CFEAR++: An Efficient Radar-inertial Semantic Odometry

Siru Li*, Yushuai Chen*, Ziyang Hong*, and Liang Hu[1]

***Abstract*— In this article, we describe the radar-inertial odometry approach CFEAR++, which is based on the CFEAR [1] method. Our method is submitted to the radar odometry competition at the workshop of Radar in Robotics[1], ICRA 2024.**

I. INTRODUCTION

Using cameras and LiDAR to perceive the surrounding environment has greatly contributed to the rapid development of autonomous driving. The localization task based on cameras and LiDAR has become quite mature. However, in extreme weather conditions such as rain, snow, and fog, the aforementioned sensors struggle to gather effective information because of the occlusion from various particles. The electromagnetic spectrum of radar typically lies in a much lower frequency (GHz) band compared to cameras and LiDAR. Relying on this characteristic, radar maintains robustness in various weather conditions, but its precision and other capabilities are weaker compared to the other two sensors. At the same time, the information obtained by radar includes various types of noise, which can affect traditional tasks such as localization.

In the past few years, many researchers have used deep learning methods to outputs clear radar images using lidar as a supervisory signal [2], [3]. Meanwhile, To address the issue of deep learning causing the loss of some features, the method of sliding windows is used which will preserve long-range sensing and penetrating capabilities [4]. Traditional methods, represented by constant false-alarm rate (CFAR) filtering, have also been used [5]. Furthermore, by projecting semantically segmented images onto the point cloud, RSS-Net [6] outputs radar data with semantic information.

In terms of sources of semantics supervision, we can leverage information from vision and LiDAR.

II. CFEAR++

The foundation of our proposed CFEAR++ method is the radar odometry approach described in [1], which we extend by integrating semantic information provided by a semantic segmentation deep neural network as illustrated in Fig. 1. We use the semantics information to filter out some dynamic objects and noises like multipath reflection and saturation. To further enhance the rotation estimation, we use the angular velocity information from the gyro to compensate the motion. This is different from the CFEAR approach since they use previous motion as prior to perform motion compensation.

*Equal Contribution.

The authors are with the Department of Automation, School of Mechanical Engineering and Automation, Harbin Institute of Technology, Shenzhen, China. Email: l.hu@hit.edu.cn.

[1]https://sites.google.com/view/radar-robotics/competition

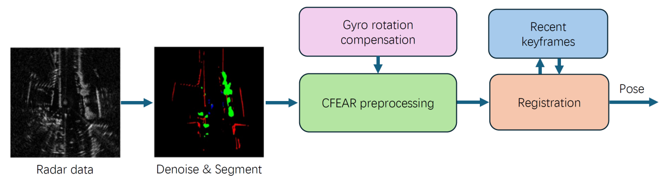

Fig. 1: Overview of CFEAR++ Radar-inertial odometry

A. Using LiDAR Data As Supervision

To generate radar semantics map, we first need to train a deep semantics network given the raw radar input image. We use a deep learning approach, with lidar information as the supervision signal and radar as input, which outputs clear radar data. Considering that the sensing range of radar is much greater than that of lidar, while lidar has a distinct advantage in precision. Therefore, similar to Kung's work [4], we only input raw radar data within the lidar sensing range. What makes it different from Kung's work is that we employ polar sliding window inference only for inferring radar data beyond the lidar sensing range. The reason is that we need lidar labels for the next step, which makes it challenging to utilize the parts that are not visible to lidar. Without ground truth, we cannot judge whether the objects preserved through polar sliding window inference have been correctly segmented.

B. Radar Segmentation Without Expert Labels

To maximize the accuracy of the generated ground truth, input the raw lidar point cloud data into the current state-of-the-art point cloud semantic segmentation model. Generate the produced output as the ground truth for supervising radar semantic segmentation. Some labels in radar data may only be represented by a few small points, such as human, at the same time, segmenting them out holds little significance for our localization purpose. We choose to uniformly classify them as noise points. Finally, considering the impact of the radar elevation beamwidth, remove some points from the point cloud.

C. Denoise and Segment

By examining the labels generated by the segmentation model, we found that the accuracy for most generating labels of objects is high and relatively stable, maintaining the same labels for nearby frames. But there is a large number of incorrect labels identifying manmade as vegetation, which significantly impacts our training results. Meanwhile, the

TABLE I: Evaluation for training sequences in the Boreas dataset. Drift is measured by the translation error [%], rotation error (degree/100m). For each sequence, The least translation error of each sequence is highlighted in bold.

