Radar-only Instantaneous Ego-motion Estimation Using Neural Networks

Zhu, Simin; Fioranelli, Francesco; Yarovoy, Alexander

DOI
10.23919/EuRAD58043.2023.10289411

Publication date
2023

Document Version
Final published version

Published in
Proceedings of the 2023 20th European Radar Conference (EuRAD)

Radar-only Instantaneous Ego-motion Estimation Using Neural Networks

Simin Zhu[1], Francesco Fioranelli[2], Alexander Yarovoy[3]
Department of Microelectronics, Delft University of Technology, The Netherlands
{[1]S.Zhu-2, [2]F.Fioranelli, [3]a.yarovoy}@tudelft.nl

***Abstract*— The problem of 2D instantaneous ego-motion estimation for vehicles equipped with automotive radars is studied. To leverage multi-dimensional radar point clouds and exploit point features automatically, without human engineering, a novel approach is proposed that transforms ego-motion estimation into a weighted least squares (wLSQ) problem using neural networks. Comparison with existing methods is done using a challenging real-world radar dataset. The comparison results show that the proposed method can achieve better performance in terms of estimation accuracy, long-term stability, and runtime performance compared to a representative approach selected from the recent literature.**

***Keywords*— Automotive Radar, Ego-motion Estimation, Radar Point Cloud, Deep Learning**

I. Introduction

Methods for ego-motion estimation are extremely important for advanced driver assistance systems (ADAS) and their performance affects many downstream applications such as mapping [1], multi-object tracking (MOT) [2], and path planning [3]. Although vehicle motion can be provided by odometry sensors such as wheel encoders, inertial measurement units (IMU), and global positioning systems (GPS), these sensors are known to be affected by wheel slippage [4], cumulative error [5], and reception conditions [6], respectively.

Therefore, perception sensors such as cameras [7], light detection and ranging (LiDAR) [8], scanning radar [9], and automotive radar [10] have been actively used for ego-motion estimation. Compared with other sensors mentioned above, automotive radar can provide unparalleled advantages. First, automotive radar has been proven to work in all weather and light conditions. Also, radar can see through non-metallic objects such as trees and is less susceptible to occlusion than cameras and LiDAR. Moreover, compared to scanning radars, automotive radars are usually compact and lightweight enough to fit behind cars' bumper.

Due to these advantages, several approaches have been proposed attempting to use automotive radars for ego-motion estimation. In summary, these previous approaches can be classified into two categories. The first category is the so-called scan-match methods [11]. They were originally developed to solve the LiDAR-based ego-motion estimation problem and some of them have been modified for radar data [12], [13], [14]. With a single radar, these methods can estimate the full 2D motion (lateral, longitudinal, and rotational velocities) of the vehicle and can exploit other object features such as radial velocity [12] and returned signal power [14]. However, they require stable object detection for data association, their result is sensitive to initialization, and at least two valid scans are needed.

The second type of approach is the instantaneous methods [15], [16], which use radial velocity and angle measurements, and only require one radar scan. However, instantaneous methods can estimate partial 2D motion only (longitudinal and rotational velocities) and are unable to exploit other object features captured by automotive radars. Besides the aforementioned disadvantages of each method, both method types implement an iterative estimation process, which makes them difficult to apply to real-time applications.

To address these gaps, this work proposes a weighted least squares (wLSQ) scheme for radar-based ego-motion estimation. Notably, this is the first time that the task of instantaneous ego-motion estimation is formulated as a wLSQ problem using neural networks (NNs). More specifically, different from previous works [15], [16], the proposed method uses NNs to predict the weight of each detection point in radar point clouds. Moreover, the proposed method is non-iterative and can provide an end-to-end solution with minimal pre-processing steps. More importantly, the proposed method is customized for processing multi-dimensional (MD) radar point clouds and extracting relevant features automatically, without feature engineering. To the best of the authors' knowledge, this is the first work that uses NNs to directly process radar point clouds for instantaneous ego-motion estimation.

The rest of this paper is organized as follows. First, the problem statement and the proposed solution will be explained in detail in Section II. Then, Section III provides the evaluation results of the proposed method. Finally, the conclusions and directions for future explorations are presented in Section IV.

II. Methodology

A. Problem Formulation

This paper considers the general problem of estimating within one measurement cycle the 2D motion of a moving vehicle equipped with at least one automotive radar with a 1D antenna array. The radar point cloud after detection at time t given by the n_{th} radar is denoted as $\mathbf{P}_{t,n}^{J\times M}$, where J is the number of detected points, and M is the number of features of each point of the cloud. Note that matrices and vectors are marked in bold in the following equations.

