like object which is defined as all the clusters of which the geometrical features are similar to vehicle. This helps to decrease the computation time of segmentation and to cope with complicated scenarios in an urban environment with dense traffic where different degrees of occlusion exists.

The 3D laser points are first projected into a planar 2D polar grid, then the average height $\overline{H}$, the height variance ${\sigma_H}^2$, and difference $\Delta H$ between maximum and minimum

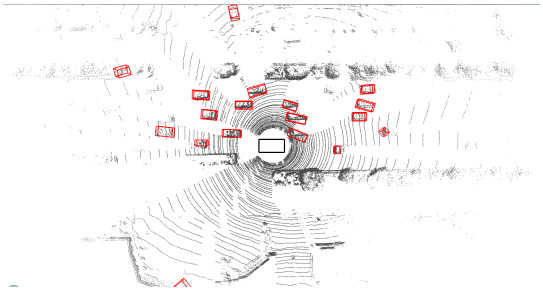

Fig. 2. Point Cloud Segmentation. The red rectangles are the segmented-out vehicle-like targets.

height of each grid cell are calculated as the grid's features which are used for classifying the grid as ground or non-ground by using Algorithm 1.

**Algorithm 1** Segmentation:Groud

IF $|\overline{H}_i - H_g| > \delta_1$
THEN Grid $i$ belongs to the non-ground object.
ELSE IF $(\sigma_H^2)_i > \delta_2$
THEN Grid $i$ belongs to the non-ground object.
ELSE IF $\Delta H_i < \delta_3$
THEN Grid $i$ belongs to the non-ground object.
ELSE Grid $i$ belongs to the ground.

where $H_g$ is the height of ground in the vehicle coordinate system which is calibrated in advance, $\delta_1$, $\delta_2$ and $\delta_3$ are thresholds.

The points which are classified as ground points are removed and other points are then clustered based on a distance criterion. Afterwards, a best fitting cube bounding box is built for each cluster. The length $l$, the width $w$ and the height $h$ of a cube bounding box is used for distinguishing whether a cluster is a vehicle-like object or not with Algorithm 2. As illustrated in the Algorithm 2,

**Algorithm 2** Segmentation:Vehicle-like Objects

IF $l_i < 1.5m$ or $l_i > 7m$
THEN Cluster $i$ belongs to the non-vehicle-like object.
ELSE IF $w_i < 1.5m$ or $w_i > 7m$
THEN Cluster $i$ belongs to the non-vehicle-like object.
ELSE IF $h_i < 1.0m$ or $h_i > 4m$
THEN Cluster $i$ belongs to the non-vehicle-like object.
ELSE Cluster $i$ belongs to a vehicle-like object.

in our assumption, the vehicle-like object is defined as the object of which the width and length are both between 1.5m and 7m, and the height is between 1.0m and 4.0m. The segmentation algorithm is deliberately designed to be simple to reduce the computation time.

If a cluster is labelled as a vehicle-like target, the centroid coordinate $(x_c, y_c)$ is calculated to form the raw measurement $Z$.

## 3. MULTIPLE HYPOTHESIS TRACKING

### *3.1 Data Association*

Data association is generally considered as the most complex problem in multiple target tracking. It identifies which

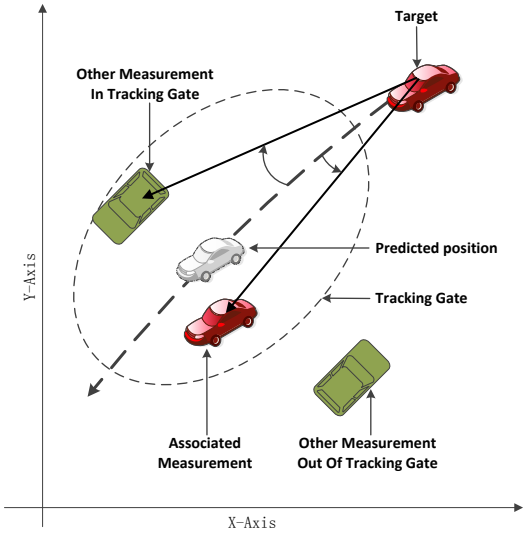


Fig. 3. Tracking Gate

measurements at different times belong to the same targets. In general, data association is considered as a 2D Linear Integer Assignment Problem (LIAP) and is mathematically represented as equation (1):

