Joint ego-motion estimation and multiple object tracking using automotive radar

Sen Yuan[#1], Taoyue Wang[#2], Alexander Yarovoy[#3], Francesco Fioranelli[#4]

[#]MS3 Group, Department of Microelectronics, Faculty of EEMCS, Delft University of Technology, The Netherlands
{[1]s.yuan-3, [3]a.yarovoy,[4]f.fioranelli}@tudelft.nl, [2]t.wang-32@student.tudelft.nl

***Abstract* — The problem of joint ego-motion estimation and multiple object tracking (MOT) in automotive multiple-input and multiple-output (MIMO) radar has been studied. The 3D ego-motion estimation is performed based on phase changes of the raw signal caused by relative movement between objects and the radar, and the ego-motion-induced velocities are compared with the detected ones to label static vs moving objects. The static objects are used for ego-motion estimation again to improve the accuracy, while the moving objects are used for MOT. The performance of the algorithm has been studied on simulated data and evaluated using different tracking algorithms, proving the feasibility of this approach.**

***Keywords* — Ego-motion estimation, Multiple object tracking, MIMO radar, signal processing.**

I. Introduction

RADAR can provide accurate and direct measurements of the range, relative velocity, and angle of multiple objects, as well as a long-range coverage of over 200 meters even in challenging weather or lighting conditions, outperforming other sensors, namely, camera and Lidar. Thus, radar has attracted significant importance for autonomous driving.

To measure the vehicle's own motion with radar, namely 'ego-motion estimation', state-of-the-art methods can be mainly divided into detection point cloud-based [1], [2] and intermediate frequency signal-based approaches [3], [4]. Performing ego-motion estimation starting from the lower signal level (i.e., the radar base-band signal before range-Doppler processing) can be beneficial in automotive scenarios. Firstly, the ego-motion estimation can be performed fast, within one frame. Secondly, using algorithms implemented directly at the signal level, it is easier to combine them with other high-resolution imaging algorithms.

For self-driving vehicles, the problem of multiple object tracking (MOT) for moving objects in proximity is a critical component to ensure situational awareness and safe planning and control [5]. Various MOT algorithms are used in different tracking applications, the most popular ones being GNN (Global Nearest-Neighbor) [6], JPDA (Joint Probabilistic Data Association) [7], MHT (Multiple Hypotheses Tracking) [8], as well as deep learning methods [9]. GNN, JPDA, and MHT are conventionally used in MOT systems. These methods utilize pruning and merging techniques to manage the growing complexity of tracking multiple objects and to ensure computational efficiency.

However, jointly estimating ego-motion and the movement of objects into a single process can be beneficial for performance improvement and efficiency. Extensive research has been conducted on Camera and LIDAR [10], [11]. However, to our knowledge, not much literature is available on radar data, especially those operating at a raw signal level in the automotive context. One of the challenges is that the estimation of the ego-motion is based on a single frame, while the MOT is done based on the processing of multiple frames. Meanwhile, ideal ego velocity estimation considers static objects, while MOT focuses on only moving objects. This paper addresses this joint problem by tackling the challenges associated with the proposed processing pipeline. The proposed pipeline is verified with numerical simulations and related metrics, showing the feasibility of this approach.

The rest of the paper is organized as follows. In Section II, the fundamental of ego-motion estimation and MOT is discussed. The proposed method and evaluation metrics are given in Section III. The simulation results are provided in Section IV. Finally, conclusions are drawn in Section V.

II. Fundamentals

A. Radar-based ego-motion estimation

The ego-motion estimation is implemented based on the approach in [3], as shown in Fig. 1. The whole process is divided into two steps, initial estimation and iterative estimation. For initial estimation, 2D FFT (Fast Fourier Transform) and 2D CA-CFAR (Cell Averaging Constant False Alarm Rate) detection are performed for the first group of chirps, locating the detected objects. Then, an optimization is performed to find the azimuth and elevation angles of such objects. The same operations are performed for another chirp group, and an initial, coarse value of ego velocity is estimated based on the information of all detected objects, including static and moving ones, as detailed in [4]. However, the moving objects introduce extra Doppler components, degrading the estimation of the ego velocity. Thus, an iterative ego-motion estimation is proposed by comparing the value of object velocity derived from the ego-motion estimation with each detected velocity, thus improving the distinction between static vs moving objects. The static objects are used in the next step of ego-motion estimation until a certain threshold is met and a final value of ego velocity is obtained.

