

What Could Go Wrong? Introspective Radar Odometry in Challenging Environments

Roberto Aldera, Daniele De Martini, Matthew Gadd, Paul Newman
Oxford Robotics Institute, Dept. Engineering Science, University of Oxford, UK.
{roberto, danielle, mattgadd, pneman}@robots.ox.ac.uk

Abstract—This paper is about detecting failures under uncertainty and improving the reliability of radar-only motion estimation. We use weak supervision together with inertial measurement fusion to train a classifier that exploits the principal eigenvector associated with our radar scan matching algorithm at run-time and produces a prior belief in the robot’s motion estimate. This prior is used in a filtering framework to correct for vehicle motion estimates. We demonstrate the system on a challenging outdoor dataset, for which current radar motion estimation algorithms fail frequently. By knowing when failure is likely, we achieve qualitatively superior motion estimates and quantitatively fewer odometry failures. Specifically, we see 24.7% fewer failures in motion estimation over the course of a 15.81 km drive through a difficult, mixed rural-and-urban scene, with lower RMSE in translational and rotational estimates during particularly challenging conditions.

Index Terms—radar, sensing, ego-motion estimation, introspection, field robotics

I. INTRODUCTION

Radar offers the field of robotics a new level of robust sensing. While many of the hurdles in navigation have been cleared using lasers or vision to great success, these tools can only take us so far. Autonomous machines that can function reliably in challenging unstructured environments need a sensor that thrives in all conditions – rain, snow, dust, fog, or direct sunlight. We also need our robots to see further if they are to travel safely at higher speeds or operate in wide open spaces where distinct features are few and far between. These requirements are met in radar, which provides a 360°-view of the scene and detects targets up to 500 m away. However, radar data is highly complex, making any sort of inference challenging. Measurements contain significant components made up of multipath reflections, speckle noise, and other artefacts in addition to the radar’s internal noise characteristics [1].

Despite this, recent work has presented precise radar-only motion estimation algorithms that consistently outperform visual odometry systems, even in conditions which are favourable to vision [2, 3]. Yet, as we present in this paper, the state-of-the-art Radar Odometry (RO) system with no means of introspection is still prone to a high failure rate. We can distinguish between a few common scenarios for failures:

- Operation in off-road environments where the robot’s motion is not truly planar causes large swathes of the scene to disappear and reappear intermittently while the robot is moving [4].

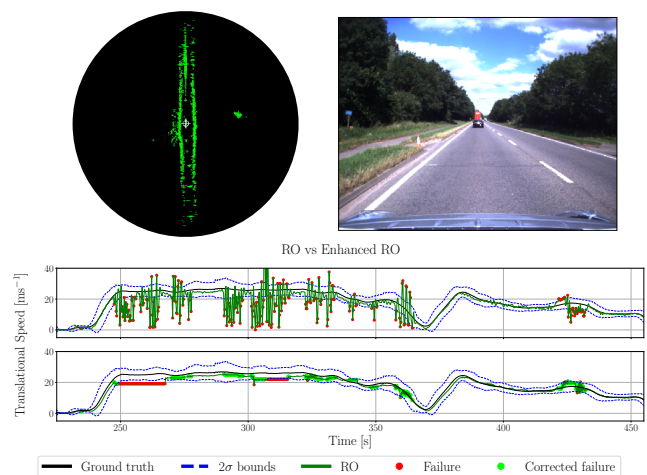


Fig. 1. RO has no means of detecting odometry failures. Here between the hedgerows the radar-only algorithm is starved of distinct features where it is difficult to estimate forward ego-motion between measurements. Our enhanced system addresses this and other common failure cases through introspection, allowing RO to function robustly under the uncertainty presented by challenging environments. Pose estimation is significantly improved, and here translational speeds are shown to closely track the baseline GPS measurements where previously RO had failed.

- Stationary or dynamic occluding obstacles (walls, buses, dense vegetation) encountered in the scene lead to sudden feature absence.
- A shortage of distinct features observable across wide baselines creates ambiguity, as shown between hedgerows in Figure 1.

In this paper, we present a system for both estimating confidence in and correcting the solution of the scan matching RO algorithm. At run-time, we use only the radar data itself to detect failures based on a weakly supervised classification procedure which is trained offline by data fusion. The classification output is passed to a motion correction system that performs an update based on the reliability of the odometry estimate.

We proceed by reviewing existing literature in Section II and discuss some preliminary concepts in Section III which will introduce notation and provide background on the relevant operational aspects of the RO system. Sections IV and V present necessary details for implementation before Section VI, where we report our results showing superior motion estimation performance under extremely challenging conditions.