$$\min \sum_{i=1}^{N} \sum_{j=1}^{N} X_{ij} C_{ij} \tag{1}$$

$$\text{s.t.} \quad \sum_{i=1}^{N} X_{ij} = 1 \quad \text{and} \sum_{j=1}^{N} X_{ij} = 1, \tag{2}$$

where $N$ is greater value between the number of trackers and the number of measurements. $C_{ij}$ is the cost for associating measurement $j$ with tracker $i$ . $X_{ij} \in 0, 1$, if measurement $j$ is assigned to track $i$, $X_{ij} = 1$, otherwise, $X_{ij} = 0$. The cost $C_{ij}$ is computed by using equations (3) and (4). If measurement $j$ is within the gate of tracker $i$

$$C_{ij} = \alpha(\Delta l \cdot \cos(\Delta\theta)) + \beta \Delta Size \tag{3}$$

Otherwise,

$$C_{ij} = \text{BIGVALUE} \tag{4}$$

where $\alpha$ and $\beta$ are the weight values ranging from 0 to 1, $\Delta l$ is the distance difference between the line from the predicted position of tracker $i$ to its filtering position of the last scan and the line from the real measurement $j$ to the filtering position of tracker $i$, $\Delta\theta$ is the slope difference between these two lines, $\Delta_{size}$ are the difference of size between measurement $j$ and the previous measurement of tracker $i$ (fig. 3). At present, the value of $\alpha$ and $\beta$ are set based on experience.

There are many algorithms to solve the LIAP, such as Auction algorithm (Bertsekas (1991)) and JVLAP algorithm (Jonker et al. (1987)). Because of the high efficiency and optimization, the JVLAP algorithm is adopted in this paper. In particular, if $X_{ij} = 1$ in the solution but the corresponding cost $C_{ij} =$ BIGVALUE, it indicates that the tracker $i$ is lost or vanishes from sensor view or the measurement $j$ is a new target.

### *3.2 Hypotheses Generation and Pruning*

A single best solution (or hypothesis) for LIAP inevitably results in some association errors. MHT effectively mitigates association errors by retaining multiple data association hypothesises until enough evidence is accumulated to resolve the assignment ambiguities occurred in the past. This is realized through Murty's algorithm (Cox and Hingorani (1996); Katta (1968)) which generates the $k$-best hypothesises of LIAP incorporating with an arbitrary optimization algorithm. In general, there is a set of hypothesises $\Omega^{t-1}$ of the previous iteration, which is also called parent hypothesis set. $Z_t$ denotes the measurements set of scan $t$. Each parent hypothesis set contains a set of trackers and spawns $k$ new child hypothesises with $Z_t$ according to Murty's algorithm. With time going on, the number of hypothesises increases exponentially, which will arouse numerous computational burden and reduce practicality of the algorithm. Therefore, it is essential to prune hypothesises. One of the two most common hypotheses pruning methods proposes to discard hypothesises with a probability lower than a pre-defined threshold, and the other one is always keeping top $k$-best hypothesises. In this paper, the $k$-best method is employed because of its high efficiency. Fig. 4 shows the process of hypothesises generating and pruning.

