Intelligent Integrated Multimodal Sensing and Communications for 6G

Fanyi Meng
fanyimeng@link.cuhk.edu.cn

June 14, 2024

1 Introduction

With the continuous advancement of wireless communication networks, B5G/6G networks also present new development trends. Among them, integrating sensing and communication (ISAC) is envisioned as a key pillar for 6G systems and is attracting increasing interest from both academia and industry [1]. This technology aims to significantly enhance spectral and energy efficiencies while reducing hardware and signaling costs by integrating radar sensing and communication into a single system [2].

However, much of the current work in ISAC remains focused on traditional optimization methods such as signal processing and resource allocation. While these efforts have demonstrated ISAC's potential, the practical application of this technology still presents significant challenges. First, despite the similarities between communication and sensing in terms of signal processing and waveform design, their underlying purposes differ, leading to conflicting requirements. Specifically, sensing signals tend to be deterministic, while communication signals are typically stochastic, so it is nearly impossible to provide sensing services without any loss of communication performance [3]. Second, many ISAC underlying theories rely on mathematical assumptions about sensing channels. In complex RF environments, ISAC signals inevitably experience significant interference, which can adversely affect sensing results. Therefore, in practical scenarios, compared to widely used sensors such as LiDAR, millimeter-wave radar, and cameras, the sensing capabilities and advantages of ISAC are still not well-proven.

This research proposal is dedicated to enhancing the integration of multimodal sensing and communication. As mentioned in [4, 5], multimodal data can provide detailed environmental information which is related to communication performance. It has been widely studied that sensors can be used to help with many problems in communication such as beam prediction, blockage prediction, and channel prediction. As sensors provide rich multipath channel information, the transmitter and receiver can avoid complex channel estimation and pilot overhead, showcasing the potential of using sensors to enhance communication performance.

2 Literature Review

Numerous studies have leveraged various sensors to enhance communication systems. This includes the use of cameras, radar, GPS, IMU, and other sensors. Here we list cameras and radar/LiDAR as representative examples. This is primarily because of their extensive application, especially within the realm of autonomous driving. Additionally, these two modalities offer rich environmental data that can substantially augment communication capabilities.

Furthermore, we will examine the representation of currently available datasets to provide a comprehensive understanding of the field.

Vision-aided communication As one of the most common sensors, the camera has been widely deployed in various mobile devices due to its universality, low cost, and high resolution. Cameras provide rich visual information, capturing detailed images and videos of the environment. This visual data is invaluable for a wide range of applications, from object detection and recognition to scene understanding. In [6], considering segmenting different vehicles in pictures is relatively easy, the authors designed a vision-aided communication scheme to help the base station efficiently achieve user matching and the corresponding resource allocation. In [7], the authors used PSPnet to extract multiple semantic features in the image, such as "building", "location", etc., and used the above-extracted features to assist in beam prediction and blockage prediction. In [8], the authors collected a dataset in the real world and validated the camera's effectiveness. A millimeter-wave TX was placed on the vehicle, while a stationary millimeter-wave RX and a camera were placed at the roadside. They trained a machine learning model where the input is the photo taken and the output is the strongest or top 5 beam indexes. In [9], the camera was used to estimate the states of the UAVs, which would help the BS to quickly build initial access to multiple UAVs.

Radar/LiDAR-aided communication Compared to cameras, LiDAR and radar are also widely used because they can provide distance information that vision alone cannot offer. Their sensing results, point clouds, reflect the signals of electromagnetic waves scattered back by environmental objects. These point clouds contain information about the positions, velocities, and RCS (Radar Cross Section) of environmental objects, making them valuable references for communication. In [10], a distributed architecture that utilized LiDAR on vehicles for line-of-sight detection and beam selection was proposed, confirming the potential of LiDAR to reduce beam training overhead. In [11], LiDAR information was analyzed to classify dynamic and static scatterers. They examined the distributions of parameters such as distance, angle, and power related to these scatterers. Based on this analysis, they achieved channel modeling informed by these parameters. In [12], they used vehicle-mounted millimeter-wave radar to estimate the vehicle's motion state and yaw and employed an Extended Kalman Filter (EKF) algorithm to fuse pilot information, achieving robust parameter estimation with just one pilot symbol, thereby reducing beam tracking overhead. In [13], similar to [8], they collected a dataset of radar and communication and then trained a machine learning model to predict the beam index.

