Precise Ego-Motion Estimation with Millimeter-Wave Radar under Diverse and Challenging Conditions

Sarah H. Cen and Paul Newman

Abstract— In contrast to cameras, lidars, GPS, and proprioceptive sensors, radars are affordable and efficient systems that operate well under variable weather and lighting conditions, require no external infrastructure, and detect long-range objects. In this paper, we present a reliable and accurate radar-only motion estimation algorithm for mobile autonomous systems. Using a frequency-modulated continuous-wave (FMCW) scanning radar, we first extract landmarks with an algorithm that accounts for unwanted effects in radar returns. To estimate relative motion, we then perform scan matching by greedily adding point correspondences based on unary descriptors and pairwise compatibility scores. Our radar odometry results are robust under a variety of conditions, including those under which visual odometry and GPS/INS fail.

I. INTRODUCTION

In order to confidently travel through its environment, an autonomous vehicle must achieve robust localization and navigation despite changing conditions and moving objects. Currently, most platforms employ lidar, vision, GPS, internal sensors, or a combination of these systems to obtain information about their surroundings and perform motion estimation. While extremely fast and high-resolution, lidar is sensitive to weather conditions, especially rain and fog, and cannot see past the first surface encountered. Vision systems are versatile and cheap but easily impaired by scene changes, like poor lighting or the sudden presence of snow. Both optical sensors only yield dependable results for short-range measurements. A typical GPS on its own guarantees at best 3-m accuracy, experiences reception difficulties near obstructions, and relies on an external infrastructure. Additionally, proprioceptive sensors, like wheel encoders and IMUs, suffer from significant drift among other detrimental effects.

In contrast, radar is a long-range, on-board system that performs well under a variety of lighting and atmospheric conditions, and it is rapidly becoming more affordable and efficient than lidar. Due to its long wavelength (which allows it to penetrate certain materials) and beam spread, radar can return multiple readings from the same transmission and generate a grid representation of its world. As a result, radar sensors detect stable, long-range features in the environment.

For these reasons, radar is a promising sensor for odometry, a task for which it is not typically utilized, and we seek to explore its capabilities via a radar-only system. In this paper, we demonstrate robust motion estimation using a frequency-modulated continuous-wave (FMCW) scanning radar alone. Our main contributions are two-fold: (1) a landmark extraction method that reliably identifies meaningful features while avoiding false detections; and (2) a robust radar-only scan matching algorithm that is effective under any rotations and
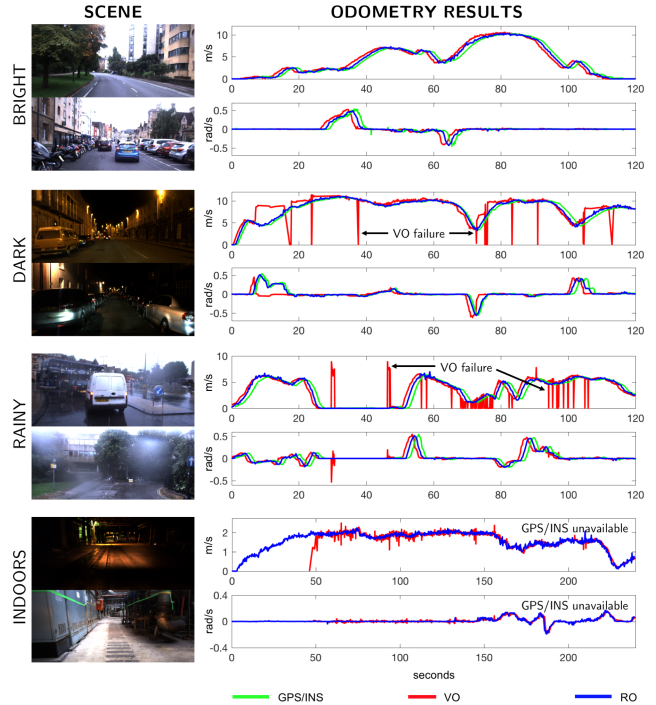

Fig. 1.   A comparison of three odometry systems: radar (RO), vision (VO), and GPS/INS (note: the curves are shifted in time for better visual comparison). Only the radar is successful across all scenes, and its motion estimates closely match those of the other sensors when they are available. VO fails in poor lighting or rain, and GPS/INS is unavailable indoors.

large translations. While radar offers many benefits over the alternatives, working with it is non-trivial as it operates more slowly than lidar, generates noisy measurements, and often returns "ghost" objects. We address these issues and show that radar-only motion estimation succeeds across diverse settings and conditions. Moreover, we show that radar does well compared against vision and GPS even under conditions that are ideal for the latter two sensors.

Sections II and III-A discuss the relevant literature on radar and introduce the FMCW scanning radar, respectively. Section III details our algorithms on landmark extraction and data association. Section IV discusses our results, and Section V summarizes our work and suggests future directions.

II. RELATED WORKS

The first step of most radar motion estimation approaches is the extraction of important features from radar scans. Some researchers [1], [2] draw from the literature on vision, creating amplitude gridmaps that transform the radar scans into grayscale images, then extracting features, such as SIFT

and FAST. Others [3] use the gridmaps to find continuous areas that are deemed interesting using DBSCAN, MSER, and Connected Components. In contrast, approaches devised specifically for the radar acknowledge its high and irregular noise floor. CFAR [4], [5], a common filtering algorithm, adapts to the variable noise floor along the received signal. Another technique [6], [7] infers the appearance of landmarks by estimating the radar's noise characteristics and exploiting temporal-spatial continuity. Marck et al. [8] avoid the filtering process altogether by recording only the range with the greatest power return per azimuth, but this simplification discards potentially relevant information. As expected, the landmark extractions tailored to radar produce the most meaningful and robust detections.

The next step for motion estimation is the data association of landmarks corresponding to the same object observed at different times. These approaches assume that the majority of the scene is static. Vision-inspired works [1], [2] pair radar landmarks that have sufficiently similar feature descriptors. Modifying this idea for radar occupancy grids, Schuster et al. [9] identify and associate landmarks using the binary annular statistic descriptor (BASD) [10] and Hamming distance. Although feature descriptors work well for images, which contain complex high-density information, they are unable to produce consistent results with radar, for which the readings are much noisier and less dense, and fail to consider the advantages of analyzing the radar scan as a whole.