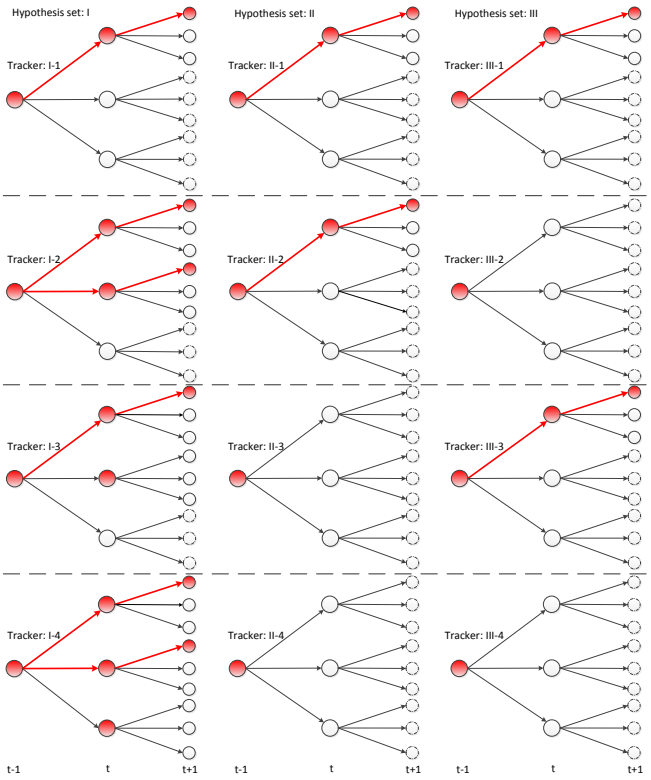


Fig. 4. Hypotheses generation and pruning. As shown in the figure, each parent hypothesis will generate three new child hypothesises at most. There will be nine hypothesises for one target at scan $t$. However, only the first three best hypotheses represented by red points in the figure with the least association cost are retained, and the others are discarded. A tracked target retains no more than three possible trajectories represented by red polylines. Four different retaining cases are shown as tracker 1,2,3,4 respectively.

### 3.3 State Estimation

The linear Kalman filter (KF) is used for each target's motion state estimation in our algorithm. The motion process and measurement process are respectively modelled as (5):

$$X(k) = AX(k-1) + v(k-1) \tag{5}$$

$$Z(k) = HX(k) + w(k), \tag{6}$$

where $X$ is the motion state vector in the static absolute coordinate system, whose specific form is $[x_{\mathrm{p}}, y_{\mathrm{po}}, x_{\mathrm{v}}, y_{\mathrm{v}}]$. $A$ and $H$ are time invariant. $v(k)$ and $w(k)$ are both white Gaussian noise sequence with covariance $Q$ and $R$ respectively. Specially, the measurement noise $w(k)$ is linked with the distance and the size.

## 4. INTEGRATION OF DPCR AND MHT

### 4.1 DPCR

DPCR is incorporated to consistently align overlapping $3D$ point clouds captured by a high-speed rotating Velodyne HDL-64 laser scanner in a dynamic environment into a static absolute coordinate system. Compared with Point Cloud Registration (PCR) for indoor or static environments, DPCR focuses on much more complicated outdoor urban streets, so it is more challenging because of the characterless ground and many potential moving objects. To accurately register point cloud in a dynamic environment, the ground and moving objects should be removed first. Then the remaining points will be processed by a fast and reliable ICP matching algorithm Borrmann et al. (2008). After ICP matching processing, the ego-motion consists of three translation parameters and three angle parameters are estimated, which can also be represented as a $4 \times 4$ matrix $Q_t^{t-1}$:

$$Q_t^{t-1} = \begin{pmatrix} R_t^{t-1} & 0 \\ T_t^{t-1T} & 1 \end{pmatrix}. \tag{7}$$

$Q_t^{t-1}$ is converted to a vector $\Gamma_t^{t-1}$ which consists of Euler angles $(\theta_x, \theta_y, \theta_z)_t^{t-1}$ and translations $(x, y, z)_t^{t-1}$. Assume that $Q_t$ is the transformation matrix from $t$ frame to the first frame.

$$Q_t = Q_t^{t-1} Q_{t-1} \tag{8}$$

Specially,

$$Q_0 = I \tag{9}$$

where $I$ is a $4 \times 4$ identity matrix.

As is well-known, the more accurate the initial value is, the higher efficient the ICP algorithm is. To accelerate the ICP matching, a 3-order polynomial regressor is adopted to predict ego-motion:

$$X(t) = a + bt + ct^2 + dt^3 \tag{10}$$

Where $X(t)$ is the ego-state in the static absolute coordinate system at scan $t$, and $(a, b, c, d)$ are coefficient vectors. When a newest ego-state is computed by DPCR, the oldest ego-state will be discarded and the newest one will be used to update the coefficients. The One-step predicted ego-state will be used as initial value for ICP matching.

