

Automated Automotive Radar Calibration With Intelligent Vehicles

Alexander Tsaregorodtsev*, Michael Buchholz*, and Vasileios Belagiannis†

**Institute for Measurement-, Control-, and Microtechnology*

Ulm University, Germany

{firstname}.{lastname}@uni-ulm.de

†*Chair of Multimedia Communications and Signal Processing*

Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

vasileios.belagiannis@fau.de

Abstract—While automotive radar sensors are widely adopted and have been used for automatic cruise control and collision avoidance tasks, their application outside of vehicles is still limited. As they have the ability to resolve multiple targets in 3D space, radars can also be used for improving environment perception. This application, however, requires a precise calibration, which is usually a time-consuming and labor-intensive task. We, therefore, present an approach for automated and geo-referenced extrinsic calibration of automotive radar sensors that is based on a novel hypothesis filtering scheme. Our method does not require external modifications of a vehicle and instead uses the location data obtained from automated vehicles. This location data is then combined with filtered sensor data to create calibration hypotheses. Subsequent filtering and optimization recovers the correct calibration. Our evaluation on data from a real testing site shows that our method can correctly calibrate infrastructure sensors in an automated manner, thus enabling cooperative driving scenarios.

Index Terms—automatic calibration, roadside infrastructure, automotive radars

I. INTRODUCTION

With the increasing availability of advanced driver-assistance system (ADAS) functionalities in commercial vehicles and the advancements in the field of automated driving, the research of assisting roadside infrastructure is slowly moving into focus. Roadside units (RSUs) already exist for tasks like traffic surveillance and monitoring, speed limit enforcement, toll collection, and construction site warning. In recent research, the ability of such infrastructure for assisted automated driving and cooperative maneuver planning is actively being investigated [1], [2]. To provide the information required for assisted driving, the RSUs need to be equipped with sensors capable of distinguishing different types of road users. Similar to automated vehicles, cameras, lidars, and radars can be used for such purposes, as all three sensor types are capable of detecting traffic participants. However, if such sensor equipment is to be deployed at real intersections, cost efficiency comes into play, as potentially hundreds of locations need to be equipped. This can become costly if

Part of this work was financially supported by the Federal Ministry for Economic Affairs and Climate Action of Germany within the program "New Vehicle and System Technologies" (project LUKAS, grant number 19A20004F).

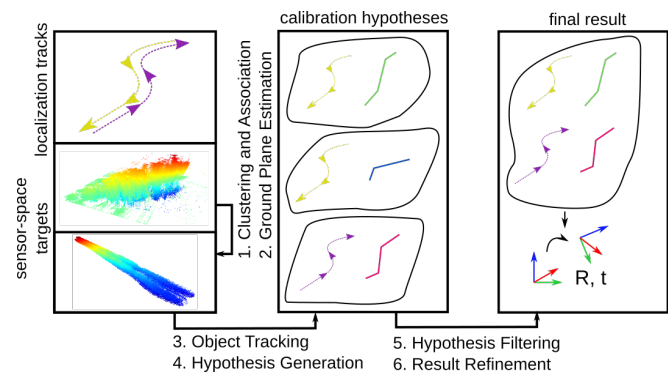


Fig. 1. Overview of the calibration pipeline. Beginning on the left, the raw data is recorded and clustered (1). Afterward, the sensor roll and pitch are estimated by determining the ground plane (2). Using this information, object trajectories are smoothed and associated with vehicle localization data to create correspondences and estimate calibration hypotheses (3-4). The hypotheses are then filtered and refined to obtain the final calibration (5-6). Steps 1-6 are covered in detail in Sec. II-B-II-F.

complex and expensive sensors are used in the field. At the same time, sensor redundancy is desired, as the quality and robustness of the environment perception can be improved by fusing information from different sensors and sensor types. Therefore, it may be beneficial to use a combination of camera and radar sensors, as both are comparatively cheap compared to lidar sensors and are also already installed in commercial vehicles and thus are widely available. When building such a proposed camera-radar sensor infrastructure setup, it has to be precisely extrinsically calibrated before use in any assisted and cooperative driving scenarios. This is usually a manual and time-consuming process. Targets need to be manually placed in the infrastructure environment, then recordings have to be made, and often annotated to obtain a precise extrinsic calibration [3]. Furthermore, there are currently no approaches for automatic and geo-referenced radar calibration. Additionally, existing approaches for camera-radar cross-calibration may not be applicable, if the sensor's field of view (FoV) is not overlapping with the FoV of other sensors or no cameras are present.

