

# Cooperative Raw Sensor Data Fusion for Ground Truth Generation in Autonomous Driving

Egon Ye<sup>1</sup>, Philip Spiegel<sup>1</sup>, and Matthias Althoff<sup>2</sup>

**Abstract**—Ground truth data plays an important role in validating perception algorithms and in developing data-driven models. Yet, generating ground truth data is a challenging process, often requiring tedious manual work. Thus, we present a post-processing approach to automatically generate ground truth data from environment sensors. In contrast to existing approaches, we incorporate raw sensor data from multiple vehicles. As a result, our cooperative fusion approach overcomes drawbacks of occlusions and decreasing sensor resolution with distance. To improve the alignment precision for raw sensor data fusion, we include mutual detections and match the jointly-observed static environment to support differential global positioning system localization. We further provide a new registration algorithm, where all point clouds are moved simultaneously, while restricting the transformation parameters to increase the robustness against misalignments. The benefits of our raw sensor data fusion approach are demonstrated with real lidar data from two test vehicles in different scenarios.

## I. INTRODUCTION

Ground truth data must be more reliable compared to online perception models, e.g., for evaluating object tracking and detection quality. However, previous approaches for offline ground truth generation (e.g., [1]–[4]) only post-process the sensor data of a single vehicle, which are subject to the same limitations as online methods regarding the sensor resolution in the distance and restricted sensing perspectives. Thus, we propose an offline cooperative raw sensor data fusion approach to generate ground truth with lidar data from multiple vehicles. The benefits of our cooperative fusion particularly include improved detection of object shapes, higher detection reliability, and increased awareness of occluded objects.

### A. Related Work

The classical approach to obtaining object reference trajectories relies on differential global positioning systems (DGPS) [5], which is limited to target vehicles equipped with high-cost localization devices. A cheaper alternative is to enhance recorded data offline: In contrast to online methods, one can use more precise methods since real-time capability is no longer required and one can exploit future measurements in the recordings.

Only few works on offline generation of ground truth data exist: The work in [1] combines classical fitting of characteristic shapes with offline smoothing. Vehicle trajectories are

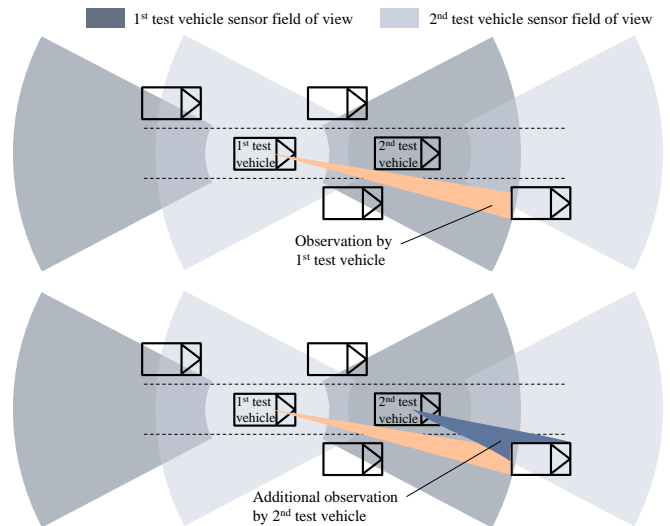


Fig. 1. Motivation for a two vehicles setup for ground truth generation based on a typical highway scenario: By fusing the additional perception of a second vehicle into the perception of the first vehicle, a more comprehensive detection of the object appearance is provided.

iteratively extracted in [2] by initializing the vehicle track at an optimal time step and extending the track both into the past and the future. An approach using a CAD model has been shown in our previous work [3], where the maximum achievable accuracy for vehicle tracking is investigated. In [4], objects are tracked forward using a standard approach based on occupancy grids. The estimates are refined using an additional backward pass with updated object size and extended using object detections in the time steps before the initialization. Since previous works suffer from occlusions and a decreasing resolution with distance when considering sensors of a single vehicle, we pursue a multiple vehicle concept instead.

Processing data from multiple vehicles has been widely investigated for data exchange between vehicles to extend their sensing range [6]. Such cooperative methods are considered as a helpful supplement to on-boards sensors [6, pp. 10–17]. Relevant research regarding cooperative perception includes the localization of communicating vehicles and sensor data fusion for detecting other traffic participants. The latter can be broadly subdivided into object-level and map-level fusion [7]. The cooperative fusion of raw sensor data [8] is an apparent extension, but not the scope of previous work due to the challenge of transmitting the required data in real time.

