

ORORA: Outlier-Robust Radar Odometry

Hyungtae Lim¹, *Student Member, IEEE*, Kawon Han¹, *Member, IEEE*, Gunhee Shin²,
Giseop Kim³, Songcheol Hong¹, *Member, IEEE*, and Hyun Myung^{1*}, *Senior Member, IEEE*

Abstract—Radar sensors are emerging as solutions for perceiving surroundings and estimating ego-motion in extreme weather conditions. Unfortunately, radar measurements are noisy and suffer from mutual interference, which degrades the performance of feature extraction and matching, triggering imprecise matching pairs, which are referred to as outliers. To tackle the effect of outliers on radar odometry, a novel outlier-robust method called *ORORA* is proposed, which is an abbreviation of *Outlier-RObust Radar odometry*. To this end, a novel decoupling-based method is proposed, which consists of graduated non-convexity (GNC)-based rotation estimation and anisotropic component-wise translation estimation (A-COTE). Furthermore, our method leverages the anisotropic characteristics of radar measurements, each of whose uncertainty along the azimuthal direction is somewhat larger than that along the radial direction. As verified in the public dataset, it was demonstrated that our proposed method yields robust ego-motion estimation performance compared with other state-of-the-art methods. Our code is available at <https://github.com/url-kaist/outlier-robust-radar-odometry>.

I. INTRODUCTION

In recent years, the demand for localization and perception technologies on mobile vehicles has been increased. To perceive surroundings, various sensors, such as light detection and ranging (LiDAR) and camera sensors, have been widely adopted [1]–[11]. However, it has been known that these sensors are severely affected by weather, illumination conditions, and interfering substances in the air, which make them difficult to be used in extreme environments [12]–[16].

To overcome the aforementioned drawbacks of those sensors, radar sensors are increasingly employed in autonomous vehicles and mobile robots [15], [17], [18]. The radar sensors are relatively robust to atmospheric and light conditions because the radar data are measured by electromagnetic waves [19]–[21]. In addition, with the recent advances in low-cost and high-performance radar sensors, several novel radar-based simultaneous localization and mapping (SLAM) frameworks were proposed [22]–[26].

In this context, numerous researchers have also studied radar odometry, which estimates the ego-motion of a robot when two consecutive radar data are given [17], [28]–[34]. Radar odometry is mainly classified into two categories.

*Corresponding author: Hyun Myung

¹Hyungtae Lim, Kawon Han, Songcheol Hong, and Hyun Myung are with the School of Electrical Engineering, KAIST (Korea Advanced Institute of Science and Technology), Daejeon, 34141, Republic of Korea. {shapelim, smh1176, schong1234, hmyung}@kaist.ac.kr

²Gunhee Shin is a research intern in Urban Robotics Lab., KAIST, Daejeon, 34141, Republic of Korea. gunmaplehee@naver.com

³Giseop Kim is with NAVER LABS, Seongnam, Gyeonggi-do, 13561, Republic of Korea. giseop.kim@naverlabs.com

This work was supported by the BK21 FOUR (Republic of Korea).

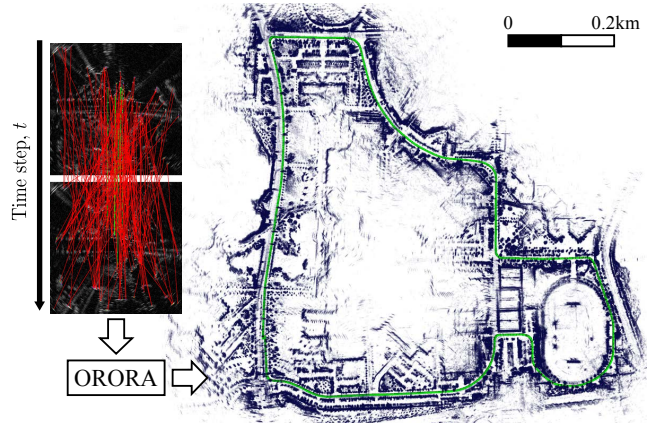


Fig. 1. Radar odometry result of our *Outlier-RObust Radar odometry (ORORA)* without any loop closing on a single loop of KAIST03 in MulRan dataset [27]. Our ORORA shows robust performance even though numerous outliers are included in the estimated correspondences. The red and green lines represent the outlier and inlier feature pairs on two consecutive radar images, respectively. Note that the radar images represent an example of catastrophic failure of feature matching, which outputs only 4% of inliers within the estimated correspondences (best viewed in color).

One is the direct method [28] and the other is the indirect method [13], [14], [17], [30]. The direct method does not use feature extraction and takes the whole radar image as an input. For instance, Park *et al.* [28] proposed PhaRaO, which is a robust direct method using phase correlation properties. Kung *et al.* [31] modeled the uncertainty of whole radar information and exploited normal distribution transform. Barnes *et al.* [30] proposed a deep learning-based method to achieve a novel radar matching by finding correlation in the latent space. The direct method efficiently estimates the relative poses, yet this method can be potentially affected by the integrated noises of raw radar images.

On the other hand, the indirect method exploits feature extraction and matching from the radar images. Accordingly, the indirect method takes advantage of being able to apply the techniques that are actively used in the computer vision or LiDAR fields [9], [14], [25], [26]. Unfortunately, the indirect method also suffers from the noise of measurements and mutual interference. That is, these phenomena degrade the performance of feature extraction and matching, thus triggering numerous outlier pairs, as shown in Fig. 1.

For example, a frequency-modulated continuous-wave (FMCW) radar sensor incorrectly detects interfering waveforms emitted by other radar signals as ghost objects. This interference also triggers speckle noise [26], which produces random noises on the radar image. Furthermore, multi-path reflection issue [35] generates one or more

targets at virtual positions in the radar image. These effects also become worse once the movement of the sensor base becomes aggressive. For these reasons, the radar data are inevitably vulnerable to noises, which induces the outliers.

To tackle these noisy measurement issues, the constant false alarm rate (CFAR) algorithm has been widely used [36]. As an extension, Cen and Newman [13], [14] presented reliable feature detection methods that allow identifying objects from the radar images by using power spectra to avoid false detection. Burnett *et al.* [29] and Barness *et al.* [37] proposed novel deep learning-based methods to extract valid feature pairs in a self-supervised manner. In addition, Burnett *et al.* [17] proposed a novel random sample consensus (RANSAC) that compensates for both motion distortion and the Doppler effect. Hong *et al.* [25], [26] utilized graph-theoretic methods to prune spurious correspondences and presented feature tracking-based approach. Adolfsson *et al.* [32] proposed a more advanced, surfel-based feature extraction method and robust point-to-line error-based optimization.

