Towards Real-World 6G Drone Communication: Position and Camera Aided Beam Prediction

Gouranga Charan, Andrew Hredzak, Christian Stoddard, Benjamin Berrey, Madhav Seth, Hector Nunez, Ahmed Alkhateeb
School of Electrical, Computer, and Energy Engineering, Arizona State University
Emails: {gcharan, ahredzak, cstodda2, bberrey, mseth2, hanunez, alkhateeb}@asu.edu

***Abstract*—Millimeter-wave (mmWave) and terahertz (THz) communication systems typically deploy large antenna arrays to guarantee sufficient receive signal power. The beam training overhead associated with these arrays, however, make it hard for these systems to support highly-mobile applications such as drone communication. To overcome this challenge, this paper proposes a machine learning based approach that leverages additional sensory data, such as visual and positional data, for fast and accurate mmWave/THz beam prediction. The developed framework is evaluated on a real-world multi-modal mmWave drone communication dataset comprising co-existing camera, practical GPS, and mmWave beam training data. The proposed sensing-aided solution achieves a top-1 beam prediction accuracy of 86.32% and close to 100% top-3 and top-5 accuracies, while considerably reducing the beam training overhead. This highlights a promising solution for enabling highly-mobile 6G drone communications.**

***Index Terms*—Millimeter wave, drone, sensing, deep learning, computer vision, position, camera, beam selection.**

I. Introduction

Future wireless communication systems in 5G-advanced, 6G, and beyond will need to reliably support higly-mobile devices such as drones and autonomous vehicles. Drones (and unmanned aerial vehicles (UAVs)) [1] are envisioned to form the basic building block of next-generation aerial networks and are the key to enable futuristic applications such as extending the coverage of mmWave/sub-THz wireless networks, supporting latency-critical applications, and enhancing the capabilties of security monitoring systems. In order to satisfy the high data rate requirements of these novel applications, the drones will need to be equipped with mmWave/THz [2] transceivers and deploy large antenna arrays. Carefully adjusting the narrow beams of these arrays at both the transmitters and receivers is essential to guarantee sufficient receive SNR. Adjusting these narrow beams, however, is typically associated with large training overhead which scales with the number of antennas. Furthermore, the three dimensional motion along with the highly mobile nature of the drone necessitates frequent update to the optimal beam index, which further increases the beam training overhead. This motivates looking for new approaches to overcome the challenges and enable highly-mobile mmWave/THz drone communication.

In recent years, several solutions [3]–[8] have been developed to overcome the beamforming/channel estimation overhead. Initial approaches focused on three main directions: (i) Beam training with adaptive beam codebook [3], [4], (ii) compressive channel estimation by leveraging channel sparsity [4], and (iii) designing beam tracking solutions [5]. In beam training, the optimal beam at the transmitter and receiver is obtained using exhaustive or adaptive beam training, incurring a large beam-training overhead and is not suitable for highly-mobile multi-user scenarios [3], [4]. By leveraging the inherent sparsity of mmWave channels, [4] formulates the mmWave channel estimation as a sparse reconstruction problem. Although the compressive channel estimation techniques help in reducing the beam training overhead, they can typically save only one order of magnitude in the training overhead. Further, for these solutions, the training overhead scales with the number of antennas, reducing the impact for systems with large antenna arrays. Next, [5] proposed an extended Kalman filter-based (EKF) channel tracking solution to maintain the communication link between the basestation and mobile user. Although such an EKF-based beam tracking approach helps in minimizing the beam training overhead, they can only predict beams for a short future time window and their performance is normally limited in NLOS scenarios. The limitations of these approaches motivate the development of more efficient beam prediction approaches.