### 4.2 Integration of MHT and DPCR

Existence of moving objects is the main factor affecting the precision of ICP matching. To align 3D point clouds of an environment comprising many moving objects, the problem of discriminating which part of point cloud is static and which part is dynamic has to be solved. Most solutions to PCR assume that the unknown environment is static, containing only rigid and stationary objects. Nonrigid or moving objects are just processed as outliers and filtered out. One of the most important and difficult issues of DPCR is to discriminate moving objects from stationary background. Our approach solves this dilemma by integrating MHT into the ICP matching process, which discriminates all the points in the scans into static and moving, by tracking clusters of points, and matching the static part of point cloud iteratively. The problem of combining MHT and DPCR is similar to the chicken-and-egg problem: before doing tracking, the ego-motion is needed for coordinate transformation; tracking helps to remove the moving objects, and DPCR calculates correct ego-motions if moving objects are removed correctly. Here, we use the one-step predicted ego-motion from the polynomial regressor to accompany this problem, that is, the one-step predicted ego-motion will be used as transformation parameters for coordinate system transformation as well as initial value for ICP matching.

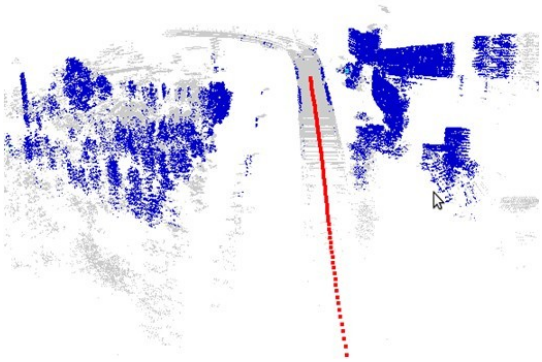

Fig. 5. DPCR Result. The red line is ego-trajectory computed by DPCR.

## 5. VALIDATION SCENARIOS AND RESULTS

The first data set for validation was collected with a Velodyne-64E laser scanner mounted on the self-driving car (Smart V-II) developed by Wuhan university. The experiment was carried out on approximate 800 meters long LuXiang Circle of LuoYu Road in Wuhan, China which is fraught with other vehicles and lasted about 1100 scans. The top part of Fig. 7 is the bird's view of the circle. and the below figures show the traffic situation acquired with a CCD camera synchronously. There are about 10 to 15 vehicles in one scan, and the number is not constant.

The left figure in Fig. 8 presents the registration result of the whole circle which contains thousand scans by the method of Point Cloud Registration based on ICP when only the ground is removed, while the right figure in Fig. 8 demonstrates the result of MHT assisted registration in which the moving points are discriminated and removed as well as the ground points. It is obvious that the quality of registration results have been significantly improved

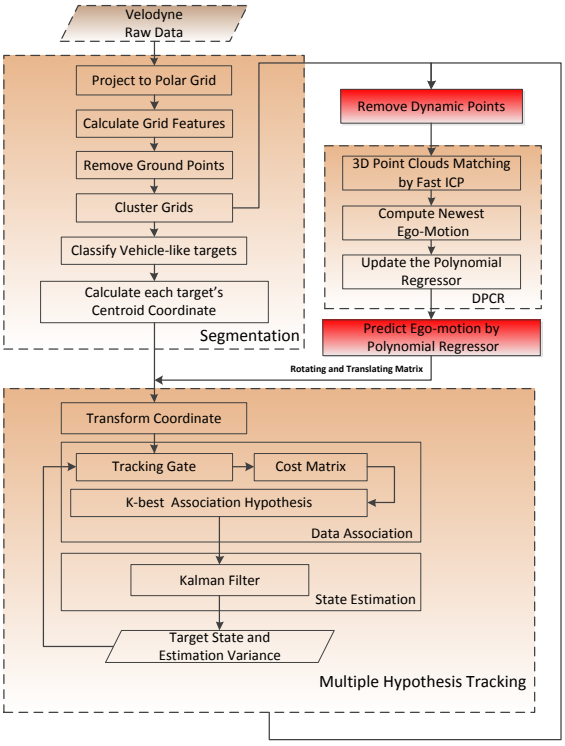


Fig. 6. The flow of proposed algorithm

Fig. 7. Bird's eye view of the LuXiang circle and typical traffic scenarios there

incorporating MHT. Fig. 9 shows the tracking process. Though the test environment was messy and full of cars, there were only 7 association errors in the entire process. The second data set we used to compare the effect of MHT and GNN was from the Karlsruhe Institute of Technology. In that data set, the ego-vehicle was static. A car stayed in front of it, which made occlusion in point cloud of each frame (Fig. 10). When other cars went into the occlusion, there would be few laser points reflected from them, which made it's impossible to abstract them from point cloud at these frames. However, we found the MHT was very robust to the occlusion. When a car went out from the occlusion, the MHT continued to track it instead of generating a new tracker, and only failed once during 160 frames time. On the contrary, the GNN was more apt to be affected by occlusion, it always started a new tracker (Fig. 11,Fig. 12).

