

## 2.2.1.2. Defining understanding requirements

### Practical guidance – cross-domain

**Author: Dr Matthew Gadd, University of Oxford**

#### What mobile robots need to understand

We consider the architecture to consist of algorithms, the sensor suite, and the mobile platform itself. This is done to stress that in tandem with unusual sensing for robustness, structured but data-driven algorithms for understanding can play their part and ease the burden on that technology.

In terms of algorithms, we phrase the requirements on intelligent understanding of the sensed world in terms of knowing:

1. How the RAS moves and where it is (“where am I?”),
2. Objects and agents in the vicinity of the RAS (“what is around me?”),
3. Likely and/or desired action of the RAS (“what should I do next?”).

We advocate that none of (1)-(3) is sufficient in isolation for safe operation and algorithms for understanding sensor data should all be designed with all three in mind.

While cameras and laser are exploited heavily in the industry and academia for good reason, this document focuses on the less well investigated scanning radar modality, which presents unique opportunities for as well as challenges to scene understanding.

#### Radar for robust understanding

The use of radar in this area can benefit safety greatly in scenarios in which traditionally exploited sensors like camera and laser will fail. There are various such difficult scenarios, including inclement weather under all driving conditions and at high-speed in motorway driving scenarios.

Radar is an interesting sensor modality in the sense that it lends itself less well to human interpretation than other sensors and is correspondingly difficult to process via hand-crafted or learned algorithms. Despite this, there is copious data inherent to the scan formation process which can - if processed properly - be used to understand the scene powerfully from recordings in this modality.

#### Implicit and explicit task understanding

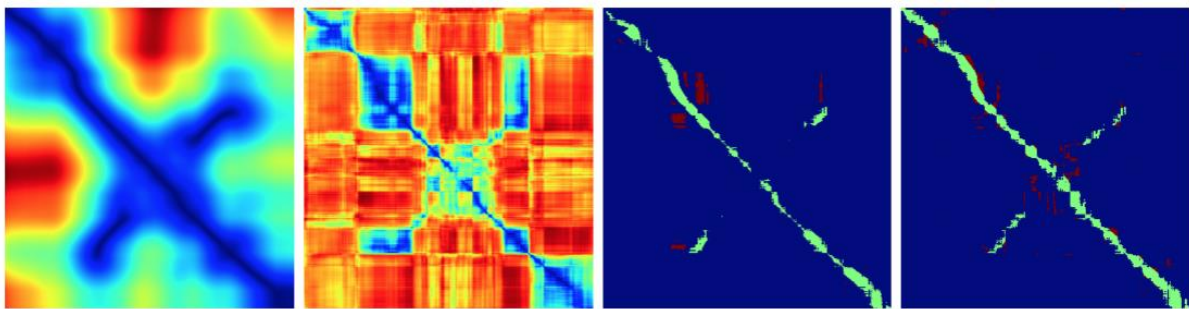
Therefore, analogously to the use of machine learning to bolster traditional sensing (e.g. raindrop removal from camera frames), our work explores how vast experience data can be leveraged to learn how to better understand the environment through this rich but tricky radar modality.

We consider lower-level autonomy-enabling tasks such as motion estimation which require implicit understanding (i.e. exact relevance of data features to the task are not understood easily by humans) as well as higher-level autonomy-enabling tasks such as semantic segmentation which provide explicit understanding of the world as sensed by the radar.

#### “Where am I?” - Motion estimation and localisation

We show that radar can understand the motion of the RAS precisely [20] and quickly [19] as well as with a degree of introspection under failure cases [14].

Precise understanding of motion (to the centimetre-level) is important for autonomous driving - we cannot deploy machines on our roads and in our public spaces with tolerances on the level of metres, accidents would abound. Understanding the scene quickly is important to safety in high-speed scenarios. Even in low-speed scenarios, the ability to process information quickly can free resources for other safety-oriented tasks.



