

**This item is the archived peer-reviewed author-version of:**

R2L-SLAM : sensor fusion-driven SLAM using mmWave Radar, LiDAR and deep neural networks

**Reference:**

Balemans Niels, Hooft Lucas, Reiter Philippe, Anwar Ali, Steckel Jan, Mercelis Siegfried.- R2L-SLAM : sensor fusion-driven SLAM using mmWave Radar, LiDAR and deep neural networks  
Proceedings of IEEE Sensors - ISSN 1930-0395 - IEEE, 2023, p. 1-4  
Full text (Publisher's DOI): <https://doi.org/10.1109/SENSORS56945.2023.10324990>  
To cite this reference: <https://hdl.handle.net/10067/2013740151162165141>

# R2L-SLAM: Sensor Fusion-Driven SLAM using mmWave Radar, LiDAR and Deep Neural Networks

Niels Balemans<sup>†‡</sup>, Lucas Hooft<sup>\*</sup>, Philippe Reiter<sup>†</sup>, Ali Anwar<sup>†</sup> (Member, IEEE), Jan Steckel<sup>†§</sup>, Siegfried Mercelis<sup>†</sup>  
Email: niels.balemans@uantwerpen.be

**Abstract**—Optical sensing modalities are extensively used in autonomous vehicles (AVs). These sensors are, however, not always reliable, particularly in harsh or difficult sensing conditions, such as with smoke or rain. This limitation can impact their application potential due to safety concerns, since optical sensors can fail to reliably perceive obstacles in such harsh conditions. To address this, it would be desirable to include other modalities, such as radar, into the perception sensor suites of these AVs. However, this is difficult because many recent state-of-the-art navigation algorithms are designed specifically for LiDAR sensors. In this work, we propose a modality prediction method that allows for the addition of a single-chip mmWave radar sensor to an existing sensor setup consisting of a 2D LiDAR sensor, without changing the current downstream applications. We demonstrate the increased reliability of our method in situations where optical sensing modalities become less accurate and unreliable.

**Index Terms**—Sensor systems, mmWave Radar, Autonomous vehicles, Navigation, Deep learning

## I. INTRODUCTION

While autonomous vehicles (AVs) have shown potential in various domains, from consumer to industrial and agricultural applications [1], [2], [3], [4], their safe and robust deployment is not always guaranteed. It is evident that the sensors used to perceive the environment play a crucial role in achieving safe and reliable AVs. To observe the surroundings and create a representation of the environment for use by navigation and detection algorithms, the majority of current state-of-the-art vehicles are mostly equipped with cameras and time-of-flight (ToF) sensors (e.g., LiDAR) [5]. By only using these optical sensors, the reliability of these AVs operations is significantly reduced in scenarios that have difficult conditions for these sensors, such as smoke, rain or sun-flare [6]. While safety and reliability are the priority, AVs have an immense application potential that is limited by only equipping these AVs with optical sensor suites [7].

Due to the widespread use of camera and LiDAR sensors, most research efforts focus on extending and improving algorithms and models designed for these sensors. In addition to these research achievements, these methodologies have also been widely adopted in various application domains, creating

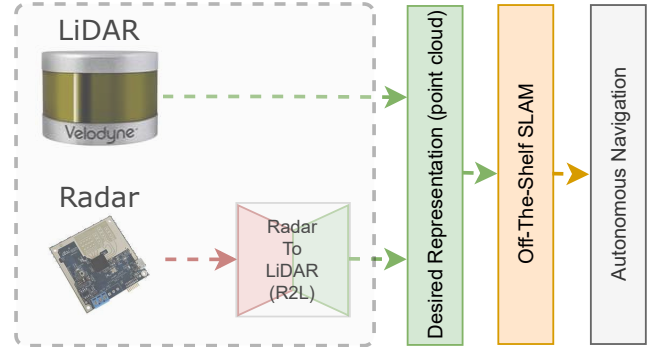


Fig. 1. An overview of the proposed method that aims to abstract the sensors used in the navigation stack. By converting measurements from a multimodal sensing system into a desired representation, the navigation stack remains static, while the best modality can be selected for a given scenario.

a large (commercial) community that has tested, deployed, and improved upon them. Therefore, it would be desirable to extend these methodologies to other sensing modalities. With the approach outlined in this paper, we aim to add single-chip mmWave radar to new and existing sensing setups, without the need to completely replace or change the processing designed for the current sensing modalities.

