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

The use of cameras and LiDAR for environmental perception has significantly advanced the development of autonomous driving. However, both sensors face challenges in adverse weather conditions such as rain, snow and fog due to occlusion from various particles. In contrast, radar which operates in a lower frequency band within the GHz range offers robust and reliable perception under such conditions. Despite this advantage, radar data generally lacks the precision and quality of camera and LiDAR data due to issues like data sparsity and sensor-specific noise. Consequently, several effective noise reduction methods have been proposed for radar denoising, particularly using convolutional neural networks (CNNs) [2]–[5].

Semantic information plays a crucial role in complementing visual inputs and has been applied in many advanced semantic visual simultaneous localization and mapping (SLAM) and navigation methods [6], [7]. However, semantic annotations remain absent in most existing radar datasets [8], [9], not to mention downstream tasks semantic radar SLAM. To fill the gap, we propose a weakly supervised learning approach to radar semantic segmentation. For odometry estimation tasks, most algorithms are based on the assumption that the environment is static, and moving objects such as cars and pedestrians which are ubiquitous in outdoor environments, if not identified and removed correctly, could cause incorrect feature matches between frames and hence degrade odometry accuracy. With our method, semantic information allows us to selectively retain stable features from static objects like buildings, improving the accuracy of odometry estimation.

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 for the motion. This is different from the CFEAR approach since they use previous motion as prior to perform motion compensation.

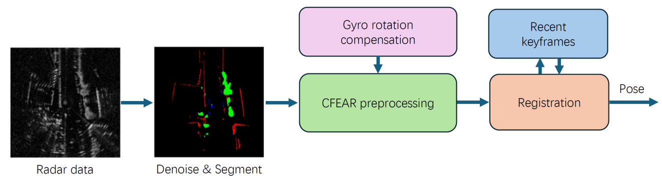

Fig. 1: Overview of CFEAR++ Radar-inertial odometry

A. LiDAR Data Preprocessing

Building on Lee's method [10] for removing ground points from LiDAR data, we first align the LiDAR coordinate system with that of the radar. Subsequently, we eliminate the corresponding LiDAR point cloud data based on the differences in the fields of view of the two sensors. We calculate the angle between the line connecting each LiDAR point and the origin of the new coordinate system with the horizontal plane. Points that fall outside the radar's vertical field of view need to be removed.

B. LiDAR Label Generalization

We first apply the ground segmentation method to remove the ground point cloud. The remaining point cloud is then fed into the deep learning model to obtain the corresponding semantic labels. To facilitate the refinement of labels, we consolidate all labels into four categories: noise, vehicle, vegetation and building. The noise category primarily includes objects appearing as small clusters in radar points, and pedestrian is classified into noises as segmenting them as an additional class increases the complexity of the semantic segmentation task while also offering limited benefits for downstream tasks. After analyzing the generated point cloud labels, we find that the semantic segmentation of vehicle is highly accurate. However, issues still arise in the segmentation of vegetation and building. We refine the wrong label based on structural features and distribution patterns. Following that, the data is projected to create a supervisory signal with specific semantic information. This enables semantic segmentation of radar for downstream tasks.

C. Implementation Details

We train the network with four Nvidia RTX4090 GPUs and a batch size of 200 over 50 epochs. For the task of radar