Fusing object lists is shown in [7], where sensor information transmitted from surrounding vehicles is treated

<sup>1</sup>Egon Ye and Philip Spiegel are with BMW Group, D-85716 Unterschleissheim, Germany. egon.ye@bmw.de

<sup>2</sup>Matthias Althoff is with the Department of Computer Science, Technical University of Munich, D-85748 Garching, Germany. althoff@in.tum.de

as virtual sensor data supporting the global object tracker for each vehicle. To handle unknown correlations between object tracks perceived by different vehicles, covariance intersection filters [9] are applied. For simulative evaluation, ideal communication and known relative positioning between communicating vehicles were assumed. Real-world deployment is challenging due to establishing temporal and spatial alignment, which are required to preprocess the sensor/object data [10]. This problem is tackled in [11] by applying point matching algorithms on object lists from different vehicles to align the perceived information.

Occupancy maps are often used as an intermediate representation in perception models between raw sensor data and object lists [12], [13]. Typically, more information is shared compared to object-level fusion requiring a higher data rate. In [14], occupancy maps of different vehicles are superimposed using Dempster-Shafer reasoning to yield a fused map. It has been further shown that erroneous estimations of vehicle states pose a major challenge in map alignment [15], [16].

Cooperative localization approaches [6, pp. 12–13] improve the pose estimation accuracy of communicating vehicles, which remedies the sensor data fusion alignment problem. In [17], a cooperative localization method is presented based on the GPS position of the ego vehicle, range sensor measurements of the distances to other vehicles, and positions communicated from detected vehicles. The fusion particularly improves the yaw estimation of the ego vehicle. A framework for cooperative localization is shown in [18] by fusing low-cost GPS, odometry, camera and lidar with data transmitted between vehicles. By sharing the polygon enclosing the ego vehicle [19], other vehicles can better localize its relative position and return the information to the ego vehicle.

### B. Contributions

This paper proposes a state estimation framework for the localization and sensor data fusion of lidars from multiple vehicles to extend the field of view (FOV) of a single vehicle. Compared to previous approaches with inter-vehicle data sharing, no limitations on the communication data rate exist for the goal of ground truth generation, which enables raw sensor data to be fused. To increase the localization accuracy for better aligning sensor data, we fuse DGPS data with measurements from mutual detections and the registration of the jointly observed static environment. For the latter, we further present a novel robust registration algorithm that matches the point clouds while preserving their initial origins. We demonstrate the performance of our fusion approach in real-world experiments with two test vehicles, for which we evaluate the alignment precision and the outcome of a tracking algorithm applied on the fused sensor data. Even though we apply object tracking, the fused sensor data can be equally used for other perception tasks like road modeling or free-space estimation.

This paper is organized as follows: Section II shows our approach for localizing the involved vehicles and for fusing

raw sensor data. In section III, we present a new registration algorithm, which restricts the transformation parameters to improve robustness. The proposed raw sensor data fusion concept is evaluated in section IV. We conclude our research work in section V.

## II. PROPOSED LOCALIZATION AND RAW SENSOR DATA FUSION FRAMEWORK

We fuse raw sensor data in a local world-fixed coordinate system as presented in Fig. 2. To this end, we require a localization of all involved vehicles, which we estimate in a centralized fusion framework. Compared to a distributed estimation system as shown in [18], centralized fusion benefits from the lapsed handling of correlations between parallel estimations. Below, we describe the transition model of our joint localization framework for an arbitrary number  $N$  of involved vehicles and give details about the considered measurements.

1) *Joint State Space Model with Dynamics Input:* We use a joint state space model with the stacked state vector

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_N \end{bmatrix}, \quad \mathbf{x}_i = \begin{bmatrix} s_x \\ s_y \\ \psi \\ v_x \\ v_y \end{bmatrix}_i. \quad (1)$$

Let us introduce the variables of the  $i^{\text{th}}$  vehicle:  $s_x$  and  $s_y$  represent the rear axle center position in a local world-fixed coordinate system (see Fig. 3),  $v_x$  and  $v_y$  are the longitudinal and lateral velocity,  $a_x$  and  $a_y$  are the longitudinal and lateral acceleration,  $\psi$  is the yaw angle, and  $\dot{\psi}$  is the yaw rate. Our nonlinear 2D kinematic model for a single vehicle  $\mathbf{x}_i$  is [20]

$$\dot{\mathbf{x}}_i = \begin{bmatrix} v_x \cos \psi - v_y \sin \psi \\ v_x \sin \psi + v_y \cos \psi \\ 0 \\ v_y \dot{\psi} \\ -v_x \dot{\psi} \end{bmatrix}_i + \begin{bmatrix} 0 \\ 0 \\ u_{\dot{\psi}} \\ u_{a_x} \\ u_{a_y} \end{bmatrix}_i + \begin{bmatrix} 0 \\ 0 \\ w_{\dot{\psi}} \\ w_{a_x} \\ w_{a_y} \end{bmatrix}_i, \quad (2)$$

with the inputs  $u_{a_x}, u_{a_y}, u_{\dot{\psi}}$  consisting of the measurements of the inertial measurement unit (IMU), and random variables  $w_{a_x}, w_{a_y}, w_{\dot{\psi}}$  accounting for the process noise of the subscribed variables. The dynamic model is discretized using the forward Euler method.