In this paper, we build upon our previous work and incorporate mmWave radar into the sensing setup for a 2D LiDAR SLAM (Simultaneous Localisation and Mapping) application, as shown in Figure 1. We believe that our contributions with this paper are twofold: first, increased applicability of mmWave radar by enabling its use with popular methodologies designed for LiDAR; and second, a framework that allows for dynamic switching between sensor measurements. Section II provides the background for the proposed approach introduced in Section III. Finally, Section IV discusses the research results and introduces potential future directions.

## II. BACKGROUND

In the following subsections, we briefly introduce our previous work on predicting LiDAR data using a 3D in-air ultrasonic sensor (eRTIS), developed within our research group [8]. We also provide an overview of the state-of-the-art in radar sensors for perception and navigation.

### A. LiDAR point cloud prediction

In previous work, we introduced a method for incorporating the eRTIS sonar sensor into an existing sensor setup without making any changes to the downstream processing and

<sup>\*</sup> Faculty of Applied Engineering, University of Antwerp, Groenenborgerlaan 171, 2020 Antwerp, Belgium

<sup>†</sup> IDLab - Faculty of Applied Engineering, University of Antwerp - imec, Sint-Pietersvliet 7, 2000 Antwerp, Belgium

<sup>‡</sup> Cosys-Lab - Faculty of Applied Engineering, University of Antwerp, 2000 Antwerp, Belgium

<sup>§</sup> Flanders Make Strategic Research Centre, Lommel, Belgium

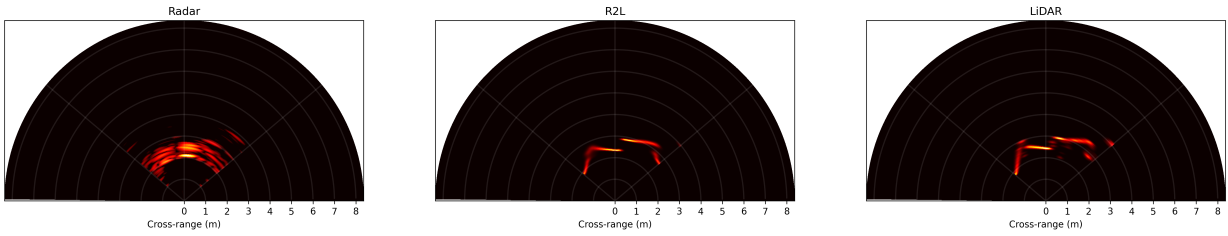


Fig. 2. Samples of the radar, R2L and LiDAR scan data, from left to right. Field-of-view is limited to that of the radar sensor, with a maximum range of 8 meters and angular field-of-view of 100 degrees.

decision-making. We showed using a stacked convolutional autoencoder to predict how a LiDAR sensor would perceive an environment based on the ultrasonic measurements. We designed and trained S2L-Net (pronounced “Sonar to LiDAR”), using a data-driven supervised learning method. These S2L-Net predictions are then converted to 2D point-clouds and used in off-the-shelf 2D LiDAR-SLAM (S2L-SLAM). We concluded that a data-driven, learning-based method is suitable for approaching an ill-posed inverse problem such as modality conversion, especially when large datasets can be easily obtained through either simulation or real-world measurements. For a detailed discussion on S2L-Net and S2L-SLAM we redirect the reader to [9] and [10] respectively.