An alternative to feature-based radar odometry uses multi-sensor fusion. These systems use the other sensors' odometry to transform the incoming radar landmark pointcloud and compare it against an existing map of landmarks. Schuster et al. [9] match each point to its nearest neighbor in the map. Diessler et al. [11] use Monte Carlo methods to select a solution from probabilistically assigned weights. The data association between the radar pointcloud and map provides a motion estimate that is then fused with the original odometry readings. While these multi-sensor methods are promising, none produce satisfactory results using only the radar, and they thus rely on the availability of the other sensors. In addition, they often require the integration of model-reliant filters (e.g. Kalman and particle) and the creation of maps, both of which introduce unnecessary complications.

On the other hand, scan matching, which aligns landmark sets in order to minimize some cost function, accomplishes data association by considering information captured across the entire radar scan and does so without other sensors. One widespread approach, called Iterative Closest Point (ICP), iteratively matches and aligns the pointclouds until the desired termination condition is met [8], [12], [13]. Chandran and Newman [14] adopt a different strategy, developing a function that quantifies the quality of a map created by superimposing radar pointclouds according to the unknown motion parameters; they then perform an exhaustive search to optimize over the motion parameters. Both works assume that the movement between scans is small, which imposes an undesirable constraint on the algorithms and prevents them from being applied to arbitrary inputs. Vivet et al. [5] design
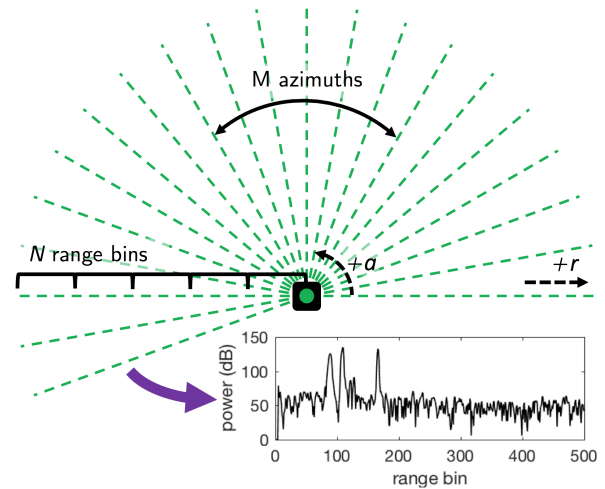

Fig. 2. Visualization of the FMCW scanning radar viewed from above. The radar (green circle), centered on the vehicle (black box), sequentially gathers power-range spectra (dotted green rays) at each azimuth. A sample signal is plotted. Variables a and r denote azimuth and range, respectively.

an innovative technique well suited for high velocities, using the radar scan distortions, which are typically regarded as a drawback of mobile radar systems, to backsolve for velocity with the help of an extended Kalman filter. Rapp et al. combine spatial scan matching with Doppler information for a joint ego-motion estimate [15].

Other scan matching algorithms do not extract features and operate directly on the radar outputs instead. Checchin et al. [16] apply the Fourier-Mellin transform to efficiently compute the vehicle's rotation and translation from the entire radar output. The Doppler radar used by Kellner et al. [17] returns the position and speed of the surrounding objects, from which the vehicle's motion is easily computed with enough detections. Both concepts are unencumbered by heavy preprocessing. Yet feature extraction is often necessary for other tasks, like object tracking, so forgoing this step reduces the system's overall versatility.

In this paper, we present robust radar-only motion estimation using our own algorithms for landmark extraction and scan matching. We adopt this approach in order to fully utilize the information captured by the radar while providing a method that identifies meaningful radar features that are useful for other tasks. In contrast to the works above, we accomplish these goals without other sensors, the creation of maps, model-reliant filters, or outlier detection.

III. OUR APPROACH

A. FMCW Scanning Radar

We employ the FMCW scanning radar, which is visualized in Figure 2. This radar rotates about its vertical axis while continuously transmitting and receiving frequency-modulated radio waves. The received power corresponding to a position in the environment indicates the reflectivity, size, and orientation of an object at that location. The radar inspects one azimuth at a time. For each, it produces a 1D signal, called the power-range spectrum, that contains
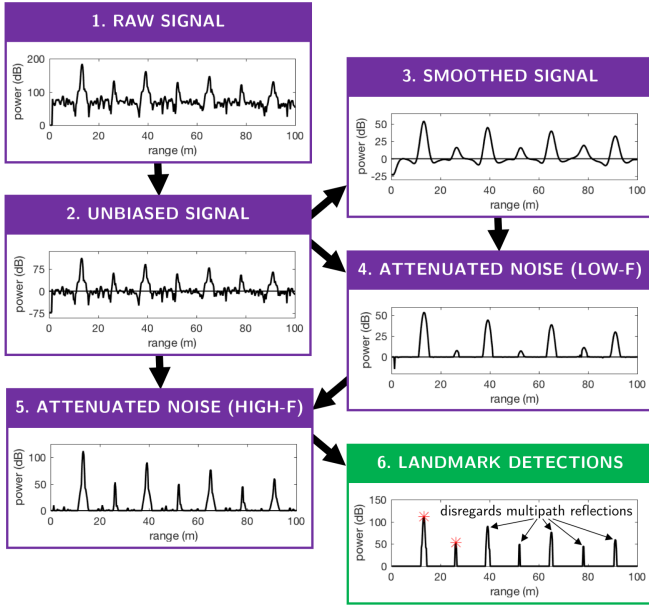
Fig. 3. Procedure for landmark extraction from a power-range spectrum. The input (raw signal) is processed from the top-left to produce the output on the bottom-right, in which the landmarks are denoted with red asterisks. Box 6 in this example highlights the ability of our approach to remove detections due to ghost objects and noise. Boxes 3 and 5 show the importance of incorporating the high-frequency signals, as using the smooth ones in boxes 2 and 4 alone would discard the high range resolution of the FMCW radar.

the power readings for a series of range bins, and one full rotation across all azimuths constitutes a scan. Let N be the number of range bins in a power-range spectrum and M the number of azimuths in one rotation.

The FMCW radar's advantages are its range-measurement accuracy, ability to gather readings at close range, and low peak power. Its disadvantages include sidelobes—radiation sent in unintended directions—and multipath reflections—which result when a wave encounters additional surfaces before returning to the receiver. These two effects cause the scan to contain "ghost" objects at locations where in fact none exist. Errors in range can also occur due to relative motion (via the Doppler effect) and the compression of 3D information into the 1D spectrum. Other issues include phase jitter, saturation, and atmospheric attenuation.

B. Landmark Extraction

Our first objective is to accurately detect objects in the radar's environment with minimal false positives. Specifically, our detector should find all landmarks perceived by the radar while minimizing the number of redundant returns per landmark and avoiding the detection of nonexistent ones, such as those due to noise, multipath reflections, harmonics, and sidelobes. In this section, we present our method for extracting landmarks while adhering to the aims above.