2) *Differential Global Positioning System:* The propagation of the dynamic model using IMU measurements alone results in a drift of pose and velocity states, which is corrected using absolute DGPS pose measurements. The corresponding measurement function is

$$\mathbf{y}_{i,\text{DGPS}} = \begin{bmatrix} s_{x,i} \\ s_{y,i} \\ \psi_i \end{bmatrix}_{\text{DGPS}} + \mathbf{f}_D(\mathbf{v}_{i,\text{DGPS}}), \quad (3)$$

where  $\mathbf{f}_D(\mathbf{v}_{i,\text{DGPS}})$  denotes the DGPS measurement noise.

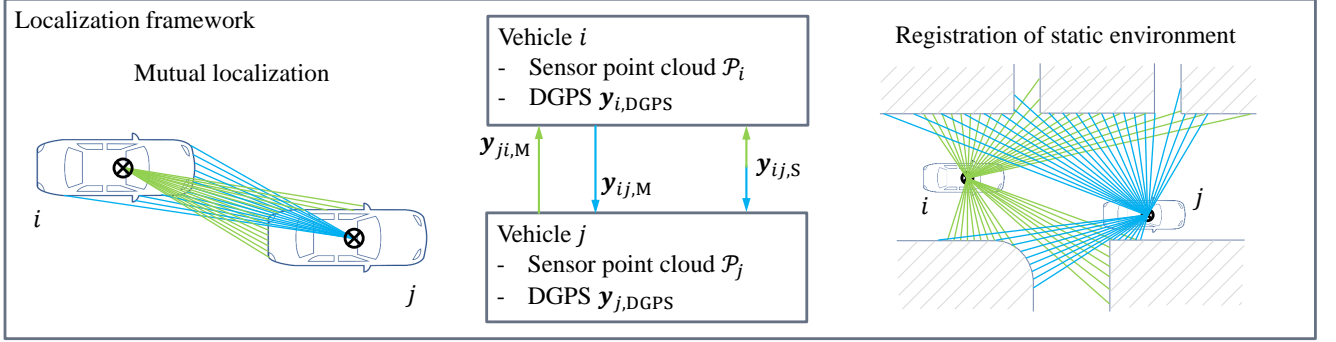


Fig. 2. Measurement relations between two vehicles  $i$  and  $j$  in our cooperative localization framework.

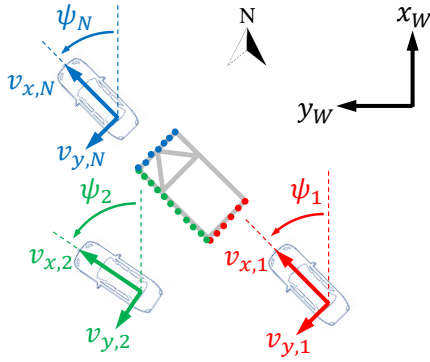


Fig. 3. Local world-fixed coordinate system.

3) *Mutual Localization*: The mutual localization (see Fig. 2) requires a direct line-of-sight between vehicles and estimates the relative poses by matching CAD models with sensor point clouds as presented in [3]. The relative pose obtained by vehicle  $i$  detecting vehicle  $j$  is considered by the measurement function

$$\mathbf{y}_{ij,M} = \begin{bmatrix} (s_{x,j} - s_{x,i}) \cos(\psi_i) + (s_{y,j} - s_{y,i}) \sin(\psi_i) \\ -(s_{x,j} - s_{x,i}) \sin(\psi_i) + (s_{y,j} - s_{y,i}) \cos(\psi_i) \\ \psi_j - \psi_i \end{bmatrix}_M + \mathbf{f}_M(\mathbf{v}_{ij,M}), \quad (4)$$

with  $i \neq j$  for  $i, j = 1, \dots, N$  and the expression  $\mathbf{f}_M(\mathbf{v}_{ij,M})$  accounting for the measurement noise. To account for erroneous registrations due to, e.g., occlusions between vehicles, measurements are discarded if the likelihood of the measurement falls below a threshold (cf. [3]).

4) *Registration of Static Environment*: A point cloud acquired by a 3D lidar usually contains many points reflected from the static environment, which is used to match point clouds from different vehicles for relative pose estimation (see Fig. 2). We use a new robust joint registration algorithm that restricts the transformation of each involved point cloud (see section III for further details), which provides the yaw angles  $\psi_i, \psi_j$  and relative position  $\Delta s_{x,ij}, \Delta s_{y,ij}$  of each

vehicle pair (without repetition) so that

$$\mathbf{y}_{ij,S} = \begin{bmatrix} s_{x,j} - s_{x,i} \\ s_{y,j} - s_{y,i} \\ \psi_i \\ \psi_j \end{bmatrix}_S + \mathbf{f}_S(\mathbf{v}_{ij,S}) \quad (5)$$

with  $i, j = 1, \dots, N$  and  $i < j$ , and noise  $\mathbf{f}_S(\mathbf{v}_{ij,S})$ . Compared to mutual localization, no line-of-sight is necessary, but an overlap of the sensor point clouds is required. Thus, both types of measurements effectively complement each other.

