Vision And Lidar Based Beamtracking

School of Engineering and Applied Science (SEAS), Ahmedabad University
Wireless communication
Kunj Kanzariya - AU2140202

***Abstract*—mmWave transmission or tera-hertz communication requires beamforming for adequate Signal to Noise ratio. It requires a clear Line-of-sight(LOS) for optimal performance. However, knowing the precise location of a receiver to maintain the Line Of Sight is a significant challenge. Current solutions use available vision, lidar, radar, or position sensory information separately to analyze the communication system and to predict the optimal beam index. However, using a single modality of data increases our dependency on a single source of information and hence decreases the preciseness. The proposed machine learning architecture explores the multi-modality of sensory information by using lidar and vision data together. The proposed Long Short Term Memory (LSTM) model is evaluated on the Deepsense 6G scenario 8 dataset, where it achieves an accuracy of 65.78%. The proposed methodology could significantly impact the tera-heartz communication by making it more reliable and efficient.**

***Index Terms*—mmWave, beamtracking, vision, lidar, multi-modality, machine learning.**

I. Introduction

A. Background

Wireless communication relies on the spectrum of electromagnetic waves to transmit information. Cellular networks, ranging from 2G to 5G, utilize a frequency bandwidth of approximately 600MHz to 20GHz. Millimeter Wave (mmWave) represents a specific band in the high-frequency spectrum, ranging from 30GHz to 300GHz. It offers increased network capacity and low latency, enabling the service of more concurrent users and facilitating significantly higher data transfer rates compared to average cellular networks. Working with such high-frequency bandwidth, however, introduces challenges such as high path loss, limited communication range, and dependence on line-of-sight. To overcome these challenges, the Beamforming method is employed. This technique focuses on directing the transmission and reception of electromagnetic waves in specific directions, thereby enhancing the signal strength.While Beamforming proves effective, it becomes even more challenging when the receiver or transmitter is in motion. Scenarios such as Vehicle-to-Everything (V2X) communications and autonomous vehicles, are considered to be the key applications for future wireless communications.Traditional methods to address the challenge of motion involve beam training over a large number of beams using a pre-defined codebook. However, this approach increases beam training overhead, resulting in high latency for predicting the optimal beam. this high latency makes this approach impractical for moving transmitter/receiver system.

B. Motivation

The dependence of mmWave communication on Line Of Sight shows us the importance of the positions of the receiver and transmitter as well as the surrounding environment. In order to make the system reliable, precise information about the location of the receiver/transmitter antennas becomes essential. Sensors such as cameras, Lidar, and radar possess diverse modalities of information about the moving receiver's position, allowing us to analyze the motion of the receiver and predict the optimal beam. Machine Learning frameworks are employed to forecast future beams based on the sensory information provided.

However, relying on a single modality of the data could risk the required preciseness of the position information, resulting in an unreliable and inefficient approach. to solve this, multiple modals of data can be used together. In [1], vision and position data are used together for better analysis of the receiver's motion and prediction of an optimal beam, this approach results in around 75% top-1 accuracy which is significant compared to the baseline and single modality approaches like [2] vision-aided or [3] lidar-aided beam tracking. These results clearly state the improvement in performance using the multimodal data. this motivates me to experiment with the unexplored combination of Lidar and Vision data.

II. Contribution

This study is build upon the fundamental work which is presented in [2]. The base paper used Gated Recurrent Unit Network to analyze the photos of the moving reciever to predict its motion and optimal beam at particular instance. In our study, we extend this approach by adding the following improvements:

- **Multimodal Approach:** We utilize vision(photos) with lidar data, to increase the precision with which our model analyzes the receiver's motion.
- **LSTM Model:** We use Long Short Term Memory machine learning model to analyze the complex dependencies of our problem with provided data.
- **Comparative analysis:** We compared the performance of our model with the existing single modal based approaches i.e [1] vision-aided and [2] Lidar-aided beam tracking.