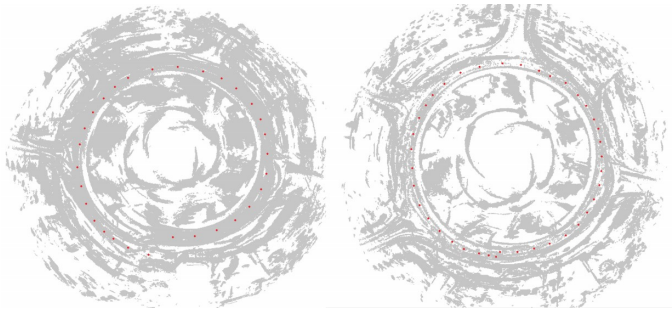

Fig. 8. DPCR Results in Luxiang Cicle

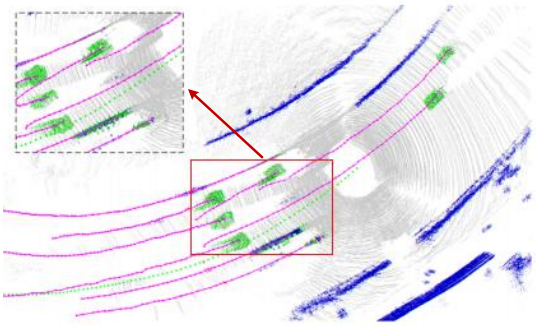

Fig. 9. A Scene of MHT tracking. The green line is ego-trajectory, and red lines are the trajectories of tracked targets.

Fig. 10. One scenario of the KIT data set

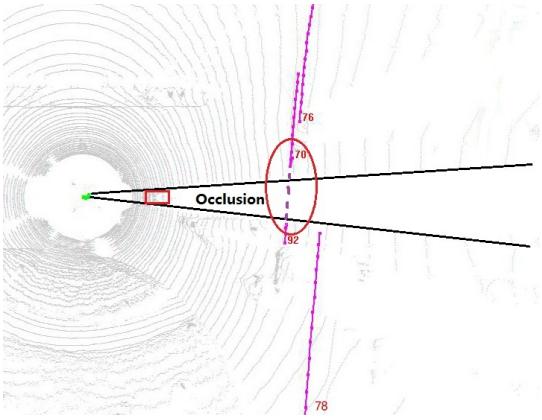


Fig. 11. GNN Tracking around Occlusion Area. When the tracker 70 went into the occlusion, it lost. And when it went out, the GNN tracking method generated a new tracker 90, which was the same car as tracker 70.

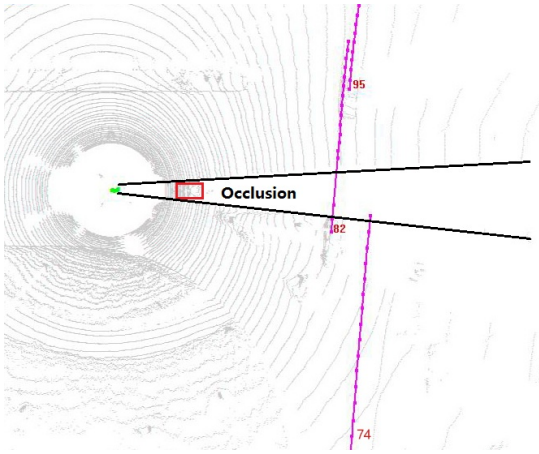


Fig. 12. MHT around Occlusion Area. The tracker 82 which was the same car as tracker 72 in Fig. 11 went into the occlusion, and then went out, the MHT continued to track it instead of starting a new tracker.

## 6. CONCLUSIONS

In this paper, we propose a MHT method incorporating with dynamic velo-registration for vehicle-like targets tracking for self-driving in urban environment. The algorithm is purely based on a Velodyne HDL64E laser-scanner. The experimental results show that while the dynamic objects being removed by MHT, the precision of DPCR is significantly improved, and in turn, the translation matrix and rotation matrix for coordinate system transformation will be more accurate. Meanwhile, the MHT adopted in our algorithm is more robust than GNN. In the future research, we will improve the performance of our algorithm in time-consuming so that it will be able to compute in real-time.