Related Datasets As mentioned before, different types of sensing data significantly improve communication performance. Many of the methods leveraging this data are based on machine learning, necessitating the use of widely recognized datasets. Similar to autonomous driving, current datasets are primarily categorized into real-world data and simulated data. For instance, Viwi [14] established a data-generating framework that simulates wireless data using REMCOM and combines it with visual data from the same scenes using Blender. Similarly, M^3SC [15] simulates camera and LiDAR data based on AirSim and wireless data using REMCOM under different weather conditions. FLASH [16] collected multimodal sensor data and mmWave radio data via an autonomous vehicle. Additionally, Deepsense [17] provides a large-scale, real-world dataset comprising co-existing and synchronized multimodal sensing and communication data from over 40 deployment scenarios, including vehicle-to-infrastructure, vehicle-to-vehicle, reconfigurable intelligent surfaces, pedestrians, and drone communication.

3 Methodology

Although extensive work has already been discussed in this direction, we believe there are two areas where improvements can be made.

1. **Enhancement of datasets and simulators:** The current datasets include a variety of scenarios and sensor data. However, in real-world measurements, it is challenging to extract detailed Channel State Information (CSI) due to hardware noise and other factors, resulting in data that is often suitable only for specific tasks. While simulated data can provide channel information, the enormous computational load of ray tracing in communication makes it difficult to perform large-scale simulations across multiple scenarios. Sionna [18], a link-level simulator produced by Nvidia, uses TensorFlow as its backend to support machine learning. One of its toolboxes, Sionna RT [19], enables ray tracing on GPUs, significantly reducing the computation time and resources required for CSI simulation. Additionally, the scenes in Sionna RT are imported through Blender, making it compatible with other perception simulation software such as Airsim and Blensor. This integration makes it possible to build a multi-modal simulator and quickly validate machine learning algorithm performance, providing a valuable opportunity for advancement in this field.

2. **Improvement of the interpretability of algorithms:** Currently, many works use sensor data as input and communication parameters as labels for training. This approach is partly because machine learning can deeply capture the underlying relationships between environmental factors and communication, and partly because feature extraction in some sensor data requires machine learning. However, the interpretability and generalization of such methods still need enhancement. In some scenarios, this approach is highly effective. For instance, in beam prediction, the strongest beam index typically corresponds to the direct path between the transmitter and receiver in LOS scenarios. However, for NLOS paths and conditions like low light, the interpretability of machine learning results remains unclear, making it challenging to ensure that trained models are applicable in different scenarios. Referencing works like NeRF [20] and NeRF2 [21], which consider optical and electromagnetic field propagation models respectively, simple neural networks can achieve excellent predictive performance. Therefore, designing neural networks that integrate prior knowledge of sensors and wireless communication can significantly enhance algorithm performance and interpretability.

4 Tools and datasets used

As mentioned before, we plan to use Sionna and Airsim to build a simulator that includes three components:

1. **Scene Import:** We will use OpenStreetMap to import existing 3D maps from the internet, bringing real-world physical environments into the simulation. Blender models will be used to simulate the movement of vehicles, drones, and other objects.

2. **Communication Simulation:** Using Sionna RT, we will simulate communication in the previously edited environments. With ray-tracing obtained CSI, we can extract any required parameters, such as beam index and occlusions. We have successfully deployed this part of the code and extended it to OFDM RF-ISAC.