We, therefore, propose an automatic method for infrastruc-

ture radar calibration, which uses cooperative and connected automated vehicles to provide a target vehicle with a known location, orientation, and dimension. An outline of our approach is shown in Fig. 1. By recording radar target data in conjunction with connected vehicle data, a timestamped calibration dataset can be created. Then, using a multi-stage pipeline, the extrinsic calibration is estimated. In particular, objects are first detected in the raw sensor data and associated over time. From the obtained tracks, the driving lanes and the sensor pitch and roll are estimated, which allows the sensor data to be projected into a 2D plane. The object tracks are then re-tracked to improved accuracy and fused with the calibration vehicle data to obtain point-to-point correspondences, which are used to obtain possible sensor positions and rotations. By applying a novel filtering scheme and further refinement procedures, a final extrinsic calibration is obtained. Our contributions are summarized as:

- An automated approach for geo-referenced infrastructure radar-only calibration based on the concept of hypothesis filtering relying on unmodified and unfiltered data from the infrastructure.
- The only requirements for the calibration vehicle are a localization unit and known vehicle dimensions. No appearance requirements for the vehicle as well as the sensor environment are imposed. Therefore other traffic participants may freely use the area observed by the sensor during the calibration process.
- Experiments on real-world data show that our method can be applied to radar sensors to be used to enhance the safety and reliability of automated and cooperative driving tasks.

A. Prior Work

Compared to radar sensors, the calibration of camera sensors is already widely explored for generic camera setups [4] as well as roadside infrastructure cameras [5]. Due to the different measurement type obtained from radars (3D target point clouds) compared to images (2D images with color and missing depth information), methods for estimating a 3D transform [6] may be adapted to work with radars, while approaches solely relying on 2D-3D correspondences [7] may not be applicable.

Classical approaches for sensor calibration use human annotations and/or measurements of remarkable targets in both the real-world environment as well as the measurement space. Remarkable targets include chessboard patterns for camera, plastic spheres for lidar, and metal objects for radar sensors [3]. Feature-based approaches use sensor-specific features to perform matching with features from the real environment. Examples include the automatic calibration algorithm from [8] based on line features and [9] based on textured models. Other methods use features obtained from neural networks, e.g. vehicle landmarks in [10], or semantic labels in [11].

More automatic calibration methods for outdoor applications are also available. The neural network presented in [12] performs a 3D point cloud registration between two

point clouds, while [13] uses neural networks to perform automatic rotational calibration between cameras and radars. [14] optimizes a Gaussian mixture model to align radar target measurements to a GPS-equipped vehicle and estimate the yaw and 2D translation of infrastructure radars.

II. CALIBRATION METHOD

We present our approach as a multi-stage pipeline. Each stage is described in its own section and shown in Fig. 1.

A. Problem Formulation

Consider the following transformation model:

$$\begin{bmatrix} x_{\text{world}} \\ y_{\text{world}} \\ z_{\text{world}} \end{bmatrix} = \mathbf{R} \begin{bmatrix} x_{\text{sensor}} \\ y_{\text{sensor}} \\ z_{\text{sensor}} \end{bmatrix} + \mathbf{t}, \quad (1)$$

where $[x_{\text{sensor}}, y_{\text{sensor}}, z_{\text{sensor}}]^T$ are the coordinates of the radar sensor measurement in the sensor reference frame and $[x_{\text{world}}, y_{\text{world}}, z_{\text{world}}]^T$ is the measurement location in a global coordinate system. The goal of our approach is to estimate the sensor rotation matrix \mathbf{R} and sensor position \mathbf{t} , where \mathbf{R} can also be expressed by $\mathbf{f}(\alpha, \beta, \gamma)$ in terms of roll α , pitch β , and yaw γ . The sensor-referenced measurements are radar targets $[x_{\text{sensor}}, y_{\text{sensor}}, z_{\text{sensor}}, v_{\text{rad}}]^T$ provided by the radar sensor, while the world location in UTM coordinates together with the roll, pitch, and yaw of the calibration vehicle $[x_{\text{world,UTM}}, y_{\text{world,UTM}}, z_{\text{world,UTM}}, \alpha_{\text{world}}, \beta_{\text{world}}, \gamma_{\text{world}}]^T$ is obtained from a combined GNSS+RTK receiver and inertial measurement unit. Furthermore, measurements originating from both the radar sensor as well as the vehicle localization unit are assumed to be timestamped to the same time reference.