## ACKNOWLEDGEMENTS

The authors would like to thank Andreas Geiger, Philip Lenz and Raquel Urtasun from Karlsruhe Institute of Technology for providing their data.

## REFERENCES

Andriyenko, A. and Schindler, K. (2011). Multi-target tracking by continuous energy minimization. In *Computer Vision and Pattern Recognition, 2011 IEEE Conference on*, 1265 –1272.

Anna, P. and Thrun, S. (2008). Model based vehicle tracking for autonomous driving in urban environments. In *Proceedings of Robotics: Science and Systems IV*.

Benfold, B. and Reid, I. (2011). Stable multi-target tracking in real-time surveillance video. In *CVPR*, 3457–3464.

Bertsekas, D.P. (1991). *Linear network optimization: algorithms and codes*. The MIT Press.

Besl, P. and McKay, N.D. (1992). A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2), 239–256.

Borrmann, D., Elseberg, J., Lingemann, K., Nüchter, A., and Hertzberg, J. (2008). Globally consistent 3d mapping with scan matching. *Journal Robotics and Autonomous Systems (JRAS)*, 56(2), 130–142.

Cox, I.J. and Hingorani, S.L. (1996). An efficient implementation of reid's multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 18(2), 138–150.

Frank, M. and Christoph, S. (2011). Velodyne slam. In *Proceedings of the IEEE Intelligent Vehicles Symposium*, 393–398.

Jonker, R., Volgenant, A., and Amsterdam (1987). A shortest augmenting path algorithm for dense and sparse linear assignment problems. *Computing*, 38, 325–340.

Katta, G.M. (1968). An algorithm for ranking all the assignments in order of increasing cost. *Operations Research*, 16(3), 682–687.

Koller, J. and Ulmke, M. (2005). Multi hypothesis tracking of ground moving targets. In *GI Jahrestagung(2)*, volume 68GI, 307–311. GI.

Kuo-Chu, C., Shozo, M., and Chee-Yee, C. (1994). Evaluating a multiple-hypothesis multitarget tracking algorithm. *IEEE Transaction on Aerospace and Electronic Systems*, 30(2), 578–590.

Lau, B., Arras, K.O., and Wagner, B. (2010). Multi-model hypothesis group tracking and group size estimation. *JInt. Journal of Social Robotics*, 2(1).

Lennart, S., Daniel, S., Marco, G., and Peter, W. (2011). Set jpda filter for multitarget tracking. *Transctions on Signal Processing, IEEE*, 59(10), 4677–4691.

Li, M., Li, W., Wang, J., Li, Q.Q., and Nüchter, A. (2012). Dynamic veloslam: Preliminary report on 3d mapping of dynamic environments. In *2012 IEEE Intelligent Vehicles Symposium*. IEEE.

Nüchter, A., Surmann, H., Lingemann, K., Hertzberg, J., and Thrun, S. (2004). 6d slam with application in autonomous mine mapping. In *IEEE 2004 International Conference Robotics and Automation*, 1998–2003.

Rudakova, V. (2010). *Probabilistic framework for multitargettracking using multi-camera: applied to fall detection*. Gjovik University College.

Ryoo, M.S. and Aggarwal, J.K. (2008). Observe-and-explain: A new approach for multiple hypotheses tracking of humans and objects. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, 1–8. IEEE.

Thomaidis, G., Spinoulas, L., Lytrivis, P., Ahrholdt, M., Grubb, G., and Amditis, A. (2010). Multiple hypothesis tracking for automated vehicle perception. In *Intelligent Vehicles Symposium (IV), 2010 IEEE*, 1122–1127.

Wang, C.C., Thorpe, C., and Thrun, S. (2003). Online simultaneous localization and mapping with detection and tracking of moving objects: Theory and results from a ground vehicle in crowded urban areas. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*.

Xu, F., Huang, C.R., Wu, Z.J., and Xu, L.Z. (2011). Video multi-target tracking based on probabilistic graphic model. *Journal of Electronics(China)*, 28(4), 548–557.