Our method accepts power-range spectra (i.e. 1D signals), as inputs and returns a set of landmarks, each specified by its range and azimuth. The core idea is to estimate the signal's noise statistics then scale the power value at each range by the probability that it corresponds to a real detection.

Algorithm 1: Landmark Extraction Method

Input: Power-range spectra $s \in \mathbb{R}^{N \times 1}$ for azimuth a
Output: Set of landmark detections $L(\mathbf{s})$
Parameters: Median filter width w_{median}; binomial filter width w_{binom}; z-value $z_{\mathbf{q}}$ to threshold noise; boolean F and threshold d_{thresh} (optional) for multipath reflection removal.

1   $\mathbf{q} \leftarrow \mathbf{s} - \text{medianFilter}(\mathbf{s}, w_{\text{median}})$
2   $\mathbf{p} \leftarrow \text{binomialFilter}(\mathbf{s}, w_{\text{binom}})$
3   $Q \leftarrow \{q_i : q_i \leq 0\}$
4   $\mathcal{N}(\mu_{\mathbf{q}}, \sigma_{\mathbf{q}}^2) \leftarrow \text{normalDistribution}(Q \cup -Q)$
5   Initialize $N \times 1$ vector $\hat{\mathbf{y}}$ to zeros.
6   **for** $i \leftarrow 1$ **to** N **do**
7   **if** $q_i > 0$ **then**
8         $\hat{y}_i \leftarrow p_i \cdot \left(1 - \frac{f(p_i | 0, \sigma_{\mathbf{q}}^2)}{f(0 | 0, \sigma_{\mathbf{q}}^2)}\right)$
9         $\hat{y}_i \leftarrow \hat{y}_i + (q_i - p_i) \cdot \left(1 - \frac{f(q_i - p_i | 0, \sigma_{\mathbf{q}}^2)}{f(0 | 0, \sigma_{\mathbf{q}}^2)}\right)$
10   Threshold $\hat{y}_i$ values below $z_{\mathbf{q}} \sigma_{\mathbf{q}}$.

11   $L(\mathbf{s}) \leftarrow \{(a, r(i)) : \hat{y}_i > 0 \cap \hat{y}_{i+1} > 0 \cap \hat{y}_{i-1} = 0\}$
12   If F, remove multipath reflections in $L(\mathbf{s})$ using d_{thresh}.

Continuous peaks in this reshaped signal are treated as objects; per peak, only the range at the center of the peak is added to the landmark set.

Let the vector $\mathbf{s}(t) \in \mathbb{R}^{N \times 1}$ be the power-range spectrum at time t such that the element s_i is the power return at the i-th range bin, and $a(t)$ is the associated azimuth. Let $r(i) = \beta(i - 0.5)$ give the range of bin $i \in \{1, 2, \ldots, N\}$, where β is the range resolution. Suppose that $\mathbf{y}(t) \in \mathbb{R}^{N \times 1}$ is the ideal signal if the environment was recorded perfectly. Then, $\mathbf{s}(t) = \mathbf{y}(t) + \mathbf{v}(\mathbf{y}(t))$, where $\mathbf{v}$ represents unwanted effects, like noise. Therefore, inferring $\mathbf{y}(t)$ from $\mathbf{s}(t)$ in order to accurately isolate the landmarks requires an approximation of $\mathbf{v}(\mathbf{y}(t))$ such that $\hat{\mathbf{y}}(t) = \mathbf{s}(t) - \hat{\mathbf{v}}(\mathbf{s}(t))$. Removing $\hat{\mathbf{v}}$ from $\mathbf{s}$ is the aim of our method. The landmark detections extracted from $\hat{\mathbf{y}}(t)$ are stored in the set $L(\mathbf{s}(t))$.

The landmark extraction method, as described next, references Figure 3 and Algorithm 1. To begin, an unbiased signal $\mathbf{q}$ that preserves high-frequency information (box 2) is acquired by subtracting the noise floor of $\mathbf{v}(\mathbf{s})$ from $\mathbf{s}$ (line 1). The result is then smoothed to obtain the underlying low-frequency signal $\mathbf{p}$ (box 3), which better exposes obvious landmark peaks (line 2). At this point, $\mathbf{q}$ is not discarded for two reasons: radar landmarks often manifest as high frequency peaks, so smoothing would dampen their presence; and smoothing muddles the peaks of landmarks that are in close proximity, making it difficult to distinguish between them. Thus, we integrate the information of both $\mathbf{q}$ and $\mathbf{p}$.

To estimate the noise characteristics, we treat the values of $\mathbf{q}$ that fall below zero as Gaussian noise with mean $\mu_{\mathbf{q}} = 0$ and standard deviation $\sigma_{\mathbf{q}}$ (line 4). Let $f(x | \mu, \sigma^2)$ be the probability density at x for the normal distribution $\mathcal{N}(\mu, \sigma^2)$. Then, for every range bin, the power values are scaled by

the probability that they do not correspond to noise in two steps. First, each value of the smoothed signal p_i is scaled by $f(p_i|0, \sigma_{\mathbf{q}}^2)$ (box 4 and line 8). This process is repeated for the high-frequency signal q_i relative to the smoothed signal p_i such that the scaling factor is $f(q_i|p_i, \sigma_{\mathbf{q}}^2)$ (box 5 and line 9). The sum of both values is stored in $\hat{y}_i$. These steps integrate high- and low-frequency information to preserve range accuracy while suppressing signal corruptions due to noise. Finally, the $\hat{y}_i$ values that are below the upper $z_{\mathbf{q}}$-value confidence bound of $\mathcal{N}(0, \sigma_{\mathbf{q}}^2)$ and therefore less likely to represent real landmarks are set to zero (box 6 and line 10).

The method extracts landmarks from $\hat{y}_i$ (the black signal in box 6) as follows. All values of $\hat{\mathbf{y}}$ are now either zero or belong to a peak. For each peak's center located at range bin i, the tuple $(a, r(i))$ is added to the landmark set $L(\mathbf{s})$ (line 11). These landmarks are then tested, and those identified as multipath reflections (MR) are removed (box 6 and line 12). Since MRs cause peaks with similar wavelet transform (WT) signatures to appear in the power-range spectrum at different ranges with amplitudes that decrease with distance, this step compares the continuous WTs $\mathbf{w}_i, \mathbf{w}_j \in \mathbb{R}^{H \times 1}$ for each set of peaks P_i and P_j where $j > i$. If $d_{ij}/H < d_{\text{thresh}}$ and the maximum power of P_i is greater than that for P_j, where $d_{ij} = |\frac{\mathbf{w}_i}{\max(\mathbf{w}_i)} - \frac{\mathbf{w}_j}{\max(\mathbf{w}_j)}|$ is a measure of dissimilarity, then P_j is considered a MR, and $(a, r(j))$ is removed from $L(\mathbf{s})$. MR removal produces good results but requires significant computation time, making it optional.