Given a vehicle c, it is common to assume that the 2D coordinate system is at the vehicle's center of gravity, and its ego-motion can be described by $\mathbf{e}_c = \left\{v_x^c, v_y^c, \omega^c\right\}$ (i.e., longitudinal, lateral, and rotational velocity, respectively). Assuming an automotive radar n is installed at the position $\{x_n^c, y_n^c, \theta_n^c\}$, then x_n^c, y_n^c, and θ_n^c are the relative distance and angle to the vehicle's origin. Therefore, the transformation between the vehicle and the radar motion can be expressed as

$$\begin{bmatrix} v_x^n \\ v_y^n \\ \omega^n \end{bmatrix} = \begin{bmatrix} cos(\theta_n^c) & sin(\theta_n^c) & 0 \\ -sin(\theta_n^c) & cos(\theta_n^c) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -y_n^c \\ 0 & 1 & x_n^c \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_x^c \\ v_y^c \\ \omega^c \end{bmatrix}. \tag{1}$$

Given the radar point cloud $\mathbf{P}_{t,n}^{J\times M}$ and assuming $M \geq 2$ containing at least radial velocity and Angle of Arrival (AoA) measurements denoted by d_j^n and α_j^n, the relationship between the radar motion and the velocity measurement can be expressed as

$$\begin{bmatrix} d_1^n \\ d_2^n \\ \vdots \\ d_J^n \end{bmatrix} = - \begin{bmatrix} cos(\alpha_1^n) & sin(\alpha_1^n) \\ cos(\alpha_2^n) & sin(\alpha_2^n) \\ \vdots & \vdots \\ cos(\alpha_J^n) & sin(\alpha_J^n) \end{bmatrix} \begin{bmatrix} v_x^n \\ v_y^n \end{bmatrix}. \tag{2}$$

(2) can be re-written in a matrix form as

$$\mathbf{D} = \mathbf{A} \cdot \mathbf{V}. \tag{3}$$

Based on (3), it is clear that the radar motion can be estimated by using the least squares method

$$\mathbf{V}^{est} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{D}. \tag{4}$$

It is important to acknowledge that since this study proposes an instantaneous ego-motion estimation method similar to [15], it inevitably assumes no lateral vehicle motion (i.e., $v_y^c = 0\ m/s$). However, this is a common assumption in many previous works in the related literature [15], [13], [14]. Therefore, given this assumption, the estimated $\mathbf{V}^{est}$, and (1), the vehicle motion can be calculated as in [15]

$$v_x^c = v_x^n cos(\theta_n^c) - v_y^n sin(\theta_n^c) + \omega^c y_n^c \tag{5}$$

$$\omega^c = \frac{v_y^n cos(\theta_n^c) + v_x^n sin(\theta_n^c)}{x_n^c}. \tag{6}$$

However, the remaining problem is how to alleviate the negative impact of outliers in real-world scenarios, such as moving objects that do not satisfy (2), while solving the LSQ problem defined in (4).

B. The Proposed Solution

Fig. 1 shows the architecture of the proposed method. In summary, the key idea behind the proposed approach is to use NNs to extract point features from input radar point clouds. These generated features are then combined and decoded to estimate the weights for each detection point, which are used by the wLSQ method.

As shown in the figure, the input radar point cloud is first processed by a shared multilayer perceptron (shared-MLP) [17]. The shared-MLP extracts point features and projects them onto a high-dimensional (HD) feature space. Then, these point-wise features are mixed by global feature extraction for average pooling (AvgPool). The output of global feature extraction is a HD feature vector which encodes the global feature of the input point cloud. It is important to note that since the combination of the shared-MLP and global feature extraction is a symmetric function, the HD feature vector is insensitive to the order of input point clouds.

After the global feature extraction, the HD feature vector is concatenated with the input point features, so that each point contains not only local features but also global features. Then, another shared-MLP is used as a decoder which mixes the local and global features and transforms them into a low-dimensional (LD) feature space. Finally, each LD feature vector is converted to a weight in the range of 0 to 1 in the output layer by using a single-layer perceptron.