Leveraging machine learning (ML) to address the beam prediction task has gained increasing interest in the last few years [6], [9]–[12]. These solution mainly focus on leveraging the additional information to provide awareness about the wireless environment. In [6], the authors propose to utilize the receive wireless signature to predict the optimal beam indices at the basestation. Such a solution, though promising, is limited to a single-user setting. Position information was leveraged in [9], [10] to predict the optimal beam index. Although the solutions can help in reducing the training overhead, relying only on location alone might result in inaccurate predictions due to the inherent errors associated with the GPS data. In [11], [12], we proposed to leverage the visual data (captured by camera) to predict the optimal beam indices. These solutions, however, are based on synthetic data and focused on scenarios with humans, vehicles, or robots as the transmitter, where the users typically move in easy to predict mobility patterns in two dimensions. In general, vehicles or robots tend to travel in the azimuthal plane without any change in the elevation during movement. Drones or UAVs have six degrees of freedom,

978-1-6654-3540-6/22

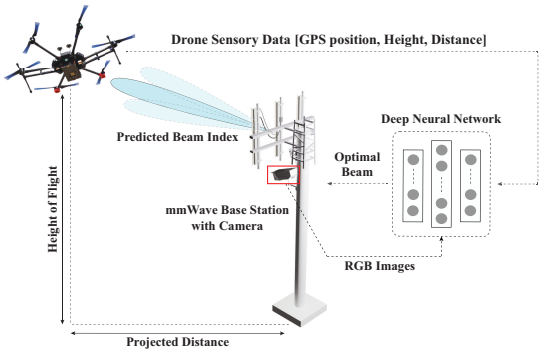


Fig. 1. An illustration of the mmWave basestation serving a drone in a real wireless environment. The basestation utilizes additional sensing data such as RGB images, GPS location of the drone, etc., to predict the optimal beam.

three translation, and three rotation, which further increases the challenge of predicting the optimal beam index. An important question that arises is whether the promising results in [11], [12] can be achieved in reality for mmWave drones?

In this paper, we attempt to answer this important question. In particular, we propose a deep learning-based sensing-aided solution to reduce the beam training overhead in mmWave/THz drone communication. The main contribution of this work can be summarized as follows:

- Formulating the sensing-aided beam prediction problem for mmWave/THz drone communication considering practical visual and communication models.
- Developing a novel deep learning-based solution for mmWave/THz drone beam prediction by utilizing different sensory data such as vision (captured at the basestation) and the position, orientation, height of the drone.
- Providing the first real-world evaluation of sensing-aided drone beam prediction based on our large-scale dataset, DeepSense 6G [13], that consists of co-existing multi-modal sensing and wireless communication data.

Based on the adopted real-world dataset, the developed solution achieves $\approx$ 86% top-1 (and close to 100% top-3) beam prediction accuracy. This highlights the capability of the proposed sensing-aided beam prediction approaches in significantly reducing the beam training overhead.

II. Sensing-Aided Beam Prediction: System Model and Problem Formulation

This work considers a communication system where a mmWave basestation is serving a drone flying at different speeds and heights in a real wireless communication environment. In this section, we first present the adopted wireless communication system model. Then, we formulate the sensing-aided beam prediction problem.

A. System Model

This paper adopts the system model illustrated in Fig. 1, where a basestation, equipped with an M-element uniform linear array (ULA) and an RGB camera, is serving a flying drone. The drone carries a single-antenna transmitter and is equipped with a GPS receiver capable of collecting real-time position information. The adopted communication system employs OFDM transmission with K subcarriers and a cyclic prefix of length D. To serve the mobile user, the basestation is assumed to employ a pre-defined beamforming codebook $\boldsymbol{\mathcal{F}} = \{\mathbf{f}_q\}_{q=1}^{Q}$, where $\mathbf{f}_q \in \mathbb{C}^{M\times 1}$ and Q is the total number of beamforming vectors. In the downlink transmission (from the basestation to the drone), if $\mathbf{h}_k[t] \in \mathbb{C}^{M\times 1}$ denotes the channel between the basestation and the drone at the kth subcarrier and time t, then the received signal at the drone can be written as

$$y_k[t] = \mathbf{h}_k^T[t]\mathbf{f}_q[t]x + v_k[t], \tag{1}$$

where $\mathbf{f} \in \boldsymbol{\mathcal{F}}$ is the optimal beamforming vector at time t and $v_k[t]$ is a noise sample drawn from a complex Gaussian distribution $\mathcal{N}_{\mathbb{C}}(0, \sigma^2)$. The transmitted complex symbol $x \in \mathbb{C}$ need to satisfy the following constraint $\mathbb{E}\left[|x|^2\right] = P$, where P is the average symbol power. The beamforming vector $\mathbf{f}^\star[t] \in \boldsymbol{\mathcal{F}}$ at each time step t is selected to maximize the average receive SNR and is defined as

