M.S. THESIS

Vision-Aided Blockage Prediction and Proactive Handover for Indoor Terahertz Communications

테라헤르츠파 통신을 위한 비전 기반 차단 예측 및 핸드오버에 관한 연구

BY

Liu Yiying

AUGUST 2023

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING
COLLEGE OF ENGINEERING
SEOUL NATIONAL UNIVERSITY

M.S. THESIS

Vision-Aided Blockage Prediction and Proactive Handover for Indoor Terahertz Communications

테라헤르츠파 통신을 위한 비전 기반 차단 예측 및
핸드오버에 관한 연구

BY

Liu Yiying

AUGUST 2023

DEPARTMENT OF ELECTRICAL AND COMPUTER
ENGINEERING
COLLEGE OF ENGINEERING
SEOUL NATIONAL UNIVERSITY

Vision-Aided Blockage Prediction and Proactive Handover for Indoor Terahertz Communications

테라헤르츠파 통신을 위한 비전 기반 차단 예측 및 핸드오버에 관한 연구

지도교수 심 병 효 교 수 님

이 논문을 공학석사 학위논문으로 제출함

2023년 8월

서울대학교 대학원

전기 정보 공학부

유 의 영

유의영의 공학석사 학위 논문을 인준함

2023년 8월

위 원 장: 김 성 철

부위원장: 심 병 효

위 원: 이 경 한

Abstract

To support extremely high data rates in 6G wireless networks, Terahertz (Thz) communications that explore the abundant spectrum resources at the Thz band have attracted great interest in recent years. However, due to the strong directivity and severe signal attenuation of THz signals, the link quality is highly sensitive to obstacles, especially when there is only a line-of-sight (LoS) path. To enable proactive handover to a transmitter with an alternative LoS link, accurate blockage prediction is essential to avoid the sudden drop in transmission quality. Unfortunately, existing methods focusing on outdoor environments often fail to predict blockages in complicated indoor environments.

In this paper, we propose a background-aware vision-aided blockage prediction framework that utilizes the sequences of historical RGB-depth (RGB-D) information and the beam indices to detect and localize users and potential blockages, predict their trajectories, and foresee the blockages in dynamic indoor scenarios. Specifically, we first model the background with the medium filter and use a deep-learning-based object detector to detect the users as well as potential blockages. We then predict the future locations of the users using an LSTM-based neural network and predict the time when the users locate behind the background. We demonstrate from numerical results that the proposed scheme outperforms conventional schemes in terms of blockage prediction and proactive handover decision-making accuracy.

Keywords: Vision-aided communications, Proactive handover, Blockage prediction, Indoor communications

Student Number: 2021-24838

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Terahertz (THz) communications have recently received considerable attention for the potential to support high data rates by exploiting abundant spectrum resources in THz frequency band ($0.1 \sim 10$ THz) [1, 2]. The link quality of THz communications relies heavily on the existence of the line-of-sight (LoS) link due to the strong directivity and severe path loss of THz signals. When the LoS path between the base station (BS) and the user equipment (UE) is blocked by obstacles (e.g., pedestrians, cars, buildings), the signal transmission is suddenly interrupted, leading to a significant degradation in link quality. To prevent such circumstances, proactive handover techniques that predict the occurrence of blockage and trigger the handover operation before the blockage occurs have been suggested [11]. Since the proactive handover is based on the prediction of blockage, too many unnecessary handovers might occur if the blockage prediction is inaccurate [10]. To avoid additional signalling and resource usage caused by unnecessary handovers, fast and accurate blockage prediction is of great importance in proactive handovers.

Over the years, various blockage prediction techniques for proactive handover have been proposed [7, 8]. In [8], a machine learning-based blockage prediction technique for mmWave systems has been proposed. Also, in [7], a deep reinforcement learning-based blockage prediction technique for unmanned aerial vehicle (UAV) communica-

tion systems has been proposed. In these schemes, sequential beam indices or reference signal received power (RSRP) are used as indicators to predict the occurrence of blockage. However, since these indicators are obtained from the control signal transmission and reception, the blockage prediction accuracy might degrade severely in a low signal-to-noise ratio (SNR) regime.

Recently, vision-aided blockage prediction techniques have been proposed [4, 6]. With the rapid development of deep learning (DL) and neural processing unit (NPU) technology, computer vision (CV) techniques have made remarkable success in object detection, semantic segmentation, and object tracking [5]. By learning and understanding the high-resolution visual information (e.g., RGB image, depth image) using DL techniques, the vision-aided proactive handover scheme can significantly improve the localization and blockage prediction accuracy significantly. In conventional vision-aided blockage prediction techniques, the bounding box (the smallest box containing the target object) of the detected object is identified using the DL-based object detector and then embedded as the input to an RNN network to determine whether the blockage will happen or not. A major drawback of these schemes is that since all objects, including UEs and obstacles, should be detected, the processing latency and computational overhead are considerable. This issue is even more serious in indoor communication scenarios due to the complex surroundings and the vast diversity of obstacles.