### B. Radar in autonomous navigation and mapping

The increased availability of low-cost, single-chip mmWave radar devices has led to increased activity in the research field, particularly for navigational purposes. Numerous works have focused on increasing the resolution of radar data [11], [12], [13], [14]. Other works have focused on obtaining odometry and mapping information using radar data [15], [16] as well as the application of scan matching techniques and sensor fusion techniques with IMU (Inertial Measurement Unit) or LiDAR data [17]. While most methods are designed to perform specific tasks using radar, our goal is to maintain reliable environment measurements for safe and robust operation of the downstream applications even in situations with difficult optical sensing conditions. We demonstrate this by using laser-scan predictions based on mmWave radar alongside LiDAR scans in 2D LiDAR-SLAM to increase reliability.

## III. R2L-SLAM

Our goal is to increase the reliability of AVs by incorporating a single-chip mmWave radar sensor into an existing sensor setup designed for LiDAR. To achieve this, we developed R2L-Net (pronounced “Radar to LiDAR”) that predicts how an environment is perceived by a LiDAR sensor based on radar measurements. In the following subsection, we explain the R2L-Net architecture more in-depth and present the design considerations that enable us to obtain high-resolution LiDAR-like point clouds as the output of our model.

### A. R2L-Net design and data considerations

When comparing the heat-maps generated from mmWave radar to the LiDAR data, in figure 2, it becomes clear that the radar measurement provides a much sparser representation of

the environment than LiDAR. This difference can be attributed to the distinct sensing modalities and physical aspects that each sensor measures. In typical man-made environments, the mmWave radar will only receive sparse (specular) reflections due to the large Helmholtz number of the used radar signals [18]. On the other hand, a LiDAR will also receive dense (diffuse) reflections by using light. Our R2L-Net model’s task is to upscale the sparse point-cloud obtained by the mmWave radar into a dense LiDAR-like point-cloud. For designing R2L-Net, we created a stacked convolutional autoencoder, inspired by S2L-Net [9] and U-Net [19], customized for the radar data. This model encodes the range-azimuth heat-map images using five encoder layers, creating a lower-dimensional latent representation consisting of multiple channels, each of which contains different features. To solve the vanishing gradient problem, most of the encoder layers feature residual connections [20]. After several intermediate residual layers, the latent representation is up-scaled to a range-azimuth image by the decoder layers. At the input of our R2L-Net model, we use a sequence of five range-azimuth heat-map images, as this represents the data in an orthogonal grid, which is best suited for use with the 2D convolution kernels throughout the model. For the kernel sizes, we opted for rectangular kernels with an aspect ratio similar to the input data and an size that becomes smaller towards the center of the network. These rectangular kernels, along with the network depth, allows far-away pixels to exchange information, which improves results significantly according to our previous work. The model is trained with a weighted combination of the mean-square-error (MSE), cross entropy loss and structural similarity index. The cross-entropy loss provides a classification for each angle across the possible ranges, determined by the minimum, maximum range and resolution. For each angle, the model is trained to predict the closest ‘range class’. If the desired range for a specific angle is out of bounds the field-of-view, the range class is set to 0 meters. Performing this classification provides an output image suitable to convert to a point-cloud, with out-of-bounds measurements detected when the range class is set to 0 meters.

We collected our own dataset to obtain training and validation data. To our knowledge, no public dataset exists in the correct format that includes time-synchronized, raw mmWave radar data and LiDAR measurements. Most publicly available datasets, such as the nuScenes dataset [21], only consist of processed radar data, which is not suitable for our purposes. To create our own dataset, we used a single-chip mmWave TI