B. Multiple object tracking (MOT)

Given a radar point cloud after detection, the goal of MOT algorithms is to determine the number of moving objects

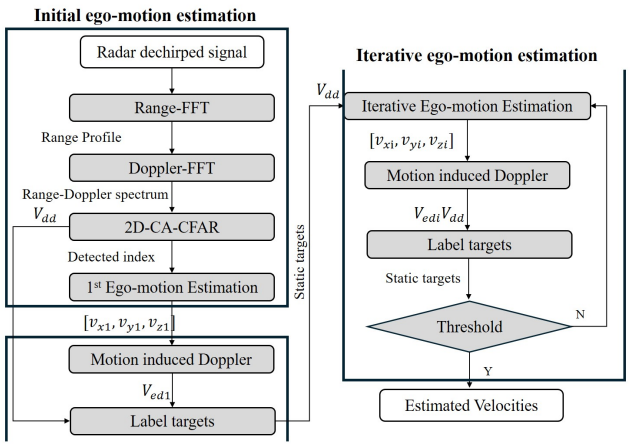

Fig. 1. Block diagram of the ego-motion estimation algorithm inspired by [3].

and their state over time. The state variable of each object includes its 3D position and velocity. The output is a set of tracking trajectories for all moving objects, each associated with a unique object identity. Fig. 2 shows the functional modules within MOT algorithms. First, clutter removal is typically implemented to reduce the contributions caused by static objects, together with a conversion of the detections from the radar coordinate to the world coordinate. Next, clustering is performed to group the detections of the same object. For each time step, gating rejects invalid detections that are too far from the existing object tracks. Only valid detections enter the data association module, which assigns detections to each track. With the prediction step, assigned tracks are updated with the corresponding detections inside the state estimation module. Unassigned tracks are kept, only performing prediction but without updates. Unassigned detections are sent to the track management module for track initialization. Tracks after state estimation are also sent to the track management module for track confirmation and deletion. Finally, the estimated states for each confirmed object track are provided in the output.

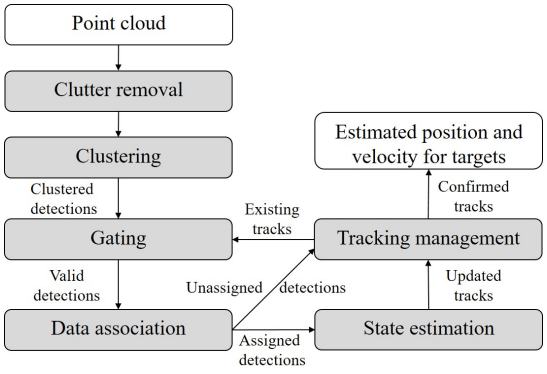

Fig. 2. Block diagram of a typical MOT algorithm

III. Proposed Method & Evaluation Metrics

As discussed in Section II, ego-motion estimation considers only static objects, while MOT considers moving objects. Thus, there is an advantage in formulating a joint approach for both problems, using all objects wisely and providing the trajectory of both the ego-vehicle and the moving objects at the output.

A. Proposed method

This paper proposes a systematic processing pipeline to jointly approach ego-motion estimation and MOT in automotive radar. The proposed pipeline is shown in Fig. 3. An initial ego-motion estimation is first implemented to estimate coarse ego velocities for each time step and generate the point cloud for the MOT algorithm.

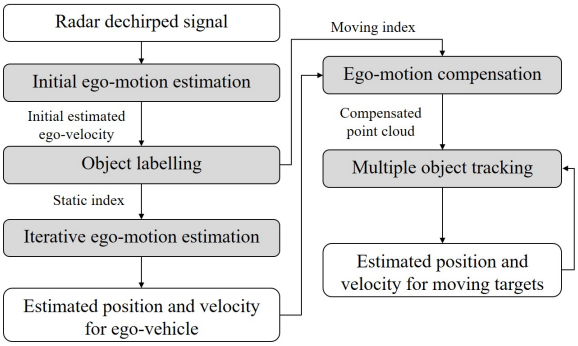

Fig. 3. Block diagram of the proposed joint algorithm for ego-motion estimation & MOT

For this joint formulation where multiple moving & static objects exist in the environment and tracking moving objects is also a part of the main task, static objects act as a major clutter source. To remove them, an object labelling step is used to distinguish moving objects from static objects, which is also used in the iterative ego-motion estimation. The decision logic is to compare the velocity difference between the detected velocity and the ego-motion-induced velocity with a fixed threshold, as follows:

$$\begin{cases} \text{Label moving object} & \text{if } |V - V_{induced}| > \text{Threshold} \\ \text{Label static object} & \text{if } |V - V_{induced}| \leq \text{Threshold} \end{cases} \tag{1}$$

where the ego-motion-induced velocity $V_{induced}$ is calculated by projecting the ego-motion estimation results to the objects' directions, and V denotes the velocity directly obtained after Doppler processing. The threshold is chosen as 1.25 m/s after empirical verifications. Using only the detections of static objects, an iterative ego-motion estimation is performed to obtain the final ego-motion estimation results.