### III. POINT CLOUD REGISTRATION WITH RESTRICTED TRANSFORMATIONS

To match point clouds from different vehicles, we propose to use a registration algorithm that transforms all involved point clouds and regularizes their transformation parameters to increase robustness. Below, we define the preliminaries for rigid transformations and introduce our new registration method based on the well-known point-to-plane iterative closest point (ICP) algorithm [21].

#### A. Preliminaries on Rigid Transformations

Let the lidar point clouds of vehicle  $i = 1, \dots, N$  be defined as  $\mathcal{P}_i = \{\mathbf{p}_k \in \mathbb{R}^3 | k = 1, \dots, N_i\}$  with  $N_i$  being the number of points in  $\mathcal{P}_i$ . The rigid transformation in SE(3) is given by  $\mathbf{p}'_k = \mathbf{R}_i \mathbf{p}_k + \mathbf{t}_i$ , where the orthogonal rotation matrix  $\mathbf{R}_i \in \mathbb{R}^{3 \times 3}$  and the translation vector  $\mathbf{t}_i \in \mathbb{R}^3$  are obtained by

$$\mathbf{R}_i = \mathbf{R}_x(\delta\phi_i) \mathbf{R}_y(\delta\theta_i) \mathbf{R}_z(\delta\psi_i), \quad (6)$$

$$\mathbf{t}_i = [\delta x_i, \delta y_i, \delta z_i],$$

with Euler angle increments  $\delta\phi_i, \delta\theta_i, \delta\psi_i$  relating to the roll, pitch, and yaw angle, and  $\delta x_i, \delta y_i, \delta z_i$  to the incremental translation in the Cartesian frame. For compactness, we further concatenate these parameters to a vector

$$\mathbf{z}_i = [\delta\psi_i, \delta\theta_i, \delta\phi_i, \delta x_i, \delta y_i, \delta z_i]^T, \quad (7)$$

and introduce the rigid transformation function

$$g(\mathbf{z}, \mathbf{p}, \mathbf{o}^P) = \mathbf{R}(\mathbf{z})(\mathbf{p} - \mathbf{o}^P) + \mathbf{t}(\mathbf{z}) + \mathbf{o}^P \quad (8)$$

to transform a point  $\mathbf{p}$  around the origin  $\mathbf{o}^P$  of point cloud  $\mathcal{P}$ .

## B. Robust ICP with Restricted Transformations

The original point-to-plane ICP algorithm [21] registers two point clouds iteratively by alternately finding point correspondences between them and minimizing their squared perpendicular distances to the local surface. Due to the unrestricted transformation (8) of one point cloud to match the other and the proneness to converge into local minima, large transformation errors can arise depending on the initial relative pose. Thus, we propose to restrict the transformation parameters in (8) and extend the registration approach to multiple point clouds, while jointly moving the point clouds of all involved vehicles at the same time.

Let us first introduce the association function  $\mathcal{A}^{\mathcal{P}_j}(\mathbf{p}_k) = \arg \min_{\mathbf{p}_l \in \mathcal{P}_j} \|\mathbf{p}_k - \mathbf{p}_l\|$  from [21], which basically returns the closest point  $\mathbf{p}_l \in \mathcal{P}_j$  to  $\mathbf{p}_k \in \mathcal{P}_i$ . We denote the normal vector of the local surface at  $\mathbf{p}_k \in \mathcal{P}_i$  by  $\mathbf{n}_k^{\mathcal{P}_i}$  and the stacked transformation parameter vector by  $\mathbf{z}_{1:N} = [\mathbf{z}_1^T, \dots, \mathbf{z}_N^T]^T$ . Our minimization problem for the joint registration over all involved point clouds is defined by

$$\mathbf{z}_{1:N}^* = \arg \min_{\mathbf{z}_{1:N}} \sum_{i=1}^N \sum_{j=i+1}^N \sum_{\mathbf{p}_k \in \mathcal{P}_i} w_k \left[ \|\mathbf{\Gamma}_i \mathbf{z}_i\|^2 + \|\mathbf{\Gamma}_j \mathbf{z}_j\|^2 + \underbrace{\left( \mathbf{n}_k^{\mathcal{P}_i} \right)^T \left( g(\mathbf{z}_i, \mathbf{p}_k, \mathbf{o}^{\mathcal{P}_i}) - g(\mathbf{z}_j, \mathcal{A}^{\mathcal{P}_j}(\mathbf{p}_k), \mathbf{o}^{\mathcal{P}_j}) \right)}_{h_{ijk}} \right]^2, \quad (9)$$

where we include  $\|\mathbf{\Gamma} \mathbf{z}\|^2$  with Tikhonov regularization matrix  $\mathbf{\Gamma}$  [22] to restrict the transformation from the initial pose and a weighting factor  $w_k$  to handle points of dynamic objects and outliers outside the overlapping area. The restriction of point cloud transformation enhances the robustness of the registration algorithm, particularly in case of small overlaps, which otherwise would increase the transformation errors in each iteration due to erroneous point associations. However, the regularization requires that the deviation between the initial and true relative poses of the point clouds are small since a higher deviation introduces a higher bias error to  $\mathbf{z}_{1:N}^*$ .