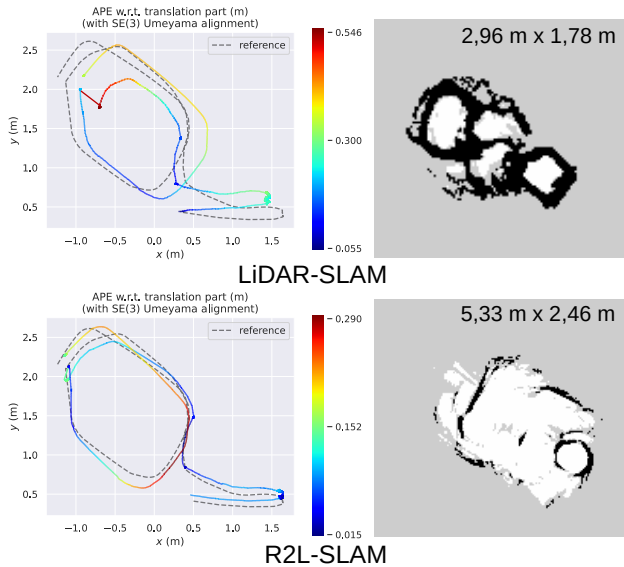


Fig. 3. Pose-graph and occupancy grid plots for comparison between SLAM on LiDAR and R2L scans, respectively, in a smoke-filled room. Reference data for the pose-graphs is obtained from the vehicles odometry data.

TABLE I  
SLAM APE W.R.T TRANSLATION (M)

Scenario	RMSE	MEAN	STD.	SSE
LiDAR [smoke]	0.285	0.267	0.100	845.00
R2L [smoke]	0.121	0.101	0.067	151.78
LiDAR + R2L [smoke]	0.113	0.087	0.072	130.84
LiDAR [no smoke]	0.249	0.214	0.126	1723.60
R2L [no smoke]	0.415	0.343	0.234	4727.24
LiDAR + R2L [no smoke]	0.235	0.208	0.107	1516.14

IWR1443 radar sensor [22] and a Hokuyo UST-20L [23] on a small robotic platform. The platform performed measurements along 15 trajectories in an indoor office and lab environment, providing approximately 10,000 synchronized samples for our training dataset and 1,500 samples for our validation dataset. The TI IWR1443 radar sensor was configured to the 77-81 GHz frequency band with an angular field-of-view of approximately 100 degrees and a maximum range of 8 meters with a 5 cm range resolution. Samples of our datasets are presented in figure 2.

#### IV. RESULTS AND DISCUSSION

We used an off-the-shelf 2D LiDAR SLAM method to verify our model predictions. We opted for Google Cartographer [24] as this is a very mature and capable method. A standard configuration is used for which odometry data is used, based on the wheel encoders of the robot, along with laser-scans to build the occupancy grid. The laser-scan data is either provided by the LiDAR sensor or predicted R2L-Net data. Experiments were performed by switching between LiDAR and R2L scans at several steps along the trajectory, as well as experiments with only R2L scans to compare against LiDAR. We created multiple validation trajectories using our platform, both in indoor office and lab environments. For one of the validation scenarios, we constructed a small smoke chamber equipped with a smoke machine having an emission volume of



Fig. 4. Office environment, filled with smoke and used as test dataset.

approximately 110 m<sup>3</sup>/min, presented in figure 4. The smoke was allowed to spread across the room, and the machine was started before the measurements were recorded. We observed that the LiDAR sensors would quickly become noisy and stop perceiving the environment, even with a limited amount of smoke. The results from this smoke test are presented in figure 3 the output pose-graph and occupancy grid are shown from Cartographer SLAM on LiDAR and R2L-Net scans along the complete trajectory. These results show that, even with the available odometry data along with relative position information obtained by scan-matching, the results of LiDAR affected by smoke become uncertain and unreliable, while the results with R2L predictions based on mmWave radar are stable and closely match the ground-truth. Table I presents absolute position error (APE) of the approximated positions obtained by Cartographer over our validation trajectories. These results demonstrate that using only R2L-Net scans performs slightly worse in situations with good conditions for optical sensors, but performs better in situations where, for example, smoke is present. The slightly less accurate results of R2L-SLAM compared to LiDAR-SLAM can be attributed due to the lower maximum range the radar sensor and the R2L-Net predictions are able to achieve. While this is something that can be mitigated by a different sensor, with our method we wanted to demonstrate the flexibility of changing between sensor measurements, and selecting the best modality given the situation, without changing the processing stack.

