A Multi-Modal Simulator for Aerial Communication with Applications to Beam Search

Jinho Kwon, Jihyeok Jung, Jeongwon Jeon, and Song Noh‡
Department of Information and Telecommunication Engineering
Incheon National University
Incheon, Republic of Korea
{jinhokwon, jihyeok.jung, jeongwon.jeon, songnoh}@inu.ac.kr

***Abstract*—Autonomous beam alignment presents a fundamental challenge for emerging drone-based aerial networks. While integrated sensing and communication offers a solution using heterogeneous onboard sensors, its development requires datasets that integrate multi-modal sensor information, such as the inertial measurement unit (IMU) and camera. To address this data requirement, we develop a multi-sensor simulator based on the Unity 6 engine. The simulator produces camera images, IMU-characterized attitude data, and ground truth vectors. A case study of autonomous beam search is presented to demonstrate a practical application of the generated data.**

I. Introduction

Aerial networks employing drones are emerging as a key technology in next-generation mobile communications, where maintaining a stable link through continuous beam alignment remains a fundamental challenge [1]. Integrated sensing and communication (ISAC) can assist autonomous beam search using onboard sensors such as cameras and inertial measurement units (IMUs), without requiring additional radio resources [2]. However, mathematical modeling of such heterogeneous sensors is still immature, which highlights the need for reliable datasets to support algorithm development.

One approach is to design simulator platforms for sensor-aided communication that can adapt to diverse scenarios. Most existing simulators, however, focus on autonomous vehicles and often neglect detailed three-axis rotational dynamics represented by Euler angles (Pitch, Yaw, Roll) or by quaternions, which are essential in drone applications [3]. Another approach is to obtain datasets from experimental setups [4]. However, these datasets often have coarse resolution, such as quantized beam directions on discrete grids, due to the high cost of experiments.

To address these challenges, we develop a new simulator for aerial sensor-aided communication. The simulator generates datasets that incorporate monocular camera images, IMU-based attitude, and ground-truth locations. As a case study, we demonstrate its use for autonomous beam search with multiple sensors mounted on a drone.

‡ Corresponding author.

II. Simulator Design for Sensor-aided Communication

This section introduces the multi-modal simulator framework and demonstrates its application to beam search.

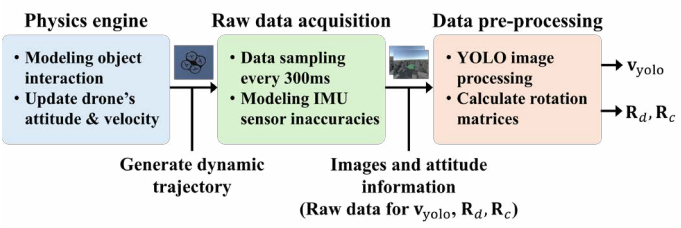

Fig. 1. Block diagram of dataset generation

A. Framework for Camera-IMU Dataset Generation

The proposed simulator consists of Unity's integrated physics engine, raw data acquisition, and data pre-processing, as shown in Fig. 1.

The physics engine models object interactions including physical disturbances, continuously updating the drone's attitude and velocity to generate a dynamic trajectory. The raw data acquisition block collects images and attitude information at discrete sampling intervals (e.g., 300 ms) from the camera and IMU. The camera's attitude is assumed to be known, while IMU errors are incorporated by adding zero-mean Gaussian noise to each axis, with statistical parameters derived from prior studies [5], [6]. The data pre-processing block computes YOLO bounding box coordinates and rotation matrices, which serve as inputs for subsequent beam search.

B. Application to Beam Search

As a case study, the developed simulator is applied to the beam search problem with two benchmark algorithms: a geometric-based method and its refined version using a fully connected neural network (FNN).

The geometric-based approach estimates the drone-to-base station (BS) direction vector from YOLO detections and rotation matrices according to

$$\hat{\mathbf{v}} = \mathbf{R}_d \mathbf{R}_c \mathbf{v}_{\text{yolo}}, \tag{1}$$

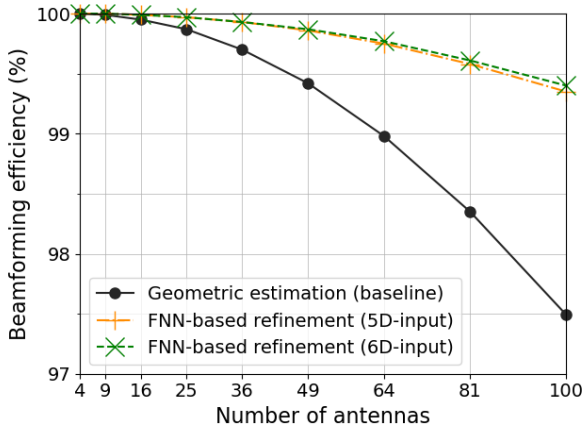

Fig. 2. Beamforming efficiency with different FNN inputs.

