Omni-directional detection and tracking of on-road vehicles using multiple horizontal laser scanners

Huijing Zhao, Chao Wang, Wen Yao, Franck Davoine, Jinshi Cui, Hongbin Zha

Abstract—This research aims at generating an omni-directional perception at the host vehicle’s surroundings, extracting accurate and continuous motion trajectories of the nearby vehicles using low cost laser scanners. A system of detecting and tracking on-road vehicles using multiple laser scanners is developed, where focuses are cast on solving data association of simultaneous measurements from multiple sensors at different viewpoints, and state estimation in case of partial observations in dense dynamic situations. Experimental results in freeways in Beijing are presented, system efficiency is demonstrated, where motion trajectories describing driving behaviors such as overtaking, lane changing and other interactions between driving objects are captured. In addition, the accuracy in vehicle detection and tracking is examined using a reference vehicle with a ground truth GPS.

I. INTRODUCTION

This research is motivated by two potential applications: 1) generate an omni-directional perception of the host vehicle’s local surroundings, detecting and tracking the moving objects to assist for safety driving; 2) obtain motion trajectories of surrounding traffic participants, which represent their driving behaviors and the interactions with traffic objects, so as to support for risk analysis and driving situation understanding. For the first task, online processing is required, high importance is put on the objects inside the host vehicle’s risk zone. However for the second one, offline processing is also acceptable depends on applications, while it stresses more on the quality of the extracted trajectories. In this research, we study the on-road traffic environment that is free of intersections, traffic signals, and pedestrians, such as a freeway. So that the algorithm for moving object detection and tracking considers on-road vehicles only.

There have been large body of research efforts on moving object detection and tracking. An extensive review to the visual-based approaches for on-road vehicle detection can be found at [1]. The techniques of capturing vehicle surrounds is studied in [2] and an omni-video-based approach is proposed to generate a panoramic surround map of the ego-vehicle’s immediate vicinity. A visualization tool with an iconic representation to the vehicle’s surrounding is developed in [3]. The state-of-the-art methods for visual-based object detection are regularly evaluated within the PASCAL Visual Object Classes Challenge [4]. Range-based approaches using laser scanners or radars are also widely studied. Laser scanner (also called LiDAR) is a popular sensor for such a purpose, which is usually set on the front bumper of a host vehicle doing horizontal scanning, moving objects are detected and tracked on a certain horizontal plane [5-6]. Multi-modal sensor fusion based approaches are also studied to improve the efficiency and save the failures from mono-modal sensor [9-10]. As the sensor platform moves, many researchers couple the compensation of the sensor’s ego motion (localization), obstacle detection (mapping) and moving object tracking in simultaneous frames, i.e., Simultaneous Localization And Mapping With Moving Object Tracking(SLAMMOT) [11-12], for which the pioneering work can be traced to [13]. A powerful 3D LiDAR, Velodyne HDL-64E, was first demonstrated to the public in the DARPA’s urban challenge [7-8]. It generates a high resolution 3D view of the surrounding with a frame rate up to 15Hz, so that the motions of nearby moving objects can be clearly captured. However, since the sensor need to extrude from the car, it is not welcomed by some of car manufacturers from the viewpoint of design. What’s more, the issue of sensor cost can not be ignored.

This research aims at generating an omni-directional perception at the host vehicle’s surroundings, extracting accurate and continuous motion trajectories of the nearby vehicles using low cost laser scanners. An example of the sensor layout can be found in Fig.1, where a SICK LMS291 and three HOKUYO UTM-30LX are installed on the body of the car, composing an omni-directional coverage at a horizontal plane. In order to generate a seamless coverage, overlaps between the scans of different sensors can not be avoided. On the other hand, overlapped coverage at highly important zones (e.g. front of the host vehicle) is also considered necessary to save the failures from a single sensor, and reduce occlusions. However, such a system setting brings challenges to data processing: in overlapped area, a single object might be measured simultaneously by multiple sensors from different viewpoints, an improper data handling might raise multiple alarms to a single object. On the other hand, in order to study the driving behaviors such as overtaking and interactive behaviors with other traffic participants, it is important to capture the trajectories of environmental vehicles as complete as possible. Thus the data from different sensors should be process in an integrated mode.