II. RELATED WORKS

Approaches to the correspondence problem using eigenvectors have been fundamental in the development of our RO system, first described in [2]. Central to this is the data association method that addresses the challenge of determining matches between sets of landmarks. Scott and Longuet-Higgins [5] present a general approach to cluster features based on the distances between them. Using an initial proximity matrix as input, the algorithm finds the clusters of features based on the eigenvectors of this matrix, ranked using the magnitude of their eigenvalues. An association matrix is computed to yield the strengths of the “bonds” between features, based on methods found in molecular physics. Building on this work, Shapiro and Brady [6] perform feature-based correspondence by incorporating shape information together with these techniques. Their algorithm is robust to translations, rotations, and scaling as it considers the structure within the given image, unlike [5] which fails when features in the first image are no longer neighbours in the next. Closest to RO’s data association implementation is the spectral technique of Leordeanu and Hebert [7], which efficiently finds the consistent correspondences between two sets of features using the principal eigenvector of a constructed adjacency matrix. This matrix is built from a graph consisting of nodes representing each potential match, and weights on the respective links that represent the pairwise agreements between these correspondences. This elegant approach sees correct assignments establish strong links between each other and is inherently robust to outliers, while being considerably faster than existing methods.

The data association algorithm implemented in RO that establishes links between landmark sets has no means of introspection. That is to say, an odometry estimate based on the proposed correspondences is made regardless of how well the landmarks have actually been matched. It is suggested in [2] that the portion of landmarks that have been matched provides a measure of uncertainty, which begins to provide indication that a given solution might be unsuitable to use in the next odometry estimate. A more rigorous approach native to the correspondence algorithm itself is discussed in [7], which looks at the difference in magnitude of the first two largest eigenvalues, known as the *eigengap*. If a proposed set of assignments between the points is not particularly distinct, the eigenvalues of the two largest eigenvectors will be similar in magnitude, as they both correspond to similarly plausible solutions. In other words, no single solution emerges above the other possible assignments, as would be the case if a distinctive set of matches could not be found. The authors in [7] note that this occurs when there is excessive symmetry in the data, too many outliers, or high deformation noise. Further to this insight, Sakar and Boyer [8] provide an extensive study on the stability of the eigenvalues and the corresponding robustness of the eigenvectors, and apply this to monitor change in how features are organised in images. In the context of ranking citation information (or web pages), Ng et al. [9] analyse the stability of the eigenvector and

provide a simple example that illustrates why a small eigen-gap corresponds to a less distinctive solution, i.e one that is sensitive to small perturbations in the graph.

To improve RO under failure, we turn to the Kalman Filter (KF), a widely-adopted algorithm that provides a means to estimate under uncertainty [10]. In [11], the authors describe a probabilistic test to detect failures based on the KF estimates, where the system model and the corresponding sensory data should always be statistically coherent. This coherence is modelled as a χ^2 -distribution where a simple threshold can be used to detect failures in the sensor measurement stream. A positive failure detection causes the KF algorithm to reject the corresponding sensor measurement.

Our proposed method makes use of an external classifier that flags motion estimation failures prior to the motion correction stage. This addresses the limitations of the approach in [11], where consecutive failures (that lead to an accumulation of uncertainties) result in a system that will be more tolerant to taking on poor sensor measurements.

III. PRELIMINARIES

We use a Frequency-Modulated Continuous-Wave (FMCW) scanning radar which rotates about its vertical axis while sensing the environment continuously through the transmission and reception of frequency-modulated radio waves. While rotating, the sensor inspects one angular portion (*azimuth*, α) of space at a time and receives a power signal that is a function of reflectivity, size, and orientation of objects at that specific azimuth and at a particular distance, ρ . The radar takes measurements along an azimuth at one of M discrete intervals and returns N power readings, or *bins*. We call one full rotation across all M azimuths a *scan* \mathcal{S} . Furthermore, let $\mathbf{s}(k) \in \mathbb{R}^{N \times 1}$ be the power-range readings at time step k , where $t(k) = t_k$ is the time value at k and $\alpha(k) \in A$ is the azimuth associated with the measurement. The element $s_i(k) \in \mathbf{s}(k)$ is the power return at the i -th range bin, with $i \in \{1, \dots, N\}$; its measurement range is given by $\rho_i(k) = \beta(i - 0.5)$, where β is the range resolution of the radar.

Our work builds upon our odometry estimation pipeline first described in [2], which we summarise here. The first aim is to extract features from the sets of power-range spectra measured by the radar as it scans the environment, and describe them as a set of landmarks characterised by a position in space relative to the radar and a power value. This is nontrivial due to the complexity of the radar data itself, which is susceptible to multipath reflections, harmonics, speckle noise, and other hindering effects [1, 3].

Given two so-obtained landmark sets, \mathbb{L}_1 and \mathbb{L}_2 , the next task is to associate each landmark from \mathbb{L}_1 with a single corresponding landmark in \mathbb{L}_2 . As in [7], we pose this data association as a pairwise problem. This leverages the assumption that the set of distances from each point to its neighbours is unique to that point in all observations, regardless of the relative translation and rotation between scans. This allows the landmark to be matched to its counterpart in any other landmark set that contains sufficient overlap.

After extracting a set of possible unary candidates B of size W we build a pairwise-compatibility matrix C , which is a non-negative symmetric matrix of size $W \times W$ containing the score for each pair of proposed matches.

