Vision-Aided Frame-Capture-Based CSI Recomposition for WiFi Sensing: A Multimodal Approach

Hiroki Shimomura*, Yusuke Koda†*, Takamochi Kanda*, Koji Yamamoto†, Takayuki Nishio‡, and Akihito Taya§
*Graduate School of Informatics, Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto, 606-8501, Japan
†Centre for Wireless Communications, University of Oulu, 90014 Oulu, Finland
‡School of Engineering, Tokyo Institute of Technology, Ookayama, Meguro-ku, Tokyo, 152-8550, Japan
§Institute of Industrial Science, The University of Tokyo, Komaba, Meguro-ku, Tokyo, 153-8505, Japan
*koda@ieee.org, kyamamot@i.kyoto-u.ac.jp

Abstract—Recomposing channel state information (CSI) from the beamforming feedback matrix (BFM), which is a compressed version of CSI and can be captured because of its lack of encryption, is an alternative way of implementing firmware-agnostic WiFi sensing. In this study, we propose the use of camera images toward the accuracy enhancement of CSI recomposition from BFM. The key motivation for this vision-aided CSI recomposition is to draw a first-hand insight that the BFM does not fully involve spatial information to recompose CSI and that this could be compensated by camera images. To leverage the camera images, we use multimodal deep learning. We conducted experiments using IEEE 802.11ac devices and revealed that the recomposition accuracy of the proposed multimodal framework is improved compared to the single-modal framework only using images or BFM.

I. INTRODUCTION

Currently, channel state information (CSI) in wireless local area networks (WLANs) has attracted increasing attention because of its fine-grained propagation characteristics. While CSI-based sensing has been widely studied, obtaining CSI from off-the-shelf WLAN devices requires customized firmware [1]. To alleviate this limitation, the study [2] utilized the beamforming feedback matrix (BFM), which is a series of right singular matrices for CSI matrices. While we can capture BFM by the frame-capturing tool because of its lack of encryption, there is a performance gap between BFM- and CSI-based sensing, e.g. the gap in the respiration estimation error [2], [3]. One promising approach to study the performance of BFM as an alternative for CSI for sensing tasks is to directly recompose the CSI from the BFM and evaluate the prediction errors [4]. In [4], they demonstrated the feasibility of the recomposing CSI amplitude from BFM. However, there still existed a recomposition error between the ground-truth and recomposed CSI, and the input information that was additionally needed to enhance the prediction accuracy was not fully determined. Answering this question leads to further understanding the BFM as an alternative feature for CSI towards wireless sensing tasks.

To this end, we develop a vision-aided CSI recomposition framework, using images alongside BFM. This is motivated by our key hypothesis that spatial information on propagation environments in camera images can aid in the compensation of the insufficiency of BFM to recompose CSI.

II. SYSTEM MODEL

Fig. 1 presents a system model of our proposed framework, which consists of an access point (AP), station (STA), red-green-blue (RGB) camera, sniffer, and estimator. The STA receives frames transmitted by the AP, measures the CSI, calculates the BFM from the CSI, and sends back the BFM to the AP by transmitting beamforming feedback frames. The sniffer captures the feedback frames to obtain the BFM. Because BFM are sent without encryption, they can be captured by frame-capturing tools, such as Wireshark. The estimator recomposes the amplitude of the CSI elements. Unlike in a previous study [4], the estimator uses not only BFM but also RGB images captured by the RGB camera.

III. MULTIMODAL LEARNING-BASED CSI RECOMPOSITION FRAMEWORK

In the vision-aided CSI recomposition framework, we use the multimodal-input (MMI) model to which RGB image data and BFM data were fed. The model is trained by computing the loss function between the output and supervised CSI data. Fig. 2 shows the structure of the MMI model, which consists of two encoders, i.e., BFM encoder and image encoder, a concatenation layer, and CSI decoder. The model is constructed as a convolutional neural network (CNN) including 0.92 million
parameters. The kernel and filter size of each layer are shown in Fig. 2.