In this research, a system of detecting and tracking on-road vehicles using multiple horizontal laser scanners on a vehicle platform is developed. Algorithms are developed with focuses on solving data association of simultaneous measurements to single objects, and state estimation in case of partial observations in dense dynamic situations. In the
following, we first give an outline to the system architecture in section 2. This paper focuses on an algorithm for vehicle detection and tracking, where a vehicle model is defined with feature parameters as well as their reliabilities to address for partial observations. Such vehicle model is used in associating the simultaneous measurements to single objects, and in estimating the parameters of vehicle detections and tracks. The algorithm details are described in section 3. Experimental results and discussions are presented in section 4, followed by conclusions and future work in section 5.

II. SYSTEM OUTLINE

The framework of detection and tracking on-road vehicles using multiple horizontal laser scanners on a vehicle platform is shown in Fig. 2. Algorithm details of each module are explained below.

1) Data integration: is to generate an integrated frame at each processing interval by collecting the scan of the nearest time stamp from each laser scanner. In order to save memory, as well as remaining the knowledge of scanning order (i.e. scanning angle), an integrated frame is recorded in the form of sequences of raw range measurements from different sensors. While each range value can be converted to a 2D coordinate (i.e. laser point) at the vehicle’s frame according to its scanning order and the calibration parameters of the laser scanner. The laser points of an integrated frame depict the visible contours of both static and mobile objects at a time instant from the host vehicle’s view point.

2) Clustering: is to segment the sequence of range measurements from each laser scanner by analyzing “distance” between subsequent valid range measurements. A distance measure is defined as the Euclidean distance between two laser points as well as their angular interval. Segmentation is conducted if a distance is larger than given thresholds on experiences. A data cluster is considered as an observation to a single object, either moving or static one, from a single laser scanner.

3) Labeling: is to discriminate the clusters into the data of static objects, moving objects, or unsure. For these, each cluster is first converted into a global reference frame using the vehicle’s pose from a GPS/IMU navigation unit. A cluster is labeled accordingly if a match can be taken with an estimation at the previous iteration to the static environment, i.e. map, or a prediction at the current frame to the dynamic environment, i.e. moving objects. A rule-based classifier can also be applied on a prior knowledge, such as the maximal reasonable size of an on-road vehicle, or a map of road geometry. If a cluster can neither be discriminated, it is labeled as unsure. The laser scan data after filtering out those labeled as moving objects, i.e. nMo: non-moving object, or as static objects, i.e. nSo: non-static object, are forwarded to the modules of map generation or moving object detection, respectively.

4) Map generation: is to generate a map representation to the static environment along the host vehicle’s motion trajectory. The map is a composition of grid cells, the value of each cell represents for the probability that the cell be occupied by obstacles. A standard incremental mapping-updating procedure is used in this research [14], where laser hits are simply counted at each grid cell.

5) Detection: is to find vehicle candidates from the current integrated frame by grouping the data clusters of the same objects, fitting on the vehicle model, and estimate their feature parameters. Here, partial observations and overlapped measurements are the two major difficulties. In order to improve reliability in feature parameter estimation as well as reduce multiple alarms, associating the observations to single objects is necessary, which is called ”grouping” to make differentiated with the data association in tracking procedure. In this research, a vehicle model is defined by simplifying the object’s horizontal contour to a rectangular shape, a grouping and model estimation algorithms are developed accordingly. In addition, a prediction to the locations of registered vehicle tracks helps in reducing grouping ambiguities, especially for those clusters apart on different sides of a vehicle.

6) Tracking: is to associate detections with the registered tracks, and update the states of each track with the parameters. For the detections that are not associated with any existing tracks, new tracks are generated for each. For the tracks that are not associated with any moving object detections for several frames, they are considered as having gone out of the host vehicle’s vision field, and removed from the track data base.

7) Validation: is to validate a track on the sequence of motion and shape parameters. If no motion is found on a track during a certain number of frames, the track is considered as a static one, the data of which is added into the map. The inter-frame change on motion and shape parameters are examined too. If irregular changes are detected, the tracks are disposed. Only the moving tracks with regularly changed motion and shape parameters are validated, and output as vehicle trajectories.