Our method requires three free parameters with an optional fourth. In general, w_{median} should represent a distance large enough to span multiple landmarks, and w_{binom} should be around the width of an average peak ($\sim \frac{w_{\text{median}}}{2}$). A greater $z_{\mathbf{q}}$ value raises the standard for peaks to be chosen as landmarks over noise, and d_{thresh} is the minimum difference between WTs for detections to be considered independent. For the following analyses, let $\mathbb{L} = \bigcup_{t \le \tau < t'} L(\mathbf{s}(\tau))$ be the set of all landmarks in one full scan from time t to t'.

C. Data Association

In this section, we present a scan matching algorithm that achieves robust point correspondences using high-level information in the radar scan. Intuitively, it seeks to find the largest subsets of two pointclouds that share a similar shape. Unlike ICP, this method functions without a priori knowledge of the scans' orientations or displacements relative to one another. Thus, our algorithm is not constrained to have a good initial estimate of the relative pose and can compare pointclouds captured at arbitrary times without a map. The only requirements are that the areas observed lie in the same plane and contain sufficient overlap.

One of the key attributes of our approach is to perform data association using not only individual landmark (i.e. unary) descriptors, but also the relationships between landmarks. For instance, imagine three landmarks that form the vertices of a scalene triangle. Then, the set of distances from each point to its neighbors is unique to that point regardless of the overall pointcloud's placement, allowing the landmark to be straightforwardly matched to its counterpart in any other
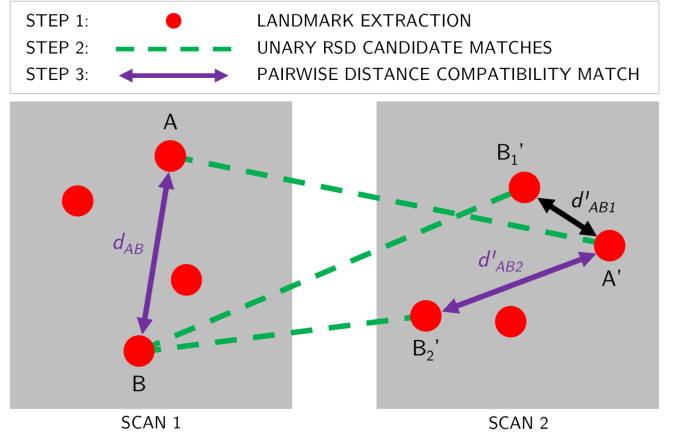

Fig. 4. The core idea behind our data association algorithm that seeks to find similar shapes within the two landmark pointclouds (in red) extracted from radar scans. The unary candidate matches (dotted green lines) are generated by comparing the points' angular characteristics. The selected matches (A, A') and (B, B_2') minimize the difference between pairwise distances $(|d_{AB} - d'_{AB2}| < |d_{AB} - d'_{AB1}|)$. In this way, we approximate a shape matching by sequentially comparing angles and side lengths.

pointcloud acquired by applying a rigid body transformation to the original triangle. The greater the number of points, the less likely it is for an individual point to have the same set of pairwise distances to its neighbors as another. Moreover, the exact position and orientation of the pointcloud does not influence the pairwise relationships within it, so great disparities between the placements and orientations of the pointclouds are inconsequential. We harness these observations to obtain reliable matches for our large landmark sets. With real data, the main challenges are that the landmark locations and detections are noisy, meaning that points do not always survive the rigid body transformation and the locations of those that do are affected by noise.

A simple example illustrating the concept behind our data association algorithm is shown in Figure 4. The method is given in Algorithm 2, which we reference in the following explanation, and its helper functions appear in the Appendix. As inputs, it accepts two pointclouds $\mathbb{L}^O$ and $\mathbb{L}^I$ for each of the two radar scans. The first $\mathbb{L}^O$ is the original set of landmarks in Cartesian coordinates. Because landmarks are detected in polar space, the resulting pointcloud will be dense at low ranges and sparse at high ones. The second $\mathbb{L}^I$ compensates for this by generating a binary Cartesian grid of resolution β that is interpolated from the binary polar grid of landmarks. The latter pointcloud is less exact and only used to sidestep the range-density bias when processing the layout of the environment while data association is performed on the former (i.e. the algorithm returns a set of matches $\mathcal{M}$ that contains tuples (i, j) such that the landmark $\mathbb{L}_1^O\{i\}$ corresponds to $\mathbb{L}_2^O\{j\}$). This distinction is a key insight. It preserves accuracy by operating on the landmarks detected in polar space while correcting for a main difficulty of scanning radars by interpreting the environment in Cartesian space.

The data association is then performed in four steps. First, for every point in $\mathbb{L}_1'$, the unaryMatches function suggests a potential point match in $\mathbb{L}_2'$ based on some unary comparison

Algorithm 2: Data Association Method

Input: Landmark sets for two scans $\mathbb{L}_1^O$, $\mathbb{L}_1^I$, $\mathbb{L}_2^O$, and $\mathbb{L}_2^I$.

Output: Set of landmark matches $\mathcal{M}(\mathbb{L}_1^O, \mathbb{L}_2^O)$.

Parameters: Percentage α of landmarks that, if matched, cause the algorithm to terminate.