An aim of this paper is to propose a novel background-aware vision-aided blockage prediction and proactive handover technique for indoor THz communication systems. To do so, the proposed technique, henceforth referred to as *background-aware vision-aided blockage prediction* (BV-BP), exploits the property that the majority of indoor movements typically are related to human activities. Using this property, BV-BP separately extracts the stationary background and the movable humans from the image and then predicts the moment when the human holding the connected UE, locates behind the background. While the conventional vision-aided blockage prediction schemes identify every object in the image including humans and other obstacles, BV-

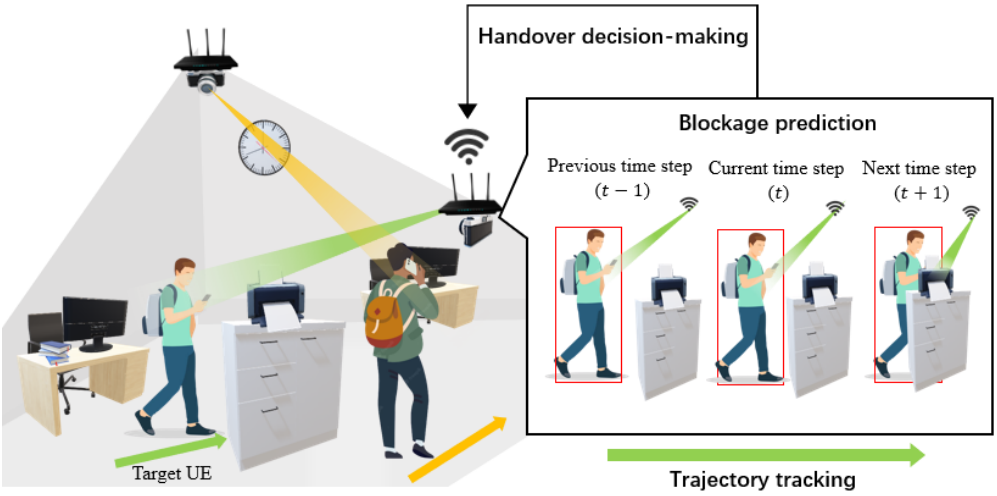

Figure 1.1: Proactive handover for indoor communications.

BP detects and tracks only the humans for the blockage prediction and obtains the observation of other possible obstacles through the background modelling, thereby achieving the processing latency reduction and the prediction accuracy improvement.

From the numerical results, we demonstrate that the proposed BV-BP scheme can predict the blockage within 0.5 s with an accuracy of 97%. We also show that BV-BP outperforms the conventional vision-aided scheme and the beam index-based scheme, achieving an increase in blockage prediction accuracy of 15% and 30%, respectively.

Chapter 2

Terahertz Indoor Communication Systems

We consider multiple-input single-output (MISO) THz indoor communication systems where a BS equipped with a uniform planar array (UPA) of $M = M_h \times M_v$ antennas serves a single-antenna target UE (see Fig 2.1), which is held by a human, called *target user*. We assume that there are multiple humans in the scenarios, and only the target user is holding the UE that communicates with the BS while the other humans act as obstacles to the communication between the BS and the target user. The BS is equipped with a $r_x \times r_y$ resolution RGB-D camera to monitor the wireless communication environment. The RGB-D camera provides sequences of RGB information $\mathbf{R} \in \mathbb{R}^{r_x \times r_y \times 3}$ and depth information $\mathbf{D} \in \mathbb{R}^{r_x \times r_y}$ to the BS. In this setting, the received downlink signal $y \in \mathbb{C}$ of the target UE is given by:

$$y = \sqrt{P_{\text{tx}}}\mathbf{h}^{\text{H}}\mathbf{f}x + n, \tag{2.1}$$

where P_{tx} is the BS transmit power, $\mathbf{h} \in \mathbb{C}^M$ is the downlink channel vector, $\mathbf{f} \in \mathbb{C}^M$ is the analog beamforming vector, x is the data symbol, and $n \sim \mathcal{CN}(0, \sigma_n^2)$ is the complex Gaussian noise. Then the data rate R of the UE is defined as

$$R = \log_2\left(1 + \frac{P_{\text{tx}}|\mathbf{h}^{\text{H}}\mathbf{f}|^2}{\sigma_n^2}\right). \tag{2.2}$$

In this work, we use the narrowband geometric channel model where the channel

vector $\mathbf{h}$ is expressed as [1]

$$\mathbf{h} = \delta\alpha^{\text{LoS}}\mathbf{a}(\theta^{\text{LoS}}, \phi^{\text{LoS}}) + \sum_{i=1}^{P} \alpha_i^{\text{NLoS}}\mathbf{a}(\theta_i^{\text{NLoS}}, \phi_i^{\text{NLoS}}), \tag{2.3}$$

where δ represents the status of LoS link:

$$\delta = \begin{cases} 1 & \text{LoS link is available} \\ 0 & \text{LoS link is blocked} \end{cases}. \tag{2.4}$$

We assume that there are two types of blockages: 1) blockage occurred by other humans when they move around and occupy the transmission link and 2) blockage occurred by the background when the target user moves around and the objects in the background obstruct the transmission signals.

Also, α^{LoS} and α_i^{LoS} are the complex path gains of the LoS path and the i-th NLoS path, respectively, $(\theta^{\text{LoS}}, \phi^{\text{LoS}})$ and $(\theta_i^{\text{NLoS}}, \phi_i^{\text{NLoS}})$ are the azimuth and elevation angles of departures of the LoS path and the i-th NLoS path, respectively, and $\mathbf{a}(\theta, \phi) = \mathbf{a}_h(\theta, \phi) \otimes \mathbf{a}_v(\phi)$ is the UPA steering vector of BS where $\mathbf{a}_h(\theta, \phi) = \frac{1}{\sqrt{M_h}}\left[1 \cdots e^{-j\frac{2\pi d_h}{\lambda}(M_h-1)\sin\theta\sin\phi}\right]^{\text{T}}$ and $\mathbf{a}_v(\phi) = \frac{1}{\sqrt{M_v}}\left[1 \cdots e^{-j\frac{2\pi d_v}{\lambda}(M_v-1)\cos\phi}\right]^{\text{T}}$ are the horizontal and vertical array steering vectors and d_h and d_v are the horizontal and vertical antenna spacings, respectively.