Below we discuss on the detection and tracking algorithm. The key issues to be solved are 1) associating the data of simultaneous measurements to the same objects, and 2) estimating their feature parameters with considerations to partial observations. The definition of a vehicle model, and making use of it in vehicle detection and tracking is the core of this research.
III. VEHICLE DETECTION AND TRACKING

As the moving objects are sensed from the viewpoints on a host vehicle, only part of the objects are observed. For a large body object, such as a car, a truck or a bus, this is even crucial. For example, if a car is only partially observed on one side, if we do not assume any strong model on the car, or we do not even know that the object is a car, estimations to its size or center point could be quite unreliable. In an on-road traffic environment, partial observations may also be caused by the occlusions from other moving objects and the dynamics of the host vehicle. Partial observation could greatly affect the reliability of feature parameter estimation and could subsequently reduce accuracy in tracking their state along frames. In order to improve the reliability in estimating vehicle parameters and tracking its states, an algorithm that addresses partial observations in estimation is a key to this reliability.

In addition, a desirable detection result is that all moving objects be detected successfully with a minimal number of false/multiple alarms. Since an object might be measured simultaneously by the laser scanners at different viewpoints, and since the contour points of an object might be spatially disconnected due to occlusions or range failures, multiple detections could arise from a single moving object, which brings more confusing to the data associations in tracking procedure too. In order to reduce multiple alarms and to have more complete data for feature parameter estimation, developing an algorithm that grouping the measurements from different laser scanners into the same object is another key to this procedure.

In order to tackle the above key issues, a vehicle model is defined with feature parameters and reliability items that address on partial observations. Such a vehicle model is used in associating the simultaneous measurements to single objects to reduce multiple alarms, and in estimating the parameters of vehicle detections and tracks to improve reliability.

A. Vehicle Model Definition

Suppose a laser scanner performs counterclockwise scanning, and the horizontal contour of a car is measured by a sequence of laser points from $s$ to $e$ (see Fig.3(a)). Simplifying the shape of a car using a rectangle, edges that represent two vertical sides of the car can be detected through a corner detector and a line fitting on the laser points. A directional vector $u_i$ that is associated with each edge is defined according to the scanning order of laser points, e.g., from a point measured later to one earlier. Let $u$ denote a directional vector that is extracted from the data of a single laser scanner after its alignment to a reference frame. It has the following properties:

1) $u$ is distinct in each side of a vehicle;
2) $u$ is equal in the measurements of different laser scanners.

Thus, through matching the directional vectors $u_i$, it is easy to associate the overlapped observations, meanwhile avoid mis-grouping the data measurements on the opposite sides of the vehicle, which have similar data appearance.

Based on the above considerations, a vehicle model is defined in this research with the feature parameters shown in Fig.3(b) and listed in Tab.I, where the contour of a vehicle is simplified into a rectangle. The directional vectors, $v_i$, $i = 1, \ldots, 4$, represent for the four sides of a vehicle, which compose a counterclockwise loop. If a $u_i$ of a cluster can be associated to a $v_i$, $v_i$ is said supported by $u_i$. In addition to the feature parameters, reliability items are defined to denote whether the corresponding features are estimated on direct observations or inferred through the assumption on vehicle model. Reliability is a binary value. For example, if a neighboring sides are both supported, the corner point is a (reliable) one, otherwise it is a guess through other feature parameters.

![Fig. 3. Definition to an object model. (a) a measurement to a car from a single laser scanner; (b) feature parameters of the vehicle model](image)

### TABLE I

<table>
<thead>
<tr>
<th>Item</th>
<th>Feature</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>directional vectors ($i = 1, \ldots, 4$)</td>
<td>$v_i$</td>
<td>$r_{v_i}$</td>
</tr>
<tr>
<td>corner points ($i = 1, \ldots, 4$)</td>
<td>$c_i$</td>
<td>$r_{c_i}$</td>
</tr>
<tr>
<td>a center point</td>
<td>$p$</td>
<td>$r_p$</td>
</tr>
<tr>
<td>lengths on two vertical edges ($i = 1, 2$)</td>
<td>$l_i$</td>
<td>$r_{l_i}$</td>
</tr>
</tbody>
</table>
parameters on the object model (unreliable). Reliable corner points are useful in finding an accurate estimation to vehicle motion. In the case of edge length, if a pair of opposite sides are both supported, or a pair of neighboring corner points are both reliable, the edge length is said (reliable). An unreliable edge length will be updated by the subsequent observations, while reliable one will not. Reliable edge lengths are also useful to infer corner points.