$$\mathbf{f}^\star[t] = \underset{\mathbf{f}_q[t]\in\mathcal{F}}{\operatorname{argmax}} \frac{1}{K}\sum_{k=1}^{K} \mathsf{SNR}|\mathbf{h}_k^T[t]\mathbf{f}_q[t]|^2, \tag{2}$$

where SNR is the transmit signal-to-noise ratio, $\mathsf{SNR} = \frac{P}{\sigma^2}$.

B. Problem Formulation

Given the system model in Section II-A, if the basestation wants to select an optimal beam $\mathbf{f}^\star[t]$ out of its codebook $\boldsymbol{\mathcal{F}}$ to serve the drone, then it can determine this optimal beam that maximizes the received power based on (2). The optimum beam is computed by either utilizing the explicit channel knowledge which is hard to acquire in mmWave/THz systems or by performing an exhaustive search over the beam codebook, which is typically associated with high beam training overhead. This makes it challenging for mmWave/THz systems to support the highly-mobile drones. In this paper, instead of following the conventional beam training approach, **we propose to predict the optimal beam index for the transmitter by utilizing the sensory data (position or vision) collected by the basestation or the drone.** In particular, this work assumes the availability of the following sensory data at the basestation: (i) the RGB images captured by a camera installed at the basestation, (ii) the GPS positional data collected by the drone and fed back to the basestation, and (iii) the height/distance of the drone in the real wireless environment. Formally, we define $\mathbf{g}[t] \in \mathbb{R}^2$ as the two-dimensional position vector of the transmitter (consisting of the latitude and longitude information) at time step t. And we define $\mathbf{X}[t] \in \mathbb{R}^{W\times H\times C}$ as the corresponding RGB image, captured by a camera installed in the basestation at time t, where W, H, and C are the width, height, and the number of color channels of the image. Further, let $d[t] \in \mathbb{R}^1$ and $v[t] \in \mathbb{R}^1$ denote the height and the distance of the transmitter from the stationary unit at time instance t. The objective of the drone beam prediction task is to find a prediction/mapping function f_Θ that utilizes the available sensory data, $\mathcal{S}[t] =$