The vector $\mathbf{m} \in \{0,1\}^W$ is a solution of the data association problem, where $m_i = 1$ for a unary match $B\{i\}$ that is considered or $m_i = 0$ otherwise. The optimal solution, i.e. the solution that maximises the overall compatibility, can be computed as:

$$\mathbf{m}^* = \arg \max_{\mathbf{m} \in \{0,1\}^W} \mathbf{m}^T C \mathbf{m} \quad (1)$$

As \mathbf{m} is discrete, the above maximisation is computationally difficult. So inline with [7], we relax this constraint to find the continuously-valued \mathbf{u}^* :

$$\mathbf{u}^* = \arg \max_{\mathbf{u} \in [0,1]^W} \mathbf{u}^T C \mathbf{u} \quad (2)$$

The elements in \mathbf{u}^* are only relevant in their respective values to each other, so we can normalise \mathbf{u}^* and from Rayleigh’s quotient theorem we iterate to obtain the \mathbf{u}^* that is the principal eigenvector of C [12]. The optimal solution \mathbf{m}^* is approximated from \mathbf{u}^* using a greedy algorithm, as described in more detail in [2]. Finally, the relative motion between the observations that yielded \mathbb{L}_1 and \mathbb{L}_2 is estimated using a singular-value decomposition algorithm [13]. In addition to the above description, we add a new component to the RO system to account for the motion during a scan that would otherwise cause significant distortions in landmark positions at higher speeds. By assuming constant velocity between motion estimates, landmarks are repositioned relative to the interpolated position of the robot at the time at which they were observed.

IV. METHODOLOGY

We begin in Sections IV-A and IV-B by describing how we use inertial measurements recorded alongside radar scans to automatically label a set of eigenvectors from RO as **good** or **bad**. Section IV-D discusses a classification procedure which makes use of the principal eigenvector from the motion estimation algorithm discussed in Section III, allowing us to predict the reliability of the odometry estimate at runtime introspectively without access to inertial measurements. Section IV-E concludes this section by discussing a motion-correcting KF implementation which uses these scores as priors for the motion of the vehicle and propagates the belief through the motion estimation pipeline.

A. Alignment of GPS and radar data

The GPS acceleration readings augmented with INS measurements are integrated twice to obtain a list of relative 6DoF poses. The 6DoF extrinsic calibration between the radar and GPS is used to express these measurements in the radar’s frame of reference. As the radar data is interpreted in two dimensions on the xy -plane, the GPS poses need to be projected onto the ground plane [14]. Finally, both radar and GPS pose sources are differentiated to obtain frame-to-frame translational and rotational velocity measurements.

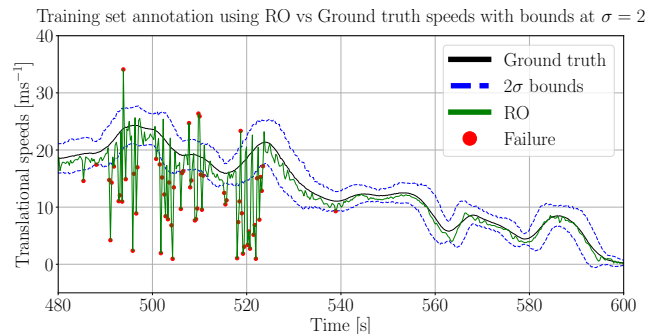


Fig. 2. Aligned instrumentation for GPS and radar sensors used in the training phase of this system. Here we show translational rates. The GPS rates are shown as a moving average, with bounds that lie two standard deviations from the mean. These thresholds are used to trigger labelling events – when the radar rate exceeds them, we produce a **bad** label.

B. Labelling odometry estimates

As a trade-off between quantity and quality of labelled data, we use a form of weak supervision known as *inaccurate supervision* [15], where information in the ground truth data presented to the model may suffer from errors. The GPS readings are preprocessed to allow direct comparison with our RO outputs. We take a moving average over the translational velocities and define an upper and lower bound two standard deviations either side of this. Any RO estimates that fall outside these bounds are labelled as **bad**, or **good** otherwise. Figure 2 shows the signals used in this comparison. After this automatic label generation, we use a median filter to improve training label consistency – during periods of failure these initial labels switch rapidly in and out of bounds, even though the associated principal eigenvectors all correspond to poor matches (evident around 520s in Figure 2). We refer to these instances in Section IV-C as **brittle** estimates, and discuss further implications in Section IV-D. Note that this framework is flexible to using another baseline signal to produce labels – RTK-corrected GPS traces [16] could have been used as a substitute for our inertial measurements.

C. Eigenvector insights

Each element in the eigenvector can be interpreted as a contribution of a particular pairwise match between two landmarks to the overall proposed alignment. As RO’s greedy algorithm from [2] selects matches based on the ordering of these elements, it is natural to want sort these in descending order of magnitude. In Figure 3, we observe that for well-posed data association problems, RO produces eigenvectors that have many similar element magnitudes. Conversely, when there is greater variance among the eigenvector elements, we generally observe a weak match that coincides with a **bad** odometry label. This is intuitive: all things being equal, the algorithm should depend relatively evenly on all available points. But if reliance on these landmarks is not particularly uniform, it indicates abnormally high uncertainty in the compatibility of two landmark sets. This could be a result of a comparison between landmark sets that simply