Sequence	20-12-01-13	21-01-26-11	21-04-29-15	21-07-20-17	21-04-15-18	21-04-20-14
CFEAR-3 [1]	0.902/0.291	0.797/0.289	0.860/0.268	0.937/0.227	0.902/0.281	0.845/0.282
None removed	0.686/0.283	0.656/0.287	0.651/0.257	0.624/0.194	**0.575**/0.264	0.611/0.272
Car removed	0.693/0.285	0.655/0.287	0.638/0.256	0.626/0.198	0.621/0.271	0.622/0.276
Car and tree removed	0.683/0.284	0.639/0.280	0.646/0.267	0.644/0.211	0.630/0.267	0.661/0.278
None removed + IMU	0.640/0.276	0.602/0.279	0.614/0.248	0.652/0.199	0.600/0.263	0.583/0.267
Car removed+ IMU	0.640/0.276	0.594/0.277	0.610/0.250	0.643/0.199	0.618/0.267	0.601/0.270
Car and tree removed + IMU (**Ours**)	**0.638**/0.276	**0.584**/0.272	**0.592**/0.249	**0.599**/0.191	0.605/0.261	**0.541**/0.258

model cannot completely segment the whole tree, resulting in some manmade labels being mixed in. We calculate the mathematical characteristics of the corresponding labels to make corrections. After corrections, we only retain the manmade labels which are the most useful for odometry. The normal vectors calculated from tree labels are not stable enough, and moving vehicles will lead to negative optimization during registration. Only retaining manmade labels with strong structual features is the most beneficial for our odometry task.

D. Implementation Details

We train the network with four Nvidia RTX4090 GPUs and a batch size b of 800 over 50 epochs. For the task of denoising, we use both Focal loss [7] and Dice loss [8], which we give the same weight. Using either loss function alone cannot complete the task. Only using Focal loss would misclassify a large number of noise points as obstacles, while only using Dice loss would fail to recognize any valid radar data. The network of the task follow the U-Net [9] architecture. We trained our model using the AdamW [10] optimizer with a learning rate of 0.0001, the weight decay of 1×10^{-6}, the momentum of 0.9. As for the registration parameters, we use the same configuration of CFEAR-3.

III. EXPERIMENTAL RESULTS

We use CFEAR-3 [1] with 10 keyframes as the baseline. As a ablation study, we conduct comparative experiments with different semantic classes removed: 1). all classes kept; 2) car class removed; and 3). both car and tree classes removed, in combination with and without IMU. As shown in Table II-B, in most sequences (5 out of 6), the method with both car and tree classes removed and with IMU achieves the best performance in terms of translation errors, the smallest 0.541% and the biggest 0.638%. The major improvement is attributed to two reasons: 1). the removal of moving car and trees facilitates more stable feature registration; and 2). the introduced gyroscope rotation information provides a good motion prior during sharp turns.

IV. CONCLUSION AND FUTURE WORK

This article presents CFEAR++, and discusses the localization improvement using semantics information and gyroscope. On Boreas sequences, we achieve a very low drift performance less than 0.6%.

REFERENCES

[1] D. Adolfsson, M. Magnusson, A. Alhashimi, A. J. Lilienthal, and H. Andreasson, "Lidar-level localization with radar? the cfear approach to accurate, fast, and robust large-scale radar odometry in diverse environments," *IEEE Transactions on robotics*, vol. 39, no. 2, pp. 1476–1495, 2022.

[2] R. Weston, S. Cen, P. Newman, and I. Posner, "Probably unknown: Deep inverse sensor modelling radar," in *2019 International Conference on Robotics and Automation (ICRA)*, pp. 5446–5452, 2019.

[3] R. Weston, O. P. Jones, and I. Posner, "There and back again: Learning to simulate radar data for real-world applications," 2020.

[4] P.-C. Kung, C.-C. Wang, and W.-C. Lin, "Radar occupancy prediction with lidar supervision while preserving long-range sensing and penetrating capabilities," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 2637–2643, 2022.

[5] H. Rohling, "Radar cfar thresholding in clutter and multiple target situations," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-19, no. 4, pp. 608–621, 1983.

[6] P. Kaul, D. de Martini, M. Gadd, and P. Newman, "Rss-net: Weakly-supervised multi-class semantic segmentation with fmcw radar," in *2020 IEEE Intelligent Vehicles Symposium (IV)*, pp. 431–436, 2020.

[7] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," 2018.

[8] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," 2016.

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015.

[10] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," 2019.