3. **Multi-Modal Sensing Simulation:** The same environment will be imported into Airsim, where we will use Airsim's toolbox to simulate cameras, LiDAR, and other sensors.

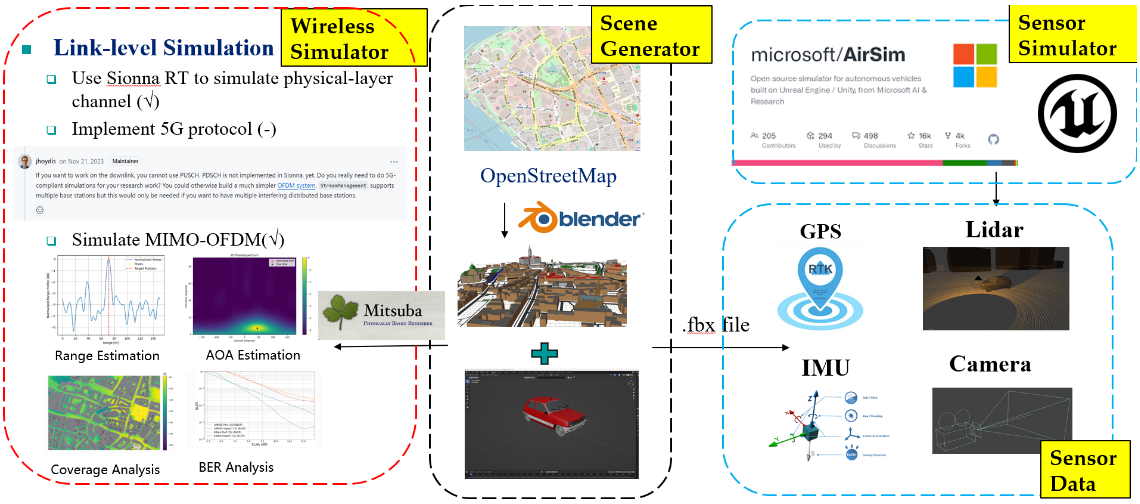

Figure 1: Workflow of the Simulator

This part has been initially deployed, but further development is needed for actual alignment and validation.

References

[1] Fan Liu, Yuanhao Cui, Christos Masouros, Jie Xu, Tony Xiao Han, Yonina C. Eldar, and Stefano Buzzi. Integrated sensing and communications: Toward dual-functional wireless networks for 6g and beyond. *IEEE Journal on Selected Areas in Communications*, 40(6):1728–1767, 2022.

[2] Zhenyao He, Wei Xu, Hong Shen, Derrick Wing Kwan Ng, Yonina C. Eldar, and Xiaohu You. Full-duplex communication for isac: Joint beamforming and power optimization. *IEEE Journal on Selected Areas in Communications*, 41(9):2920–2936, 2023.

[3] Yifeng Xiong, Fan Liu, Yuanhao Cui, Weijie Yuan, Tony Xiao Han, and Giuseppe Caire. On the fundamental tradeoff of integrated sensing and communications under gaussian channels. *IEEE Transactions on Information Theory*, 69(9):5723–5751, 2023.

[4] Xiang Cheng, Haotian Zhang, Jianan Zhang, Shijian Gao, Sijiang Li, Ziwei Huang, Lu Bai, Zonghui Yang, Xinhu Zheng, and Liuqing Yang. Intelligent multi-modal sensing-communication integration: Synesthesia of machines. *IEEE Communications Surveys & Tutorials*, 26(1):258–301, 2024.

[5] Umut Demirhan and Ahmed Alkhateeb. Integrated sensing and communication for 6g: Ten key machine learning roles. *IEEE Communications Magazine*, 61(5):113–119, 2023.

[6] Weihua Xu, Feifei Gao, Yong Zhang, Chengkang Pan, and Guangyi Liu. Multi-user matching and resource allocation in vision aided communications. *IEEE Transactions on Communications*, 71(8):4528–4543, 2023.