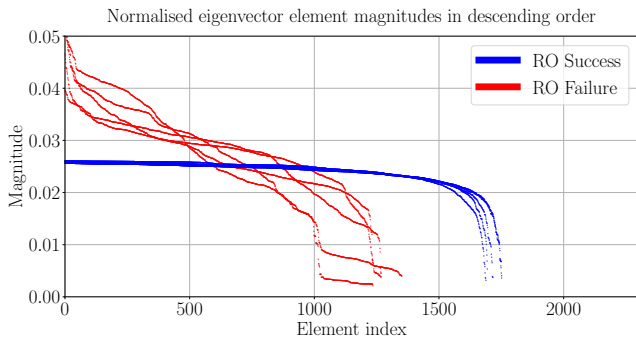


Fig. 3. A sample of eigenvectors produced during the data association step of the RO system. We sort each eigenvector’s elements in descending order of magnitude. Odometry estimates labelled **good** and **bad** correspond to blue and red curves respectively. Here, we observe that **bad** eigenvectors typically start off steeper and vanish to zero quicker than their **good** counterparts – this is the decision boundary that the SVM classifier must learn.

do not match well, or more interestingly, it could indicate a **brittle** solution that is well matched in that it produces plausible odometry, but is likely to fail imminently. In our robotics application, these cases are highly susceptible to failure in the presence of any rolling, pitching, occlusion, or other environmental factors that could cause those crucial landmarks to be absent from the next measurement.

D. Training the classifier

Having discussed the origin of the principal eigenvector in Section III and provided insights into what it encodes in Section IV-D, we now turn to training of a classifier. As described in Section IV-B, we have annotated these eigenvectors as belonging to the **good** or **bad** class. To predict the class of a given eigenvector, we train a SVM classifier [17] with an RBF kernel which we found in practice to outperform other commonly used classifiers like random forests and Convolutional Neural Networks in this instance. However, the eigenvectors are not of consistent length as the compatibility matrix is populated by pairwise unary matches for candidate landmarks which vary in number from scan to scan. Along with sorting the elements in descending order, we found that performing a simple feature extraction step on the eigenvector before training improved performance. Other approaches included padding eigenvectors with zeros or naïvely generating a linearly spaced array of the elements to force them to a standard equal length. The selected features are now described and substantiated in detail.

- 1) **Magnitude of maximum eigenvector element:** the greedy algorithm responsible for approximating \mathbf{m}^* from \mathbf{u}^* uses the pairing related to the highest element magnitude of the eigenvector first:

$$F_1 = \max_i \mathbf{u}_i^*$$

This feature gives an idea as to how strongly the pairwise match of highest confidence contributes to a particular solution. Higher values tend to suppress other pairwise matches and result in fewer pairs being used in the odometry estimate.

- 2) **Number of candidates:** the number of possible matches, i.e. the cardinality of the unary matches set, also encodes information on the robustness of a solution based on how many pairs of landmarks are available to be associated.

$$F_2 = \|\mathbf{u}^*\| = \min(\|\mathbb{L}_1\|, \|\mathbb{L}_2\|)$$

- 3) **Area below the curve:** the area bounded by the eigenvector, defined as $\sum_i \mathbf{u}_i^*$, encodes the shape of the curve itself in a compact way. Note that the \mathbf{u}^* vector is defined as a unit vector, such that:

$$F_3 = \sum_i \mathbf{u}_i^* \neq \sqrt{\sum_i \mathbf{u}_i^{*2}} \equiv 1$$

Since the area of this constrained shape is maximised for a broad rectangle created by relatively similar element magnitudes, overdependence on a smaller subset of matches with higher contributions will lead to a smaller area value.

In addition to these features from the present eigenvector, it is possible to add to the feature space the same three features for N previous odometry estimates, which allows the current eigenvector to be placed in context with what has happened in the recent past. This should aid inference – sudden fluctuations in feature values tend to correspond to **bad** RO estimates. Section V discusses the implications of these additional features and details performance in Figure 4 with different values of N .

Prior to training, features are scaled in a standardisation step to have zero mean and unit variance to handle the range of magnitudes and minimise the effects of outliers [18]. We elect to use the tried-and-tested SVM for classification as it is easy to train on our number of features and performs satisfactorily for our application. As with the motion correction module of the pipeline, the choice of classifier for this framework is also flexible. In practice, the SVM learns a class boundary that classifies the majority of curves that look like they lead to a poor solution as **bad**. This is not always the case: sometimes a match that is excessively dependent on a subset of landmarks, or that has other concerning signs already discussed, will in fact lead to a **good** RO estimate (and sometimes be labelled as such, being acceptable to the GPS baseline and slipping past the median filter). This is a significant insight – the classifier will detect these **brittle** solutions, which do not always coincide with **bad** ones. We desire this in the motion correction phase outlined in Section IV-E as it is preferable to defer to a model when high uncertainty in a solution is present, regardless of whether the proposed RO estimate would classify as **good** or **bad**.

Due to the nature of this failure detection problem, many more examples of **good** RO are available in the training data. We select *average precision* as our evaluation metric from the Precision-Recall curve as it gives the clearest insight into binary classifier performance on imbalanced datasets [19]. At run-time, we use the SVM to classify the latest eigenvector from the data association step and treat the

binary classification output as a prior belief in the reliability of the odometry estimate. This failure detector forms the core of our contribution – providing RO with a means of introspection.