The beamforming vector $\mathbf{f}$ is chosen from the DFT-based beam codebook $\mathcal{F} = \{\mathbf{f}_i \otimes \mathbf{f}_j \mid i = 1, \cdots, M_h O_h, j = 1, \cdots, M_v O_v\}$ with oversampling ratios O_h and O_v:

$$\mathbf{f} = \mathbf{f}_{\hat{i},\hat{j}} = \arg\max_{\mathbf{f}_{i,j} \in \mathcal{F}} \left|(\mathbf{f}_i \otimes \mathbf{f}_j)^{\text{H}}\mathbf{h}\right|^2, \tag{2.5}$$

where

$$\delta = \begin{cases} \mathbf{f}_i = \frac{1}{\sqrt{M_h}}\left[1\ e^{j\frac{2\pi}{M_h O_h}i} \cdots e^{j\frac{2\pi}{M_h O_h}(M_h-1)i}\right]^{\text{T}} \\ \mathbf{f}_j = \frac{1}{\sqrt{M_v}}\left[1\ e^{j\frac{2\pi}{M_v O_v}} \cdots e^{j\frac{2\pi}{M_v O_v}(M_v-1)j}\right]^{\text{T}} \end{cases}. \tag{2.6}$$

Note that due to the strong directivity and severe path loss of the THz signal, the power gap between the LoS and NLoS path signals is significant (the power of LoS path signal is almost 100 times stronger than that of NLoS path signals). Thus, when

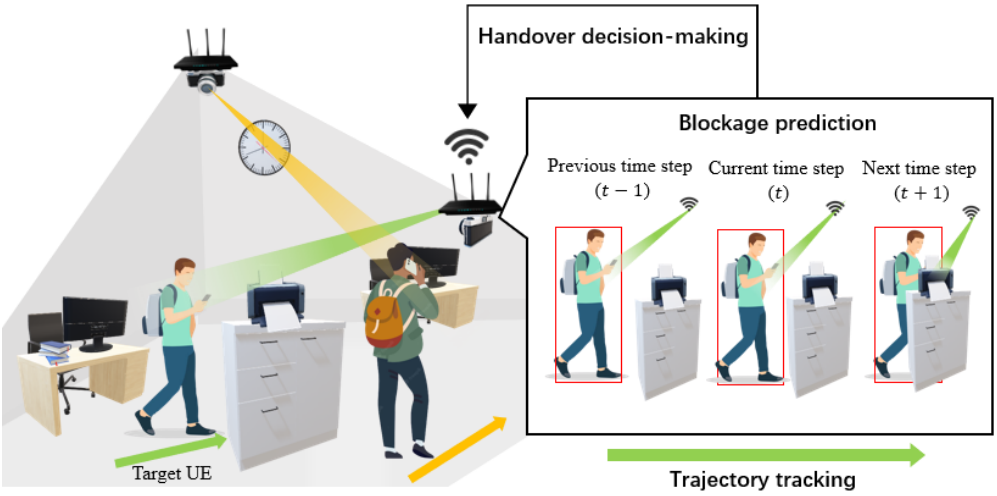

Figure 2.1: Proactive handover for indoor communications.

the LoS link is stable (i.e., $\delta = 1$), the beamforming vector $\mathbf{f}$ that is aligned with the LoS channel component $\alpha^{\text{LoS}}\mathbf{a}(\theta^{\text{LoS}}, \phi^{\text{LoS}})$ will be chosen from the beam codebook, thereby achieving the maximum beamforming gain. Also, using the chosen indices $(\hat{i}, \hat{j})$ of the beam codebook, the BS can acquire rough estimates of the azimuth and elevation angles:

$$(\hat{\theta}, \hat{\phi}) = \left(\arccos\left(\frac{\lambda}{O_h d_h \sin\hat{\phi}} \hat{i} \right), \arccos\left(\frac{\lambda}{O_v d_v} \hat{j} \right) \right). \tag{2.7}$$

When the LoS link is blocked by the obstacles, however, $\mathbf{f}$ that is aligned with the NLoS channel component $\alpha_i^{\text{LoS}}\mathbf{a}(\theta_i^{\text{NLoS}}, \phi_i^{\text{NLoS}})$ will be chosen from the beam codebook. Since the power of the NLoS channel component is much smaller than that of the LoS channel component, the degradation of beamforming gain would be significant in this case. To avoid such circumstances, it is of great importance to predict the blockage and then handover the UE to another BS that can guarantee a reliable LoS link.

Chapter 3

Vision-aided Blockage Prediction and Proactive Handover

Main goal of the proposed BV-BP scheme is to predict k sequential blockage statuses $\{\delta[t]\}_{t=1}^{k}$ in complicated indoor scenarios using the r RGB and depth information $\{\mathbf{R}[t], \mathbf{D}[t]\}_{t=1-r}^{0}$ and beam codeword index information $\{\hat{\theta}[t], \hat{\phi}[t]\}_{t=1-r}^{0}$. As mentioned, the conventional vision-aided blockage prediction schemes detect every object including humans and obstacles so the processing latency and computational complexity are considerable. To address this issue, BV-BP separately extracts the stationary background and the movable humans and then models the stationary background and tracks the trajectories of the movable humans to obtain a complete observation of the environment. After that, the LoS status at each time slot is obtained by predicting if the target user locates behind its background. The blockage prediction task can be expressed as

$$\{\delta[t]\}_{t=1}^{k} = f(\{\mathbf{R}[t], \mathbf{D}[t], \hat{\theta}[t], \hat{\phi}[t]\}_{t=1-r}^{0}; \mathbf{\Lambda}), \tag{3.1}$$

where f is the mapping function and $\mathbf{\Lambda}$ is the network parameters. Overall process of BV-BP is as follows:

- **Background extraction**: extracting the background from the depth images using the medium filtering technique