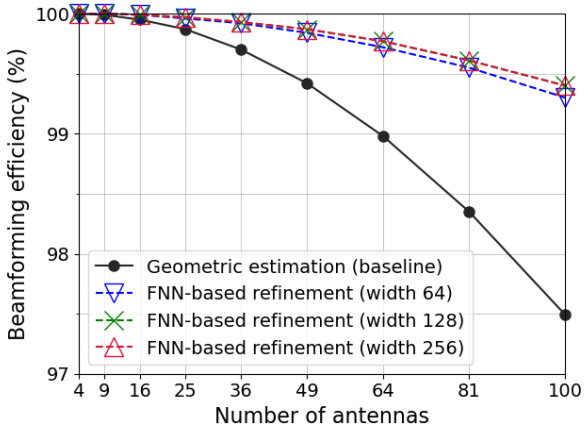

Fig. 3. Beamforming efficiency with varying FNN model widths.

where $\hat{\mathbf{v}}$ denotes the estimated direction vector in the global coordinate system; $\mathbf{R}_d$ and $\mathbf{R}_c$ are rotation matrices from the drone and camera attitudes [7]; and $\mathbf{v}_{\text{yolo}}$ is the unit direction vector to the BS in the camera's local coordinate system, obtained from the YOLO bounding box coordinates.

Since $\mathbf{R}_d$ and $\mathbf{v}_{\text{yolo}}$ are affected by noisy attitude data and estimation error [8], we also consider a refined approach that leverages an FNN [9]. Two input configurations are evaluated: (i) a 5D input combining $\hat{\mathbf{v}}$ with raw YOLO bounding box coordinates and (ii) a 6D input combining $\hat{\mathbf{v}}$ with $\mathbf{v}_{\text{yolo}}$.

III. Numerical Results

The proposed framework is evaluated by comparing a baseline geometric estimation against the FNN-based refinement model, with both its 5D and 6D input configurations. All FNNs have a depth of 4 layers. Performance is measured by the beamforming efficiency versus the number of drone's uniform planar array (UPA) antennas, with the beamforming efficiency representing the achieved gain as a percentage for the maximum gain obtained using the true direction vector.

Fig. 2 compares the geometric baseline against two FNN-based refinement models using a width of 128. Both FNNs significantly outperform the baseline and demonstrate the ability to correct errors inherent in $\mathbf{R}_d$ and $\mathbf{v}_{\text{yolo}}$. Furthermore, the higher performance of the 6D input configuration over the 5D input indicates that the pre-processed direction vector $\mathbf{v}_{\text{yolo}}$ is a more effective refinement feature than raw bounding box coordinates.

Fig. 3 shows the impact of FNN width using the 6D input configuration. Performance improves when increasing the model width from 64 to 128. However, a further increase in width to 256 yields only marginal improvement. This suggests a width of 128 offers a sufficient model capacity under our experimental conditions.

IV. Conclusions

In this paper, we developed a simulator for aerial sensor-aided communication systems. Its effectiveness was demonstrated through sensor-aided beam search, evaluated against two benchmark algorithms. Future work will enhance the simulator to support a wider range of scenarios and operating conditions.

Acknowledgment

This work was supported in part by Electronics and Telecommunications Research Institute (ETRI) grant funded by ICT R&D program of IITP (No. 2018-0-00218, Speciality Laboratory for Wireless Backhaul Communications based on Very High Frequency) and in part by the Institute of Information & Communications Technology Planning & Evaluation(IITP)-ITRC(Information Technology Research Center) grant funded by the Korea government(MSIT)(IITP-2025-RS-2023-00259061).

References

[1] Z. Xiao, L. Zhu *et al.*, "A survey of millimeter-wave beamforming enabled UAV communications and networking," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 1, pp. 557 – 610, 1st Quart. 2022.

[2] X. Cheng, H. Zhang *et al.*, "Intellignet multi-modal sensing-communication integration: synesthesia of machines," *IEEE Commun. Surveys Tuts.*, vol. 26, no. 1, pp. 258 – 301, 1st Quart. 2024.

[3] M. Alrabeiah, A. Hredzak *et al.*, "ViWi: A deep learning dataset framework for vision-aided wireless communications," in *Proc. IEEE Veh. Technol. Conf.*, Antwerp, Belgium, May. 2020, pp. 1 – 5.

[4] A. Alkhateeb, G. Charan *et al.*, "Deepsense 6G: A large-scale real-world multi-modal sensing and communication dataset," *IEEE Commun. Mag.*, vol. 61, no. 9, pp. 122 – 128, Sep. 2023.

[5] R. Munguia and A. Grau, "A practical method for implementing an attitude and heading reference system," *Int. J. Adv. Robot. Syst.*, vol. 11, no. 4, p. 62, 2014.

[6] Y. C. Lai, S. S. Jan *et al.*, "Development of a low-cost attitude and heading reference system using a three-axis rotating platform," *Sensors*, vol. 10, no. 4, pp. 2472 – 2491, Apr. 2010.

[7] H. A. Jlailaty and M. M. Mansour, "Efficient attitude estimators: A tutorial and survey," *J. Signal Process. Syst.*, vol. 94, no. 11, pp. 1309 – 1343, Nov. 2022.

[8] X. Ma, Y. Zhang *et al.*, "Delving into localization errors for monocular 3D object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, Jun. 2021, pp. 4721 – 4730.

[9] K. Seo, J. Lee *et al.*, "Deep learning-based direction finding in the presence of direction-dependent mutual coupling," *ICT Express*, vol. 9, no. 4, pp. 670 – 676, Aug. 2023.