*Fig 1: An illustration of how radar scans can be used to understand the geo-spatial location of mobile robots. Left: GPS similarity between two trajectories through the same part of the world. Middle left: Similarity of learned radar features, matching the GPS similarity well. Middle Right and Right: Mostly true place matches.*

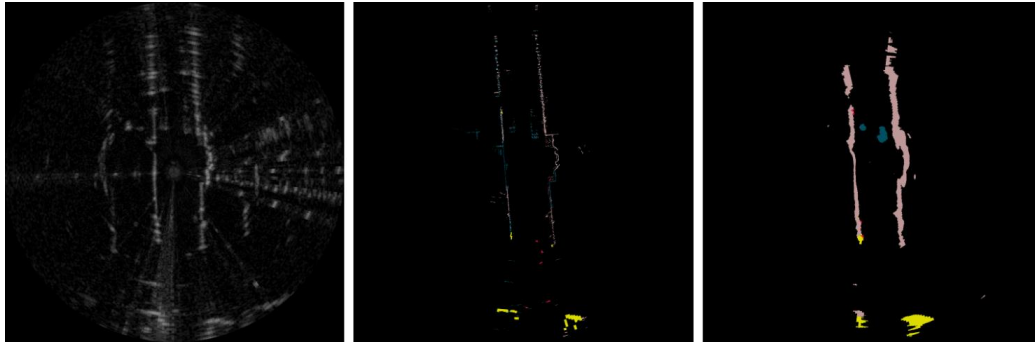
Shown in Fig. 1, we also show that radar can understand the global location of the RAS through exact supervision [12] and without explicit supervision at all [3, 21]. Furthermore, we show that imbuing understanding of video rather than static frames boosts localisation [10].

We consider introspection to have strong bearing on understanding, where we should expect our machines to as first prize anticipate failures but even if unable to, to be able to understand (and explain) the reason for failures.

#### “What surrounds me?” - Object detection and categorisation

Understanding the objects and actors in the scene is crucial for safety.

Increasingly, this task is performed by neural networks supervised directly by human annotation. Radar is a difficult domain to annotate by hand. However, supervision via other modalities is an efficient alternative.



*Fig 2: Radar (right) can be used to understand the content of scenes down to an object-category level.*

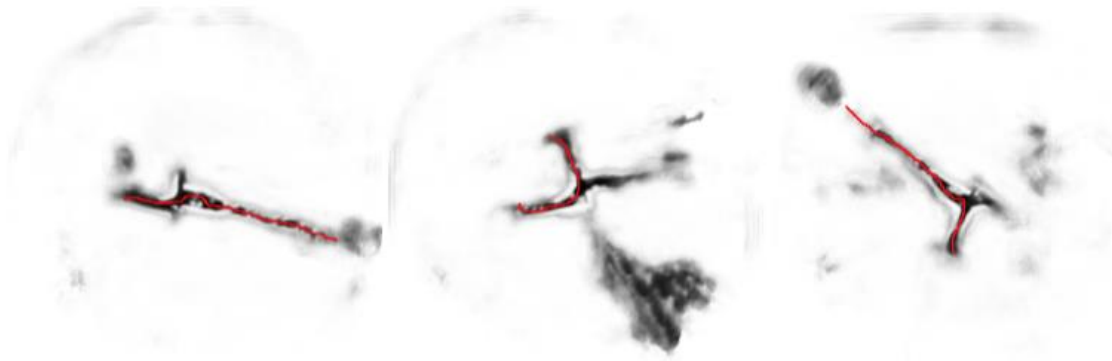
Indeed, as per Fig. 2, we show that radar, while limited in understanding materials and shapes, can nevertheless be used to richly decompose the scene into components [8].

### “What should I do next?” - Segmentation of driveable surface

There is a case to be made for autonomy which does not require prior mapping of an environment - so-called “mapless autonomy”. For this, the RAS must understand from its perceptions which parts of the scene are safe and legal to drive upon.

A powerful way to approach this problem is to learn from demonstration, and to generalise to unseen environments.