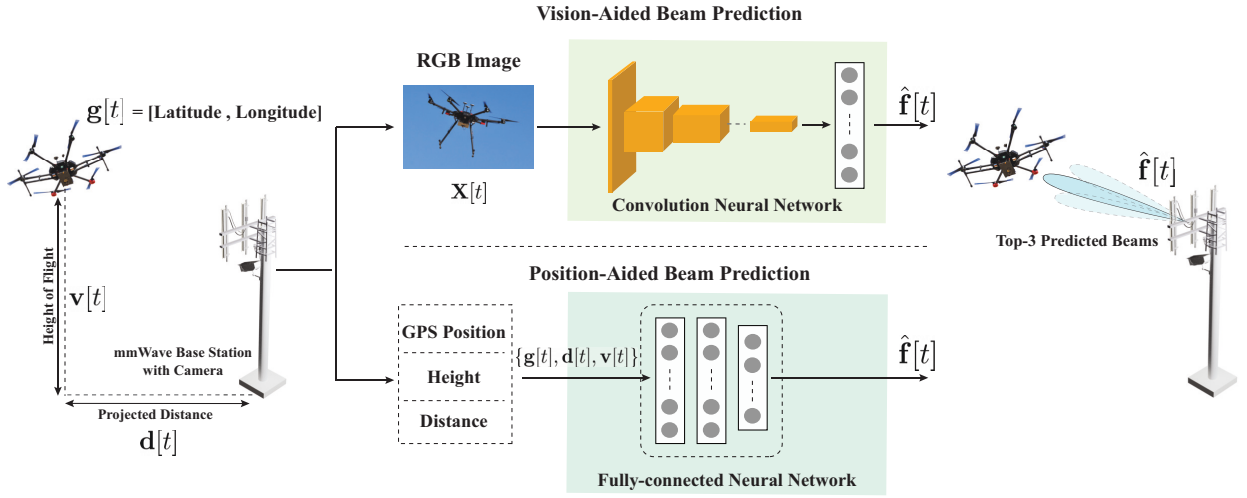


Fig. 2. A block diagram showing the proposed solution for both the vision and position-aided beam prediction task. As shown in the figure, the camera installed at the basestation captures real-time images of the drone in the wireless environment. A CNN is then utilized to predict the optimal beam index. The basestation receives the information for the other three sensing data, which is then provided to a fully-connected neural network to predict the beam.

$\{\mathbf{g}[t], \mathbf{X}[t], d[t], v[t]\}$ to predict (estimate) the optimal beam index $\hat{\mathbf{f}}[t] \in \boldsymbol{\mathcal{F}}$ with high fidelity. The mapping function can be formally expressed as

$$f_\Theta : \mathcal{S}[t] \rightarrow \hat{\mathbf{f}}[t]. \tag{3}$$

In this work, we develop a machine learning model to learn this prediction function f_Θ. Let $\mathcal{D} = \{(\mathcal{S}_u, \mathbf{f}_u^*)\}_{u=1}^U$ represent the available dataset consisting of sensing data-beam pairs is collected from the real wireless environment, where U is the total number of samples in the dataset. Then, the objective is to maximize the number of correct predictions over all the sample in $\mathcal{D}$. This can be formally written as

$$f_{\Theta^\star}^\star = \underset{f_\Theta}{\operatorname{argmax}} \prod_{u=1}^{U} \mathbb{P}\left(\hat{\mathbf{f}}_u = \mathbf{f}_u^\star | \mathcal{S}_u\right), \tag{4}$$

where the product in (4) is due to the implicit assumption that the samples in the dataset $\mathcal{D}$ are drawn from an independent and identically distribution (i.i.d.). The prediction function is parameterized by a set Θ representing the model parameters and learned from the dataset $\mathcal{D}$ of labeled data samples. Next, we present our proposed machine learning model for sensing-aided mmWave/THz drone beam prediction.

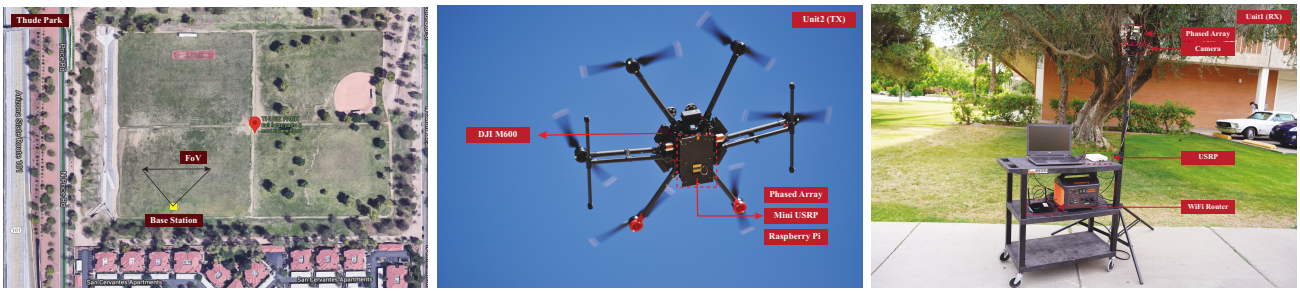
III. Sensing-Aided Beam Prediction: A Deep Learning Solution

In this section, we present an in-depth overview of the proposed beam prediction solution. First, we present the key idea in Section III-A and then explain the details of our proposed solution in Section III-B.