With this work we wanted to underline the benefit of using multi-modal sensing setups to dynamically select the most suitable sensing modality without changing the processing stack. Based on the results presented in this paper, we conclude that our modality prediction framework has the potential to succeed in its aim of abstracting the sensors used from the processing stack, allowing for addition of different sensors and dynamic sensor fusion. In an extension on this work, we will investigate improving resolution and maximum achievable range, as well as the addition of range-Doppler information as extra input for our model. We anticipate that this will improve predictions with moving objects in dynamic environments.

#### ACKNOWLEDGMENT

This work was supported by the Research Foundation Flanders (FWO) under Grant Number 1S75622N and the Flemish Government (AI Research Program).

## REFERENCES

- [1] A. Gilchrist, *Industry 4.0*. Apress, 2016. [Online]. Available: <http://link.springer.com/10.1007/978-1-4842-2047-4>
- [2] T. Zonta, C. A. da Costa, R. da Rosa Righi, M. J. de Lima, E. S. da Trindade, and G. P. Li, "Predictive maintenance in the industry 4.0: A systematic literature review," *Computers and Industrial Engineering*, vol. 150, p. 106889, 12 2020. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0360835220305787>
- [3] A. M. Farid, M. Alshareef, P. S. Badhessa, C. Boccaletti, N. A. A. Cacho, C.-I. Carlier, A. Corriveau, I. Khayal, B. Liner, J. S. Martins, F. Rahimi, R. Rossett, W. C. Schoonenberg, A. Stillwell, and Y. Wang, "Smart city drivers and challenges in urban-mobility, health-care, and interdependent infrastructure systems," *IEEE Potentials*, vol. 40, pp. 11–16, 1 2021. [Online]. Available: <https://ieeexplore.ieee.org/document/9307293/>
- [4] C. Kyrkou, S. Timotheou, P. Kolios, T. Theocharides, and C. Panayiotou, "Drones: Augmenting our quality of life," *IEEE Potentials*, vol. 38, pp. 30–36, 1 2019. [Online]. Available: <https://ieeexplore.ieee.org/document/8598647/>
- [5] C. Chen, B. Wang, C. X. Lu, N. Trigoni, and A. Markham, "A survey on deep learning for localization and mapping: Towards the age of spatial machine intelligence," *arXiv*, 6 2020. [Online]. Available: <http://arxiv.org/abs/2006.12567>
- [6] M. Bijelic, T. Gruber, F. Mannan, F. Kraus, W. Ritter, K. Dietmayer, and F. Heide, "Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather," *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2020*, 2 2019. [Online]. Available: <http://arxiv.org/abs/1902.08913>
- [7] M. Milford, S. Anthony, and W. Scheirer, "Self-driving vehicles: Key technical challenges and progress off the road," *IEEE Potentials*, vol. 39, pp. 37–45, 1 2020.
- [8] R. Kerstens, D. Laurijssen, and J. Steckel, "ertis: A fully embedded real time 3d imaging sonar sensor for robotic applications." *IEEE*, 5 2019, pp. 1438–1443. [Online]. Available: <https://ieeexplore.ieee.org/document/8794419/>
- [9] N. Balemans, P. Hellinckx, and J. Steckel, "Predicting lidar data from sonar images," *IEEE Access*, vol. 9, pp. 57 897 – 57 906, 4 2021. [Online]. Available: <https://ieeexplore.ieee.org/document/9400352/>