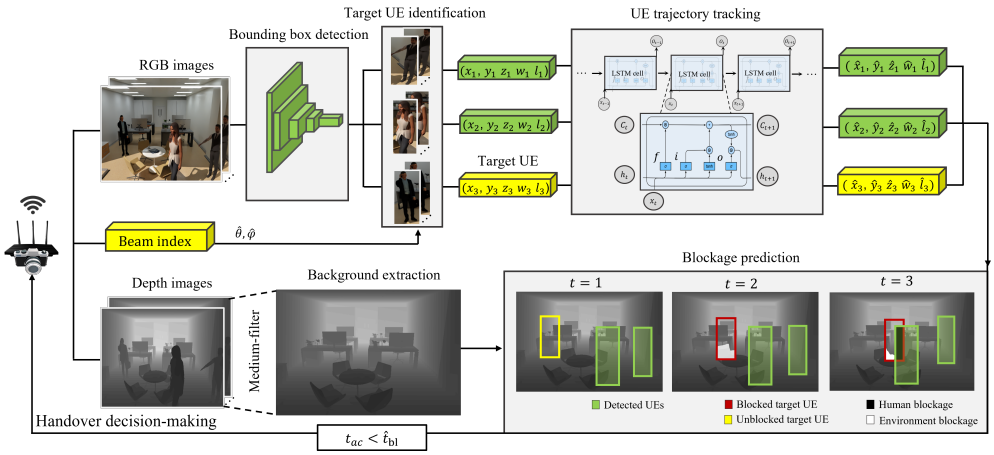

Figure 3.1: Illustration of vision-aided blockage prediction and proactive handover system.

- **Target user identification**: detecting the 3D bounding boxes of all humans using the DL-based object detector and identifying the target user by finding out the bounding box closest to the beam direction in each time slot
- **Trajectory tracking**: tracking the trajectories of all users using long-short term memory (LSTM) network
- **Blockage prediction and handover decision-making**: predicting the moment of blockage occurrence by comparing the depths of estimated human trajectories and the extracted background

3.1 Background Extraction

In vision-aided blockage prediction, precise identification of the target user and the surrounding environment is crucial to obtain a complete observation for further estimation of blockage. In complex indoor scenarios, however, due to a variety of randomly-arranged furniture, the pre-defined classification labels of object detectors cannot describe the overall environment with an acceptable time and computational complexity.

To overcome this, we exploit a temporal medium filter med to calculate the median value over a temporal window for each pixel (x, y) in the depth image, through which the transient objects or changes in the scenario can be filtered out. Then the entire background can be expressed as

$$\hat{\mathbf{D}}(x, y) = \text{med}\{\mathbf{D}[t](x, y)\}_{t=1-n}^{0}, \tag{3.2}$$

where n is the size of the temporal window. [1] In doing so, we can capture stable background information while mitigating the noise effect and transient variations caused by moving objects.

By assuming a Gaussian distribution $\mathcal{N}(\mu_S(x, y), \sigma_S^2(x, y))$ for $\mathbf{D}[t](x, y)$, we compute the probability that $\mathbf{D}[t](x, y)$ does not follow the same distribution as the previous estimation with the following formula:

$$P(D[t](x, y)) = 1 - \Phi\left(\frac{|\mathbf{D}[t](x, y) - \mu S(x, y)|}{\sigma_S(x, y)}\right) \tag{3.3}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. If the probability of $\mathbf{D}[t](x, y)$ not following the background distribution is higher than a predefined threshold, denoted as p_{th}, we update the background model with the past n time slot depth images. Empirically, we choose $p_{th} = 0.0013$, which corresponds to a probability of 99.87% of the pixel not following the background distribution, which is equivalent to:

$$|\mathbf{D}[t](x, y) - \mu_S(x, y)| > 3\,\sigma_S(x, y). \tag{3.4}$$

3.2 Target User Identification

Since the variations of the indoor environment are mainly caused by human activities including human blockage and device displacement, accurate blockage prediction

[1] Note that the temporal window size n is selected based on the object duration. For example, a larger n is required if the moving objects stay in the scenario for a long duration to capture their complete trajectory.

highly relies on the prediction of the human's locations in the upcoming time slots. To do so, we first detect and localize the bounding boxes of all humans using a DL-based object detector and then match the bounding box of individual humans in each time slot through similarity matching. After that, by combining the bounding box information and the quantized transmission angles, we identify the target user while considering other humans as blockages.

3.2.1 Bounding Box Detection

To identify the target user, we use the DL-based object detector to analyze the RGB images and determine the bounding box for each human. To be specific, the object detector generates a set of anchor boxes with different sizes, using which it predicts the confidence scores (the likelihood of the pixel being the centre of the object) and the bounding box coordinates. Based on confidence scores, the object detector selects the appropriate anchor boxes to obtain 2D coordinates of the detected objects.

Let g be the object detector, the bounding box detection process can be expressed as

$$\{\mathbf{b}_i[t]\}_{i=1}^{N} = g(\mathbf{R}[t]), \tag{3.5}$$

where g is the mapping function, N is the number of detected humans and $\mathbf{b}_i[t] = (x_i[t], y_i[t], w_i[t], l_i[t])$ is the bounding box vector with $(x_i[t], y_i[t])$ and $(w_i[t], l_i[t])$ being the top-left corner and width-height pair of the i-th detected bounding box in the t-th time slot, respectively. By combining the bounding boxes and the depth information, we then form a 3D bounding box vector $\mathbf{b}_i^d[t] = (x_i[t], y_i[t], w_i[t], l_i[t], d_i[t])$ representing the 3D location and size of each detected human.

3.2.2 Bounding Box Pairing

After the bounding box detection, we group the detected bounding boxes of the same human in sequential images to estimate the trajectory using the cosine similarity.