E. Motion correction

A KF technique [10] is used to propagate the belief through the motion estimation pipeline. The system model is given by:

$$\begin{aligned} x_k &= A_k x_{k-1} + \varepsilon_k \\ z_k &= C_k x_k + \delta_k \end{aligned} \quad (3)$$

where x_k and z_k are the state vector and the sensor readings at the k -th instant respectively. We denote the motion and sensor models as A_k and C_k , with corresponding error terms ε_k and δ_k . We exclude the control input matrix, and we consider it as included in the ε_k error term. In this work, we select a constant velocity model for A_k (which could easily be substituted for a more complex model) and use C_k to sample the relative poses from RO directly.

While the prediction process is carried out at each step – yielding the motion estimate μ_m and motion covariance Σ_m – the sensor update is computed only if the eigenvector that produced the proposed match has not been flagged by the failure detector. In that case, the intermediate results μ_s and Σ_s are calculated.

A second stage determines if the proposed sensor update is consistent with the expected motion. Note that although the SVM classifier output is not dependent on the motion model, this additional test relies on it largely. We use the technique described in [11] to calculate a probabilistic index and define the scalar quantity λ as:

$$\lambda = (\mu_m - \mu_s)^T (\Sigma_s - \Sigma_m)^{-1} (\mu_m - \mu_s) \quad (4)$$

It has been demonstrated that λ follows a χ^2 probability distribution. We set a threshold k on the value such that if $\lambda \geq k$ we reject the sensor reading and update the model with μ_m and Σ_m instead of μ_s and Σ_s .

V. EXPERIMENTAL SETUP

Test data is collected on the *Oxford RobotCar* platform [20], with the addition of our radar sensor.

A. Hardware

We employ a CTS350-X Navtech FMCW scanning radar without Doppler information, mounted on top of the platform with an axis of rotation perpendicular to the driving surface. It is characterised by an operating frequency of 76 GHz to 77 GHz, capturing 400 azimuth readings per rotation at 4 Hz, with a beam spread of 2° . Each azimuthal slice contains 3768 range measurements at a bin resolution of 0.1752 m. For inertial and GPS measurements, the vehicle is equipped with a NovAtel SPAN-CPT Single Enclosure GNSS/INS Receiver¹ that provides fused GPS and INS measurements.

¹<https://www.novatel.com/products/span-gnss-inertial-systems/span-combined-systems/span-cpt/>

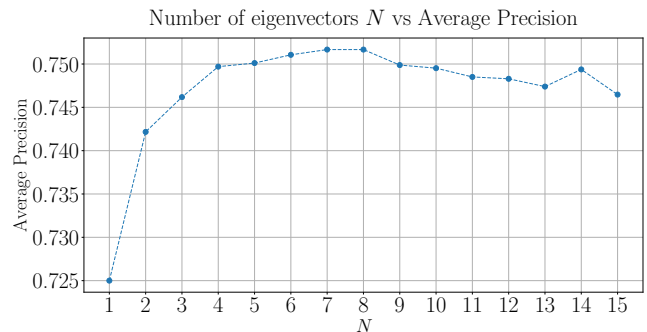


Fig. 4. Plot motivating the choice of hyperparameter N showing the effects on average precision for varying N values. As the number of preceding eigenvectors increases, the classifier performance also rises, until a maximum value where $N = 7$. These average precision values are taken from a mean of the scores obtained on the training set during cross validation using 5 folds.

B. Dataset curation

RO performs excellently in central Oxford where it estimates precise odometry with ease. However, as mentioned in Section I, failures are common when traversing uneven ground or driving between hedgerows at higher speeds. For this reason, we seek to train and evaluate our enhanced RO system on this challenging rural Oxfordshire data. For training, we use radar and GPS data from a single run of a 15.99 km route. At test time, we produce odometry estimates using only radar data as input on a different 15.81 km route in similar surroundings. Training and evaluation were performed in scikit-learn [21] prior to C++ implementation in the enhanced RO system.

C. Training details

We train on a rural Oxfordshire dataset consisting of 5956 radar scans captured during 15.99 km of driving. This produces 5955 labels, of which 21.24% are labelled **bad** RO. After feature extraction, the dataset is shuffled and split by a 80% to 20% training-testing ratio. Using k -fold cross-validation, a grid search that maximised the classifier’s average precision yielded a linear kernel with a C value of 1. A further search over values of hyperparameter N determined the optimal number of preceding eigenvectors to use including the latest candidate was 7, as shown in Figure 4. Over the training set, average precision was 75.3%. The precision was 63.3% with a recall of 95.5%.

The testing set sees an average precision of 75.9%, for a corresponding precision and recall of 62.3% and 95.7% respectively. Due to the nature in which RO fails, labels generated automatically using inaccurate supervision as mentioned in Section II tend to fluctuate rapidly between **good** and **bad** during a period of consecutive failures. As already discussed in Section IV-D, the SVM learns a decision boundary that will classify these **brittle** matches as **bad**. Figure 5 shows the effect of passing the labels through a median filter of width 5 where we enjoy a 8.93% performance boost in our average precision metric.