In the meanwhile, Zhou *et al.* [38] showed that graduated non-convexity (GNC), which gradually increases the non-linearity of the kernel as the iteration progresses to reject the effect of outliers, outperforms RANSAC [39] in terms of both robustness and computation speed in LiDAR fields. Empirically, GNC robustly endures up to 70-80% of outliers and is much faster than RANSAC [40]. Accordingly, GNC has been widely employed to achieve both robust and fast pose estimation [40]–[44]. In addition, Yang *et al.* [42] and Lim *et al.* [40] demonstrated that decoupling rotation and translation estimation shows better performance when gross outliers are included in the estimated correspondences, tolerating up to 99% of outliers.

Sharing the philosophy of GNC and the decoupling method, we propose a robust radar odometry method called *ORORA*, which is an abbreviation of *Outlier-ROBust Radar odometry*. To the best of our knowledge, this is the first attempt to introduce the GNC-based decoupling method in the radar fields.

In summary, the contribution of this paper is threefold.

- A novel decoupling-based outlier-robust odometry is proposed, which consists of GNC-based rotation estimation and anisotropic component-wise translation estimation (A-COTE).
- To this end, the anisotropic characteristics of radar measurements, each of whose uncertainty along the azimuthal direction is larger than that along the radial direction, are mathematically modeled in the manifold space.
- In experiments, our ORORA showed promising performance compared with the state-of-the-art methods even though the imprecise feature correspondences are given, which demonstrates the robustness of our ORORA against gross outliers, as presented in Fig. 1.

II. ORORA: OUTLIER-ROBUST RADAR ODOMETRY

In this section, the pipeline of our radar odometry is explained, as illustrated in Fig. 2(a). Our ORORA mainly consists of four parts: a) Doppler distortion compensation, b) graph-based outlier pruning, c) GNC-based rotation estimation, and d) A-COTE.

A. Radar Image Preprocessing and Data Association

First, we briefly explain how to estimate correspondences between two consecutive radar images. Originally, the most widely used 360° FMCW radar sensors output a radar image whose height and width are N_h and N_w , respectively, and each axis along the height and width denotes the radial and azimuthal direction with respect to the sensor frame, respectively. Note that the resolution in the radial direction is much better than that of the azimuthal direction, so $N_h \gg N_w$ (e.g., 3371 \gg 400 in MulRan dataset [27]).

Next, the radar image is taken as an input of the feature extraction (in this study, the methods proposed in [13] and [14] were used). Then, the extracted feature point on the pixel (h, w) in the polar coordinates is transformed into the Cartesian coordinates as $\mathbf{p} = [hc_r \cos \theta, hc_r \sin \theta]^T$. Here, \mathbf{p} , c_r , and θ denote a 2D feature point in the Cartesian coordinates, scaling factor whose unit is meter per pixel, and radial angle, which is defined as $\theta = \frac{2\pi \cdot w}{N_w}$, respectively. Consequently, we can get the $(t-1)$ -th and t -th feature point sets, each of which is defined as ${}^{t-1}\mathbf{P}$ and ${}^t\mathbf{P}$, respectively.

Next, the correspondence set between ${}^{t-1}\mathbf{P}$ and ${}^t\mathbf{P}$, which is denoted as \mathcal{A} , is estimated. To this end, ORB descriptor [45] is employed as proposed in [17] to generate a descriptor of each point on the radar image transformed into the Cartesian image, as shown in Fig. 2(a). Then, brute-force matching is performed, which is followed by the ratio test [46] to initially filter out the definite outliers.

This whole process finally outputs \mathcal{A} . Further details of feature extraction and matching can be found in [17].

B. Problem Definition of Radar Odometry

Given these putative correspondences, the goal of radar odometry is to estimate relative pose of the sensor frame between two consecutive time steps, i.e. on t and $t-1$, which is expressed as $\mathbf{T}_t^{t-1} \in \text{SE}(2)$. However, using all the correspondences in \mathcal{A} may cause imprecise pose estimation because the false matchings inevitably exist due to some noisy measurements, such as speckle noise [26] that produces random noises due to the interference of different radar waves, multi-path reflections [25], or the effect of moving objects [35].

Therefore, our ultimate goal can be summarized as estimating the relative pose while minimizing the influence of the outliers as follows:

$$\hat{\mathbf{R}}, \hat{\mathbf{t}} = \underset{\mathbf{R} \in \text{SO}(2), \mathbf{t} \in \mathbb{R}^2}{\operatorname{argmin}} \sum_{(i,j) \in \mathcal{A} \setminus \hat{\mathcal{O}}} \rho(r(\mathbf{q}_j - \mathbf{R}\mathbf{p}_i - \mathbf{t})) \quad (1)$$

where $\mathbf{R} \in \text{SO}(2)$ and $\mathbf{t} \in \mathbb{R}^2$ are the relative rotation and translation that correspond to \mathbf{T}_t^{t-1} , respectively; (i, j) be the index pair where i and j denote the indices of a point in

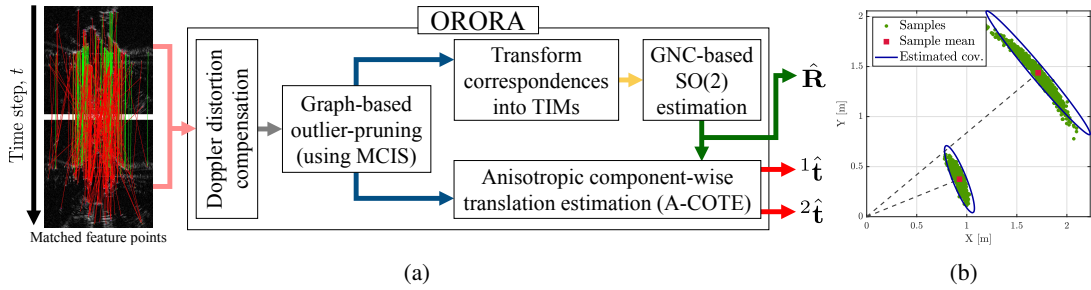


Fig. 2. (a) Overall pipeline of our proposed radar odometry method. Even though the putative correspondences that include numerous outliers are given as an input (pink arrow), our ORORA robustly estimates the relative rotation, $\hat{\mathbf{R}}$, and component-wise translation, i.e. ${}^1\hat{\mathbf{t}}$ for the x -axis and ${}^2\hat{\mathbf{t}}$ for the y -axis, respectively. (b) Visual description of anisotropic characteristics of radar feature points. Samples (green dots) indicate that the uncertainty along the azimuthal direction is relatively larger than that along the radial direction. The blue ellipse represents the estimated covariance by our proposed mathematical modeling (best viewed in color).