The expression denoted by  $h_{ijk}$  in (9) is adopted from the point-to-plane ICP algorithm, which corresponds to the squared perpendicular distance of the associated point  $\mathcal{A}^{\mathcal{P}_j}(\mathbf{p}_k)$  to the local surface at  $\mathbf{p}_k \in \mathcal{P}_i$ . Note that normal vectors  $\mathbf{n}_k^{\mathcal{P}_i}$  are only taken from one point cloud in each pair, which also applies to the original point-to-plane ICP algorithm. Analogously to [21], we solve (9) using the Gauss–Newton algorithm [23, pp. 396–399] with the weighting factor  $w_k$  computed using the Welsch robust criterion function [24].

To provide an initial relative pose for our registration algorithm, we take the predicted state estimates  $s_{x,i,pred}$ ,  $s_{y,i,pred}$ , and  $\psi_{i,pred}$  with  $i = 1, \dots, N$  from section II. The results of the point cloud registration method are returned to the localization framework as relative position and yaw angle measurements of each vehicle pair (cf. section II-4),

which are defined as

$$\begin{aligned} \Delta s_{x,ij} &= (s_{x,i,pred} + \delta x_i^*) - (s_{x,j,pred} + \delta x_j^*), \\ \Delta s_{y,ij} &= (s_{y,i,pred} + \delta y_i^*) - (s_{y,j,pred} + \delta y_j^*), \\ \psi_i &= \psi_{i,pred} + \delta \psi_i^*, \\ \psi_j &= \psi_{j,pred} + \delta \psi_j^*. \end{aligned} \quad (10)$$

## IV. EXPERIMENTAL RESULTS

In our experiments, we first visualize the performance of our approach and quantify its performance using the crispness measure [25]. Secondly, we apply a simple multi-object tracker on the fused and individual point clouds from the involved vehicles to compare the tracking performance.

We acquire real sensor data using two test vehicles, each equipped with a Hesai Pandar laser scanner on a roof rack. The sensor rotates at 10Hz and provides a 3D point cloud with 40 vertical layers, a vertical resolution of  $0.33^\circ$  and  $1^\circ$  depending on the angle ranges, and a horizontal resolution of  $0.2^\circ$ . DGPS and IMU data are provided by RTK-supported DGPS systems from OxTS. We also record the DGPS and IMU data from a third test vehicle, which we consider as the target to be tracked. Each vehicle is a BMW 740Li, for which a point cloud is obtained by scanning the CAD model from 5 sides with a resolution of  $2\text{cm}$  (cf. [3]). For the evaluation, the sensor data timestamps are synchronized between the vehicles using the GPS timestamps of the DGPS-system.

DGPS information was always available during our experiments. However, short DGPS outages of a single vehicle could be bridged using the measurements from the mutual localization and the registration of the static environment.

### A. Point Cloud Alignment Quality with Localization Framework

To visually assess the performance of our raw sensor data fusion framework, Fig. 4 provides an exemplary alignment comparison of the point clouds of both vehicles. It can be seen that the alignment based solely on DGPS and IMU exhibits small heading errors leading to large position errors in the distance. Such errors are reduced using the predicted poses obtained from our localization approach. A visually perfect match of both point clouds is produced by the proposed registration algorithm. Compared to conventional registration algorithms, which transform one point cloud to match the other, the origins of both point clouds are approximately preserved.

We further quantify the performance using the crispness measure [25], which is a well-known metric in sensor calibration problems. Therein, the alignment quality of point clouds from different platform poses is determined by maximizing the smoothness of the superposed point cloud. We adapt the original metric to the matching problem by only considering point pairs between both point clouds, yielding

$$E = \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \exp \left( -\frac{\|\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j\|^2}{2\sigma_d^2} \right) \quad (11)$$