[7] Yuwen Yang, Feifei Gao, Xiaoming Tao, Guangyi Liu, and Chengkang Pan. Environment semantics aided wireless communications: A case study of mmwave beam prediction and blockage prediction. *IEEE Journal on Selected Areas in Communications*, 41(7):2025–2040, 2023.

[8] Gouranga Charan, Muhammad Alrabeiah, Tawfik Osman, and Ahmed Alkhateeb. Camera based mmwave beam prediction: Towards multi-candidate real-world scenarios. *arXiv preprint arXiv:2308.06868*, 2023.

[9] Cui Yanpeng, Zhang Qixun, Feng Zhiyong, Qin Wen, Zhou Ying, Wei Zhiqing, and Zhang Ping. Sensing-assisted accurate and fast beam management for cellular-connected mmwave uav network. *China Communications*, pages 1–19, 2024.

[10] Marcus Dias, Aldebaro Klautau, Nuria González-Prelcic, and Robert W. Heath. Position and lidar-aided mmwave beam selection using deep learning. In *2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 1–5, 2019.

[11] Ziwei Huang, Lu Bai, Mingran Sun, and Xiang Cheng. A lidar-aided channel model for vehicular intelligent sensing-communication integration. *arXiv preprint arXiv:2403.14185*, 2024.

[12] Cen Liu, Guangxu Zhu, Fan Liu, Yuanwei Liu, and Kaibin Huang. Successive pose estimation and beam tracking for mmwave vehicular communication systems. In *2023 IEEE Globecom Workshops (GC Wkshps)*, pages 13–19. IEEE, 2023.

[13] Umut Demirhan and Ahmed Alkhateeb. Radar aided 6g beam prediction: Deep learning algorithms and real-world demonstration. In *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 2655–2660, 2022.

[14] Muhammad Alrabeiah, Andrew Hredzak, Zhenhao Liu, and Ahmed Alkhateeb. Viwi: A deep learning dataset framework for vision-aided wireless communications. In *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, pages 1–5, 2020.

[15] Xiang Cheng, Ziwei Huang, Lu Bai, Haotian Zhang, Mingran Sun, Boxun Liu, Sijiang Li, Jianan Zhang, and Minson Lee. M3sc: A generic dataset for mixed multi-modal (mmm) sensing and communication integration. *China Communications*, 20(11):13–29, 2023.

[16] Batool Salehi, Jerry Gu, Debashri Roy, and Kaushik Chowdhury. Flash: Federated learning for automated selection of high-band mmwave sectors. In *IEEE INFOCOM 2022 - IEEE Conference on Computer Communications*, pages 1719–1728, 2022.

[17] Ahmed Alkhateeb, Gouranga Charan, Tawfik Osman, Andrew Hredzak, Joao Morais, Umut Demirhan, and Nikhil Srinivas. Deepsense 6g: A large-scale real-world multi-modal sensing and communication dataset. *IEEE Communications Magazine*, 61(9):122–128, 2023.

[18] Jakob Hoydis, Sebastian Cammerer, Fayçal Ait Aoudia, Avinash Vem, Nikolaus Binder, Guillermo Marcus, and Alexander Keller. Sionna: An open-source library for next-generation physical layer research. *arXiv preprint*, Mar. 2022.

[19] Jakob Hoydis, Faycal Ait Aoudia, Sebastian Cammerer, Merlin Nimier-David, Nikolaus Binder, Guillermo Marcus, and Alexander Keller. Sionna rt: Differentiable ray tracing for radio propagation modeling. In *2023 IEEE Globecom Workshops (GC Wkshps)*, pages 317–321, 2023.

[20] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020.

[21] Xiaopeng Zhao, Zhenlin An, Qingrui Pan, and Lei Yang. Nerf2: Neural radio-frequency radiance fields. In *Proc. of ACM MobiCom '23*, pages 1–15, 2023.