In this area, we show that radar, while limited in understanding texture, can nevertheless be used to understand the driveability of surfaces in urban [5] (Fig. 3) and off-road [7] scenarios.



*Fig 3: Radar is used to understand the driveability of the scene (black and white), giving us representations through which the RAS plans its motion (red).*

### Cross-modal understanding

In large part we advocate for novel technologies developed alongside algorithms for more robust yet interpretable scene understanding. However, we stop short of promoting one technology to the exclusion of others. There are certain scenarios in which one modality will perform best. There will be some scenarios in which a range of choices are available for which technology to deploy.

Ultimately, we advocate a mixed approach which is informed by experience. We need to learn to infer from contextual information to perform sensor scheduling and/or sensor fusion.

Our work in this area has so far focused on leveraging other modalities to learn in the radar domain [5,8,14], as radar data is inherently tricky to label.

## Summary

In summary, we have found that scanning radar, despite its age as a technology, is a tool that can rival vision and laser in scene understanding across all crucial autonomy-enabling tasks. It is a challenging modality in the sense that it is not always readily interpretable by humans and therefore difficult to label. Despite this, however, alongside its robustness it is information-rich and lends itself to perception best when knowledge is learned from strong training signals rather than handcrafted.

## References

- [3] M. Gadd, D. De Martini, P. Newman, "Unsupervised Place Recognition with Deep Embedding Learning over Radar Videos", Workshop on Radar Perception for All-Weather Autonomy at the IEEE International Conference on Robotics and Automation (ICRA), 2021
- [5] M. Broome, M. Gadd, D. De Martini, and P. Newman, "On the Road: Route Proposal from Radar Self-Supervised by Fuzzy LiDAR Traversability," *AI*, vol. 1, no. 4, pp. 558–58, 2020.
- [7] D. Williams, D. De Martini, M. Gadd, L. Marchegiani, and P. Newman, "Keep off the Grass: Permissible Driving Routes from Radar with Weak Audio Supervision," in *IEEE Intelligent Transportation Systems Conference (ITSC)*, Rhodes, Greece, 2020.
- [8] P. Kaul, D. De Martini, M. Gadd, and P. Newman, "RSS-Net: Weakly-Supervised Multi-Class Semantic Segmentation with FMCW Radar," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV)*, Las Vegas, NV, USA, 2020.
- [10] M. Gadd, D. De Martini, and P. Newman, "Look Around You: Sequence-based Radar Place Recognition with Learned Rotational Invariance," in *IEEE/ION Position, Location and Navigation Symposium (PLANS)*, Portland, OR, USA, 2020.
- [12] S. Saftescu, M. Gadd, D. De Martini, D. Barnes, and P. Newman, "Kidnapped Radar: Topological Radar Localisation using Rotationally-Invariant Metric Learning," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Paris, 2020.
- [14] R. Aldera, D. De Martini, M. Gadd, and P. Newman, "What Could Go Wrong? Introspective Radar Odometry in Challenging Environments," in *IEEE Intelligent Transportation Systems (ITSC) Conference*, Auckland, New Zealand, 2019.
- [19] R. Weston, M. Gadd, D. De Martini, P. Newman, and I. Posner, "Fast-MbyM: Leveraging Translational Invariance of the Fourier Transform for Efficient and Accurate Radar Odometry Conference on Robotics and Automation (ICRA), (Philadelphia, PA, USA), September 2022.

[20] R. Aldera, M. Gadd, D. De Martini, and P. Newman, "Leveraging a Constant-curvature Motion Constraint in Radar Odometry," in IEEE International Conference on Robotics and Automation (ICRA), (Philadelphia, PA,USA), September 2022.

[21] M. Gadd, D. De Martini, and P. Newman, "Contrastive Learning for Unsupervised Radar Place Recognition," in IEEE International Conference on Advanced Robotics (ICAR), (Ljubljana, Slovenia), December 2021