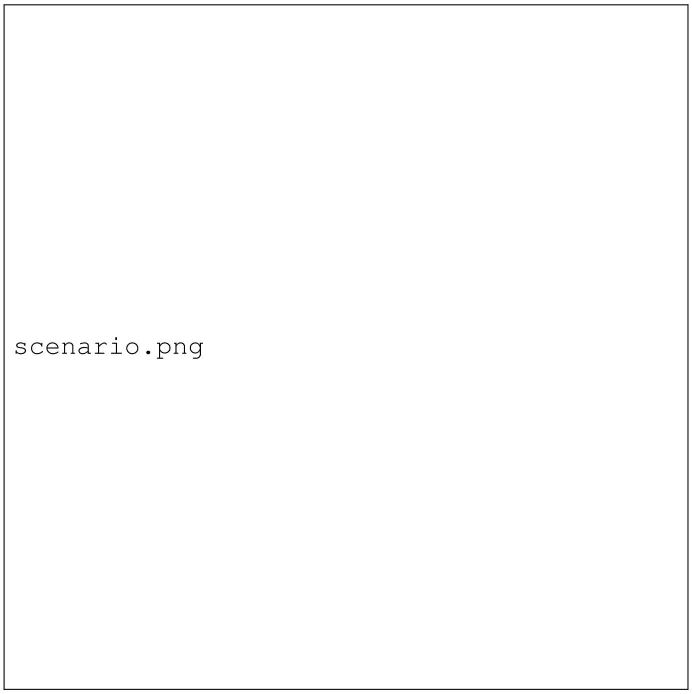

Fig. 1. Experiment Setup

III. System model and Problem formulation

Fig. 1 shows the considered mmWave communication scenario for this study, it includes a stationary Base station equipped with a RGB camera and a Lidar sensor which provides photos and Lidar_scr file respectively. Base Station is also equipped with a predefined codebook $\boldsymbol{f}$ containing 64 beam vectors. Moving receiver contains a single antenna and is moving in front of the Base Station.

A. System Model

Suppose a Complex signal $\boldsymbol{B[t]}$ is transmitted by the Base Station through the channel $\boldsymbol{h}$ towards moving receiver.

$$y[t] = h^H[t]f[t]B[t] + n[t] \tag{1}$$

Where, $\boldsymbol{f[t]}$ and $\boldsymbol{n[t]}$ are the beam forming vector and noise at the receiver's end for the instance $\boldsymbol{t}$ respectively.

B. Problem Formulation

This study focuses on finding optimal beam at particular instance given sensory information about the position of transmitter/receiver and the environment around.

Predicting optimal beams at the base station involves a probability-based approach using sensory data. The task is to find the most probable future beam index, considering information available up to a certain point. The goal is to maximize the likelihood that the predicted beam index matches the actual one, subject to predefined constraints. The equation mentioned below, highlights the data-driven nature of this probabilistic beam prediction task, emphasizing its significance in addressing the challenges of mmWave communication.we can now formulate the above problem as:

$$\begin{aligned} &\max_{p^\wedge[t]} \mathbb{P}\{p[t] = p_\star[t] | O_{t-\xi}\} \\ &\qquad \text{s.t. } p^\wedge[t] \in [1, |F|], \end{aligned} \tag{2}$$

where $p^\wedge[t]$ is the predicted optimal beam index for time step t. $O_{t-\xi}$ is any auxiliary information obtained before time step $t - \xi + 1$ that contains partial information of the optimal beam at time step t and $p^\star[t]$ is the optical beam index.

The optimal beam is defined as the beam vector from the codebook which gives the highest beamforming gain. This can be mathematically modelled as:

$$f^*[t] = \underset{f[t] \in F}{\operatorname{argmax}} |h_t[t] f[t]|^2 \tag{3}$$

C. Multimodal Approach

The analytical analysis of the beamtracking problem gives us the equation suggesting that the probability of finding the optimal beam index depends upon the information given. In equation (2), the $O_{t-\xi}$ defines the sensory information of the t - ξ instances available to us. In [2], images of the mobile receiver is taken as $O_{t-\xi}$, while [3] uses Lidar_data. Multimodality approach uses combination of different modality of data available to us, which will provide more information of the position of the transmitter/receiver and the environment around it. The use of combination of vision and position information which performs significantly higher in [1] justifies this approach. motivated from this, we propose the multimodality LSTM network which works on Vision and Lidar data.

IV. Methodology

We propose a Long short Term Memory machine learning architecture to solve the classification problem of the beamtracking which trains on the Images of the mobile receiver and lidar output. We used real world dataset to train our model and compared it with the single modal architectures.

A. Data Preprocessing

The LSTM model is trained over vision and Lidar data which needs to be processed because of the variable dimension. Vision data includes images of the mobile receiver which is passed through the advanced YOLOv4 machine learning architecture to detect the receiver's coordinates and length-width, hence it is 4×1 array for each time instance. whereas the Lidar is the combination of 216×2 and 460×2 arrays.

V. Experiments

A. Dataset

The LSTM model for beam tracking is trained over Deepspsense 6g scenario-8 dataset which include Lidar, position, vision and wireless sensor at the Base station and single antenna moving receiver. This real world dataset is processed and used as training dataset for the model. I used vision and Lidar data. Vision data inlcudes images of mobile reciever of 960×540 dimensions which are captured at 30 frames per preprocess.drawio.png

Fig. 2. Experiment Setup

second. Lidar data includes Lidar and Lidar_SCR mat file for each instances

References

[1] G. Charan, T. Osman, A. Hredzak, N. Thawdar and A. Alkhateeb, “Vision-Position Multi-Modal Beam Prediction Using Real Millimeter Wave Datasets,” In Proc. of IEEE Wireless Communications and Networking Conference (WCNC), 2022.

[2] S. Jiang and A. Alkhateeb, “Computer vision aided beam tracking in a real-world millimeter wave deployment.” in IEEE Globecom workshops, 2022.

[3] S. Jiang, G. Charan, and A. Alkhateeb, “LiDAR Aided Future Beam Prediction in Real-World Millimeter Wave V2I Communications.” in IEEE Wireless Communications Letters, 2022.