B. Data Structure and Estimation

As shown in Fig.4, three kinds of data are addressed in the system.

- **cluster**: a sensor measurement of an object;
- **detection**: an estimation of a vehicle on a number of simultaneously measured clusters;
- **track**: a dynamic estimation of the state of a vehicle up to the current frame;

1) *Estimation of a cluster*: After receiving a scan from a laser scanner, clustering is conducted by segmenting the scan into clusters, and extracting feature parameters subsequently. Due to partial observations, a vehicle might be scanned on one or two sides, or even by few laser points. So that a cluster can be broadly divided into three types according to the number of axes that is detected by a KL transform. For a two-axes cluster, a corner detector is first conducted to divide the laser points into two sequential parts, and line fittings are then conducted on each part to extract the directional vectors \((u_1, u_2)\) and edge lengths \((L_1, L_2)\). \(P_i\) are the laser points that are associated to the edge. For a zero- and one-axis cluster, a subset of observed parameters are estimated.

2) *Estimation of a vehicle detection*: A vehicle model \(m\) is used to record the estimations to a vehicle detection \(g\), i.e. \(g = (m)\). Given a set of clusters \(\{c_j\}\), estimating a vehicle model \(m\) of a detection \(g\) is to fit parameters that are listed in Tab.I. Normally, an edge is longer, estimation of the directional vector is more reliable. We chose the cluster, which has the longest edge length, to initialize \(m\). \(m\) is then iteratively updated using other clusters, where matching between \(u_1\) of \(c_j\) with \(v_1, ..., 4\) of \(m\) is taken, so as to associate the axes \(\{u_i, L_i, P_i\}\) of \(c_j\) to the edges of \(m\). Feature parameters and their reliabilities of \(m\) are updated subsequently using the corresponding observations of the cluster.

3) *Estimation of a track*: A vehicle track contains both static (i.e. appearance) and dynamic (i.e. motion) factors. A vehicle model is used to record the estimations to the object’s appearance at the moment, while the parameters such as speed, directional vector are recorded in addition for the vehicle motion. Given a vehicle detection \(g_k = \{gm_k\}\) at the current frame \(k\), the objective is to find a posterior estimation \(t_k = (tm_k, td_k)\) to a vehicle track from its previous estimation \(t_{k-1} = (tm_{k-1}, td_{k-1})\) to the current frame, where both \(gm_k\) and \(tm_k\) are the vehicle models recording the parameter estimations on object appearance, \(td_k = (speed_k, dirv_k)\) records the vehicle motion.

C. Data Association

Data association has two fold meanings: associating the data clusters of the measurements to a single object within an integrated frame (called “grouping”); associating the vehicle candidates that are detected from the current integrated frame with a set of previously registered tracks.

1) *Grouping*: As mentioned previously, the clusters of a single vehicle might be apart far from each other, e.g. on different sides of a car, which brings big challenge to associate the clusters. In this research, predictions to the locations of registered tracks at the current frame are used to find the clusters that could be the measurements to the same vehicles. For each registered track, a region is predicted telling where measurements to the vehicle are most probably happen, the clusters that overlap with the region are extracted and grouped to fit on a vehicle model. If a vehicle model is successfully estimated (with reasonable fitting residuals), the grouping of the clusters is proved. In addition to track prediction, from the rest of clusters, a cluster is chosen to initialize a vehicle detection, it will then iteratively absorb other clusters in its local vicinity. A cluster \(c'\) can be merged to a vehicle detection, which contains a set of clusters \(\{c_i| i = 1, ...\}\), only if a vehicle model can be successfully estimated on the merged set of clusters \(\{c', c_i| i = 1, ...\}\). The procedure continues until no more clusters can be grouped.