Specifically, we extract the feature vector $\mathbf{v}_i$ from the last layer of the object detector to denote the visual feature of each detected human. We then calculate the cosine similarity $l_{i,j}[t]$ between the j-th human at time slot t and the i-th human at time slot $t-1$:

$$l_{i,j}[t] = \frac{\mathbf{v}_j[t]^{\mathrm{T}}\mathbf{v}_i[t]}{\|\mathbf{v}_j[t-1]\|_2\|\mathbf{v}_i[t-1]\|_2}. \tag{3.6}$$

Next, we formulate the bounding box pairing task as an optimization problem where we match the j-th detected human at time slot t with the i-th detected human at time slot $t-1$ using a match function $i = \gamma[t](j)$ where $\gamma[t] = \arg\max \sum_{i=1}^{N} l_{i,\gamma[t](i)}$ such that the cosine similarity is maximized. To find the optimal matching function $\gamma[t]$ for each time slot t, we use the Hungarian algorithm that iteratively adjusts task-resource assignments in a weighted bipartite graph to maximize the summed cosine similarity, $l_{i,\gamma[t]}$.

3.2.3 User Identification

After detecting each human, we need to identify the target user who holds the UE connected to the corresponding BS from the detected human. To do so, we first convert the quantized transmission angles $(\hat{\theta}, \hat{\phi})$ to target two-dimensional coordinates in RGB images. We utilize the interval $[\theta_{\text{min}}, \theta_{\text{max}}]$ to denote the field of view (FOV) that determines the angular range within which the visual information can be captured by the camera, and thus the target coordinate in the RGB image is

$$(\hat{u}_x, \hat{u}_y)[t] = (r_x \frac{\cot\hat{\phi}[t] - \cot\phi_{\text{min}}}{\cot\phi_{\text{max}} - \cot\phi_{\text{min}}}, r_y \frac{\tan\hat{\theta}[t] - \tan\theta_{\text{min}}}{\cot\theta_{\text{max}} - \tan\theta_{\text{min}}}). \tag{3.7}$$

By selecting the detected human with the least average distance with the target coordinates over r observed time slots, we find out the target user who holds the connected UE:

$$\hat{i}_{\text{target}} = \arg\min_i \sum_{t=0}^{r} \|((\hat{u}_x[t], \hat{u}_y[t]) - (x_i[t] + \frac{w_i[t]}{2}, y_i[t] + \frac{l_i[t]}{2})\|^2. \tag{3.8}$$

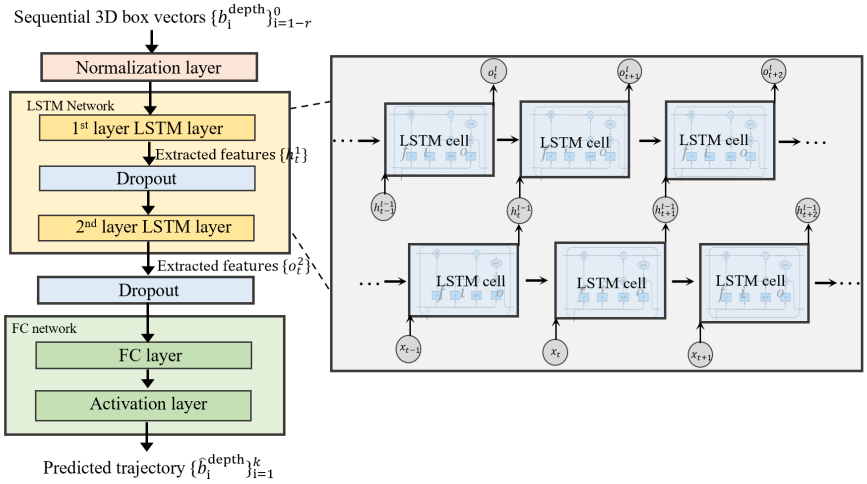

Figure 3.2: Structure of LSTM-based trajectory tracking network.

3.3 LSTM-based Trajectory Tracking

We learn the nonlinear mapping between the past 3D box vectors $(\mathbf{b}^d[t-r+1], \cdots, \mathbf{b}^d[t])$ and the subsequent k future 3D box vectors $(\mathbf{b}^d[t+1], \cdots, \mathbf{b}^d[t+k])$ through an LSTM-based network, l:

$$\{\hat{\mathbf{b}}^d[t+1], \cdots, \hat{\mathbf{b}}^d[t+k]\} = l(\mathbf{b}^d[t-r+1], \cdots, \mathbf{b}^d[t]|\Theta), \tag{3.9}$$

where Θ is the set of network parameters. The overall block diagram network is depicted in Fig 3.2.

In the proposed LSTM network, each layer consists of a sequence of LSTM cells, and each LSTM cell includes a cell state, a hidden state, and three gates: the input gate $\mathbf{i}^l \in \mathbb{R}^{N_l \times 1}$, forget gate $\mathbf{f}^l \in \mathbb{R}^{N_l \times 1}$, and output gate $\mathbf{o}^l \in \mathbb{R}^{N_l \times 1}$, where N_l is the number of hidden units in l-th LSTM layer. Specifically, the cell state $\mathbf{c} \in \mathbb{R}^{N_l \times 1}$ stores information from previous inputs and the gates control the information flow by determining which information to incorporate, discard, and obtain from the cell state. Then the cell state at l-th layer at time slot t is

$$\mathbf{c}_t^l = \mathbf{f}_t^l \odot \mathbf{c}_{t-1}^l + \mathbf{i}_t^l \odot \tanh(\mathbf{W}_{ch^{l-1}}\mathbf{h}_t^{l-1} + \mathbf{W}_{ch^l}\mathbf{h}_{t-1}^l + \mathbf{b}_c^l), \tag{3.10}$$

where $\mathbf{W}_{ch^{l-1}} \in \mathbb{R}^{N_l \times N_{l-1}}$ and $\mathbf{W}_{ch^l} \in \mathbb{R}^{N_l \times N_l}$ are weight matrices and $\mathbf{b}_c^l \in \mathbb{R}^{N_l \times 1}$ is the bias.