1   $B \leftarrow$ unaryMatches($\mathbb{L}_1^O, \mathbb{L}_1^I, \mathbb{L}_2^O, \mathbb{L}_2^I$) and $W \leftarrow |B|$
2   $C_{W \times W} \leftarrow$ pairwiseCompatibilities($B, \mathbb{L}_1^O, \mathbb{L}_2^O$)
3   $\mathbf{u}^* \leftarrow$ normalizedMaxEigenvector(C)
4   Initialize the empty set $\mathcal{M}$.
5   Initialize the $W \times 1$ vector *unsearched* to all *True*.
6   **while** *(any True in unsearched)* $\cap$ *($|\mathcal{M}| < \alpha W$)* **do**
7      $(max\_match, max\_reward) \leftarrow$
       $(i, u_i^{*2}) : u_i^{*2} \geq u_j^{*2} \; \forall \, i, j \in$ *unsearched*)
8   Terminate function if ($max\_reward \cdot W < 1$).
9   Add the match $B\{max\_ind\}\}$ to $\mathcal{M}$.
10     $searched \leftarrow \{i : B\{max\_match, 1\} = B\{i, 1\} \cup$
           $B\{max\_match, 2\} = B\{i, 2\}\}$
11     $unsearched_{i \in searched} \leftarrow$ *False*

method (line 1). We discuss the *unaryMatches* function in Section III-D. Next, the non-negative compatibility score for each pair of proposed matches $g = (i, i')$ and $h = (j, j')$ is computed and assigned to the elements (g, h) and (h, g) of the $W \times W$ matrix C such that it is symmetric and diagonally dominant (line 2). If the landmark matches g and h are correct, then the relationship between i and j in the first radar scan is similar to that between i' and j' in the second; the compatibility score reflects this pairwise similarity. In our method, the value is computed from the distances between corresponding pairs of points in the two scans. It reflects the understanding that real, correctly identified landmarks are the same distance apart in any two radar scans.

The optimal set of matches $\mathcal{M}$ maximizes the overall compatibility, or reward. Suppose that $\mathbf{m} \in \{0, 1\}^W$ such that (1) $m_i = 1$ if the unary match $B\{i\}$ is deemed plausible and $m_i = 0$ otherwise; and (2) the selected matches do not conflict (i.e. a point in one pointcloud cannot correspond to two in the other). Then, the optimal solution $\mathbf{m}^*$ satisfies

$$\mathbf{m}^* = \underset{\mathbf{m} \in \{0,1\}^W}{\arg\max} \; \mathbf{m}^\top C \mathbf{m}.$$

Due to the discretization of $\mathbf{m}$, this maximization is computationally difficult, so we relax the aforementioned constraint to seek the continuously-valued $\mathbf{u}^*$ such that

$$\mathbf{u}^* = \underset{\mathbf{u} \in [0,1]^W}{\arg\max} \; \mathbf{u}^\top C \mathbf{u}.$$

Under these conditions, $\mathbf{u}^*$ is the normalized eigenvector of the maximum eigenvalue of the positive semi-definite matrix C. The optimal solution $\mathbf{m}^*$ is then be approximated from $\mathbf{u}^*$ using the greedy approach shown in lines 3-11.

In short, the greedy method iteratively adds satisfactory matches to the set $\mathcal{M}$. On each iteration, the remaining valid matches are evaluated (line 7), that which returns the
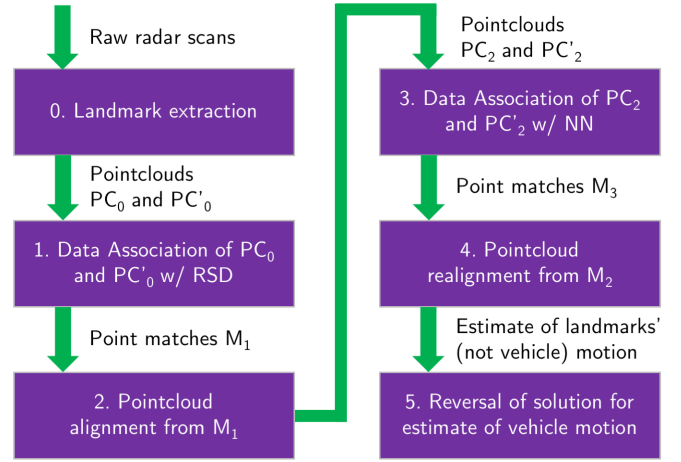

Fig. 5. The relative motion estimation pipeline.

maximum reward is accepted (line 9), and those that conflict with it are removed from further consideration (lines 10 and 11). When the most recently selected match yields a reward less than the that if all matches were valued equally (i.e. is a weak match) or more than α percent of the landmarks in either set are matched, the algorithm terminates (lines 6 and 8). Note that α is the only free parameter in this method, and no outlier removal is required.

D. Relative Motion Estimation

Given two sets of corresponding points, Challis [18] presents a method for finding the rigid body motion that optimally aligns them in the least-squares sense using singular value decomposition (SVD). We apply this technique to $\mathcal{M}$ to estimate the relative motion between two radar scans.

Our motion estimation algorithm, summarized in Figure 5, first performs a data association (step 1) that is capable of aligning pointclouds separated by any distance and rotation. The relative motion is then estimated with Challis' method (step 2). The second round of data association (step 3), which serves only to refine the motion estimate, makes use of the fact that the new pointclouds are now approximately aligned when generating unary matches. The final motion calculation is performed on the new set of dense point matches (step 4). Two examples of this process are shown in Figure 7.

The only difference between steps 1 and 3 is the choice of unary matching function. In step 1, each point is associated with its closest match in the other pointcloud according to what we term the *Radial Statistics Descriptor* (RSD). This descriptor specifies each landmark ℓ by the radial statistics of its neighboring points along every angular slice centered at ℓ (Algorithm 3 in the Appendix). In step 3, the unary match of a point is its nearest neighbor in the other pointcloud.

Recall that the overarching intuition for our scan algorithm is to find and match the largest subsets of points in two pointclouds that share a similar shape. Generally, a shape is uniquely defined by two characteristics: its side lengths and angles. While shape matching is computationally demanding, our algorithm efficiently approximates this operation by
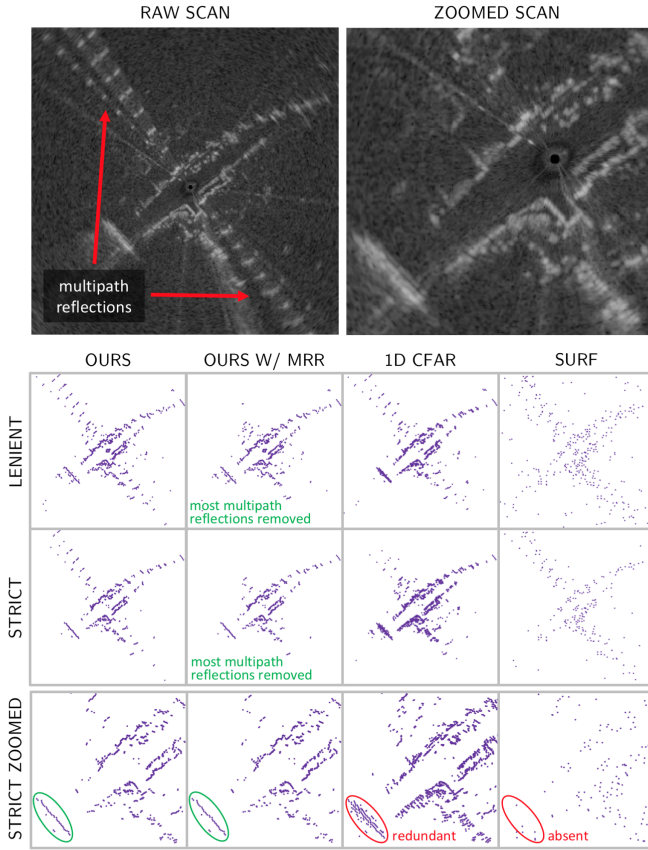
Fig. 6. A comparison of our landmark extraction methods to 1D CFAR [4], [5] and SURF [1], [2]. The raw radar images are at the top, and each table column features a method. The first and second rows give the pointclouds for lenient (i.e. more detections with the potential for more noise) and strict parameters, respectively. The bottom row zooms in on the latter. Only our method with multipath reflection removal (MRR) successfully disregards MRs. 1D CFAR contains redundant returns (e.g. for the wall at the bottom left), and SURF does not consistently find real objects.