B. Clustering and Association

At the beginning, data from both the sensor as well as the vehicle's localization unit is recorded, while the vehicle passes through the sensor's field of view multiple times. This process is required to filter out false positives in the next steps. During the first data pre-processing step, the recorded radar targets are filtered in order to remove unwanted clutter detections. As the radar sensor itself is stationary, this is done by removing all targets which have a radial velocity of less than 0.1km/h and thus are not considered to be moving. The remaining targets are then clustered using an adapted DBSCAN algorithm [15], which uses the euclidean distance between two data points to decide its association with an existing cluster.

By applying DBSCAN, multiple targets belonging to the same object can be grouped together. As this can cause problems with objects with low radar cross-section (RCS) values, e.g. pedestrians, as well as objects, which pass by each other at a close distance, some adjustments have been made to the standard DBSCAN approach. To address the first problem, we set the minimum cluster size to 3 while accumulating the targets of three consecutive measurements. This allows us to identify moving objects which only produce one single target during a measurement step. To solve the second issue, we expanded the euclidean distance measure to also include the

radial velocity v_{rad} of each target. This increases the measured distance between two radar point clouds in close proximity if they are moving at different speeds and/or in different directions.

By using this modified approach, object clusters can be estimated for the entire calibration recording. However, the clusters between two consecutive time steps are still not associated with each other. Therefore, an association is implemented by exploiting the information gathered from the previously mentioned buffering procedure, where the cluster identifier of targets from previous timestamps is propagated to the current timestamp. Subsequently, the calculated associations are verified using the Hungarian Algorithm [16], where the distance between two cluster centers is used as the association cost. By introducing a gating step to prevent associating new object clusters with previous clusters that may already be outside the sensors' field of view, incorrect associations are removed. After this procedure, each clustered object in the target point cloud has an associated track it belongs to.

C. Ground Plane Estimation

In the next step, the pitch and roll of the sensor are estimated. As the sensor is installed on infrastructure, it may have a significant downward pitch in order to cover the desired road area. Therefore, the previously obtained object clusters can not be directly projected into the ground plane by removing the z-coordinate. Instead, we estimate the ground plane normal vector, which is then rotated to be the z-axis unit vector. The rotation required to do so contains both the pitch and the roll of the sensor. To estimate the 2D plane, we combine every target from the calibration sequence into one merged point cloud. Then, the point cloud density is estimated for each point by counting the number of neighboring points in a radius around the point. We discard all points which lie in low-density regions (in our case, we removed all points with less than 20% of the maximum observed density) to obtain the most-occupied spaces which lie right above the lanes of the observed road area. Using the filtered point cloud, we can estimate its principal axes by performing a singular value decomposition (SVD) [17] of the entire point cloud and using the left-hand base vector belonging to the smallest eigenvalue as the plane normal vector \mathbf{n} . By defining the desired normal vector to be $\mathbf{n}_{\text{ref}} = [0, 0, 1]^T$, we can calculate a rotation matrix $\mathbf{R}_{\alpha\beta}$ such that $\mathbf{n}_{\text{ref}} = \mathbf{R}_{\alpha\beta}\mathbf{n}$, which contains the roll and pitch of the sensor. The yaw is still unknown, as the ground plane only defines two orthogonal base vectors, against which the pitch and roll are estimated.

D. Object Tracking

After determining the roll and pitch of the sensor, the recorded targets can be transformed into a new sensor coordinate system with zero roll and pitch. By dismissing the z-coordinate we can transform the sensor data into a 2D bird's eye view (BEV) coordinate system. This allows us to track the objects using planar kinematic models. The tracking step is required, as the extent of the target cluster

as well as the number of targets contained within a cluster changes with every measurement. This leads to extremely noisy object trajectories if the cluster center is used as a reference point. We use an Unscented Kalman filter [18] with a CTRA transition model [19] to smooth the cluster center trajectories. As we also have data from future time steps available for each measurement, we employ the Rauch-Tung-Striebel (RTS) Kalman smoother [20] to further smooth the resulting trajectory by applying backward filtering. At the end of the tracking procedure, each object cluster in the recorded sequence is smoothed down to a trajectory containing the x- and y-positions of the object centers.