A. Sensing-Aided Drone Beam Prediction: Key Idea

The mmWave/THz communication systems require large antenna arrays and use narrow directive beams to guarantee sufficient signal power gain. This is primarily to overcome the severe path loss associated with the high-frequency signals. Selecting the optimal beams in these systems is typically associated with large beam training overhead, which becomes more challenging in high-mobility dynamic wireless environments with moving transmitters, reflectors, and scatters. The highly mobile nature with very high flying speeds of the drones and the added capability of hovering or traveling in a three-dimensional space further increases the challenges faced in the mmWave drone communication system. Instead of relying on conventional beam training, this work selects the optimal beam index by utilizing additional sensory data. In this paper, the task of selecting the optimal beam index from a pre-defined codebook, $\boldsymbol{\mathcal{F}}$, at any coherence time, is defined as the **beam prediction** task.

High-frequency systems suffer from large path-loss, which makes line-of-sight (LOS) a preferable setting. This dependence of both high-frequency communication on LOS operation forms the building block of sensing-aided solution. In general, the beamforming vectors provide directional information that summarizes dominant signal direction for well-calibrated antenna arrays. The beam vectors divide the scene (spatial dimensions) into multiple (possible overlapping) sectors, where each sector is associated with a particular beam value. Therefore, given a pre-defined codebook, the beam prediction task can be transformed into a classification task, where depending on the user location in the wireless environment, a beam index from the codebook is assigned. With this motivation, we plan to utilize visual and positional data to predict the optimal beam indices. The recent advancements in computer vision and object detection have provided the capability to accurately detect the different objects and extract the user's relative position in the visual scene. Similarly, for any outdoor location, advanced positioning systems such as GPS can be used to accurately (with some error margin) locate a user in the scene. At every time step, the basestation captures a visual image of its environment and receives the



(a) (b) (c)

Fig. 3. This figure presents the overview of the DeepSense 6G testbed and the location used in this scenario. (a) Shows the google map top-view of Thude Park utilized for this data collection. Fig. (b) and (c) present the different components of the drone (acting as the transmitter) and the basestation. In (b), we highlight the mmWave phased array attached to the drone transmitting signals to the basestation on the 60 GHz band.

sensing data such as GPS location, the height, and distance of the transmitter. Instead of performing beam training at every step, the basestation utilizes machine learning models and the additional sensory data to predict the optimal beamforming vector from a pre-defined codebook.

B. Proposed Solution

In this section, we describe the proposed machine learning-based solutions for sensing-aided beam prediction. First, we present the detailed overview of the proposed position-aided beam prediction task followed by the details of the vision-aided beam prediction task. In Fig. 2, we present the block diagram of the proposed beam prediction ML model.

1) Proposed Position-aided Solution: This sub-task entails the prediction of optimal beam index by leveraging the positional information of the transmitter. A Multi-Layer Perceptron (MLP) network is adopted to perform the position-aided beam prediction task. The inputs to the MLP network are the normalized Latitude and Longitude values. The MLP network is designed to have two hidden layers with 512 hidden units each and an output layer consisting of Q units. ReLU activation function is applied to the output of the hidden layers in order to introduce non-linearity. Since the position-aided beam prediction task is posed as a classification problem as presented in Section II-B, the softmax function is applied to the final output layer.

In order to perform a comparative evaluation of the different sensing modalities, we extend the solution to utilize the other sensing data, such as the height and the distance of the drone from the basestation. For this, along with the normalized GPS data, we provide the normalized height and distance information to the proposed ML model. The rest of the architecture is similar to the solution proposed for the position-alone beam prediction.

2) Proposed Vision-aided Solution: In this subsection, we present our proposed deep learning model for the vision-aided beam prediction task. The objective is to learn the prediction function $f_{\Theta}(\mathbf{X}[t])$ by utilizing only visual data. The ideal choice of the deep learning model for this task, as mentioned above, is the CNNs. The idea is to utilize CNN to perform this classification task, i.e., the model learns to map an image to a beam index. The CNN in the proposed solution needs to meet two essential requirements: (i) an accurate and generalizable classifier for the vision-based classification task and (ii) low computational footprint. The residual neural network (ResNet) [14] has proven to be highly efficient for image classification tasks and, most importantly, addresses the two major requirements mentioned above. For this particular task, a ResNet-50 [14] model has been selected as the primary choice of CNN. However, instead of training the ResNet model from scratch on the single candidate beam prediction dataset, we select an ImageNet2012 pre-trained ResNet model as the initial architecture. The model is further modified by removing the last fully connected layer and replacing it with a layer consisting of Q neurons. The fully connected layer parameters are initialized randomly following a normal distribution with zero mean and unit variance. Unlike conventional transfer learning, the ResNet-50 model is fine-tuned end-to-end in a supervised fashion, using a labeled dataset.