${}^t\mathbf{P}$ and ${}^{t-1}\mathbf{P}$, i.e. $\mathbf{p}_i \in {}^t\mathbf{P}$ and $\mathbf{q}_j \in {}^{t-1}\mathbf{P}$, respectively; $\hat{\mathcal{O}}$ denotes the estimated outlier correspondences; $\rho(\cdot)$ denotes surrogate function [41] to reduce the unintended effect caused by outliers and $r(\cdot)$ denotes the squared residual function, i.e. $\|\cdot\|^2$.

C. Anisotropic Uncertainty Modeling of Radar Data

Before we explain our proposed method, the uncertainty of each point is modeled to achieve more radar-friendly ego-motion estimation. Unlike the LiDAR measurements whose uncertainties are isotropic, which means their variances along the axes are likely to be equal to each other [40], [42], radar measurements have anisotropic characteristics. That is, the uncertainty of each point along the azimuthal direction is relatively larger than that along the radial direction; thus, the variances along the x -axis and y -axis are not equal anymore. This is due to the longer wavelength of radar signal compared with that of laser signal of a LiDAR sensor [15].

To model the uncertainty of a radar feature point, the measurement noise is expressed as the superposition of a) radial uncertainty along the direction in which the radar signal is emitted and b) azimuthal uncertainty whose direction is perpendicular to the radial direction. Accordingly, the uncertainty is presented as a banana-like shape (green dots in Fig. 2(b)).

Based on these observations, let us denote the k -th \mathbf{p} as $\mathbf{p}_k = \gamma_k \mathbf{u}_k = \gamma_k [\cos \theta_k \ \sin \theta_k]^\top$ where $\gamma_k \in \mathbb{R}$, $\mathbf{u}_k \in \mathbb{R}^2$, and θ_k denote the measured distance, a normalized direction vector satisfying $\|\mathbf{u}_k\|^2 = 1$, and $\theta_k = \text{atan2}(y_k, x_k)$. Here, x_k and y_k denote the x and y values of \mathbf{p}_k , respectively. Then, the relation between the true range value, γ_k^{true} , and γ_k is defined as follows:

$$\gamma_k = \gamma_k^{\text{true}} + \delta_{\gamma_k} \quad (2)$$

where $\delta_{\gamma_k} \sim \mathcal{N}(0, \sigma_{\gamma_k}^2)$ is a noise term and σ_{γ_k} denotes its standard deviation.

Second, the azimuthal uncertainty is modeled. Let \mathbb{S} be the tangential space that is perpendicular to the radial direction and $\delta_{\mathbf{u}_k}$ be noise term in \mathbb{S} which follows $\delta_{\mathbf{u}_k} \sim \mathcal{N}(0, \sigma_{\mathbf{u}_k}^2)$, where $\sigma_{\mathbf{u}_k}$ denotes the standard deviation in terms of the azimuthal direction. Because $\delta_{\mathbf{u}_k}$ induces unintended perturbation of rotation, the relation between the true direction vector, $\mathbf{u}_k^{\text{true}}$, and \mathbf{u}_k is expressed as follows:

$$\mathbf{u}_k = \mathbf{u}_k^{\text{true}} \boxplus_{\mathbb{S}} \delta_{\mathbf{u}_k} = \exp(\delta_{\mathbf{u}_k} \mathbf{B}) \mathbf{u}_k^{\text{true}} \simeq (\mathbf{I}_2 + \delta_{\mathbf{u}_k} \mathbf{B}) \mathbf{u}_k^{\text{true}} \quad (3)$$

where $\boxplus_{\mathbb{S}}$ denotes the boxplus operation which is an equivalent operator of summation between 2D dimensional vector space and its manifold, i.e. $\boxplus_{\mathbb{S}}: \mathbb{R}^2 \times \mathbb{S} \rightarrow \mathbb{R}^2$; $\exp(\cdot)$ denotes matrix exponential, which maps $\mathfrak{so}(2)$ to $\text{SO}(2)$ space; \mathbf{I}_2 denotes 2×2 identity matrix; $\mathbf{B} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ denotes skew-symmetric matrix, which maps scalar into $\mathfrak{so}(2)$ space.

Finally, the uncertainty and covariance of each point is derived by approximating $\gamma_k \mathbf{u}_k \simeq (\gamma_k^{\text{true}} + \delta_{\gamma_k})(\mathbf{I}_2 + \delta_{\mathbf{u}_k} \mathbf{B}) \mathbf{u}_k^{\text{true}}$, based on (2) and (3). Consequently, the uncertainty, $\delta_{\mathbf{p}_k}$, and its covariance, $\mathbf{C}_{\mathbf{p}_k}$, are expressed respectively as follows:

$$\delta_{\mathbf{p}_k} = \begin{bmatrix} \mathbf{u}_k & \gamma_k \mathbf{B} \mathbf{u}_k \end{bmatrix} \begin{bmatrix} \delta_{\gamma_k} \\ \delta_{\mathbf{u}_k} \end{bmatrix}, \quad (4)$$

$$\mathbf{C}_{\mathbf{p}_k} = \mathbf{A}_k \begin{bmatrix} \sigma_{\gamma_k}^2 & 0 \\ 0 & \sigma_{\mathbf{u}_k}^2 \end{bmatrix} \mathbf{A}_k^\top \quad (5)$$

where $\mathbf{A}_k \in \mathbb{R}^{2 \times 2}$ represents $[\mathbf{u}_k \ \gamma_k \mathbf{B} \mathbf{u}_k]$ for simplicity. Visual description of $\mathbf{C}_{\mathbf{p}_k}$ is represented by the blue ellipse in Fig. 2(b). The modeled $\mathbf{C}_{\mathbf{p}_k}$ is directly utilized in the initialization of rotation estimation and A-COTE (see Sections II.E and II.F).

D. Overview of ORORA

In this subsection, the procedure of ORORA is briefly explained. After the feature extraction and matching, Doppler distortion is compensated by using the estimated relative pose in the previous step, i.e. $\hat{\mathbf{T}}_{t-1}^{t-2}$. Based on the constant velocity assumption, Doppler-compensated range is obtained by adding $\Delta \gamma_k^{\text{DPLR}} = \beta (v_{k,x} \cos(\theta_k) + v_{k,y} \sin(\theta_k))$ where β denotes a constant parameter depending on the modulating frequency of a radar sensor [17]; $v_{k,x}$ and $v_{k,y}$ denote the velocity terms of the previous movement along the x - and y -axes, respectively. In our study, we only deal with autonomous vehicles which follow the non-holonomic motion, so it can be assumed to be $v_{k,y} \simeq 0$. Finally, the range of each point included in \mathcal{A} is corrected by adding $\Delta \gamma_k^{\text{DPLR}}$.