capturing angular information in the RSD unary matches and side lengths in the pairwise compatibility calculation, and this combination ensures robust scan matching.

IV. RESULTS

We utilize the Navtech CTS350-X, a FMCW scanning radar without Doppler information. For this radar, $M = 399$, $N = 2000$, and $\beta = 0.25$ m. The beam spread is 2 degrees in azimuth and 25 degrees in elevation. The radar operates at 4 Hz, and our algorithm (not fully optimized) operates at approximately 3 Hz. The radar is placed on the roof of a ground vehicle with an axis of rotation perpendicular to the driving plane. We adopt the usual odometry assumptions that the environment is mostly static and non-deformable. We also assume that the instantaneous motion of the vehicle is planar. We utilize the following parameters, chosen empirically: $w_{median} = 200$, $w_{binom} = 50$, $z_q = 2.5$, $d_{thresh} = 0.1$, and $\alpha = 0.5$ with MR removal. When driving, the vehicle typically travels between 5 and 10 m/s; when turning, up to 0.6 rad/s (see Figures 1 and 8). The vehicle is driven through various parts of downtown Oxford, UK.
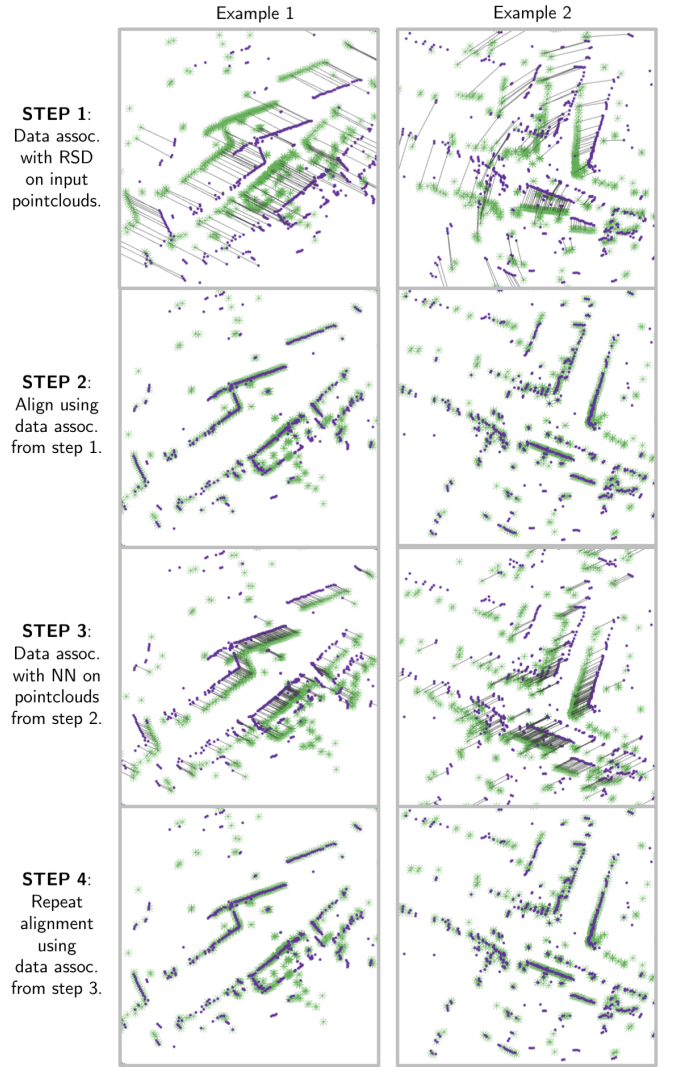

Fig. 7. Illustration of the data associations and pointcloud alignments corresponding to steps 1 through 4 in Figure 5. The displacement between the two pointclouds (purple and green) in step 3 is inserted to better visualize the point matches (black lines). Note that step 1 operates on pointclouds separated by any distance and rotation since it only uses the pointclouds' shapes. Steps 2 and 4 show the alignments produced by the matches in steps 1 and 3, respectively. Step 3 refines the alignment of step 2, thus improving the motion estimate, as evidenced by the denser matches.