IV. Testbed Description and Development Dataset

In order to evaluate the performance of the proposed sensing-aided mmWave drone prediction solution, we utilize the **DeepSense 6G** [13] dataset. DeepSense 6G is a real-world multi-modal dataset dedicated to sensing-aided wireless communication applications. It contains co-existing multi-modal data such as vision, mmWave wireless communication, GPS data, LiDAR, Radar collected in a real-wireless environment. This section presents a brief overview of the scenario adopted from the DeepSense 6G dataset, followed by the analysis of the final development dataset utilized for the sensing-aided beam prediction study.

A. DeepSense 6G: [Scenario 23]

This study adopts Scenario 23 of the DeepSense 6G dataset specifically designed to study high-frequency wireless communication applications with drones. The hardware testbed and the exact location used for collecting these data are shown in Fig. 3. The DeepSense testbed 4 is utilized for this data collection consisting of a stationary and a mobile unit. The testbed is deployed at the Southwest corner of the park, as shown in Fig. 3(a). The stationary unit $\{$unit1 (RX)$\}$ is equipped with a standard-resolution RGB camera,

TABLE I
DEEPSENSE 6G SCENARIO 23: DEVELOPMENT DATASET

Task	Modality	Number of Samples	
		Training	Test
Sensing-Aided Beam Prediction	GPS Position	8402	3602
	GPS Position + Height	8402	3602
	GPS Position + Height + Distance	8402	3602
	RGB Image	8402	3602

TABLE II
BEAM PREDICTION: DESIGN AND TRAINING HYPER-PARAMETERS

Parameters	Vision	Position/Combined
ML Model	ResNet-50	2-layered MLP
Batch Size	32	32
Learning Rate	1×10^{-4}	1×10^{-2}
Learning Rate Decay	epochs 4, 8 and 12	epochs 20, 40 and 80
LR Reduction Factor	0.1	0.1
Total Training Epochs	20	100

and mmWave Phased array. The stationary unit adopts a 16-element ($M = 16$) 60GHz-band phased array and it receives the transmitted signal using an over-sampled codebook of 64 pre-defined beams ($Q = 64$). The mmWave phased array and the RGB camera is placed on a table at the height of ≈ 1.5 meters from the ground level. Both the camera and the phased array are facing towards the sky, which helps increase the basestation's field-of-view (FoV). In this data collection scenario, the mobile unit $\{$unit2 (TX)$\}$ is the RC drone equipped with a mmWave transmitter, GPS receiver, and inertial measurement units (IMU). The transmitter consists of a quasi-omni antenna constantly transmitting (omnidirectional) at 60 GHz band. In order to increase the diversity of the dataset, the drone is flown at varying heights and distances from the basestation with different speeds of flight. For more information regarding the data collected testbed and setup, please refer to [13].

B. DeepSense 6G: [Development Dataset]

The evaluation of the proposed sensing-aided beam prediction solution requires data collected in a real wireless environment with a drone as the mmWave transmitter. In this work, we utilize the publicly available scenario 23 of the DeepSense 6G dataset. The different data modalities collected are the RGB images, real-time GPS location, distance, height, speed, and orientation of the drone, and a 64×1 vector of mmWave received power. The first step involves downsampling the 64×1 power vector to 32×1 ($Q = 32$) by selecting every alternate sample in the vector. Since the basestation receives the mmWave signal using an oversampled codebook of 64 pre-defined beams, the downsampling does not affect the total area covered by the beams. In order to compute the updated optimal beam index for a particular sample, we select the index of the beam with maximum received power in the downsampled power vector. The final step in the processing pipeline is dividing the dataset into training and test set following a 70-30% split. In Table I, we present the details of the development datasets for the sensing-aided beam prediction task.

Fig. 4. This figure plots the top-k accuracies ($k \in (1, 2, 3, 5)$) for the proposed sensing-aided beam prediction solution. It is observed the vision-aided beam prediction solution outperforms the other approaches.