The hidden state $\mathbf{h} \in \mathbb{R}^{N_l \times 1}$ is the primary output of each cell that stores the summarised information and acts as the input to the subsequent LSTM layer, given by

$$\mathbf{h}_t^l = \mathbf{o}_t^l \tanh(\mathbf{c}_t^l). \tag{3.11}$$

By exploiting a $L = 2$-layer LSTM network, we capture the complex spatio-temporal patterns and higher-level abstractions in the input sequence as $\mathbf{h}_t^L \in \mathbb{R}^{N_l \times 1}$.

Followed by the LSTM network, we use a fully connected (FC) layer to convert the extracted features to the future 3D box vectors $\mathbf{z}[t] = (\mathbf{b}^d[t+1], \cdots, \mathbf{b}^d[t+k]) \in \mathbb{R}^{k|\mathbf{b}| \times N_2}$, given by

$$\mathbf{z}[t] = f_{\text{ReLU}}(\mathbf{W}\mathbf{h}^L[t] + \mathbf{b}), \tag{3.12}$$

where $\mathbf{W} \in \mathbb{R}^{k|\mathbf{b}| \times N_2}$ is the weight matrix, $\mathbf{b} \in \mathbb{R}^{k|\mathbf{b}| \times 1}$ is the bias, and $f_{\text{ReLU}}(x) = \max(0, x)$ is the rectified linear unit (ReLU) activation layer [**?**]. Next, we add the dropout layers that randomly disable a fraction of the neurons during training to improve the generalization ability and prevent over-fitting. To estimate the future locations and sizes during k time slots, we apply the MSE-based loss function J during the training process:

$$J_{\text{MSE}} = \frac{1}{N} \sum_{n=1}^{N} \|z_i[t] - \hat{z}_i[t]\|_F^2, \tag{3.13}$$

where N is the batch size and $\hat{z}[t]$ is the ground truth vector.

3.4 Blockage Prediction and Proactive Handover

In this subsection, we predict the blockage status of the target user at the following k time slots using the obtained trajectory of each human and the background model with pixel-wise depth information. To be specific, we first reconstruct the extracted background with respect to the target user: $\hat{\mathbf{D}}_u[t](x_i, y_i) = \min(\hat{\mathbf{D}}(x_i, y_i), \hat{d}_i[t])$ where

$i \in \{1, \cdots, u-1, u+1, \cdots, N\}$ and $(x_i, y_i) \in \{(x, y) | \hat{x}_i \leq x \leq \hat{x}_i + \hat{w}_i, \hat{y}_i \leq y \leq \hat{y}_i + \hat{l}_i\}$. We then estimate whether the target user moves behind the background or not to predict the LoS status at time slot t:

$$\hat{\delta}[t] = \begin{cases} 0, & \hat{d}_u[t] < \hat{\mathbf{D}}_u[t](x_u, y_u) \\ 1, & \text{otherwise} \end{cases} . \tag{3.14}$$

After that, we initiate the handover process based on the predicted blockage status in k following time slots. Specifically, we choose the maximal acceptable blockage period t_{ac} enduring the performance degradation for reduced power consumption if the predicted continuous blockage time $t_{\text{bl}} < t_{\text{ac}}$. We initiate the handover if $\tau = \lfloor \frac{t_{\text{ac}}}{\Delta} \rfloor$ continuous time slots is predicted to be blocked where Δ is the prediction interval. In practice, t_{ac} and τ can be properly determined based on the application scenarios and the camera characteristics.

Finally, by predicting the occurrence and duration of the blockage on the transmission link, we initiate the handover process. Therefore, we can effectively avoid the sudden drop in throughput, achieving seamless and reliable wireless transmission in complex indoor scenarios.

Chapter 4

Numerical Evaluations

4.1 Communication Scenario and Dataset Generation

We consider a single-user indoor communication environment built using a game engine that depicts a typical indoor environment with various elements (i.e., humans, tables, and computers). We use the human folding phone to represent a UE, and there are 3 humans included in the scenario. To imitate the walking trajectory of humans, we randomly set reachable destinations and speeds, and add extra noise to the motion trajectory at each time slot. The dimensions of the scenario are $20\,\mathrm{m}$ in width and $30\,\mathrm{m}$ in length, with a 720×480 resolution RGB-D camera capturing images from a bird's-eye view with different periods ranging from $0.25\,\mathrm{s}$ to $0.02\,\mathrm{s}$. Furthermore, based on the location of the BS and the target UE that is connected to the BS, we calculated the transmission angles and correspondingly build a virtual transmission channel to obtain the optimal beam indices, i and j, as shown in (5). We generated a total of 8750 RGB-D images combined with labelled LoS status, the locations and sizes of humans, and the optimal beam indices to form the dataset for training and evaluating the blockage prediction and proactive handover system.

To evaluate the blockage prediction performance of the proposed BV-BP scheme, we compared it with 3 benchmark schemes in the generated dataset: 1) Beam index-

Figure 4.1: Generated indoor communication scenarios.

based scheme [**?**] that utilizes an LSTM network to estimate the historical beam index without vision information; 2) Object detector-based scheme [**?**] that detects and tracks every object without background extraction; 3) Kalman filter-based scheme that employs Kalman filter for human trajectory tracking.