After extracting all detections labelled as moving objects, the clutter caused by static objects is removed. Before using these point clouds in the MOT algorithm, ego-motion compensation is needed. Since the point cloud measures objects in the moving radar coordinates, the detections corresponding to moving objects should be compensated for the radar ego-motion to convert the detections into the world coordinates. In Equation (2), detections after ego-motion compensation are obtained with known ego-motion estimation results in $[V_x^{ego}, V_y^{ego}, V_z^{ego}]$.

$$\begin{bmatrix} dx \\ dy \\ dz \\ v_d \\ p_o \end{bmatrix} = \begin{bmatrix} R * \cos(\theta)\cos(\phi) + V_x^{ego} * \text{Time} \\ R * \sin(\theta)\cos(\phi) + V_y^{ego} * \text{Time} \\ R * \sin(\phi) + V_z^{ego} * \text{Time} \\ |V - V_{induced}| \\ p \end{bmatrix} \quad (2)$$

Specifically, each detection consists of the 3D position $[dx, dy, dz]$, the Doppler velocity v_d and the reflected power p_o. It is worth noting that object labelling and ego-motion compensation link jointly the ego-motion estimation algorithm and MOT. The performance of ego-motion estimation will directly affect the performance of MOT. In this paper, two different MOT algorithms are implemented for comparisons: GNN [6] based on hypotheses pruning, and JPDA [7] based on hypotheses merging.

B. Evaluation metrics

Two metrics are used to evaluate the performance of ego-motion estimation algorithms in 3D space: Absolute Pose Error (APE) and Relative Trajectory Error (RTE).

APE measures the pose difference between the estimation and the true motion for each frame. As shown in (3), it calculates the RMSE between the estimated poses and the actual poses.

$$\epsilon_{APE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m} \|P_{est}(i) - P_{actual}(i)\|^2} \quad (3)$$

where m is the total number of frames, P_{est} and P_{actual} are the estimated velocities and actual velocities, respectively. L-2 norm is used to calculate the Euclidean distance.

While APE focuses on the instantaneous velocity estimation error, RTE measures the long-term localization error. It measures the relative position error over short segments. This is the RMSE of the differences between the relative displacements over a small period in the estimated trajectory and the actual trajectory. As in Equation (4), T_{est} and T_{actual} are the estimated positions and actual positions, respectively. N is the segment interval, which is set to 10 frames (0.05 seconds) in this study.

$$\epsilon_{RTE} = \sqrt{\frac{1}{m-N}\sum_{i=1}^{m-N} \begin{pmatrix} T_{est}(i+N) - T_{est}(i) \\ -T_{actual}(i+N) - T_{actual}(i) \end{pmatrix}^2} \quad (4)$$

To evaluate the MOT results, the Generalized Optimal Sub-Pattern Assignment (GOSPA) [12] metric is used. A list of estimated object 3D position vectors is obtained from the MOT method. Each element represents the tracking position of one object. The GOSPA metric measures the difference or error between the output list and the ground truth list, as:

$$GOSPA(\mathbf{X}, \mathbf{Y}) = \left[\min_{\gamma \in \Gamma} \left(\sum_{(i,j)\in\gamma} \begin{matrix} d(x_i, y_j)^p + \\ \frac{c^p}{2}\left(|\mathbf{X}| - |\gamma| + |\mathbf{Y}| - |\gamma|\right) \end{matrix} \right) \right]^{1/p} \quad (5)$$

where X and Y are two data lists; γ is the global association hypothesis which assigns elements from X to Y. Γ is the set of all possible global association hypotheses. There are three hyperparameter choices for the metric: the order of distance p, the distance metric $d(x, y)$, and the maximum allowable localization error c. The hyperparameter p is set to 2 to make the localization error component the same as the RMSE. The Euclidean distance is chosen as distance metric d. The hyperparameter c determines the trade-off between the localization error component and the missed detection and false alarm number component, and can be considered as the distance where the designer wants to penalize a false or missing estimate. In automotive vehicle perception applications, c is commonly set to 10 meters [13].