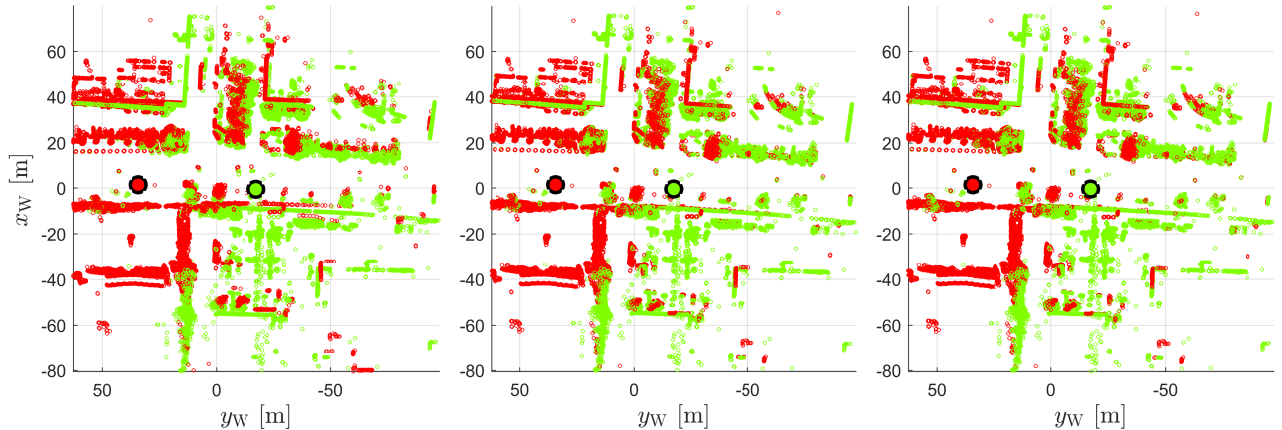


Fig. 4. Exemplary comparison of the raw sensor data fusion performance. Left: Alignment only based on the DGPS and IMU data of both vehicles. Mid: Alignment using predicted poses of our localization framework. Right: Alignment after applying the proposed registration algorithm to match both point clouds. The positions of both vehicles are marked using DGPS (left) and predicted (mid and right) poses.

with the number of points  $N_1$  and  $N_2$  of both point clouds and distance variance parameter  $\sigma_d$ . Note that this metric can be extended to more than two point clouds using pairwise computation, which is neglected here for clarity.

The results of the crispness measure are shown in Fig. 5 for a dataset of approximately 7 minutes, recorded on mixed urban and rural roads. One can see that the alignment based on predicted poses provides a higher crispness measure than solely using DGPS poses, while the jointly matched point clouds perform even better. In the time range  $100s < t < 120s$ , higher crispness measures are obtained for all three configurations due to the close distance between both vehicles, consequently leading to a higher overlap between both point clouds. Large occlusions by the environment or other traffic participants reduce the overlap between the sensor point clouds, which can be observed at  $t \approx 180s$ , where a truck drives between both test vehicles. Since ground points are removed in our input data, open fields (at  $t \approx 245s$ ) lead to a small number of lidar points, which in turn causes a dip in the crispness measure for all configurations.

### B. Evaluation based on Object Tracking Application

To demonstrate the benefits of our approach as a pre-processing step for perception algorithms, we evaluate its impact using a simple multi-object-tracker (MOT). To this end, we apply a distance-based segmentation algorithm from [26] on sensor point clouds, from which ground points are removed beforehand. The global nearest neighbor method [27] associates the centroids from the clustered point clouds to the MOT, for which we deploy the object management concept from [28] with a constant turn rate and velocity model [29]. The MOT provides an object list, from which we further extract the relevant tracks corresponding to the tracked test vehicle by associating the detected tracks to the recorded DGPS positions.

We perform the MOT on a test drive of about 14 minutes,

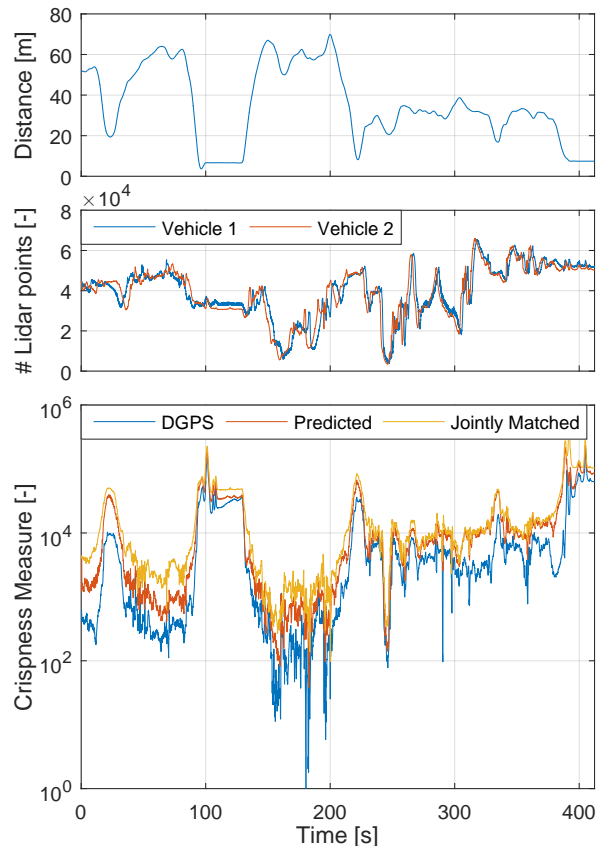


Fig. 5. Evaluation of a real test drive section. Top: DGPS Euclidean distance between both test vehicles. Middle: Number of points recorded from the lidars of both vehicles (excluding ground points). Bottom: Comparison of the crispness measure (with  $\sigma_d = 0.1m$ ) for alignments based on DGPS, predicted state estimates, and using a new registration algorithm.