Figure 6 compares our landmark extraction algorithm to other common approaches. It examines the use of image processing features for radar localization and mapping [1], [2] via SURF. Designed to detect sharp gradients, SURF is highly susceptible to unwanted radar artifacts. As shown, it returns numerous false positives, and many landmarks do not correspond to meaningful objects in the scene, demonstrating the need for radar-specific landmark detection in order to perform robust motion estimation. In contrast, 1D CFAR yields a better depiction of the surroundings, from which structures (e.g. walls and buildings) are easily discernible. However, we maintain that the landmark set produced by our algorithm is preferable for the following reasons. Our pointclouds are qualitiatively clean and crisp with few false detections, especially with multipath reflection removal (MRR). The
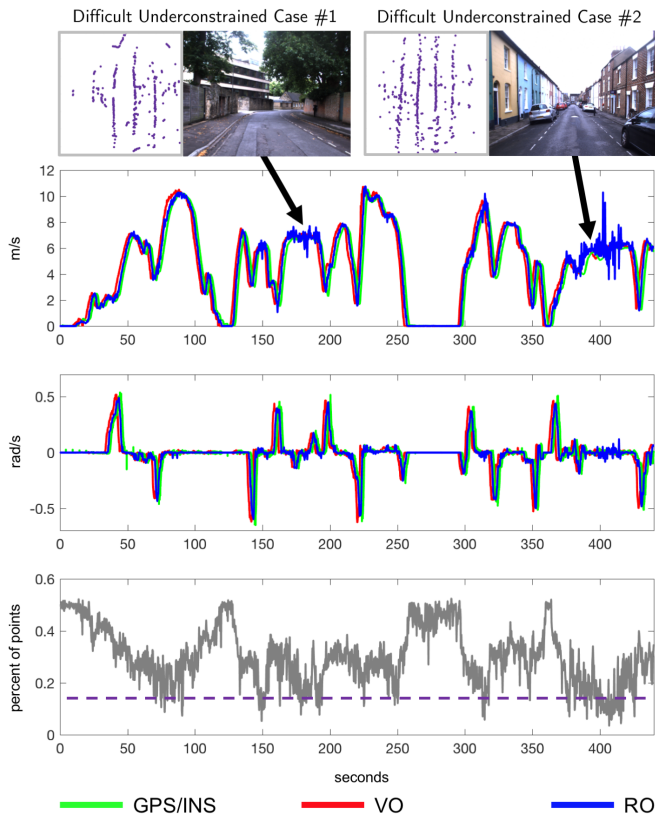
Fig. 8. A comparison of odometry using the radar (RO), vision (VO), and GPS/INS under favorable conditions for vision and GPS/INS (note: the curves are slightly shifted in time for better visual comparison). This route illustrates the success of RO (it closely matches the estimates of VO and GPS/INS) while including portions of its worst performance. The images and landmark sets when RO is noisiest show that this occurs in featureless narrow environments where the pointclouds are underconstrained between scans. The top and middle plots show the estimated translational and angular velocities, respectively. The bottom plot tracks the portion of the landmark pointcloud that is matched during the data association step using RSD. This value appears to provide an uncertainty measure, as it is approximately inversely correlated with RO performance, an added benefit of our approach.

Translational Velocity Error (m/s)			
Quantile	Diverse setting	Busy city center	Narrow backstreets
0.25	0.0293	0.0420	0.0495
0.50	0.1057	0.1134	0.1218
0.75	0.2116	0.2094	0.2321
Rotational Velocity Error (deg/s)			
Quantile	Diverse setting	Busy city center	Narrow backstreets
0.25	0.0824	0.0801	0.2108
0.50	0.3210	0.3188	0.5165
0.75	0.7118	0.7017	0.9589

sensor systems. RO is robust across a variety of settings and conditions. Importantly, even when visual odometry (VO) [19] and GPS/INS are available, they are closely matched by RO. In the dark and rain, VO periodically fails while RO gives a clean and smooth result. The data for the bottom example of Figure 1 was captured indoors in a dark, crowded, and enclosed area. As a result, GPS/INS is unavailable and VO is sporadic. Due to the presence of large nearby metallic objects, which cause the scans to be noisy, the RO curves are less smooth than the outdoor ones, but the radar still produces the most reliable motion estimation. Figure 8 shows a route chosen because it contains areas in which our RO performs suboptimally. Our RO system experiences difficulties when the landmark pointclouds contain insufficient information to conclusively deduce the vehicle's motion (i.e. are underconstrained). Specifically, narrow corridors appear as parallel lines in consecutive radar scans; due to the lack of variability in the scene, our algorithm generates visually satisfactory alignments that do not reflect the actual movement between the scans. We can evaluate the uncertainty of our motion estimate by examining the portion of the landmark pointcloud that is matched using RSD during data association. Since our approach seeks the largest number of matches that are mutually consistent, a low number indicates either that the two pointclouds are dissimilar or multiple solutions exist.

Table I confirms that the error for RO is greater when driving through narrow backstreets than in the busy city center. Over a 10 km route through Oxford, UK (the diverse setting), the median RO error is about 0.106 m/s in translation and 0.321 deg/s in rotation. We compare against VO because it provides accurate and fine-grained odometry while that from GPS/INS odometry is smoothed.

V. CONCLUSION AND FUTURE WORK

In this paper, we introduced a robust radar-only motion estimation system that rivals the performance of VO (even under conditions optimal for vision) and demonstrates the importance of radars for autonomous vehicles. As an on-board sensor that operates under diverse conditions, the successful implementation of RO improves the reliability and versatility of mobile systems. Our method stands out because it is not only dependable and accurate, but also straight-forward and intuitive with few free parameters, no outlier-detection methods, and no model-reliant filters. Although multi-sensor fusion is beneficial, we show that RO can stand

landmarks accurately reflect the coherent structure of the surroundings, and our approach avoids returning redundant points corresponding to the same object. Finally, adjusting z_q in our algorithms tunes the leniency in an intuitive manner (i.e. raising z_q reduces the number of landmarks, beginning with those that appear to be noise). Yet adjusting the parameters for 1D CFAR deteriorates the pointcloud quality in some places and improves it in others. This property makes the CFAR filter parameters difficult to tune.

Figure 7 displays the outcomes of our data association algorithm through two scan matching examples. The steps 1 and 3 show point matches that are dense, accurate, and consistent with one another without any outlier removal. These correspondences provide the highly precise and clean scan matches in the steps 2 and 4. Because step 1 can be applied to pointclouds separated by any displacement as long as they contain sufficient scene overlap, the matches in step 3, which exploit the approximate alignment from step 2, are more dense, thus producing the refined alignment in step 4.

Figures 1 and 8 illustrate the odometry results across three

alone if the other sensors drop out. Our landmark extraction algorithm produces sparse yet meaningful detections with minimal false positives. It feeds into our data association algorithm, which performs scan matching using a greedy approach that, intuitively, seeks to find the largest subsets of the two pointclouds that share similar shapes. The resulting radar-only motion estimation is accurate and robust under conditions for which other common sensor systems fail.

In the future, we intend to address the scenarios in which RO encounters difficulties and to develop a technique that accurately quantifies the uncertainty of our component estimates. We also aim to include comprehensive comparative studies against other RO methods.

REFERENCES

[1] J. Callmer, D. Törnqvist, F. Gustafsson, H. Svensson, and P. Carlbom, "Radar slam using visual features," *EURASIP Journal on Advances in Signal Processing*, vol. 2011, no. 1, p. 71, 2011.

[2] F. Schuster, C. G. Keller, M. Rapp, M. Haueis, and C. Curio, "Landmark based radar slam using graph optimization," in *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on*. IEEE, 2016, pp. 2559–2564.

[3] K. Werber, J. Klappstein, J. Dickmann, and C. Waldschmidt, "Interesting areas in radar gridmaps for vehicle self-localization," in *Microwaves for Intelligent Mobility (ICMIM), 2016 IEEE MTT-S International Conference on*. IEEE, 2016, pp. 1–4.