IV. Results and Discussion

To validate the performance and show the feasibility of the proposed method, results based on simulations are presented.

An automotive MIMO radar with eight virtual array elements for azimuth and eight for elevation estimation is considered here. The omnidirectional antenna pattern is considered for the transmitter and receiver. A typical parameter setting is used for the FMCW waveform, with 77.5 GHz centre frequency, $20\mu s$ pulse repetition time and 1 GHz bandwidth. The range resolution under this setting is equal to 0.15 m, and the Doppler resolution is 0.378 m/s.

The simulated scenario consists of a car equipped with a side-looking radar moving at a constant speed, alongside multiple static and moving objects. The ego-vehicle is assumed to travel forward at a constant velocity of 12 m/s (43.2 km/h), with zero velocity in both the cross-forward and elevation directions, an assumption that aligns with real-world driving conditions on well-maintained roads. Within the radar's field of view, 15 static objects and 3 moving objects are generated. Each object is modeled as a collection of scatterers randomly distributed in 3D space. The range of these scatterers is selected from $[0, 35]$ meters, the azimuth angle from $[-60, 60]$ degrees, and the elevation angle from $[-30, 30]$ degrees. The amplitude of all scatterers is drawn from the uniform distribution $\alpha_o \sim \mathcal{U}(0, 300)$. According to the Swerling Model I the amplitude can be seen as constant during one coherent processing interval. The scatterers are also assumed to be isotropic and provide constant amplitude and phase during the processing period. The static objects in this scenario represent common roadside elements such as trees, traffic signs, or parked vehicles. Additionally, three moving objects are simulated, representing a car, a bicycle, and a pedestrian. These objects move at constant speeds of 9.11 m/s (33 km/h), 4.12 m/s (15 km/h), and 1.41 m/s (5 km/h), respectively. Their trajectories are designed to capture diverse interactions with the ego-vehicle: the car moves in a nearly parallel direction, the bicycle approaches from the opposite direction, and the pedestrian crosses perpendicularly to the vehicle's path.

To evaluate the precision and robustness of the proposed joint algorithm, simulations were performed in four independent scenarios at SNR levels ranging from +20 dB

to -5 dB, and the results were averaged at each level. In each scenario, the positions of static and moving objects were randomly generated and each simulation lasted 2.56 s (equivalent to 500 consecutive frames). The quantitative performance metrics defined in Section III-B for different SNR levels are shown in Table 1. With the proposed joint method, both the estimation of ego-motion and the tracking of multiple objects can perform a decent estimation even under high noise condition, i.e., SNR = 0 dB. It can be observed that while both the estimation of ego-motion and the performance of MOT decrease with decreasing SNR as expected, the estimation of ego-motion is more stable compared with MOT. The GNN algorithm outperforms JPDA in high-SNR scenarios, but performs worse in low-SNR scenarios. Since the ego-motion estimation remains consistent over consecutive frames and the clutter removal step effectively identifies most static objects, the data association problem is simplified, and both GNN and JPDA MOT algorithms yield good performance.

Table 1. Performance evaluation with different SNR in 3D space

SNR (dB)	APE (m/s)	RTE (m)	Mean GOSPA (GNN)	Mean GOSPA (JPDA)
+20	0.3058	0.0090	2.868	2.995
+15	0.3204	0.0094	3.239	3.681
+10	0.3349	0.0102	3.413	3.283
+5	0.3557	0.0102	3.765	4.698
0	0.4740	0.0123	4.582	5.024
-5	0.8418	0.0213	15.474	13.303

For visual evaluation, the estimated trajectories for the ego-vehicle and moving objects from an example scenario at the 20 dB SNR level are shown in Fig. 4. The figure only displays results in the X and Y dimensions for simplicity in visualization, while all quantitative metrics are computed in the full 3D space. Since the GNN and JPDA algorithms provide similar results, only the results of GNN are visualized here. One can observe the trajectories of the ego-vehicle and three moving objects are all accurately estimated, with small deviations from the ground truth and no false tracks.