E. Hypothesis Generation and Filtering

In order to choose the correct object track, we create calibration hypotheses to then confirm or reject them and find the final calibration. A calibration hypothesis is defined by a tuple containing an object track as well as the localization data from the vehicle. For each state from the object track, its corresponding timestamp is obtained from the original radar measurement. Afterward, the localization measurement with the closest timestamp is taken as the 3D correspondence, where the elevation of the object data can either be set to 0 or obtained from the targets prior to Kalman smoothing. Therefore, it is important that the time synchronicity of both data sources is ensured. Using these correspondences, a 3D transform can be estimated, e.g. by employing convex global registration [6]. By repeating this step for each separate hypothesis, we obtain a transform for each one. These hypotheses can then be grouped by their estimated translation vector \mathbf{t} using DBSCAN. As the calibration vehicle passed the sensor multiple times and thus is responsible for multiple separate object tracks, a cluster of correct sensor positions should exist. However, other location clusters may exist. To filter out incorrect location clusters, the angle of the rotation error matrix $\mathbf{R}_j^T \mathbf{R}_i$ converted to an axis-angle representation is calculated for every possible pair of transforms inside of the hypothesis cluster, as ideally \mathbf{R}_i and \mathbf{R}_j are the same, thus $\mathbf{R}_j^T \mathbf{R}_i = \mathbf{I}$. The cluster with the lowest average angle can then be chosen as the accepted calibration hypothesis by re-estimating the transform using correspondences merged from each track in the cluster, as the relative rotations of each hypothesis pair in the chosen cluster are closest to 0. The final calibration is then found by concatenating this estimated transform with the previously estimated $\mathbf{R}_{\alpha\beta}$.

F. Result Refinement

To further refine the result, the object tracking step can be repeated with the final transformation already applied to the target clusters. However, as the sensor location is already known, instead of using the cluster center, the target closest to the sensor can be chosen instead. By then incorporating the offset from the localization unit to the car point closest to the sensor, the dimensions of the vehicle are taken into account, and the noise caused by varying cluster sizes and extents is partially suppressed. This improves the quality of the

correspondences and provides a better result. In the final step, the overall transform can be further refined by maximizing the number of targets contained in the ground area polygon \mathbf{a} covered by the vehicle. For this purpose, we define a loss $\mathcal{L}(p) = \{\min_i |\mathbf{a}_i - p| \text{ if } p \in \mathbf{a}, 0 \text{ else}\}$, where \mathbf{a}_i is the i -th polygon vertex. Then, by accumulating the loss for all correspondences and using a gradient-free optimization method like [21] to minimize the loss, we estimate a 2D offset vector, which, when applied after the calibration transform, corrects the calibration in such a way that most targets lie on the calibration vehicle.

III. EVALUATION

We present an experimental evaluation based on recorded sensor data from our real-world test site.

A. Data Sources and Evaluation Procedure

To gather data for evaluation, we use the sensor setup of our real-world test site located in Ulm-Lehr [1]. The test site contains a total of 7 light poles equipped with camera and automotive radar sensors and attached processing units. The radars mounted at the intersection consist of multiple long-range radars capable of outputting target lists with points in spherical coordinates in conjunction with the radial velocity and additional information about each point with a frequency of 15Hz. The calibration vehicle is equipped with a highly precise localization unit containing both a 4-band GNSS receiver and RTK correction support as well as an inertial measurement unit, whose position is calibrated with respect to the vehicles' rear axle position and the vehicle dimensions to model an occupancy 3D bounding box of the vehicle. The localization data is recorded with a frequency of 50 Hz and reaches a positional accuracy of up to 2cm. Both the radar and localization data are timestamped with the sensor clocks being synchronized to GPS time, with measured clock offsets below 2ms. For the evaluation protocol, two sequences of around 15 minutes in length are recorded for each sensor in parallel while the calibration vehicle passes all sensors' fields of view multiple times.

We run our calibration approach with the recorded data, where one sequence is used for calibration and the second one for evaluation. Only the object tracking step is performed on the evaluation set. Employing the previously used data association by synchronized timestamps, each object track is associated with a localization track and transformed into world coordinates by applying the previously estimated transform. To remove tracks of other vehicles, associations with mean offsets between localization and radar data of over 5m are removed. The remaining tracks are merged into a single track. Due to our object clustering approach not estimating any object bounding boxes, we report the mean outlier error δ_{op} , which is calculated by averaging the distance of all outlier targets to the vehicle polygons. We also introduce the metric r_i , which calculates the fraction of radar targets in a cluster lying inside the calibration vehicle's bounding box. r_i captures the quality of the estimated alignment and considers the random nature of the occurring

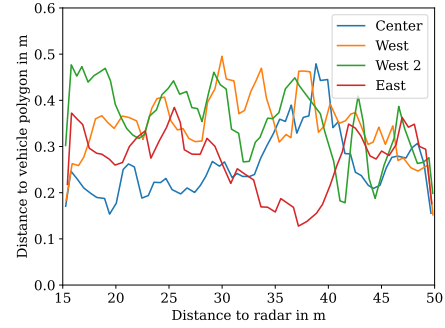


Fig. 2. Distance δ_{op} between vehicle polygon and targets not contained in the vehicle polygon, which are considered outlier targets, plotted over the distance between target and sensor.