If there are J detection points in the input radar point cloud, then J point weights are estimated and formed into a diagonal matrix $\mathbf{W}^{est}$. Also, the matrix $\mathbf{A}$ and $\mathbf{D}$ in (3) can be constructed given the input data. Therefore, the ego-motion of the radar platform $\mathbf{V}^{est}$ can be estimated based on the weighted least squares method as follows

$$\mathbf{V}^{est} = (\mathbf{A}^T\mathbf{W}^{est}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{W}^{est}\mathbf{D}. \tag{7}$$

Finally, the vehicle motion can be calculated using (5) and (6) from this estimate.

C. Implementation Details

The shared-MLP consists of three fully-connected (FC) layers. For the encoder, the number of neurons in the three FC layers are 128, 256, and 512, respectively. For the decoder, the three FC layers have the same number of neurons as the encoder but in reverse order. Note that each FC layer is followed by a batch normalization (BN) layer and an activation function of rectified linear unit (ReLU). For the output layer, a shared-MLP with a single neuron is used. To predict a weight between 0 and 1, the output layer uses a sigmoid activation function.

The proposed NN is optimized using the mini-batch gradient descent method. The batch size is set to 512 and the learning rate is initialized as $1e-3$. To optimize the objective function, the root mean squared propagation (RMSProp) is used, and it stops optimization when the validation loss stops decreasing for more than 50 epochs. Due to space limitations, readers interested in more details of data preprocessing, network architecture & training, and performances can refer to [18].

III. Results

A. Dataset

To evaluate the proposed method, this work uses the RadarScenes dataset [19]. The dataset contains 158 recordings

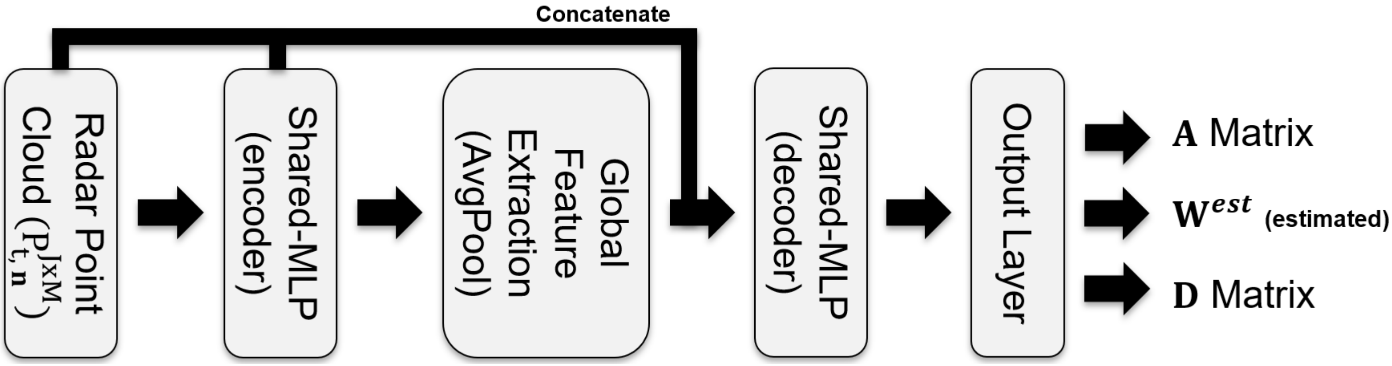

Fig. 1. The architecture of the proposed method. The input data is the radar point cloud with J number of points and M number of features. Note that $M \geq 2$ is required in this work because radial velocity and Angle of Arrival (AoA) measurements are essential object features for instantaneous ego-motion estimation. The outputs are matrices $\mathbf{A}$ and $\mathbf{D}$, which can be constructed from the input data (as shown in (3)) and estimated weights $\mathbf{W}^{est}$ which will be used in (7) for the wLSQ method.

from four automotive radars mounted on a moving vehicle. In total, 64 recordings are selected for performance analysis. It is important to note that the equivalent drive length of the selected recordings is over 79 kilometers. Due to space limitations, readers are referred to [19] for more details, such as the radar geometry, radar specifications, and data collection process for this dataset.

B. *Evaluation Metrics*

The evaluation metrics used in this study are the absolute pose error (APE), relative trajectory error (RTE), and runtime performance (RTP). APE measures the L2 distance between the estimated vehicle motion and the ground truth motion. It reflects how accurate (or biased) the motion estimation achieved by the tested method can be. Unlike APE, RTE measures the mean squared error between the estimated trajectory and the true trajectory. Therefore, it can indicate the long-term stability (variance) of the method under test. Finally, RTP provides the update rate of the tested method.