- [10] N. Balemans, P. Hellinckx, S. Latre, P. Reiter, and J. Steckel, "S2l-slam: Sensor fusion driven slam using sonar, lidar and deep neural networks." *IEEE*, 10 2021, pp. 1–4. [Online]. Available: <https://ieeexplore.ieee.org/document/9639772/>
- [11] H. Yamada, T. Kobayashi, Y. Yamaguchi, and Y. Sugiyama, "High-resolution 2d sar imaging by the millimeter-wave automobile radar," vol. 2018-January. *IEEE*, 12 2017, pp. 149–150. [Online]. Available: <http://ieeexplore.ieee.org/document/8273384/>
- [12] Y. Cheng, J. Su, M. Jiang, and Y. Liu, "A novel radar point cloud generation method for robot environment perception," *IEEE Transactions on Robotics*, vol. 38, pp. 3754–3773, 12 2022.
- [13] A. Prabhakara, T. Jin, A. Das, G. Bhatt, L. Kumari, E. Soltanaghaei, J. Bilmes, S. Kumar, and A. Rowe, "High resolution point clouds from mmwave radar," 6 2022. [Online]. Available: <http://arxiv.org/abs/2206.09273><http://dx.doi.org/10.1109/ICRA48891.2023.10161429>
- [14] A. Geiss and J. C. Hardin, "Radar super resolution using a deep convolutional neural network," *Journal of Atmospheric and Oceanic Technology*, vol. 37, pp. 2197–2207, 11 2020. [Online]. Available: <https://journals.ametsoc.org/view/journals/atot/37/12/jtech-d-20-0074.1.xml>
- [15] C. X. Lu, S. Rosa, P. Zhao, B. Wang, C. Chen, J. A. Stankovic, N. Trigoni, and A. Markham, "See through smoke: Robust indoor mapping with low-cost mmwave radar." *ACM*, 6 2020, pp. 14–27. [Online]. Available: <http://arxiv.org/abs/1911.00398><https://dl.acm.org/doi/10.1145/3386901.3388945>
- [16] Z. Hong, Y. Petillot, A. Wallace, and S. Wang, "Radar slam: A robust slam system for all weather conditions," 4 2021. [Online]. Available: <https://arxiv.org/abs/2104.05347v1>
- [17] J. Michalczyk, C. Schoffmann, A. Fornasier, J. Steinbrener, and S. Weiss, "Radar-inertial state-estimation for uav motion in highly agile manoeuvres," *2022 International Conference on Unmanned Aircraft Systems, ICUAS 2022*, pp. 583–589, 2022.
- [18] G. Schouten, W. Jansen, and J. Steckel, "Simulation of pulse-echo radar for vehicle control and slam," *Sensors 2021, Vol. 21, Page 523*, vol. 21, p. 523, 1 2021. [Online]. Available: <https://www.mdpi.com/1424-8220/21/2/523/htm><https://www.mdpi.com/1424-8220/21/2/523>
- [19] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," 5 2015. [Online]. Available: <http://arxiv.org/abs/1505.04597>
- [20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," vol. 2016-Decem. *IEEE*, 6 2016, pp. 770–778. [Online]. Available: <http://ieeexplore.ieee.org/document/7780459/http://image-net.org/challenges/LSVRC/2015/>
- [21] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuscenes: A multimodal dataset for autonomous driving," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 11 618–11 628, 3 2019. [Online]. Available: <https://arxiv.org/abs/1903.11027v5>
- [22] K. Liu, Y. Li, N. Xu, and P. Natarajan, "Learn to combine modalities in multimodal deep learning," 5 2018. [Online]. Available: <https://arxiv.org/abs/1805.11730v1>
- [23] Hokuyo, "Ust-20lx (uust004) specification," pp. 4–9, 2014.
- [24] W. Hess, D. Kohler, H. Rapp, and D. Andor, "Real-time loop closure in 2d lidar slam," vol. 2016-June. *IEEE*, 5 2016, pp. 1271–1278. [Online]. Available: <http://ieeexplore.ieee.org/document/7487258/>