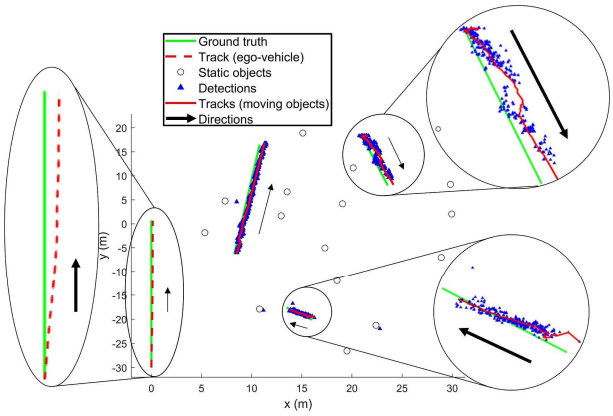

Fig. 4. Ground truth, detections and estimated trajectories for the ego-vehicle and moving objects from an example scenario at the +20 dB SNR level

V. Conclusion

A processing pipeline is proposed to solve the problem of joint ego-motion estimation and MOT in automotive MIMO radar. This pipeline links ego-motion estimation with MOT by labeling static and moving objects to perform ego-motion compensation. Static objects are used for ego-motion estimation, while moving ones are used for MOT algorithms. The performance of the proposed approach has been studied in simulations and the evaluation has been implemented on GNN and JPDA algorithms for feasibility. Promising results are shown for the feasibility of the proposed method.More detailed results can be found in [14]. Future work will involve adopting more realistic motion models, as well as expanding the pipeline to support multiple extended object tracking.

References

[1] D. Kellner, M. Barjenbruch, J. Klappstein, J. Dickmann, and K. Dietmayer, "Instantaneous ego-motion estimation using doppler radar," in *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*. IEEE, 2013, pp. 869–874.

[2] M. Rapp, M. Barjenbruch, M. Hahn, J. Dickmann, and K. Dietmayer, "Probabilistic ego-motion estimation using multiple automotive radar sensors," *Robotics and Autonomous Systems*, vol. 89, pp. 136–146, 2017.

[3] S. Yuan, D. Wang, F. Fioranelli, and A. Yarovoy, "Improved accuracy for 3d ego-motion estimation using automotive fmcw mimo radar," in *2024 IEEE Radar Conference (RadarConf24)*. IEEE, 2024, pp. 1–6.

[4] S. Yuan, S. Zhu, F. Fioranelli, and A. G. Yarovoy, "3-d ego-motion estimation using multi-channel fmcw radar," *IEEE Transactions on Radar Systems*, vol. 1, pp. 368–381, 2023.

[5] K. Granström and M. Baum, "A tutorial on multiple extended object tracking," *Authorea Preprints*, 2023.

[6] P. D. Konstantinova, A. Udvarev, and T. Semerdjiev, "A study of a target tracking algorithm using global nearest neighbor approach." in *Compsystech*, vol. 3, 2003, pp. 290–295.

[7] Y. Bar-Shalom, F. Daum, and J. Huang, "The probabilistic data association filter," *IEEE Control Systems Magazine*, vol. 29, no. 6, pp. 82–100, 2009.

[8] S. S. Blackman, "Multiple hypothesis tracking for multiple target tracking," *IEEE Aerospace and Electronic Systems Magazine*, vol. 19, no. 1, pp. 5–18, 2004.

[9] Z. Pan, F. Ding, H. Zhong, and C. X. Lu, "Ratrack: Moving object detection and tracking with 4d radar point cloud," *arXiv preprint arXiv:2309.09737*, 2023.

[10] J. Zhang, M. Henein, R. Mahony, and V. Ila, "Robust ego and object 6-dof motion estimation and tracking," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, pp. 5017–5023.

[11] X. Tian, Z. Zhu, J. Zhao, G. Tian, and C. Ye, "Dl-slot: Tightly-coupled dynamic lidar slam and 3d object tracking based on collaborative graph optimization," *IEEE Transactions on Intelligent Vehicles*, vol. 9, no. 1, pp. 1017–1027, 2023.

[12] A. S. Rahmathullah, Á. F. García-Fernández, and L. Svensson, "Generalized optimal sub-pattern assignment metric," in *2017 20th International Conference on Information Fusion*. IEEE, pp. 1–8.

[13] Á. F. García-Fernández, J. L. Williams, K. Granström, and L. Svensson, "Poisson multi-bernoulli mixture filter: Direct derivation and implementation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 4, pp. 1883–1901, 2018.

[14] T. Wang, "Combined ego-motion estimation and multiple extended object tracking with automotive radar," Master's thesis, Delft University of Technology, 2025.