4.2 Simulation Results

In Fig. 4, we compare the prediction accuracy of the proposed BV-BP scheme over the benchmark schemes with a prediction time slot of 0.1s. We observe that the proposed BV-BP scheme generally outperforms other proactive handover schemes, which is because it obtains more accurate and complete observation of the environment through background extraction. For example, when predicting LoS status in 0.5s, the BV-BP scheme demonstrates superior performance compared to the beam-index-based and object-detector-based schemes, with blockage prediction accuracy improvements of over 10% and 25% respectively. Furthermore, compared to BV-BP with the Kalman filter, the proposed LSTM network better tracks humans and achieves improved prediction accuracy, especially for long-term prediction. This is because the Kalman filter is better suited for capturing the linear trajectory and fails to capture the irregular mo-

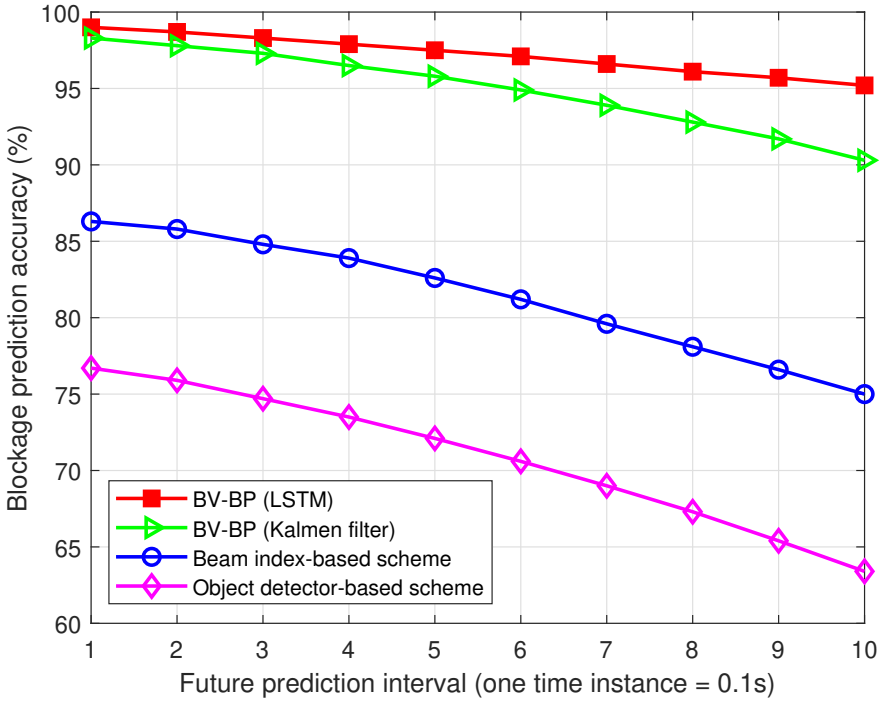

Figure 4.2: Blockage prediction accuracy v.s. prediction interval

tion manners of humans.

In Fig. 5, we plot the blockage prediction accuracy as a function of the prediction interval when the time slot $t = 0.5\,\mathrm{s}$. We see that the prediction accuracy decreases when the time instance increases because the prediction time horizon becomes longer and there is a higher degree of uncertainty and variability involved. In contrast, the prediction accuracy improves with shorter prediction intervals as the time span is shorter, thereby capturing the immediate changes and dynamics in the object's movement more accurately. We also investigate the combined influence of these factors on prediction accuracy at a fixed future time. We see that for a fixed prediction time, the accuracy is improved when the prediction interval is increased. This is because a shorter time span limits the network's ability for capturing longer-term dynamics.

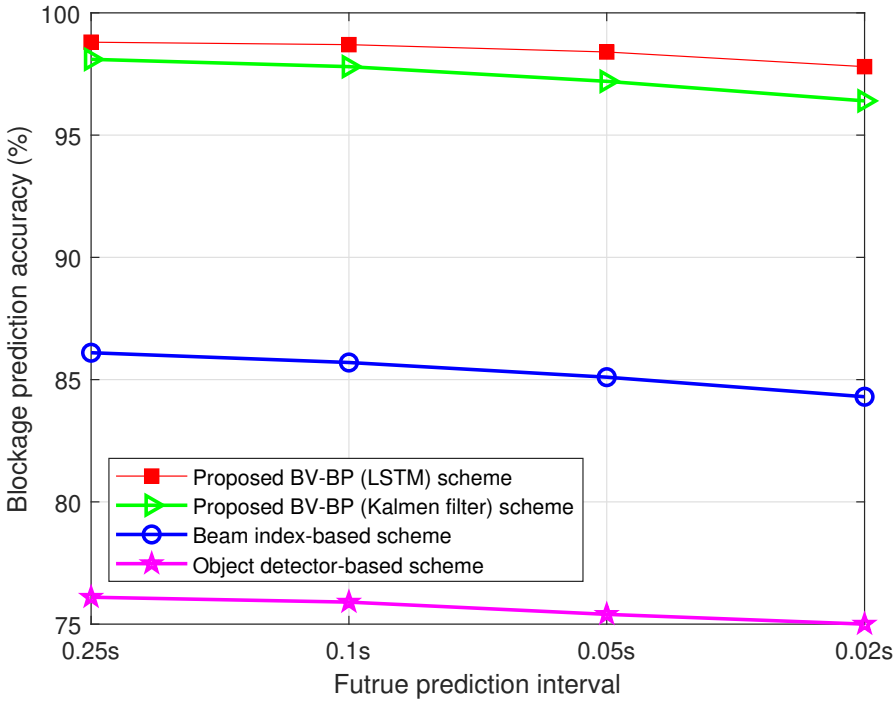

Figure 4.3: Blockage prediction accuracy v.s. prediction interval

Chapter 5

Conclusion

In this paper, we proposed a novel BV-BP scheme that utilizes the sequential RGB-D images and the optimal beam indexes to predict the blockage on the LoS transmission link and initiate handover in complicated indoor scenarios proactively. To do so, the proposed BV-BP scheme separately extracts the stationary background and the movable humans from the image and then predicts the time slots when the target user who holds the UE moves behind the constructed background. From the simulation results, we demonstrated that the proposed BV-BP scheme significantly improves the blockage prediction accuracy for indoor scenarios and reduces the channel efficiency drop caused by the blockage through proactive handover.