which is recorded on mixed highway, urban, and rural roads. The dataset further includes multiple approaching, overtaking, and separating maneuvers of the involved vehicles. Table I shows the results of the tracking performance based on the

fused and individual sensor point clouds of both vehicles. Therein, the detection time ratio relates the duration the reference vehicle is tracked by the MOT to the time it lies in the theoretic sensor FOV (100 meters range from the sensor without consideration of occlusions). The root mean square errors (RMSE) of the velocity, yaw angle, and yaw rate are computed using the filtered estimates of the extracted vehicle tracks with respect to DGPS. The position errors are not included since our simple MOT only tracks the center of gravity as the reference point, which is prone to changes in the observed perspective.

The results show a much higher detection time ratio for the fused point cloud since a second vehicle is able to compensate both occlusions and blind spots from a single vehicle perspective. Regarding the RMSE of the state, the results using the fused sensor data performs better, except for the RMSE of the yaw angle and yaw rate compared to vehicle 1. However, one must take into account that more estimates are included for the computation of the results for the fused sensor data, which also includes inaccurate estimates from time steps when the object is not tracked by vehicle 1 at all. Note that the results mainly provide a simple qualitative comparison between the fused and individual sensor data as input to the MOT. For a more detailed comparison, more sophisticated tracking algorithms have to be used.

TABLE I  
TRACKING PERFORMANCE COMPARISON BETWEEN FUSED AND  
INDIVIDUAL POINT CLOUDS OF VEHICLE 1 AND 2.

Metric	Fused	Vehicle 1	Vehicle 2
Detection time ratio	92.1% <sup>1</sup> /94.3% <sup>2</sup> / 92.2% <sup>3</sup>	68.4%	68.7%
Yaw angle RMSE	1.03 °	0.91 °	1.45 °
Velocity RMSE	0.39 m/s	0.43 m/s	0.46 m/s
Yaw rate RMSE	2.87 °/s	2.27 °/s	3.48 °/s

<sup>1</sup>superposed theoretic FOV of both vehicles

<sup>2</sup>relates to theoretic FOV of vehicle 1

<sup>3</sup>relates to theoretic FOV of vehicle 2

## V. CONCLUSIONS

This paper presents a concept of fusing raw sensor data from multiple test vehicles for the offline generation of ground truth data. To provide accurate spatial alignments of sensor data, we refine the IMU/DGPS localization data by exploiting mutual detections of the test vehicles and pose parameters obtained from a new robust registration method that matches the lidar points from the static environment. Based on experiments with real sensor data, we demonstrated the benefits of our approach, which includes a high sensor data alignment accuracy and an improved object detection quality compared to solely using the sensor data of a single test vehicle.

Our proposed method provides an intermediate step for automatic ground truth generation, which allows developers to

subsequently apply arbitrary perception algorithms according to their needs. In the future, we would like to generate reference object trajectories using more sophisticated tracking algorithms, which more effectively exploit the availability of multiple perspectives on objects at the same time.