2) *Detection-track association*: Given a detection \(g = (gm)\) and a track \(t = (tm, td)\), two vehicle models are estimated: 1) \(tm_1\) is a vehicle model generated by updating \(tm\) on \(gm\), 2) \(tm_2\) is another vehicle model by predicting the state of \(t = (tm, td)\) at the current frame on a linear dynamic model. A distance measure evaluating the likelihood of associated \(g = (gm)\) with \(t = (tm, td)\) is defined by matching \(tm_1\) with \(tm_2\).

IV. EXPERIMENTAL RESULTS

A number of experiments have been conducted in the freeways, e.g. the third to fifth ring roads in Beijing, which are free of traffic signals and intersections. As shown in Fig.1, four horizontal laser scanners are used. Three are Hokuyo UTM-30LX, which are set at the front-left, front-right, middle-rear of the vehicle, generating an omni-directional horizontal coverage to the host vehicle’s local vicinity. However, as UTM-30LX has a short range measurement, normally up to 25m in an outdoor traffic environment, a
SICK LMS291 is supplemented at the middle-front bumper to cover a semi-circle area with a radius up to 45m in front of the host vehicle. Be noted that the above range values are given according to the authors experience.

A sequence of online data processing results is shown in Fig.6. Interactions between the host (briefly "HV") with other surrounding vehicles are clearly captured. For example, the speed of HV is faster than that of vehicle #115. It gradually catches up, and finally overpasses #115 at frame 298. On the other hand, vehicle #118 is even faster. It enters the HV’s laser coverage at frame #298, overpasses the vehicle #115 at frame #309, catches up HV at frame #319, starts to make right lane change at frame #340, overtakes the vehicle #114 and successfully finishes lane change at frame #383, finally exits the HV’s laser coverage at frame #414. Such tracking results can be used to analyze driving behaviors. However, online processing contains erroneous results due to occlusions and instantaneous observations. For the applications that allow offline processing and require accurate trajectories, a two-steps procedure - map generation and moving object detection tracking - can greatly reduce erroneous trajectories.

In order to evaluate vehicle tracking results, a reference vehicle (briefly "RV") is used in the experiment, which has a GPS receiver to record its trajectory as the ground truth. Pink dots can be found at almost all frames in Fig.6-7, which denote the locations given by the ground truth GPS. The reference vehicle does not only run in front of the host vehicle as shown in Fig.6, interactions are also taken in between of two vehicles as shown in Fig.7, where HV catches up RV at frame #7234, overpasses it at frame #7360, makes lane change at frame #8190, finishes lane change and runs ahead of RV at frame #8342. During these period, the reference vehicle has been continuously tracked as the vehicle #379. The trajectory of vehicle #379 is compared with that of the ground truth GPS. The speed values between laser-based processing with those from the ground truth GPS is compared, where the difference between the values are shown in Fig.5(left). It can be found that laser-based processing results have more variations than GPS measurements. The level of variations is about 1 m/s, with an average about 0.196606 m/s. In addition, locations are also compared. Difference between the XY coordinates (within a local coordinate system) is shown in Fig.5(right). The difference might be traced to three errors: the host vehicle’s localization error, ground truth GPS error, and the vibration in laser-based data processing.

V. CONCLUSIONS

This research aims at generating an omni-directional perception at the host vehicle’s surroundings, extracting accurate and continuous motion trajectories of the nearby vehicles using low cost laser scanners. A system of detecting and tracking on-road vehicles using multiple laser scanners is developed, where focuses are cast on solving data association of simultaneous measurements from multiple sensors at different viewpoints, and state estimation in case of partial ob-

servations in dense dynamic situations. Experimental results in the freeways in Beijing are presented, where a reference vehicle with a ground truth GPS is used to examine the accuracy in vehicle detection and tracking. It is demonstrated that an omni-directional perception is generated by using a number of 2D laser scanners, and interactive behaviors between vehicles are captured. Future studies will be addressed on making use of the system in the applications of online perception and behavior analysis.

REFERENCES

Fig. 6. A sequence of data processing frames

Fig. 7. Interactions between the reference and host vehicle