Next, max clique inlier selection (MCIS) [47] is exploited, which is a graph pruning-based outlier rejection

method [25], [40], to estimate initial outlier pairs, $\hat{\mathcal{O}}_{\text{init}}$. Then, these filtered correspondences, i.e. $\mathcal{A} \setminus \hat{\mathcal{O}}_{\text{init}}$, are taken as inputs for the decoupled rotation and translation estimation, respectively (blue arrows in Fig. 2(a)).

Note that there are two main differences between the previous work [25], which also uses MCIS, and ours in pose estimation procedure. First, the authors [25] assumed that there are no outliers in $\mathcal{A} \setminus \hat{\mathcal{O}}_{\text{init}}$, so they estimate relative pose via singular value decomposition [48]. In contrast, our proposed method estimates the outliers again during the rotation and translation estimation, respectively. Second, our method also considers the anisotropic characteristics of radar measurements during the pose estimation.

E. Graduated Non-Convexity-Based Rotation Estimation

Our method is a decoupled method, so $\hat{\mathbf{R}}$ is firstly estimated, which is followed by A-COTE. To estimate $\hat{\mathbf{R}}$, two consecutive pairs (i, j) and $(i', j') = (i + 1, j + 1)$ in $\mathcal{A} \setminus \hat{\mathcal{O}}_{\text{init}}$ are subtracted to offset the effect of translation, i.e. \mathbf{t} . By doing so, the feature pairs are expressed as translation invariant measurements (TIMs) [42] as follows: $\alpha_m = \mathbf{p}_i - \mathbf{p}_i$ for the m -th TIM of ${}^t\mathbf{P}$ and $\beta_m = \mathbf{q}_{j'} - \mathbf{q}_j$ for that of ${}^{t-1}\mathbf{P}$. Note that the effect of translation is cancelled only if both two pairs are true correspondences. Otherwise, the inherent error becomes much larger (of course, the effect of these incorrect measurements is rejected by the following procedure).

Next, GNC with a truncated least square [40], [41] is exploited. To this end, the objective function is first written by leveraging Black-Rangarajan duality as follows [40], [42]:

$$\hat{\mathbf{R}} = \underset{\mathbf{R} \in \text{SO}(2), w_m \in [0,1]}{\operatorname{argmin}} \sum_{m=1}^M [w_m r(\alpha_m, \beta_m, \mathbf{R}) + \Phi_{\rho_\mu}(w_m)] \quad (6)$$

where w_m denotes the weight for the m -th TIM pair, $r(\cdot)$ is the squared residual function, and M is equal to the cardinality of $\mathcal{A} \setminus \hat{\mathcal{O}}_{\text{init}}$. $\Phi_{\rho_\mu}(w_m) = \frac{\mu(1-w_m)}{\mu+w_m} \bar{c}^2$ is a penalty term [44] where μ denotes the control parameter to increase non-convexity and \bar{c} denotes a truncation parameter.

However, solving the equation while simultaneously weighting the measurement pairs does not guarantee the optimality [38]. For this reason, (6) is iteratively solved by using alternating optimization as follows:

$$\hat{\mathbf{R}}^{(n)} = \underset{\mathbf{R} \in \text{SO}(2)}{\operatorname{argmin}} \sum_{m=1}^M \hat{w}_m^{(n-1)} r(\alpha_m, \beta_m, \mathbf{R}), \quad (7)$$

$$\hat{\mathbf{W}}^{(n)} = \underset{w_m \in [0,1]}{\operatorname{argmin}} \sum_{m=1}^M [w_m r(\alpha_m, \beta_m, \hat{\mathbf{R}}^{(n)}) + \Phi_{\rho_\mu}(w_m)] \quad (8)$$

where the superscript n denotes the n -th iteration and $\hat{\mathbf{W}}^{(n)} = \operatorname{diag}(\hat{w}_1^{(n)}, \hat{w}_2^{(n)}, \dots, \hat{w}_M^{(n)})$. Each $\hat{w}_m^{(n)}$ can be solved in a truncated closed form as follows:

$$\hat{w}_m^{(n)} = \begin{cases} 0 & \text{if } \hat{r}_m \in \left[\frac{\mu+1}{\mu} \bar{c}^2, +\infty \right) \\ \bar{c} \sqrt{\frac{\mu(\mu+1)}{\hat{r}_m}} - \mu & \text{if } \hat{r}_m \in \left[\frac{\mu}{\mu+1} \bar{c}^2, \frac{\mu+1}{\mu} \bar{c}^2 \right) \\ 1 & \text{otherwise} \end{cases} \quad (9)$$

where \hat{r}_m denotes $r(\alpha_m, \beta_m, \hat{\mathbf{R}}^{(n)})$ for simplicity. For each iteration, μ is updated as $\mu^{(n)} \leftarrow \kappa \cdot \mu^{(n-1)}$ where $\kappa > 1$ is a factor that gradually increases the magnitude of non-convexity. The iteration ends if the differential of $\sum_{m=1}^M \hat{w}_m^{(n)} \hat{r}_m$ becomes sufficiently small.

The main difference between ours and the previous works [40], [42], [44] is that our method initializes the control parameter and weights, i.e. $\mu^{(0)}$ and $\hat{w}_m^{(0)}$, by considering the uncertainty. The initialization consists of four steps. First, by considering the uncertainty of the j and $(j+1)$ -th points, let the approximated prior weight be $\hat{w}_m^{(-1)} \simeq 1/(\eta \sqrt{\gamma_j^2 + \gamma_{j+1}^2})$ based on (5), which means that the farther, the more uncertain. Here, η denotes a normalization term and the superscript (-1) indicates that the weight is on the preceding step of the initialization. Then, by solving (7) with $\hat{w}_m^{(-1)}$, $\hat{\mathbf{R}}^{(0)}$ is estimated. Third, $\mu^{(0)}$ is set as $\mu^{(0)} = \bar{c}^2 / (2 \cdot \max(r(\alpha_m, \beta_m, \hat{\mathbf{R}}^{(0)})) - \bar{c}^2)$ [42]. Finally, $w_m^{(0)}$ is set by (9) taking $\hat{\mathbf{R}}^{(0)}$ and $\mu^{(0)}$ as inputs.

F. Anisotropic Component-Wise Translation Estimation

Finally, the relative x and y translations are estimated respectively by A-COTE. Let the discrepancy of the relative translation be $\mathbf{v}_{ij} = \mathbf{q}_j - \hat{\mathbf{R}}\mathbf{p}_i$, then the covariance of \mathbf{v}_{ij} , \mathbf{C}_{ij} , is equal to $\mathbf{C}_{\mathbf{q}_j} + \hat{\mathbf{R}}\mathbf{C}_{\mathbf{p}_i}\hat{\mathbf{R}}^\top$, where $\mathbf{C}_{\mathbf{q}_j}$ and $\mathbf{C}_{\mathbf{p}_i}$ are calculated by (5), respectively. Accordingly, each uncertainty of \mathbf{v}_{ij} along the x - and y -axes is equal to the 1st and 2nd diagonal elements of \mathbf{C}_{ij} , which are denoted as ${}^1\sigma_{ij}$ and ${}^2\sigma_{ij}$, respectively. Note that ${}^1\sigma_{ij}$ and ${}^2\sigma_{ij}$ have different magnitudes due to the aforementioned anisotropic characteristics of radar points. For simplicity, the property that corresponds to the l -th element is expressed as ${}^l(\cdot)$, where $l = 1, 2$.