[4] H. Rohling, "Ordered statistic cfar technique-an overview," in *Radar Symposium (IRS), 2011 Proceedings International*. IEEE, 2011, pp. 631–638.

[5] D. Vivet, P. Checchin, and R. Chapuis, "Localization and mapping using only a rotating fmcw radar sensor," *Sensors*, vol. 13, no. 4, pp. 4527–4552, 2013.

[6] E. Jose and M. D. Adams, "An augmented state slam formulation for multiple line-of-sight features with millimetre wave radar," in *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on*. IEEE, 2005, pp. 3087–3092.

[7] ——, "Relative radar cross section based feature identification with millimeter wave radar for outdoor slam," in *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, vol. 1. IEEE, 2004, pp. 425–430.

[8] J. W. Marck, A. Mohamoud, E. vd Houwen, and R. van Heijster, "Indoor radar slam a radar application for vision and gps denied environments," in *Radar Conference (EuRAD), 2013 European*. IEEE, 2013, pp. 471–474.

[9] F. Schuster, M. Wörner, C. G. Keller, M. Haueis, and C. Curio, "Robust localization based on radar signal clustering," in *Intelligent Vehicles (IV) Symposium, 2016 IEEE*. IEEE, 2016, pp. 839–844.

[10] M. Rapp, K. Dietmayer, M. Hahn, F. Schuster, J. Lombacher, and J. Dickmann, "Fscd and basd: Robust landmark detection and description on radar-based grids," in *Microwaves for Intelligent Mobility (ICMIM), 2016 IEEE MTT-S International Conference on*. IEEE, 2016, pp. 1–4.

[11] T. Deissler and J. Thielecke, "Uwb slam with rao-blackwellized monte carlo data association," in *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*. IEEE, 2010, pp. 1–5.

[12] P. J. Besl, N. D. McKay *et al.*, "A method for registration of 3-d shapes," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

[13] E. Ward and J. Folkesson, "Vehicle localization with low cost radar sensors," in *Intelligent Vehicles (IV) Symposium, 2016 IEEE*. IEEE, 2016, pp. 864–870.

[14] M. Chandran and P. Newman, "Motion estimation from map quality with millimeter wave radar," in *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*. IEEE, 2006, pp. 808–813.

[15] 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.

[16] P. Checchin, F. Gérossier, C. Blanc, R. Chapuis, and L. Trassoudaine, "Radar scan matching slam using the fourier-mellin transform," in *Field and Service Robotics*. Springer, 2010, pp. 151–161.

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

[18] J. H. Challis, "A procedure for determining rigid body transformation parameters," *Journal of biomechanics*, vol. 28, no. 6, pp. 733–737, 1995.

[19] W. Churchill, "Experience based navigation: Theory, practice and implementation," Ph.D. dissertation, University of Oxford, Oxford, United Kingdom, 2012.

APPENDIX

Algorithm 3: Compute RSD for Landmark

Input: Landmark sets $\mathbb{L}^O$ and $\mathbb{L}^I$; index i of landmark.
Output: The $3 \times M$ matrix RSD. Element (i, j) is the i-th statistic of the points in j-th azimuth slice, which are ordered CCW, starting with the slice that contains the most points.

---

1 Initialize $3 \times M$ matrix RSD to zeros.
2 Separate space around $\mathbb{L}_i^O$ into M equal azimuth slices.
3 **for** azimuth slice $j \leftarrow 1$ **to** M **do**
4 $\quad$ Find set S of all points in $\mathbb{L}^I$ that lie in slice j.
5 $\quad$ $RSD[:, j] \leftarrow [|S|, \text{arithmMean}(S), \text{harmonMean}(S)]$
6 Normalize all rows in RSD.
7 Shift RSD columns s.t. highest-density slice is the first.

Algorithm 4: Generate Unary Matches based on RSD

Input: Landmark sets $\mathbb{L}_1^O$, $\mathbb{L}_1^I$, $\mathbb{L}_2^O$, and $\mathbb{L}_2^I$.
Output: Set of unary match indices B.

---

1 Initialize the empty set B.
2 $RSD1 \leftarrow \text{getAllPointsRSD}(\mathbb{L}_1^O, \mathbb{L}_1^I)$
3 $RSD2 \leftarrow \text{getAllPointsRSD}(\mathbb{L}_2^O, \mathbb{L}_2^I)$
4 **for** point $i \leftarrow 1$ **to** $|\mathbb{L}_1^O|$ **do**
5 $\quad$ Find point j in $\mathbb{L}_2^O$ to minimize $||RSD2_j - RSD1_i||$.
6 $\quad$ Add unary match (i, j) to set B.

Algorithm 5: Generate Unary Matches based on NN

Input: Landmark sets $\mathbb{L}_1^O$ and $\mathbb{L}_2^O$.
Output: Set of unary match indices B.

---

1 Initialize the empty set B.
2 **for** point $i \leftarrow 1$ **to** $|\mathbb{L}_1^O|$ **do**
3 $\quad$ Find point j in $\mathbb{L}_2^O$ to minimize $||\mathbb{L}_2^O\{j\} - \mathbb{L}_1^O\{i\}||$.
4 $\quad$ Add unary match (i, j) to set B.

Algorithm 6: Compute Pairwise Compatibility

Input: Set of unary matches B; landmark sets $\mathbb{L}_1^O$ and $\mathbb{L}_2^O$; indices of unary matches i and j.
Output: Pairwise compatibility score C_{ij} for (i, j) pair.

---

1 **if** i equals j **then**
2 $\quad$ $C_{ij} \leftarrow |B|$ and return.
3 $\boldsymbol{\ell}_{1,i}, \boldsymbol{\ell}_{1,j} \leftarrow \mathbb{L}_1^O\{B\{i, 1\}\}, \mathbb{L}_1^O\{B\{j, 1\}\}$
4 $\boldsymbol{\ell}_{2,i}, \boldsymbol{\ell}_{2,j} \leftarrow \mathbb{L}_2^O\{B\{i, 2\}\}, \mathbb{L}_2^O\{B\{j, 2\}\}$
5 $d_1 \leftarrow ||\boldsymbol{\ell}_{1,i} - \boldsymbol{\ell}_{1,j}||^2$ and $d_2 \leftarrow ||\boldsymbol{\ell}_{2,i} - \boldsymbol{\ell}_{2,j}||^2$
6 $C_{ij} \leftarrow (1 + |d_1 - d_2|)^{-1}$