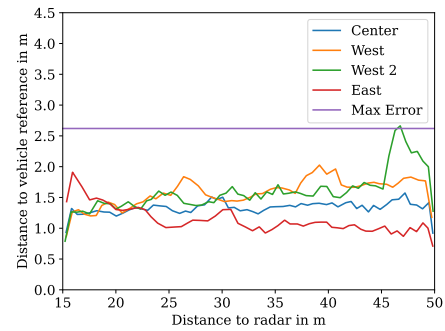


Fig. 3. Distance δ_p between vehicle corner and target closest to the sensor plotted over the distance between sensor and vehicle corner. As a comparison, the distance from the vehicle corner to the vehicle geometric center (“Max error”) is visualized alongside the data to provide an upper bound for the acceptable error.

target point clouds. We furthermore choose the object cluster target closest to the camera and report the observed offset to the vehicle corner closest to the estimated sensor position δ_p . However, it contains a random offset due to the randomness of the target location on the vehicle.

B. Evaluation results

The quantitative evaluation results for 4 sensors from our test site are presented in Table I.

TABLE I
QUANTITATIVE EVALUATION RESULTS FROM REAL-WORLD SEQUENCES.

Radar name	West	Center	East	West 2
r_i (%)	67.69	69.92	79.06	60.36
δ_{op} (m)	0.35	0.26	0.29	0.39
δ_p (m)	1.58	1.34	1.10	1.71

The results show the effectiveness of our approach despite evaluating on unstructured radar data. When considering both the inlier ratio r_i and outlier distance δ_{op} , we can observe ratios between 60% and 70%, while the remaining 30-40% of

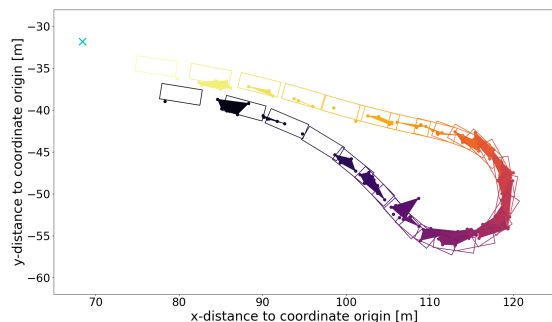


Fig. 4. Bird's eye view of a sample test track. The boxes denote the calibration vehicle, while the points represent the associated target clusters. Clusters with at least three points are visualized using their convex hull for better clarity. Earlier measurements are colored black, while recent measurements are shown in yellow. The estimated sensor position is the cyan cross in the top left corner of the image.

targets have an average distance of below 50cm to the vehicle bounding box across all evaluated sensors with different fields of view, as shown in Fig. 2. As the raw radar data may still contain clutter targets due to unwanted reflections and measurement noise, parts of the outliers may be caused by clutter measurements and not caused by calibration errors. Additionally, average errors below 50cm provide enough precision for cooperative driving tasks.

While the average errors δ_p are higher compared to the outlier error δ_{op} , they are mostly caused by the missing vehicle reference of the target cluster data, as a single target can originate from any reflective part of the vehicle. This is also apparent in Fig. 3, where the distance between the chosen reference target and vehicle corner is visualized. As the error is similar between all evaluated sensors and nearly constant over the entire sensor range, systematic errors in the estimated sensor rotation are unlikely, as an incorrectly determined rotation causes offsets proportional to the sensor distance. Fig. 4 provides a sample visual result of the estimated calibration. As the sensor is located in the top left corner of the image, it mostly captures the left or back side of the vehicle, resulting in an increased probability of targets originating from the left or back side of the vehicle. Additionally, we can see that target clusters overlap well with the bounding box in most situations, excluding clutter targets.