IV. EXPERIMENTAL EVALUATION

A. Experimental Setup

We developed an experimental system to obtain simultaneously corresponding CSI and BFM using off-the-shelf devices for creating the datasets. Fig. 3 illustrates the experimental equipment layout, where an AP, STA, camera, and measurement device (MD) are used. The MD captures frames transmitted by the AP to calculate the CSI. Thus, the CSI represents the channel characteristic between the MD and AP. To obtain the corresponding CSI and BFM, the BFM of the input data is not obtained by the sniffer but obtained from the MD by applying singular value decomposition to the CSI. We used ASUS RT-AC86U as AP, STA and MD. The equipment was all off-the-shelf, except for the firmware of the MD, which is customized according to [1]. To experiment in a dynamic environment, a pedestrian as an obstacle moved along with the paths indicated in Fig. 3(a). There are four paths perfectly included in the field of view (FOV) of the camera. We conducted a total of four experiments using each path.

B. Baselines

To evaluate the performance of the MMI model, we used two single-modal input (SMI) baselines: SMI-BFM and SMI-image, whose inputs are the BFMs and images, respectively. SMI-BFM is consisted of BFM encoder and CSI decoder and SMI-image is consisted of image encoder and CSI decoder in Fig. 2. They do not have the concatenation layer, but their components are the same as the corresponding components in the MMI model. The SMI-BFM and SMI-image model include 0.78 and 0.82 million parameters, respectively.

In ML algorithm, we set the values of batch size and epoch to 64 and 100, respectively. We used Adam with a learning rate of 0.001 for the optimizer and mean squared error for the loss function. To obtain the results without depending on the model initialization, all CNNs were trained based on five random seeds for initializing parameters.

C. Results

Fig. 4 depicts an example of the frequency series of the amplitude of the CSI element. The recomposition accuracy is visually confirmed by this figure. The CSI recomposed by the MMI matches the ground truth most closely compared to that of SMI-BFM and SMI-image. TABLE I shows the recomposition root-mean-squared error (RMSE) with one standard deviation for five random seeds. The recomposition error in the MMI is smaller than that in the baselines. Especially, the comparison between the MMI and SMI-BFM suggests that the RGB images include the spatial information on the propagation environments to fill the gap between the BFM and CSI.

V. CONCLUSIONS

In this study, we proposed a CSI recomposition framework leveraging camera images alongside the BFMs. To leverage the camera images, we used multimodal deep learning, where the input of the BFMs and camera images were integrated to recompose the CSI amplitude. We conducted experiments using IEEE 802.11ac devices and demonstrated that the recomposition accuracy of the proposed multimodal framework was improved rather than that of the comparison single-modal frameworks.

ACKNOWLEDGMENT

This research and development work was supported in part by the MIC/SCOPE #JP196000002 and JSPS KAKENHI Grant Number JP22H03575.

REFERENCES

[1] F. Gringoli, M. Schulz, J. Link, and M. Hollick, “Free your CSI: A channel state information extraction platform for modern Wi-Fi chipsets,” in Proc. Int. Workshop Wireless Netw. Testbeds, Exp. Eval. & Characterization, Los Cabos, Mexico, Oct. 2019, pp. 21–28.
[2] T. Kanda, T. Sato, H. Awano, S. Kondo, and K. Yamamoto, “Respiratory rate estimation based on Wi-Fi frame capture,” in Proc. IEEE Consumer Commun. & Netw. Conf., Online, Jan. 2022, pp. 881–884.
[3] Y. Zeng, D. Wu, J. Xiong, E. Yi, R. Gao, and D. Zhang, “Farsense: Pushing the range limit of wifi-based respiration sensing with CSI ratio of two antennas,” Proc. ACM Interactive, Mobile, Wearable and Ubiquitous Technol., vol. 3, no. 3, pp. 1–26, 2019.
[4] R. Hanahara, S. Iihara, K. Yamashita, Y. Koda, A. Taya, T. Nishio, and K. Yamamoto, “Frame-capture-based CSI recomposition pertaining to firmware-agnostic Wi-Fi sensing,” in Proc. the IEEE Annu. Consumer Commun. & Netw. Conf., Online, Jan. 2022, pp. 1–5.