V. PERFORMANCE EVALUATION

This section studies the performance of the proposed solutions for the sensing-aided beam prediction task. In the first sub-section, we will present the details of the experimental setup. Next, we discuss the performance of the proposed solution for the different sub-tasks presented in Section III.

A. Experimental Setup:

Network Training: In this work, we propose to utilize different sensing data modalities to perform the beam prediction task, i.e., position-alone, position and height combined, position, height and distance combined, and visual data. As presented in Section III, different modality-specific deep learning models are proposed to perform the sensing-aided beam prediction task. For the position-alone and the combined data modalities, we develop 2-layered fully-connected neural networks. For the vision-aided approach, the proposed solution adopts a ResNet-50 model to predict the optimal beam indices. The proposed ML models are trained and validated on the task-specific dataset as presented in Section IV-B. The cross-entropy loss with Adam optimizer is used to train the models. The details of the hyper-parameters used to fine-tune the models are presented in Table II.

Evaluation Metric: The primary metric adopted to evaluate the proposed solution is the top-k accuracy. Note that the top-k accuracy is defined as the percentage of the test samples where the optimal ground-truth beam is within the top-k predicted beams. This work presents the top-1, top-2, top-3, and top-5 accuracies to evaluate the proposed solutions comprehensively.

B. Numerical Results:

With the experimental setup described in Section V-A, in this subsection, we study the beam prediction performance of the proposed solution. In the first set of studies, we evaluate the performance of the proposed solutions from a machine learning perspective, i.e., beam prediction accuracy per approach, number of samples required for training, etc.

Beam Prediction Accuracy Comparison: The UE in this study is a drone, which brings its own set of challenges, such as the six degrees of freedom in motion, the variability of

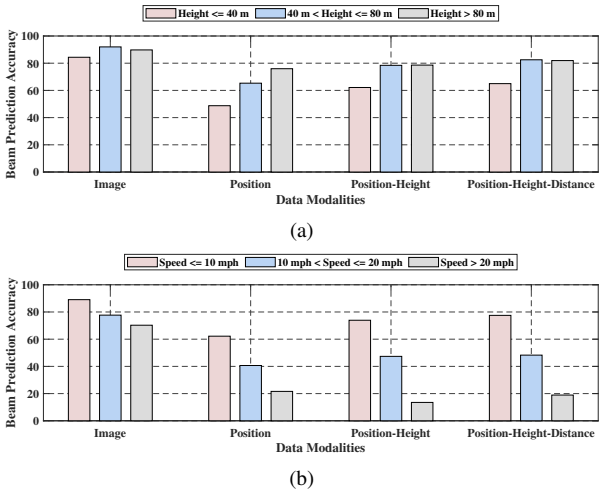


Fig. 5. This figures studies the impact of speed and height on the beam prediction performance of the proposed solutions.

orientation, etc. In Fig. 4 we compare the performance of the different proposed solutions for the mmWave drone beam prediction task. It is observed in Fig. 4 that the position-alone approach achieves only $\approx 59\%$ top-1 accuracy. This is an interesting result as **it highlights that for mmWave communication using drones, position alone might not be sufficient in predicting the optimal beam indices.** The combined modalities achieve an improvement of $\approx 10-14\%$ over the position-alone beam prediction solution. These results highlight the need for additional sensory data such as the height and distance of the drone from the basestation. Images can successfully capture the orientation and location of the object in the visual field. This is reflected in the performance of the vision-aided solution; it achieves a top-1, top-3, and top-5 accuracy of 86.32%, 99.41%, and 99.69%.

Impact of height and speed on beam prediction accuracy: As observed previously, the beam prediction accuracy improved significantly from the position-only solution with additional sensing data such as the height and the distance of the drone. Here, we consider two important data modalities, i.e., (i) height and (ii) speed, of the drone and study the impact of these data modalities on beam prediction accuracy. For both the data modalities, we first divide the test dataset into three sub-groups. For example, for the accuracy analysis based on speed, we divide the dataset into three groups: slow, medium, and fast. Next, we calculate the beam prediction accuracy for each of the sub-groups. In Fig. 5(a) and Fig. 5(b), we present the beam prediction accuracy versus the height and speed of the drone, respectively. Fig. 5(a) highlights an interesting fact that all the four ML-based solution makes the most mistakes in prediction when the drone is flying low, i.e., the height is less than 40 meters. For the speed-based analysis, we observe that the beam prediction performance starts degrading for higher traveling speeds of the drone.