Then, the component-wise relative translation is estimated in the following four steps. First, a boundary interval set is defined as ${}^l\mathcal{E}$ which is $2M$ -tuples. That is, ${}^l\mathcal{E}$ consists of M lower bounds, ${}^l\mathbf{v}_{ij} - {}^l\sigma_{ij}$, and M upper bounds, ${}^l\mathbf{v}_{ij} + {}^l\sigma_{ij}$. It is assumed that all the elements of ${}^l\mathcal{E}$ are sorted in ascending order. Second, consensus sets are assigned by using these noise bound values. That is, the g -th consensus set is defined as ${}^l\mathcal{I}_g = \{(i, j) | ({}^l\phi_g - {}^l\mathbf{v}_{ij})^2 \leq {}^l\sigma_{ij}^2\}$, where ${}^l\phi_g \in \mathbb{R}$ is any value that satisfies ${}^l\mathcal{E}(g) < {}^l\phi_g < {}^l\mathcal{E}(g+1)$ for $g = 1, 2, \dots, (2M-1)$.

Third, the optimal relative translation value for the g -th consensus set, ${}^l\hat{\mathbf{t}}_g$, is estimated by the weighted average if ${}^l\mathcal{I}_g$ is non-empty as follows:

$${}^l\hat{\mathbf{t}}_g = \left(\sum_{(i,j) \in {}^l\mathcal{I}_g} \frac{1}{{}^l\sigma_{ij}^2} \right)^{-1} \sum_{(i,j) \in {}^l\mathcal{I}_g} \frac{{}^l\mathbf{v}_{ij}}{{}^l\sigma_{ij}^2}. \quad (10)$$

Finally, by letting the set consisting of all the ${}^l\hat{\mathbf{t}}_g$ be \mathbf{H} whose size is up to $2M-1$ and the complement of ${}^l\mathcal{I}_g$ be ${}^l\mathcal{O}_g$, i.e. ${}^l\mathcal{I}_g = \mathcal{A} \setminus (\hat{\mathcal{O}}_{\text{init}} \cup {}^l\mathcal{O}_g)$, the global optimum, ${}^l\hat{\mathbf{t}}$, is estimated by selecting $\xi \in \mathbf{H}$ and ${}^l\mathcal{I}_g$ that minimize the following objective function as follows:

$${}^l\hat{\mathbf{t}}, {}^l\hat{\mathcal{I}}_g = \underset{\xi \in \mathbf{H}, {}^l\mathcal{I}_g}{\operatorname{argmin}} \sum_{(i,j) \in {}^l\mathcal{I}_g} \left(\frac{\xi - {}^l\mathbf{t}_{ij}}{{}^l\sigma_{ij}} \right)^2 + \sum_{(i,j) \in {}^l\mathcal{O}_g} {}^l\sigma_{ij}. \quad (11)$$

It is noticeable that (11) blends the best of consensus maximization [49] and weighted average. That is, the second summand is usually much larger than the first summand, so ${}^l\mathcal{I}_g$ with the most cardinality is likely to be chosen as ${}^l\hat{\mathcal{I}}_g$, which behaves like consensus maximization. In the meanwhile, even though some outliers, each of whose uncertainty is sufficiently large, are unintentionally included in ${}^l\hat{\mathcal{I}}_g$, their larger ${}^l\sigma_{ij}$ naturally suppresses the effect of the outliers on the component-wise relative translation estimation.

III. EXPERIMENTS

A. Dataset

In our experiments, most of the sequences of MulRan dataset¹ [27] were used. Note that two sequences, i.e. KAIST01 and Sejong01, were not used because the authors [27] officially do not provide radar data of these two sequences in the Oxford radar image format [12].

B. Error Metrics

To quantitatively evaluate our proposed method and baseline methods, the relative translation error, t_{rel} , and rotation error, r_{rel} , are used [50], which are renowned criteria in odometry test. Note that these terms are normalized because each ego-motion has a different amount of movement from each other. Consequently, the units of t_{rel} and r_{rel} are % and deg/100 m, respectively.

C. Implementation Details and Parameter Setting

First, before the feature matching, ${}^t\mathbf{P}$ is voxel-sampled with the size of ν to let these points be more evenly distributed. Then, the parameters presented in [17] were employed for feature extraction and matching procedure.

Empirically, it was found that optimal parameters can be changed depending on the environments (see Section IV.A). For this reason, we set $\nu = 0.6$ m, $\bar{c} = 0.75$, and $\sigma_{\mathbf{u}_k} = 10.8^\circ$ in widely distributed environments, whereas we set $\nu = 0.8$ m, $\bar{c} = 1.0$, and $\sigma_{\mathbf{u}_k} = 1.8^\circ$ in more partially distributed environments. Other parameters are set to be $\kappa = 1.4$ and $\sigma_{\gamma_k} = 0.1$ m.

IV. EXPERIMENTAL RESULTS

In our experiments, we compared our ORORA with other feature-based state-of-the-art methods. To this end, three types of RANSAC proposed in [17] are used: the original RANSAC [39] without any compensation, MC-RANSAC whose motion distortion is compensated [17], and MC-RANSAC + DPLR that both motion and Doppler distortions are compensated [17]. For a fair comparison, we used the official open-source implementations. To closely examine the performance of ego-motion estimation depending on the quality of feature extraction, two feature extraction methods proposed in [13] and [14] were tested, which are referred to as Cen2018 and Cen2019, respectively.

¹<https://sites.google.com/view/mulran-pr/home>

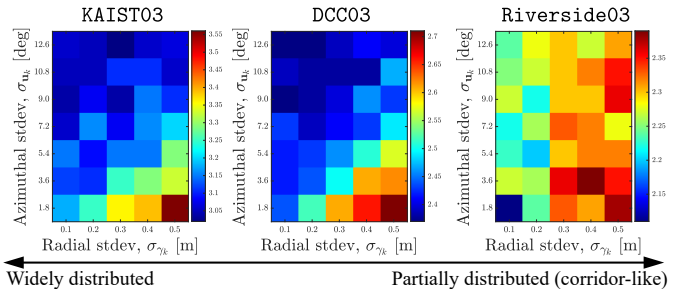


Fig. 3. Translation error with respect to the size of the radial and azimuthal uncertainties of each feature point. The double arrow indicates the level of distribution of feature points on the xy -plane (best viewed in color).

