Vision-aided Soft Handover and Blockage Prediction

Yiying Liu and Byonghyo Shim
Department of Electrical and Computer Engineering, Seoul National University
Email : {liuyiying, bshim}@islab.snu.ac.kr

. Introduction

Terahertz (THz) and millimeter wave communications have received much attention for supporting the high data rates by exploiting the plentiful spectrum resources. However, one notable feature of the high-frequency band signal is that due to the high directivity and severe path loss, the transmission is highly dependent on the line-of-sight link. The throughput sharply decreases if an object, such as a pedestrian, blocks the LOS path between the user equipment (UE) and a base station (BS). In order to prevent transmission degradation, multiple base stations can be equipped to ensure at least one LOS path is available, and then the control systems will conduct handover based on communication quality factors. However, we can hardly avoid the period when the throughput extremely decreases while the handover process is ongoing.

To cope with the problem, a soft handover process has been proposed recently to proactive the handover process based on future blockage prediction. To be specific, the control system uses past visual data to predict whether the link would be blocked in a few seconds, and conducts handover proactively to reduce the latency for the conventional reactive approach, which takes more than 200ms. A successful proactive handover process involves the time consumption of random access (about 11ms) only and significantly reduces the latency.

I. System Model

Consider a system where multiple Base Stations, operating at both sub-6GHz and mm-Wave bands, are communicating with a user, as depicted in Fig. 1. The BSs are assumed to be equipped with RGB-D cameras.

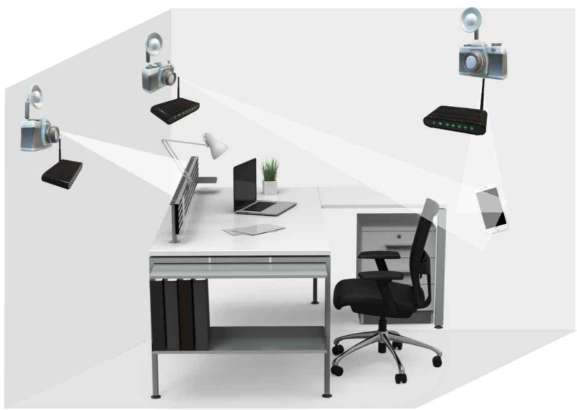

Fig . A downlink communication scenario where three candidate BSs are serving one user (mobile phone), and the other objects are potential blockages.

II.Related Work

Recent researches divide the proactive handover process into two phases: the object detection process and the blockage prediction process. Object detection includes the classification and localization of objects, and a pre-trained object detection model (such as YOLOv3) can be fine-tuned with a tiny dataset regarding target scenarios to cope with the task. The bounding box information obtained from the previous object detection process (sometimes together with beam information) is inputted to RNN followed by a classifier to output the future blockage status.

However, the prediction based on object detection and localization suffer from severe problems. Firstly, the accuracy of the pre-trained object detection model without fine-tuning is not satisfactory. whereas the fine-tuning is specific to the target scenarios, which means it cannot be generally applied, and the training process should be repeated for varied environment settings. Secondly, the localization of objects with bounding boxes ignores the edge information, which leads to inaccurate blockage prediction, especially for indoor scenarios where potential blockages are mostly large and with irregular edges, such as pedestrians. Thirdly, most researchers use a larger object as the UE in the dataset, such as a car, which is much easier to detect but is not suitable for indoor scenarios where a mobile phone should be detected instead.

IV. Proposed Scheme

The proposed scheme uses the sequential color and depth information of the varied environment to conduct the blockage prediction. Considering that the UE location (including angle and distance from BS) is known information for connected devices, the position of UE in the RGB image can be estimated through this information. Also, the classification of each object is not required for blockage prediction because we care only about whether the UE is going to be blocked, rather than what object is going to block the UE. Therefore, pixel-wise prediction involving accurate environment information would be a better solution compared to the prediction of bounding box positions. In addition, with the RGB-D camera, the depth-based pixel-wise classification through depth image reconstruction for each time frame becomes much easier and less time-consuming compared to conventional segmentation process. Therefore, we proposed a three-phase pixel-wise prediction to improve the accuracy and generality of blockage prediction.

(1) Depth image reconstruction through the fusion of RGB color information and depth information: CNN with dual-input sources.

In this research, because of the significant importance of depth information in blockage prediction (objects behind UE cannot block it, and only objects with smaller depth than UE are considered as potential blockages), we reconstruct low-resolution depth images with the help of high-resolution RGB images. To be specific, both the RGB images and depth images go through a down-sampling CNN layer separately to extract features, and then are concatenated to input to an up-sampling CNN layer for depth image reconstruction.

(2) Moving Object Detection, which is the pixel-wise classification into still objects and moving objects: Background modeling and Foreground detection.

Based on recent research, pixel-wise background modeling, which detects the still background, can be achieved with a short training period, when the variation and its frequency of pixel information are used for estimation. Then the moving objects are estimated based on the change of pixels in color and depth. After this step, each pixel in the image is labeled with color information, depth information, position information, and state information, which indicated whether it belongs to the still background or moving objects.

(3) Blockage Prediction based on the moving objects detected: LSTM.

By stacking the depth information for the pixels detected included in moving objects at time t, we obtain the representation matrix as D_t, where each element indicates the depth information of the corresponding pixel. Then with the sequential information D_{t-N} , $D_{t-(N-1)}$, ... , D_{t-1} as input, we predict future information D_{t+1} ,... , D_{t+M+1} with a LSTM. By combining the position information of UE and predicted pixel information, we can predict whether a blockage would happen.

V. Conclusion

We expect to achieve a robust blockage prediction with high accuracy, and reduce the latency of the conventianl reactive process. Even though the blockage prediction may fail for special cases, we expect an overall reduction in latency of more than 70%. With the help of the RGB-D camera, the soft handover would significantly improve the robustness and efficiency of indoor communication.

Reference

[1] L. Sun, K. Yang, X. Hu, W. Hu and K. Wang, "Real-Time Fusion Network for RGB-D Semantic Segmentation Incorporating Unexpected Obstacle Detection for Road-Driving Images," inIEEE Robotics and Automation Letters, vol. 5, no. 4, pp. 5558-5565, Oct. 2020, doi: 10.1109/LRA.2020.3007457.

[2] Cholakov, Radostin, and Todor Kolev. "Transformers predicting the future. Applying attention in next-frame and time series forecasting."arXiv preprint arXiv:2108.08224(2021).

[3] Gao, Yin, et al. "Deep Learning based Location Prediction with Multiple Features in Communication Network."2021 IEEE Wireless Communications and Networking Conference (WCNC). IEEE, 2021.

[4] Gao, Yin, et al. "Deep Learning based Location Prediction with Multiple Features in Communication Network."2021 IEEE Wireless Communications and Networking Conference (WCNC). IEEE, 2021.

[5] Vukotić, Vedran, et al. "One-step time-dependent future video frame prediction with a convolutional encoder-decoder neural network."International conference on image analysis and processing. Springer, Cham, 2017.