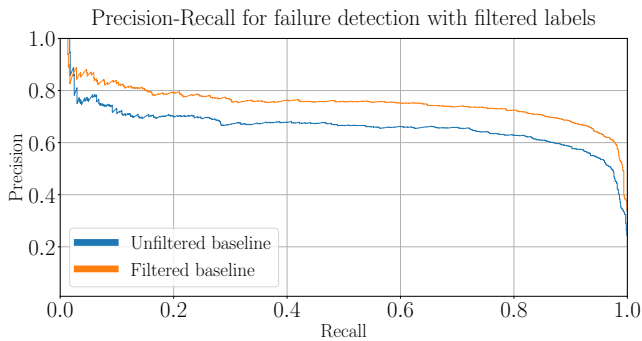


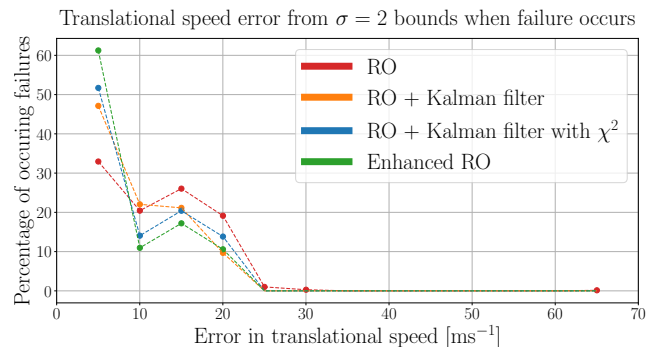
Fig. 5. Two Precision Recall curves are shown. Classifier performance using the labels generated automatically by comparing RO and GPS speeds is improved by passing these labels through a median filter. The orange curve shows this improvement, in line with what has been mentioned regarding the classifier’s tendency to flag **brittle** RO as failure in Section IV-D.

VI. RESULTS

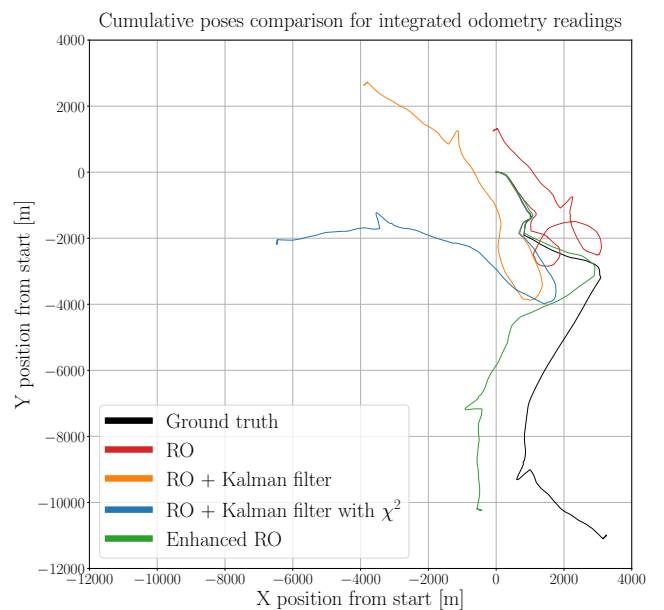
Figure 6(a) motivates our addition of each new section of the system and shows the key result of our overall introspective addition to RO. The combination of the failure detection and motion correction enhancements sees a reduction in the total number of failures over 6635 matches from 695 to 531 (24.7%), where failures that are still present have reduced from a translational RMSE of 11.112 m s^{-1} to 8.256 m s^{-1} and rotational RMSE of $7.740^\circ \text{ s}^{-1}$ to $2.418^\circ \text{ s}^{-1}$. Table I gives a detailed overview of estimation performance improvements. As illustrated in Figure 6(b), the standard RO system is insufficient in this challenging setting, as the recovered global pose is subject to serious drift and some outright failures. The global pose available from our system more closely matches the shape traced by the GPS readings, which we show here for a qualitative comparison. Figure 7 gives insight into the system’s improved performance in recovering translational speeds for each additional component, plotted alongside the GPS baseline and the corresponding bounds.

Despite significant improvement over standard RO performance, the enhanced system’s ability to recover a motion estimate under large uncertainties is limited. This is especially prevalent during periods of prolonged failure, where the system is forced to depend excessively on the most recent reliable motion estimate. By deferring to a constant velocity model during periods of likely failure, more complex trajectories can stretch the motion correction process beyond its limits. This is evident in Figures 7(b) to 7(d) where a handful of failures are introduced.

By making use of the principal eigenvectors to detect failures at run-time, our complete system enhances the KF with χ^2 method by rejecting **bad** matches before they can be considered by the χ^2 -test. Without this introspection, the filter’s tolerance on uncertainty expands to the point of eventually accepting a **bad** estimate, evident in Figure 7(c) between 250 s to 300 s. Note that the addition of the failure detector does introduce a few minor additional failures, as it can be overly sensitive to **brittle** solutions. However, we assert that the system’s ability to know when these failures are occurring could reduce future failures if we were able to



(a) Improvements to RO show that the large majority of remaining failures that do occur will be at lower magnitudes, with our complete system outperforming previous iterations.