TABLE I: Comparison of translation error, t_{rel} , between component-wise translation estimation (COTE) [40] and our A-COTE. COTE considers the uncertainty of measurements as being isotropic. It was observed that A-COTE is more significant in obstructed environments where more buildings are located in the lateral side of the sensor frame (unit: %).

Module	KAIST03	DCC03	Riverside03
COTE [40]	3.21	2.40	2.09
A-COTE (Ours)	3.04	2.37	2.11

A. Performance Changes With Respect to Uncertainty

First, performance changes in terms of radial and azimuthal uncertainties were analyzed. Interestingly, the different tendency was observed depending on the environments, as shown in Fig. 3. It was shown that larger azimuthal uncertainty leads to significant performance improvement in KAIST03, which includes more buildings so that feature points are widely distributed on the xy -plane. The experimental results can be interpreted that the larger azimuthal uncertainty enables to suppress the potential errors caused by the feature points located distantly on the lateral side, i.e. along the y -axis of the sensor frame. That is, these points are likely to have larger uncertainty along the x -axis due to their azimuthal uncertainties.

In contrast, smaller azimuthal uncertainty showed better performance in Riverside03. This is because the features are likely to be distributed along the x -axis. Accordingly, the effect of azimuthal uncertainty less affects the translation estimation. For this reason, our A-COTE shows comparable performance with the case where anisotropic modeling is not employed, i.e. COTE [40], in Riverside03, as shown in Table I.

Therefore, it was demonstrated that our A-COTE is more effective in complex urban environments.

B. Comparison With the State-of-the-Art Methods

In general, the state-of-the-art methods showed precise odometry results, overcoming the effect of outliers. However, our ORORA showed more significant performance, as shown in Fig. 4 and Table II. In particular, it is noticeable that no matter what feature extraction method was used, our ORORA outputs the most precise odometry performance. Therefore, it can be concluded that our ORORA is less sensitive to the performance of feature extraction methods.

Furthermore, it was observed that our ORORA is more robust against the featureless scenes. In MulRan dataset,

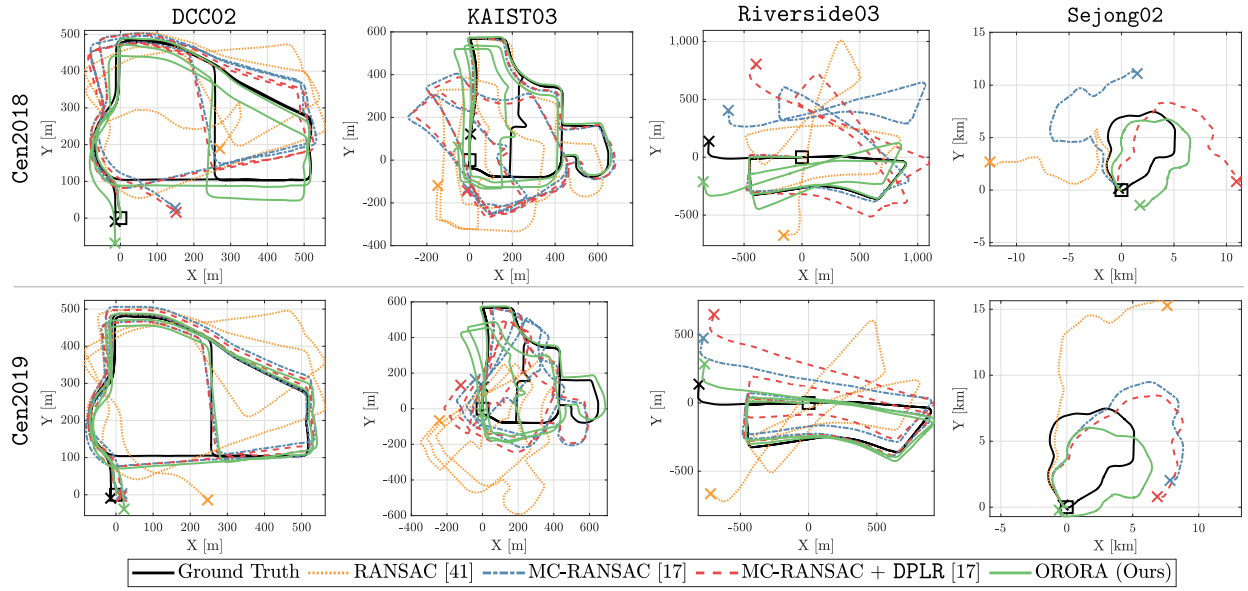


Fig. 4. Qualitative comparison with the state-of-the-art methods depending on the feature extraction methods, namely, Cen2018 [13] and Cen2019 [14], in MulRan dataset [27]. \square and \times marks denote start and end points, respectively (best viewed in color).

TABLE II: Comparison of odometry test with the state-of-the-art methods on MulRan dataset [27]. Given the feature points by Cen2018 [13] or Cen2019 [14], our ORORA outperforms the state-of-the-art methods. The bold and the gray highlight indicate the best performance among the whole method and that for each feature extraction in each sequence, respectively. All errors are represented in the form of relative translation error [%] / relative rotation error [deg/100 m].

	Method	Sequence									
		DCC01	DCC02	DCC03	KAIST02	KAIST03	RIV01	RIV02	RIV03	Sejong02	Sejong03
Cen2018	RANSAC [39]	5.11/1.10	5.39/1.83	4.43/1.43	4.80/1.16	4.94/1.42	5.19/1.27	7.46/1.87	7.31/2.21	10.23/1.71	9.17/1.95
	MC-RANSAC [17]	4.79/1.27	3.76/1.00	4.39/1.40	6.60/1.74	4.67/1.19	4.76/1.08	6.49/1.73	7.72/2.19	9.18/2.14	8.55/1.74
	MC-RANSAC + DPLR [17]	4.49/1.12	3.64/1.00	4.37/1.39	6.32/1.68	4.42/1.12	5.92/1.43	8.38/2.08	6.61/1.88	8.83/2.32	12.01/2.46
	ORORA (Ours)	3.18/0.66	2.38/0.57	2.77/0.79	3.12/0.78	2.53/0.57	3.51/0.76	3.31/0.76	2.79/0.64	3.76/0.67	5.07/0.86
Cen2019	RANSAC [39]	5.41/1.39	4.16/1.07	3.66/1.06	4.60/1.39	6.13/1.56	3.99/0.96	3.54/0.98	4.23/1.11	4.83/1.18	8.65/1.66
	MC-RANSAC [17]	4.17/0.95	3.14/0.66	2.70/0.67	3.46/1.00	3.31/0.99	3.83/1.07	3.61/1.03	3.94/0.96	4.43/1.01	6.62/1.12
	MC-RANSAC + DPLR [17]	4.01/0.89	2.87/0.55	2.70/0.66	3.59/1.08	3.60/1.05	4.12/1.13	3.93/1.11	3.81/0.92	4.51/1.01	6.55/1.13
	ORORA (Ours)	3.12/0.67	2.60/0.51	2.37/0.57	3.28/0.82	3.04/0.70	3.53/0.84	2.67/0.64	2.11/0.49	3.27/0.75	4.05/0.70

sequences captured on the riverside and Sejong city have fewer geometrical characteristics because these scenes are wide-opened environments, so the feature extraction and matching results output less number of feature pairs and a larger ratio of outliers within the correspondences. Under that circumstance, the state-of-the-art methods showed larger performance degradation than ORORA. In addition, it was shown that once the odometry estimation becomes imprecise, the Doppler compensation occasionally rather deteriorates the odometry performance (see the performance difference of MC-RANSAC and MC-RANSAC + DPLR).