C. *Comparison Results*

This sub-section provides a performance comparison between the proposed method and the method by Kellner [15]. To the best of our knowledge, Kellner's approach is the most representative work in the current literature that considers 2D instantaneous ego-motion estimation using a single automotive radar. Therefore, it is reasonable and fair to compare the two approaches.

As shown in Table 1, two performance metrics are reported and each method is evaluated separately using 64 radar recordings captured by four automotive radars. Moreover, the leave-one-out (L1O) method is used to train and test the proposed method. The L1O method takes one radar recording out for testing, and the remaining 63 records are used for model training. This can help measure the generalization ability of the proposed method to 'unseen' data.

The results show that the proposed method outperforms Kellner's method by a large margin, even with strict performance measures. For example, for APE in longitudinal velocity estimation, the proposed method achieved a performance gain of **47.7%** on average. In addition, for APE in rotational velocity estimation, the proposed method improved Kellner's method by **36.5%** on average. Moreover, the proposed method shows a performance improvement of **20.7%** in RTE metrics, on average. Notably, the proposed method is about 3 times faster than Kellner's method in terms of RTP due to the lightweight NN architecture and non-iterative estimation process.

Table 1. A full-scale performance comparison between the proposed method and Kellner's method [15]. For each radar, results are averaged over 64 data sequences. For the proposed method, the leave-one-out approach is used during model training, in order to measure the estimation error for the testing sequence that is unseen to the trained model.

APE in terms of V_x^c, unit m/s					
Methods	Radar #1	Radar #2	Radar #3	Radar #4	*Mean*
Kellner's	0.193	0.132	0.253	0.363	0.235
Proposed	0.116	0.104	0.122	0.150	0.123
Improve.	**+39.9%**	**+21.1%**	**+51.8%**	**+58.7%**	**+47.7%**
APE in terms of ω^c, unit deg/s					
Methods	Radar #1	Radar #2	Radar #3	Radar #4	*Mean*
Kellner's	0.802	2.172	2.934	0.882	1.697
Proposed	0.670	1.604	1.289	0.745	1.077
Improve.	**+16.4%**	**+26.1%**	**+56.1%**	**+15.6%**	**+36.5%**
RTE, unit m^2					
Methods	Radar #1	Radar #2	Radar #3	Radar #4	*Mean*
Kellner's	11.7	18.8	21.2	6.1	14.5
Proposed	8.0	14.6	17.0	6.2	11.5
Improve.	**+31.6%**	**+22.5%**	**+19.6%**	**-1.53%**	**+20.7%**

However, it is important to note that even though the proposed method has a high score in APE, the improvement in RTE is not as significant as in APE. This can be attributed to the fact that the proposed method has high accuracy but a relatively large variance. As a result, large estimation errors may occur, which can affect the RTE results.

D. *wLSQ Visualization*

Unlike Kellner's approach [15], which uses the random sample consensus (RANSAC) algorithm to solve the LSQ problem (shown in (4)) iteratively, the proposed method formulates the task of ego-motion estimation as a wLSQ

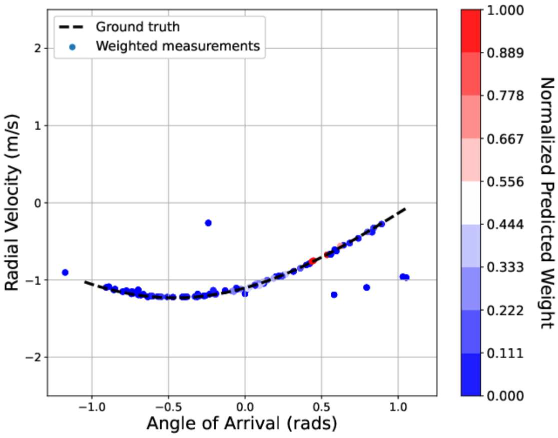

Fig. 2. Example of Velocity-AoA plot. The black-dashed line represents the expected radial velocity measurements given the ground truth ego-motion of the vehicle. Radar measurements are represented by dots, and their associated color indicates the magnitude of the predicted weight by the proposed approach. Note that the weights are normalized by the maximum weight value since the maximum weight is not guaranteed to be 1.

problem and uses NNs to estimate weights for the input radar point cloud directly.

Fig. 2 provides an example of the radial velocity-AoA plot. It illustrates the relationship between the predicted weights and radar measurements. Based on the plot, it is clear that the proposed method can automatically locate keypoints and assign them high weights. Conversely, irrelevant measurements, especially outliers caused by noise and moving objects, are given lower weight values. Nevertheless, it is important to note that not all inliers have high weights, which means the performance of the proposed method relies on its accuracy in searching keypoints.