(b) Global pose comparison gives a qualitative result in cumulative odometry improvement in the proposed introspective system from standard RO. By knowing when to reject poor solutions, the enhanced RO system recovers more of the ground trace shape than the naive KF with χ^2 , despite the total number of failures being slightly higher.

Fig. 6. Comparison between experiments showing quantitative reductions in the distributions of failure case magnitudes and qualitative improvements in the recovered global pose.

take action more appropriate to the current state, perhaps in some form of feedback loop. We leave this for future work.

VII. CONCLUSIONS

We have presented in this paper a system for online introspective analysis and correction of a radar-only motion estimation scan matching algorithm. The analysis is bootstrapped in a weakly supervised fashion using data fusion, but relies solely on the principal eigenvector associated with the radar motion estimate at run-time. We have shown that the resulting classifier can be used to maintain an evolving prior belief in the motion that the robot undergoes which is exploited by a statistical filter to optimise the quality of the recovered motion. We demonstrated and evaluated our system on an outdoor dataset which challenges the

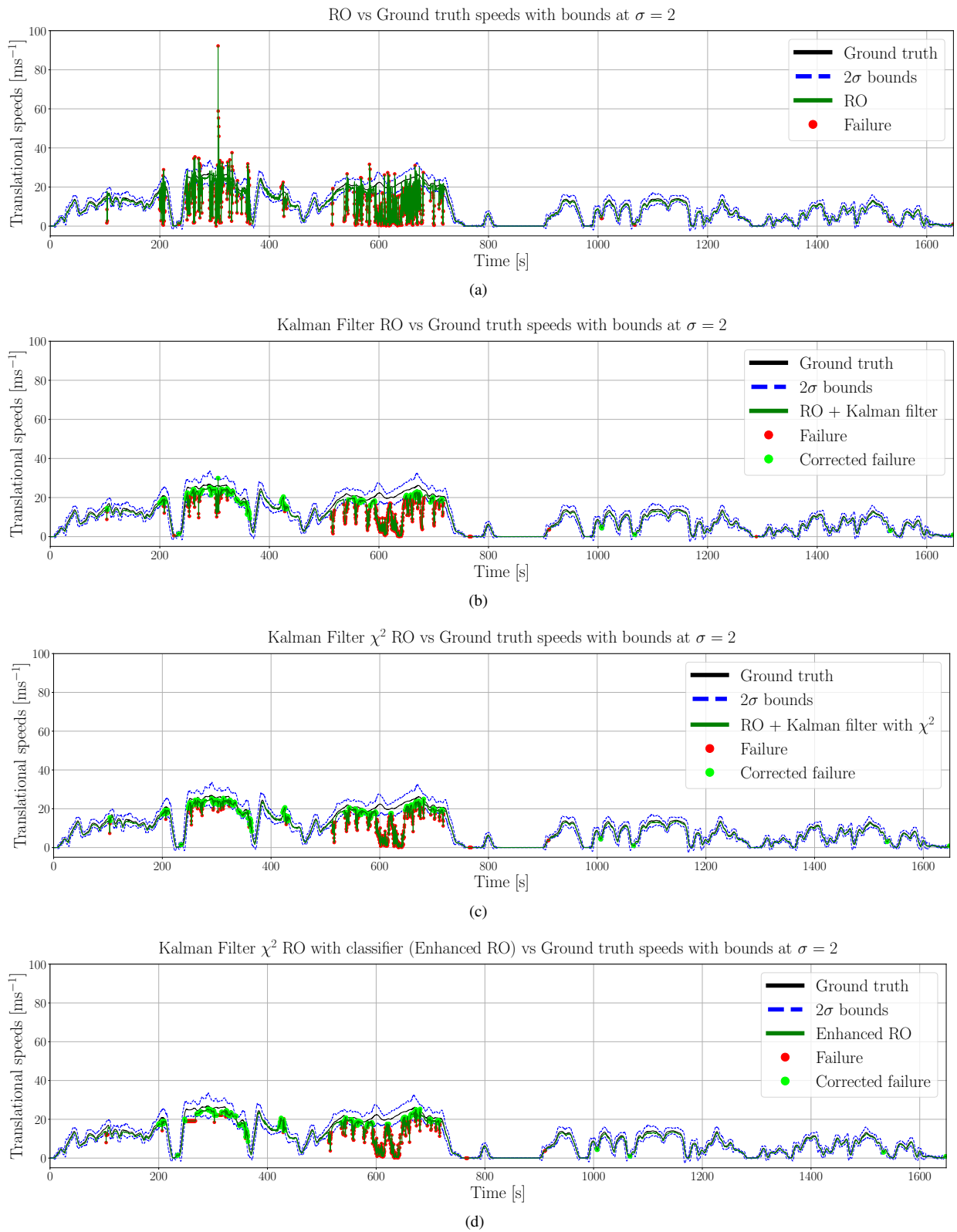


Fig. 7. Comparison of experiments for performance in estimating translational speed over 6635 matched frames. The standard RO system in Figure 7(a) sees 695 failures. Adding a KF in Figure 7(b) brings this down to 560 failures, with an addition of the χ^2 -test in Figure 7(c) reducing this further to 442 detections. Finally, the incorporation of introspection in the form of the classifier output shown in Figure 7(d) adds a small percentage of failures (1.34%), but these have a lower RMSE during failure and improve performance under prolonged periods of difficulty.