Therefore, these experimental results demonstrate that our ORORA is the most robust method in that our proposed method tolerates the effect of outliers and the performance of ORORA is more invariant to the quality of the estimated correspondences as well.

C. Algorithm Speed

Furthermore, our ORORA is available in real-time, as presented in Fig. 5. The speed of ORORA depends on the number of feature pairs, and it just took 5.63 msec on average. The whole pipeline runs at 11.5 Hz, which is sufficiently faster than the frequency at which a radar image is acquired, i.e. 4 Hz.

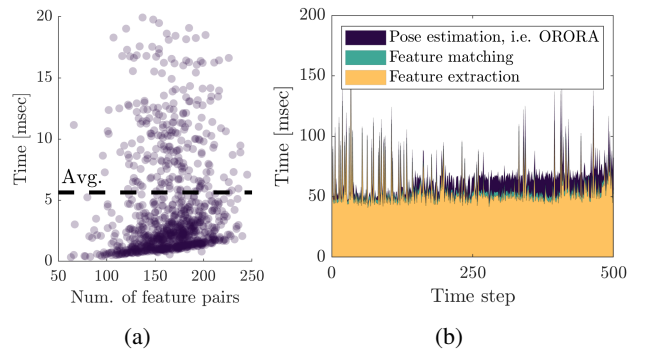


Fig. 5. (a) Time taken for pose estimation with respect to the number of correspondences. The average time is indicated by the black dashed line (Avg: 5.63 msec). (b) Area plot of computation time of each module on Intel(R) Core i7-7700K in KAIST03 of the MulRan dataset [27]. As a feature extraction method, Cen2019 [14] was exploited (best viewed in color).

V. CONCLUSION

In this study, an outlier-robust odometry method, *ORORA*, has been proposed. The experimental results demonstrate that our ORORA successfully estimates ego-motion while rejecting the effect of outliers. In future works, we plan to study the radar-based SLAM framework by adding loop detection and loop closing modules.