*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

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

Methods	20-12-01-13		21-01-26-11		21-04-15-18		21-04-20-14		21-04-29-15		21-07-20-17	
	Trans.	*Rot.*	*Trans.*	*Rot.*	*Trans.*	*Rot.*	*Trans.*	*Rot.*	*Trans.*	*Rot.*	*Trans.*	*Rot.*
CFEAR-3 [1]	0.902	0.291	0.797	0.289	0.902	0.281	0.845	0.282	0.860	0.268	0.937	0.227
None removed	0.686	0.283	0.656	0.287	**0.575**	0.264	0.611	0.272	0.651	0.257	0.624	0.194
Vehicle removed	0.693	0.285	0.655	0.287	0.621	0.271	0.622	0.276	0.638	0.256	0.626	0.198
Only building	0.683	0.284	0.639	0.280	0.630	0.267	0.661	0.278	0.646	0.267	0.644	0.211
IMU	0.866	0.284	0.735	0.279	0.886	0.278	0.801	0.274	0.825	0.260	0.966	0.233
None removed + IMU	0.640	0.276	0.602	0.279	0.600	0.263	0.583	0.267	0.614	**0.248**	0.652	0.199
Vehicle removed+ IMU	0.640	0.276	0.594	0.277	0.618	0.267	0.601	0.270	0.610	0.250	0.643	0.199
Only building + IMU	**0.638**	**0.276**	**0.584**	**0.272**	0.605	**0.261**	**0.541**	**0.258**	**0.592**	0.249	**0.599**	**0.191**

segmentation, we use a combination of Focal loss [11] and Dice loss [12], both with the same weight. The network of the task mainly follows the U-Net [13] architecture. We train our model using the AdamW [14] optimizer with a learning rate of 0.0001, the weight decay of 1×10^{-6}, and the momentum of 0.9.

III. EXPERIMENTAL RESULTS

we use CFEAR-3 [1] with 10 keyframes as the baseline. As an ablation study, we conduct comparative experiments with different semantic classes removed: 1). none removed; 2). vehicle removed; and 3). only building (both the vehicle and vegetation are removed), in combination with and without IMU. As shown in Table I, in most sequences (5 out of 6), the method with both vehicle and vegetation removed and with IMU achieves the best performance in terms of translation errors. The major improvement is attributed to two reasons: 1). the removal of moving cars and trees facilitates more stable feature registration; and 2). the introduced gyroscope rotation information provides a good motion prior during sharp turns. Moreover, in all sequences, the odometry accuracy achieved using only our semantic data surpasses the results obtained by solely incorporating IMU data. This further demonstrates the reliability and effectiveness of our approach.

In the 2024 Radar in Robotics competition, We use the method that achieved the best results in Table I. The evaluation reaches as low as 0.51% drift in the test set with ground truth held out.

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 of 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, IEEE, 2019.

[3] R. Weston, O. P. Jones, and I. Posner, "There and back again: Learning to simulate radar data for real-world applications," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 12809–12816, IEEE, 2021.

[4] H. Yin, Y. Wang, L. Tang, and R. Xiong, "Radar-on-lidar: metric radar localization on prior lidar maps," in *2020 IEEE International Conference on Real-time Computing and Robotics (RCAR)*, pp. 1–7, IEEE, 2020.

[5] 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.

[6] R. F. Salas-Moreno, R. A. Newcombe, H. Strasdat, P. H. Kelly, and A. J. Davison, "Slam++: Simultaneous localisation and mapping at the level of objects," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1352–1359, 2013.

[7] P. Roth, J. Nubert, F. Yang, M. Mittal, and M. Hutter, "Viplanner: Visual semantic imperative learning for local navigation," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5243–5249, IEEE, 2024.

[8] K. Burnett, D. J. Yoon, Y. Wu, A. Z. Li, H. Zhang, S. Lu, J. Qian, W.-K. Tseng, A. Lambert, K. Y. Leung, *et al.*, "Boreas: A multi-season autonomous driving dataset," *The International Journal of Robotics Research*, vol. 42, no. 1-2, pp. 33–42, 2023.

[9] D. Barnes, M. Gadd, P. Murcutt, P. Newman, and I. Posner, "The oxford radar robotcar dataset: A radar extension to the oxford robotcar dataset," in *2020 IEEE international conference on robotics and automation (ICRA)*, pp. 6433–6438, IEEE, 2020.

[10] S. Lee, H. Lim, and H. Myung, "Patchwork++: Fast and robust ground segmentation solving partial under-segmentation using 3d point cloud," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 13276–13283, IEEE, 2022.

[11] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318–327, 2020.

[12] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*, pp. 565–571, Ieee, 2016.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pp. 234–241, Springer, 2015.

[14] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *International Conference on Learning Representations*, 2019.