TABLE I

METRICS FOR SYSTEMS UNDER FAILURE COMPARED TO GROUND TRUTH

System	Failure [%]	Translation* [m s ⁻¹]		Rotation* [° s ⁻¹]	
		Failure	Overall	Failure	Overall
RO	10.473	11.112	4.755	7.740	3.899
RO, KF	8.440	8.739	3.567	2.271	2.826
RO, KF, χ^2	6.661	9.167	3.263	2.580	2.791
Enhanced RO	8.003	8.257	3.284	2.418	2.794

*RMSE values while under failure, and on the overall performance for this particularly challenging dataset. Please see [2, 3, 4] for performance metrics on other less challenging datasets. We remind the reader that datasets used in this paper were intentionally selected to expose where our previously existing RO system fails and highlight the benefits of the introspection addition.

current state-of-the-art in radar-only motion estimation and produced qualitatively and quantitatively superior RO ego-motion estimation.

ACKNOWLEDGEMENTS

Roberto Aldera and Daniele De Martini are supported by the UK Engineering and Physical Sciences Research Council (EPSRC) programme grant EP/M019918/1. Matthew Gadd is supported by Innovate UK under CAV2 – Stream 1 CRD (DRIVEN). Paul Newman is supported by EPSRC Leadership Fellowship Grant EP/J012017/1. The authors would also like to thank Paul Murcutt for his invaluable software engineering support.

REFERENCES

- [1] E. J. M. Adams, J. Mullane and B. Vo, *Robot Navigation and Mapping with Radar*. Artech House, 2012.
- [2] S. Cen and P. Newman, “Precise ego-motion estimation with millimeter-wave radar under diverse and challenging conditions,” in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, May 2018.
- [3] R. Aldera, D. De Martini, M. Gadd, and P. Newman, “Fast radar motion estimation with a learnt focus of attention using weak supervision,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Montreal, Canada, 2019*.
- [4] S. Cen and P. Newman, “Radar-only ego-motion estimation in difficult settings via graph matching,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Montreal, Canada, 2019*.
- [5] G. L. Scott and H. C. Longuet-Higgins, “Feature grouping by “relocalisation” of eigenvectors of the proximity matrix,” in *BMVC*, 1990, pp. 1–6.
- [6] L. S. Shapiro and J. M. Brady, “Feature-based correspondence: an eigenvector approach,” *Image and vision computing*, vol. 10, no. 5, pp. 283–288, 1992.
- [7] M. Leordeanu and M. Hebert, “A spectral technique for correspondence problems using pairwise constraints,” in *Tenth IEEE International Conference on Computer Vision (ICCV’05) Volume 1*, vol. 2. IEEE, 2005, pp. 1482–1489.
- [8] S. Sarkar and K. L. Boyer, “Quantitative measures of change based on feature organization: Eigenvalues and eigenvectors,” *Computer vision and image understanding*, vol. 71, no. 1, pp. 110–136, 1998.
- [9] A. Y. Ng, A. X. Zheng, and M. I. Jordan, “Link analysis, eigenvectors and stability,” in *International Joint Conference on Artificial Intelligence*, vol. 17, no. 1. Lawrence Erlbaum Associates Ltd, 2001, pp. 903–910.
- [10] G. Welch and G. Bishop, “An introduction to the kalman filter,” Chapel Hill, NC, USA, Tech. Rep., 1995.
- [11] B. D. Brumback and M. Srinath, “A chi-square test for fault-detection in kalman filters,” *Automatic Control, IEEE Transactions on*, vol. 32, pp. 552 – 554, 07 1987.
- [12] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 1990.
- [13] J. H. Challis, “A procedure for determining rigid body transformation parameters,” *Journal of biomechanics*, vol. 28, no. 6, pp. 733–737, 1995.
- [14] C. Karney, “Geographiclib,” *online at http://geographiclib.sourceforge.net*, 2015.
- [15] Z.-H. Zhou, “A brief introduction to weakly supervised learning,” *National Science Review*, vol. 5, no. 1, pp. 44–53, 2017.
- [16] G. S. Gurusinge, T. Nakatsuji, Y. Azuta, P. Ranjitkar, and Y. Tanaboriboon, “Multiple car-following data with real-time kinematic global positioning system,” *Transportation Research Record*, vol. 1802, no. 1, pp. 166–180, 2002.
- [17] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [18] A. Géron, *Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems*. O’Reilly Media, Inc., 2017.
- [19] T. Saito and M. Rehmsmeier, “The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets,” *PLoS one*, vol. 10, no. 3, p. e0118432, 2015.
- [20] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, “1 Year, 1000km: The Oxford RobotCar Dataset,” *The International Journal of Robotics Research (IJRR)*, vol. 36, no. 1, pp. 3–15, 2017.
- [21] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.