## REFERENCES

- [1] M. Spencer, R. Katz, and U. Lages, "Forward-backward object tracking for generation of reference scenarios based on laser scan data," in *Proc. of the 21st Intelligent Transport Systems World Congress*, 2014, pp. 3099–3105.
- [2] J. E. Stellet, L. Walkling, and J. M. Zöllner, "Post processing of laser scanner measurements for testing advanced driver assistance systems," in *Proc. of the 19th International Conference on Information Fusion*, 2016, pp. 1999–2006.
- [3] E. Ye and M. Althoff, "Model-based offline vehicle tracking in automotive applications using a precise 3D model," in *Proc. of the 22nd IEEE International Conference on Intelligent Transportation Systems*, 2019, pp. 1128–1135.
- [4] S. Hoermann, P. Henzler, M. Bach, and K. Dietmayer, "Object detection on dynamic occupancy grid maps using deep learning and automatic label generation," in *Proc. of the IEEE Intelligent Vehicles Symposium*, 2018, pp. 826–833.
- [5] M. Brahmi, "Reference systems for environmental perception," in *Automotive Systems Engineering*. Springer, 2013, pp. 205–221.
- [6] F. de Ponte Müller, "Survey on ranging sensors and cooperative techniques for relative positioning of vehicles," *Sensors*, vol. 17, no. 2, p. 271, 2017.
- [7] M. Ambrosin, I. J. Alvarez, C. Buerkle, L. L. Yang, F. Oboril, M. R. Sastry, and K. Sivanesan, "Object-level perception sharing among connected vehicles," in *Proc. of the 22nd IEEE International Conference on Intelligent Transportation Systems*, 2019, pp. 1566–1573.
- [8] W. Feiten, S. A. Bagues, M. Fiegert, F. Zhang, D. Gulati, and T. Tiedemann, "A new concept for a cooperative fusion platform," in *Proc. of the IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2016, pp. 128–133.
- [9] S. J. Julier and J. K. Uhlmann, "A non-divergent estimation algorithm in the presence of unknown correlations," in *Proc. of the IEEE American Control Conference*, 1997, pp. 2369–2373.
- [10] A. Rauch, F. Klanner, R. Rasshofer, and K. Dietmayer, "Car2X-based perception in a high-level fusion architecture for cooperative perception systems," in *Proc. of the IEEE Intelligent Vehicles Symposium*, 2012, pp. 270–275.
- [11] A. Rauch, S. Maier, F. Klanner, and K. Dietmayer, "Inter-vehicle object association for cooperative perception systems," in *Proc. of the 16th IEEE International Conference on Intelligent Transportation Systems*, 2013, pp. 893–898.
- [12] S. Steyer, G. Tanzmeister, and D. Wollherr, "Grid-based environment estimation using evidential mapping and particle tracking," *IEEE Transactions on Intelligent Vehicles*, vol. 3, no. 3, pp. 384–396, 2018.
- [13] S. Steyer, C. Lenk, D. Kellner, G. Tanzmeister, and D. Wollherr, "Grid-based object tracking with nonlinear dynamic state and shape estimation," *IEEE Transactions on Intelligent Transportation Systems*, 2019.
- [14] F. Camarda, F. Davoine, and V. Cherfaoui, "Fusion of evidential occupancy grids for cooperative perception," in *Proc. of the 13th IEEE Annual Conference on System of Systems Engineering*, 2018, pp. 284–290.
- [15] H. Li and F. Nashashibi, "A new method for occupancy grid maps merging: Application to multi-vehicle cooperative local mapping and moving object detection in outdoor environment," in *Proc. of the 12th IEEE International Conference on Control Automation Robotics & Vision*, 2012, pp. 632–637.
- [16] Y. Li, D. Duan, C. Chen, X. Cheng, and L. Yang, "Occupancy grid map formation and fusion in cooperative autonomous vehicle sensing," in *Proc. of the IEEE International Conference on Communication Systems*, 2018, pp. 204–209.
- [17] A. H. Sakr and G. Bansal, "Cooperative localization via DSRC and multi-sensor multi-target track association," in *Proc. of the 19th IEEE International Conference on Intelligent Transportation Systems*, 2016, pp. 66–71.

- [18] E. Hery, P. Xu, and P. Bonnifait, "Distributed asynchronous cooperative localization with inaccurate GNSS positions," in *Proc. of the 22nd IEEE International Conference on Intelligent Transportation Systems*, 2019, pp. 1857–1863.
- [19] —, "Pose and covariance matrix propagation issues in cooperative localization with LiDAR perception," in *Proc. of the IEEE Intelligent Vehicles Symposium*, 2019, pp. 1219–1224.
- [20] J. Farrelly and P. Wellstead, "Estimation of vehicle lateral velocity," in *Proc. of the IEEE International Conference on Control Technology and Applications*, 1996, pp. 552–557.
- [21] Y. Chen and G. Medioni, "Object modelling by registration of multiple range images," *Image and Vision Computing*, vol. 10, no. 3, pp. 145–155, 1992.
- [22] A. N. Tikhonov, "Solution of incorrectly formulated problems and the regularization method," *Soviet Mathematics Doklady*, vol. 4, no. 4, pp. 1035–1038, 1963.
- [23] Björck Ake, *Numerical Methods in Matrix Computations*, ser. Texts in Applied Mathematics. Springer, 2015, vol. 59.
- [24] P. Bergström and O. Edlund, "Robust registration of point sets using iteratively reweighted least squares," *Computational Optimization and Applications*, vol. 58, no. 3, pp. 543–561, 2014.
- [25] M. Sheehan, A. Harrison, and P. Newman, "Self-calibration for a 3D laser," *The International Journal of Robotics Research*, vol. 31, no. 5, pp. 675–687, 2012.
- [26] K. Klasing, D. Wollherr, and M. Buss, "A clustering method for efficient segmentation of 3D laser data," in *Proc. of the IEEE International Conference on Robotics and Automation*, 2008, pp. 4043–4048.
- [27] Y. Bar-Shalom and T. E. Fortmann, *Tracking and data association*, ser. Mathematics in Science and Engineering. Academic Press, 1988, vol. 179.
- [28] S. Pietzsch, "Model-based sensor data fusion of laserscanner and radar for the perception of complex vehicle environments," Ph.D. dissertation, Technische Universität München, 2014.
- [29] R. Schubert, E. Richter, and G. Wanielik, "Comparison and evaluation of advanced motion models for vehicle tracking," in *Proc. of the 11th International Conference on Information Fusion*, 2008, pp. 1–6.