VI. Conclusion

This paper develops a novel approach that leverages sensory data, such as visual and position data, for fast and accurate beam prediction in mmWave/THz drone communication systems. To evaluate the efficacy of the proposed solution, we adopt a real-world multi-modal drone communication scenario from the DeepSense 6G dataset. We perform an in-depth evaluation of different sensory modalities and compare the impact of different sensing data on the beam prediction accuracy. The proposed vision-aided solution achieves top-1 and top-5 accuracies of 86.32% and 99.69%, respectively. This highlights the promising gains of leveraging sensory data to reduce the beam training overhead in mmWave/THz drone communication systems.

VII. Acknowledgment

This work is supported in part by the National Science Foundation under Grant No. 2048021.

References

[1] L. Bariah, L. Mohjazi, S. Muhaidat, P. C. Sofotasios, G. K. Kurt, H. Yanikomeroglu, and O. A. Dobre, "A prospective look: Key enabling technologies, applications and open research topics in 6g networks," 2020.

[2] T. S. Rappaport, Y. Xing, O. Kanhere, S. Ju, A. Madanayake, S. Mandal, A. Alkhateeb, and G. C. Trichopoulos, "Wireless communications and applications above 100 ghz: Opportunities and challenges for 6g and beyond," *IEEE Access*, vol. 7, pp. 78 729–78 757, 2019.

[3] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Multilevel millimeter wave beamforming for wireless backhaul," in *Proc. of 2011 IEEE GLOBECOM Workshops (GC Wkshps)*, Houston, TX, 2011, pp. 253–257.

[4] A. Alkhateeb, O. El Ayach, G. Leus, and R. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[5] S. Jayaprakasam, X. Ma, J. W. Choi, and S. Kim, "Robust beam-tracking for mmwave mobile communications," *IEEE Communications Letters*, vol. 21, no. 12, pp. 2654–2657, 2017.

[6] A. Alkhateeb, S. Alex, P. Varkey, Y. Li, Q. Qu, and D. Tujkovic, "Deep learning coordinated beamforming for highly-mobile millimeter wave systems," *IEEE Access*, vol. 6, pp. 37 328–37 348, 2018.

[7] M. Saquib Khan, Q. Sultan, and Y. Soo Cho, "Position and machine learning-aided beam prediction and selection technique in millimeter-wave cellular system," in *ICTC*, 2020, pp. 603–605.

[8] M. Alrabeiah, A. Hredzak, and A. Alkhateeb, "Millimeter wave base stations with cameras: Vision-aided beam and blockage prediction," in *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, 2020, pp. 1–5.

[9] Y. Wang, M. Narasimha, and R. W. Heath, "Mmwave beam prediction with situational awareness: A machine learning approach," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2018, pp. 1–5.

[10] S. Rezaie, C. N. Manchón, and E. de Carvalho, "Location- and orientation-aided millimeter wave beam selection using deep learning," in *IEEE International Conference on Communications (ICC)*, 2020, pp. 1–6.

[11] G. Charan, T. Osman, A. Hredzak, N. Thawdar, and A. Alkhateeb, "Vision-position multi-modal beam prediction using real millimeter wave datasets," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, 2022, pp. 2727–2731.

[12] G. Charan, M. Alrabeiah, and A. Alkhateeb, "Vision-aided 6G wireless communications: Blockage prediction and proactive handoff," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 10, pp. 10 193–10 208, 2021.

[13] A. Alkhateeb, G. Charan, T. Osman, A. Hredzak, and N. Srinivas, "DeepSense 6G: Large-scale real-world multi-modal sensing and communication datasets," *to be available on arXiv*, 2022. [Online]. Available: https://www.DeepSense6G.net

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.