IV. CONCLUSION

We presented an approach for automated calibration of automotive radar sensors that is based on hypothesis filtering. Our method makes minimal assumptions about the calibration vehicle and the sensor environment, while it does not depend on human interaction. The calibration is performed by recording data from both the radar and the target vehicle by driving through the sensor's FoV and extracting object tracks from the recording. The sensor and vehicle tracks are combined during hypothesis generation. Subsequent filtering recovers the correct hypothesis, which is then further optimized to take the random nature of target locations into account. Our evaluation on real-world data shows the applicability of our approach.

REFERENCES

- [1] M. Buchholz *et al.*, "Handling occlusions in automated driving using a multiaccess edge computing server-based environment model from infrastructure sensors," *IEEE Intelligent Transportation Systems Magazine*, vol. 14, no. 3, pp. 106–120, 2022.
- [2] M. B. Mertens *et al.*, "An extended maneuver coordination protocol with support for urban scenarios and mixed traffic," in *2021 IEEE Vehicular Networking Conference (VNC)*, 2021, pp. 32–35.
- [3] J. Oh, K.-S. Kim, M. Park, and S. Kim, "A comparative study on camera-radar calibration methods," in *2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, 2018, pp. 1057–1062.
- [4] R. Hartley and A. Zisserman, "Computation of the camera matrix P," in *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004, p. 178–236.
- [5] S. R. E. Datondji, Y. Dupuis, P. Subirats, and P. Vasseur, "Rotation and translation estimation for a wide baseline fisheye-stereo at crossroads based on traffic flow analysis," in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, 2016, pp. 1534–1539.
- [6] J. Briaies and J. Gonzalez-Jimenez, "Convex global 3d registration with lagrangian duality," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.
- [7] V. Lepetit, F. Moreno-Noguer, and P. Fua, "EPnP: An accurate o(n) solution to the PnP problem," *International journal of computer vision*, vol. 81, no. 2, pp. 155–166, 2009.
- [8] M. Dubská, A. Herout, R. Juránek, and J. Sochor, "Fully automatic roadside camera calibration for traffic surveillance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 3, pp. 1162–1171, 2014.
- [9] E. Ataer-Cansizoglu, Y. Taguchi, S. Ramalingam, and Y. Miki, "Calibration of non-overlapping cameras using an external SLAM system," in *2014 2nd International Conference on 3D Vision*, 2014, pp. 509–516.
- [10] V. Bartl, J. Špaňhel, P. Dobeš, R. Juránek, and A. Herout, "Automatic camera calibration by landmarks on rigid objects," *Machine Vision and Applications*, vol. 32, no. 1, pp. 1–13, 2021.
- [11] A. Tsaregorodtsev, J. Müller, J. Stroheck, M. Herrmann, M. Buchholz, and V. Belagiannis, "Extrinsic camera calibration with semantic segmentation," in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2022, pp. 3781–3787.
- [12] M. Horn, N. Engel, V. Belagiannis, M. Buchholz, and K. Dietmayer, "Deepplr: Correspondence-less architecture for deep end-to-end point cloud registration," in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2020, pp. 1–7.
- [13] C. Schöller *et al.*, "Targetless rotational auto-calibration of radar and camera for intelligent transportation systems," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, 2019, pp. 3934–3941.
- [14] X. Li, Y. Liu, V. Lakshminarasimhan, H. Cao, F. Zhang, and A. Knoll, "Globally optimal robust radar calibration in intelligent transportation systems," *IEEE Transactions on Intelligent Transportation Systems*, 2023.
- [15] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu, "DbSCAN revisited, revisited: Why and how you should (still) use dbSCAN," *ACM Trans. Database Syst.*, vol. 42, no. 3, 2017. [Online]. Available: <https://doi.org/10.1145/3068335>
- [16] D. F. Crouse, "On implementing 2d rectangular assignment algorithms," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 4, pp. 1679–1696, 2016.
- [17] G. H. Golub and C. Reinsch, "Singular value decomposition and least squares solutions," *Linear algebra*, vol. 2, pp. 134–151, 1971.
- [18] E. A. Wan and R. Van Der Merwe, "The unscented kalman filter for nonlinear estimation," in *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No. 00EX373)*, 2000, pp. 153–158.
- [19] R. Schubert, E. Richter, and G. Wanielik, "Comparison and evaluation of advanced motion models for vehicle tracking," in *2008 11th International Conference on Information Fusion*, 2008, pp. 1–6.
- [20] S. Särkkä, "Unscented rauch-tung-striebel smoother," *IEEE transactions on automatic control*, vol. 53, no. 3, pp. 845–849, 2008.
- [21] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The computer journal*, vol. 7, no. 4, pp. 308–313, 1965.