Bibliography

[1] J. Wu, S. Kim and B. Shim, ”Energy-Efficient Power Control and Beamforming for Reconfigurable Intelligent Surface-Aided Uplink IoT Networks,” *IEEE Transactions on Wireless Communications*, vol. 21, no. 12, pp. 10162-10176, Dec. 2022.

[2] Chen H, Sarieddeen H, Ballal T, Wymeersch H, Alouini MS, Al-Naffouri TY. ”A tutorial on terahertz-band localization for 6G communication systems.” *IEEE Communications Surveys & Tutorials.* 2022 May.

[3] Han, Chong, et al. ”Terahertz wireless channels: A holistic survey on measurement, modeling, and analysis.” *IEEE Communications Surveys & Tutorials*, vol 24, no. 3, pp. 1670-1707, 2022.

[4] Charan, Gouranga, Muhammad Alrabeiah, and Ahmed Alkhateeb. ”Vision-aided 6G wireless communications: Blockage prediction and proactive handoff.” *IEEE Transactions on Vehicular Technology*, vol 70, no. 10, pp. 10193-10208, 2021.

[5] Y. Ahn et al., ”Towards Intelligent Millimeter and Terahertz Communication for 6G: Computer Vision-aided Beamforming,” *IEEE Wireless Communications*, pp. 1-18, 2022.

[6] Xu, Weihua, Feifei Gao, Xiaoming Tao, Jianhua Zhang, and Ahmed Alkhateeb. ”Computer Vision Aided mmWave Beam Alignment in V2X Communications.” *IEEE Transactions on Wireless Communications*, 2022.

[7] Jang Y, Raza SM, Kim M, Choo H. ”Proactive handover decision for UAVs with deep reinforcement learning.”, *Sensors*, vol 22, no.3, pp.1200, 2022 Feb 5.

[8] Alkhateeb, Ahmed, Iz Beltagy, and Sam Alex. ”Machine learning for reliable mmwave systems: Blockage prediction and proactive handoff.” *IEEE Global conference on signal and information processing (GlobalSIP)*, pp. 1055-1059, 2018.

[9] Sun, Peng, Noura AlJeri, and Azzedine Boukerche. ”An energy-efficient proactive handover scheme for vehicular networks based on passive RSU detection.” *IEEE Transactions on Sustainable Computing*, vol 5, no.1, pp. 37-47, 2018.

[10] Liu, Ziyue, et al. ”A Double-Beam Soft Handover Scheme and Its Performance Analysis for Mmwave UAV Communications in Windy Scenarios.” *IEEE Transactions on Vehicular Technology*, 2022.

[11] Moon, Jihoon, et al. ”Energy-Efficient User Association in mmWave/THz Ultra-Dense Network via Multi-Agent Deep Reinforcement Learning.” *IEEE Transactions on Green Communications and Networking*, 2023.

[12] Zhao, Hongmei, Qian Wang, and Kunfeng Shi. ”Analysis on human blockage path loss and shadow fading in millimeter-wave band.” *International Journal of Antennas and Propagation 2017*, 2017.

[13] Alrabeiah, Muhammad, and Ahmed Alkhateeb. ”Deep learning for mmWave beam and blockage prediction using sub-6 GHz channels.” *IEEE Transactions on Communications*, vol 68, no. 9 pp. 5504-5518, 2020.

[14] Jocher, G., Stoken, A., Borovec, J., Changyu, L., Hogan, A., Diaconu, L., ... & Yu, L.. ”ultralytics/yolov5: V3. 1—Bug Fixes and Performance Improvements.”

초 록

6G 무선 네트워크에서 매우 높은 데이터 전송률을 지원하기 위해 mmWave 및 테라헤르츠 (THz) 밴드와 같은 고주파 대역을 활용한 통신은 최근 큰 관심을 받고 있습니다. 그러나 고주파 신호 (예: THz 신호)의 강한 직접성과 신호 감쇠로 인해 링크 품질은 장애물에 매우 민감하게 반응하며, 특히 시야 경로 (LoS)만 있는 경우에 그 영향을 크게 받습니다. 대체 LoS 링크를 가진 송신기로 선제적인 핸드오버를 가능하게 하기 위해서는 정확한 차단 예측이 필수적으로 요구되며, 이는 전송 품질의 갑작스러운 하락을 피하기 위한 것입니다. 유감스럽게도, 기존의 야외 환경에 초점을 맞춘 기법들은 복잡한 실내 환경에서의 차단 예측에 실패하는 경우가 많습니다.

본 논문에서는 역사적인 RGB-깊이 (RGB-D) 정보와 빔 인덱스의 순열을 활용하여 사용자와 잠재적인 차단물을 감지하고 위치를 추적하며, 동적인 실내 시나리오에서의 차단을 예측하는 시각지원형 차단 예측 프레임워크를 제안합니다. 구체적으로, 우리는 먼저 중간 필터로 배경을 모델링하고, 딥러닝 기반의 객체 탐지기를 사용하여 사용자 및 잠재적인 차단물을 감지합니다. 그런 다음 LSTM 기반의 신경망을 사용하여 사용자의 미래 위치를 예측합니다. 숫자적인 결과로써, 우리가 제안한 방법이 차단 예측 및 선제적 핸드오버 결정 정확도 측면에서 기존 방법보다 우수함을 보여줍니다.

주요어: 적극적 핸드오버, 실내 통신, 비전 지원 통신

학번: 2021-24838