IV. Conclusion and Future Work

This paper presents a novel 2D instantaneous ego-motion estimation method for vehicles equipped with automotive radars. For the first time, the proposed method formulates the task of ego-motion estimation as a weighted least squares problem and uses a neural network to predict the weights of the input data so that negative effects caused by moving objects and false positives can be mitigated. Based on the evaluation results, the proposed method outperforms the most representative work in the literature [15] by a large margin, but without a significant benefit in terms of RTE metric.

For future study directions, it is important to perform more comprehensive performance comparisons with previous works in the literature, some of which are reported in [18]. Moreover, it is also worth investigating how to improve performances for the RTE metrics.

References

[1] H. Durrant-Whyte and T. Bailey, “Simultaneous localization and mapping: part i,” *IEEE robotics & automation magazine*, vol. 13, no. 2, pp. 99–110, 2006.

[2] T. Miyasaka, Y. Ohama, and Y. Ninomiya, “Ego-motion estimation and moving object tracking using multi-layer lidar,” in *2009 IEEE intelligent vehicles symposium*. IEEE, 2009, pp. 151–156.

[3] D. Fethi, A. Nemra, K. Louadj, and M. Hamerlain, “Simultaneous localization, mapping, and path planning for unmanned vehicle using optimal control,” *Advances in Mechanical Engineering*, vol. 10, no. 1, p. 1687814017736653, 2018.

[4] J. Yi, J. Zhang, D. Song, and S. Jayasuriya, “Imu-based localization and slip estimation for skid-steered mobile robots,” in *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2007, pp. 2845–2850.

[5] J. Borenstein and L. Feng, “Measurement and correction of systematic odometry errors in mobile robots,” *IEEE Transactions on robotics and automation*, vol. 12, no. 6, pp. 869–880, 1996.

[6] Y. Gu, Y. Wada, L. Hsu, and S. Kamijo, “Vehicle self-localization in urban canyon using 3d map based gps positioning and vehicle sensors,” in *2014 International Conference on Connected Vehicles and Expo (ICCVE)*. IEEE, 2014, pp. 792–798.

[7] K. Yamaguchi, T. Kato, and Y. Ninomiya, “Vehicle ego-motion estimation and moving object detection using a monocular camera,” in *18th International Conference on Pattern Recognition (ICPR’06)*, vol. 4. IEEE, 2006, pp. 610–613.

[8] K. Zindler, N. Geiß, K. Doll, and S. Heinlein, “Real-time ego-motion estimation using lidar and a vehicle model based extended kalman filter,” in *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2014, pp. 431–438.

[9] S. H. Cen and P. Newman, “Radar-only ego-motion estimation in difficult settings via graph matching,” in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 298–304.

[10] M. Hoffmann, L. Krabbe, C. Schüßler, P. Gulden, and M. Vossiek, “Instantaneous ego-motion estimation using a coherent radar network,” in *2022 19th European Radar Conference (EuRAD)*. IEEE, 2022, pp. 321–324.

[11] F. Lu and E. Milios, “Robot pose estimation in unknown environments by matching 2d range scans,” *Journal of Intelligent and Robotic systems*, vol. 18, no. 3, pp. 249–275, 1997.

[12] M. Rapp, M. Barjenbruch, K. Dietmayer, M. Hahn, and J. Dickmann, “A fast probabilistic ego-motion estimation framework for radar,” in *2015 European Conference on Mobile Robots (ECMR)*. IEEE, 2015, pp. 1–6.

[13] 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.

[14] P.-C. Kung, C.-C. Wang, and W.-C. Lin, “A normal distribution transform-based radar odometry designed for scanning and automotive radars,” in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 14 417–14 423.

[15] 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.

[16] ——, “Instantaneous ego-motion estimation using multiple doppler radars,” in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 1592–1597.

[17] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, “Pointnet: Deep learning on point sets for 3d classification and segmentation,” *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 652–660, 2017.

[18] S. Zhu, F. Fioranelli, and A. Yarovoy, “Deepego: Deep instantaneous ego-motion estimation using automotive radar,” in *IEEE Transactions on Radar System*, May 2023.

[19] O. Schumann, M. Hahn, N. Scheiner, F. Weishaupt, J. Tilly, J. Dickmann, and C. Wöhler, “RadarScenes: A Real-World Radar Point Cloud Data Set for Automotive Applications,” Mar. 2021. [Online]. Available: https://doi.org/10.5281/zenodo.4559821