REFERENCES

- [1] H. Lim, S. Hwang, and H. Myung, "ERASOR: Egocentric ratio of pseudo occupancy-based dynamic object removal for static 3D point cloud map building," *IEEE Robot. Autom. Lett.*, vol. 6, no. 2, pp. 2272–2279, 2021.
- [2] T. Shan, B. Englot, D. Meyers, W. Wang, C. Ratti, and D. Rus, "LIO-SAM: Tightly-coupled LiDAR inertial odometry via smoothing and mapping," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, 2020, pp. 5135–5142.
- [3] Y. Kim, B. Yu, E. M. Lee, J.-h. Kim, H.-w. Park, and H. Myung, "STEP: State estimator for legged robots using a preintegrated foot velocity factor," *IEEE Robot. Automat. Lett.*, vol. 7, no. 2, pp. 4456–4463, 2022.
- [4] J. Behley and C. Stachniss, "Efficient surfel-based SLAM using 3D laser range data in urban environments," in *Robot. Sci. Syst.*, 2018.
- [5] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, and J. Gall, "SemanticKITTI: A dataset for semantic scene understanding of LiDAR sequences," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 9297–9307.
- [6] M. Oh, E. Jung, H. Lim, W. Song, S. Hu, E. M. Lee, J. Park, J. Kim, J. Lee, and H. Myung, "TRAVEL: Traversable ground and above-ground object segmentation using graph representation of 3D LiDAR scans," *IEEE Robot. Automat. Lett.*, 2022, DOI: 10.1109/LRA.2022.3182096.
- [7] S. Jung, D. Choi, S. Song, and H. Myung, "Bridge inspection using unmanned aerial vehicle based on HG-SLAM: Hierarchical graph-based SLAM," *MDPI Remote Sens.*, vol. 12, no. 18, pp. 3022–3041, 2020.
- [8] D.-U. Seo, H. Lim, S. Lee, and H. Myung, "PaGO-LOAM: Robust ground-optimized LiDAR odometry," in *Proc. Int. Conf. Ubiquiti. Robot.*, 2021, [Online]. Available: arXiv preprint arXiv:2206.00266.
- [9] T. Qin, P. Li, and S. Shen, "VINS-Mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Trans. Robot.*, vol. 34, no. 4, pp. 1004–1020, 2018.
- [10] A. J. Lee, Y. Cho, Y.-s. Shin, A. Kim, and H. Myung, "ViViD++: Vision for visibility dataset," *IEEE Robot. Automat. Lett.*, vol. 7, no. 3, pp. 6282–6289, 2022.
- [11] H. Lim, J. Jeon, and H. Myung, "UV-SLAM: Unconstrained line-based SLAM using vanishing points for structural mapping," *IEEE Robot. Automat. Lett.*, vol. 7, no. 2, pp. 1518–1525, 2022.
- [12] D. Barnes, M. Gadd, P. Murcutt, P. Newman, and I. Posner, "The Oxford radar robotcar dataset: A radar extension to the Oxford robotcar dataset," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2020, pp. 6433–6438.
- [13] S. H. Cen and P. Newman, "Precise ego-motion estimation with millimeter-wave radar under diverse and challenging conditions," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2018, pp. 6045–6052.
- [14] —, "Radar-only ego-motion estimation in difficult settings via graph matching," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2019, pp. 298–304.
- [15] K. Burnett, Y. Wu, D. J. Yoon, A. P. Schoellig, and T. D. Barfoot, "Are we ready for radar to replace LiDAR in all-weather mapping and localization?" *IEEE Robot. Automat. Lett.*, 2022, DOI: 10.1109/LRA.2022.3192885.
- [16] Y. Zhou, L. Liu, H. Zhao, M. López-Benítez, L. Yu, and Y. Yue, "Towards deep radar perception for autonomous driving: Datasets, methods, and challenges," *Sensors*, vol. 22, no. 11, pp. 4208–4252, 2022.
- [17] K. Burnett, A. P. Schoellig, and T. D. Barfoot, "Do we need to compensate for motion distortion and Doppler effects in spinning radar navigation?" *IEEE Robot. Automat. Lett.*, vol. 6, no. 2, pp. 771–778, 2021.
- [18] K. Burnett, D. J. Yoon, Y. Wu, A. Z. Li, H. Zhang, S. Lu, J. Qian, W.-K. Tseng, A. Lambert, K. Y. Leung, *et al.*, "Boreas: A multi-season autonomous driving dataset," *arXiv preprint arXiv:2203.10168*, 2022.
- [19] K. Han and S. Hong, "Differential phase Doppler radar with collocated multiple receivers for noncontact vital signal detection," *IEEE Trans. Microw. Theory Tech.*, vol. 67, no. 3, pp. 1233–1243, 2018.
- [20] —, "Vocal signal detection and speaking-human localization with MIMO FMCW radar," *IEEE Trans. Microw. Theory Tech.*, vol. 69, no. 11, pp. 4791–4802, 2021.
- [21] —, "Detection and localization of multiple humans based on curve length of I/Q signal trajectory using MIMO FMCW radar," *IEEE Microw. Wirel. Compon. Lett.*, vol. 31, no. 4, pp. 413–416, 2021.
- [22] P. Checchin, F. Gérossier, C. Blanc, R. Chapuis, and L. Trassoudaine, "Radar scan matching SLAM using the Fourier-Mellin transform," in *Proc. Field and Serv. Robot.*, 2010, pp. 151–161.
- [23] M. Holder, S. Hellwig, and H. Winner, "Real-time pose graph SLAM based on radar," in *Proc. IEEE Intell. Vehicles Symp.*, 2019, pp. 1145–1151.
- [24] Y. S. Park, Y.-S. Shin, J. Kim, and A. Kim, "3D ego-motion estimation using low-cost mmWave radars via radar velocity factor for pose-graph SLAM," *IEEE Robot. Automat. Lett.*, vol. 6, no. 4, pp. 7691–7698, 2021.
- [25] Z. Hong, Y. Petillot, and S. Wang, "RadarSLAM: Radar based large-scale SLAM in all weathers," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, 2020, pp. 5164–5170.
- [26] Z. Hong, Y. Petillot, A. Wallace, and S. Wang, "RadarSLAM: A robust simultaneous localization and mapping system for all weather conditions," *Int. J. Robot. Res.*, 2022, DOI: <https://doi.org/10.1177/02783649221080483>.
- [27] G. Kim, Y. S. Park, Y. Cho, J. Jeong, and A. Kim, "MulRan: Multimodal range dataset for urban place recognition," in *IEEE Int. Conf. Robot. Automat.*, 2020, pp. 6246–6253.
- [28] Y. S. Park, Y.-S. Shin, and A. Kim, "PhaRaO: Direct radar odometry using phase correlation," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2020, pp. 2617–2623.
- [29] K. Burnett, D. J. Yoon, A. P. Schoellig, and T. D. Barfoot, "Radar odometry combining probabilistic estimation and unsupervised feature learning," *arXiv preprint arXiv:2105.14152*, 2021.
- [30] D. Barnes, R. Weston, and I. Posner, "Masking by moving: Learning distraction-free radar odometry from pose information," *arXiv preprint arXiv:1909.03752*, 2019.
- [31] P.-C. Kung, C.-C. Wang, and W.-C. Lin, "A normal distribution transform-based radar odometry designed for scanning and automotive radars," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2021, pp. 14417–14423.
- [32] D. Adolfsson, M. Magnusson, A. Alhashimi, A. J. Lilienthal, and H. Andreasson, "CFEAR radarodometry-conservative filtering for efficient and accurate radar odometry," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, 2021, pp. 5462–5469.
- [33] R. Aldera, D. De Martini, M. Gadd, and P. Newman, "What could go wrong? Introspective radar odometry in challenging environments," in *Proc. IEEE Intell. Transport. Syst. Conf.*, 2019, pp. 2835–2842.
- [34] E. B. Quist, P. C. Niefeldt, and R. W. Beard, "Radar odometry with recursive-RANSAC," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 4, pp. 1618–1630, 2016.
- [35] K. Han and S. Hong, "Phase-extraction method with multiple frequencies of FMCW radar for human body motion tracking," *IEEE Microw. Wirel. Compon. Lett.*, vol. 30, no. 9, pp. 927–930, 2020.
- [36] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 1, pp. 208–216, 1992.
- [37] D. Barnes and I. Posner, "Under the radar: Learning to predict robust keypoints for odometry estimation and metric localisation in radar," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2020, pp. 9484–9490.
- [38] Q.-Y. Zhou, J. Park, and V. Koltun, "Fast global registration," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 766–782.
- [39] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [40] H. Lim, S. Yeon, S. Ryu, Y. Lee, Y. Kim, J. Yun, D. Lee, and H. Myung, "A single correspondence is enough: Robust global registration to avoid degeneracy in urban environments," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2022, [Online]. Available: arXiv preprint arXiv:2203.06612.
- [41] V. Tzoumas, P. Antonante, and L. Carlone, "Outlier-robust spatial perception: Hardness, general-purpose algorithms, and guarantees," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, 2019, pp. 5383–5390.
- [42] H. Yang, J. Shi, and L. Carlone, "TEASER: Fast and certifiable point cloud registration," *IEEE Trans. Robot.*, vol. 37, no. 2, pp. 314–333, 2020.
- [43] L. Sun, "IRON: Invariant-based highly robust point cloud registration," *arXiv preprint arXiv:2103.04357*, 2021.
- [44] H. Yang, P. Antonante, V. Tzoumas, and L. Carlone, "Graduated non-convexity for robust spatial perception: From non-minimal solvers to global outlier rejection," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 1127–1134, 2020.

- [45] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2011, pp. 2564–2571.
- [46] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2, 1999, pp. 1150–1157.
- [47] R. A. Rossi, D. F. Gleich, A. H. Gebremedhin, M. M. A. Patwary, and M. Ali, "Parallel maximum clique algorithms with applications to network analysis and storage," *arXiv preprint arXiv:1302.6256*, 2013.
- [48] J. H. Challis, "A procedure for determining rigid body transformation parameters," *J. biomech.*, vol. 28, no. 6, pp. 733–737, 1995.
- [49] T. Probst, D. P. Paudel, A. Chhatkuli, and L. V. Gool, "Unsupervised learning of consensus maximization for 3D vision problems," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 929–938.
- [50] Z. Zhang and D. Scaramuzza, "A tutorial on quantitative trajectory evaluation for visual(-inertial) odometry," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, 2018.