WiFi CSI-Based Device-free Multi-room Presence Detection using Conditional Recurrent Network

Fang-Yu Chu, Chun-Jie Chiu, An-Hung Hsiao, Kai-Ten Feng and Po-Hsuan Tseng*
Department of Electrical and Computer Engineering, National Yang Ming Chiao Tung University, Hsinchu, Taiwan
*Department of Electronic Engineering, National Taipei University of Technology, Taipei, Taiwan
Email: fychu.cm08g@nctu.edu.tw, jack0502801.cm02g@nctu.edu.tw, eric28200732.cm07g@nctu.edu.tw,
ktfeng@mail.nctu.edu.tw and phtseng@ntut.edu.tw

2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring) | 978-1-7281-8964-2/20/$31.00 ©2021 IEEE | DOI: 10.1109/VTC2021-Spring51267.2021.9448848

***Abstract*—Human presence detection via camera-based monitoring systems has been well-adopted in various applications including smart homes, factories, and hospitals. However, its privacy concerns have been raised in many occasions such as daycare centers and homes with elderly living alone. In recent years, literatures adopting wireless signals were proposed to resolve privacy issues for presence detection; nevertheless, existing works can only be applied in a single room scenario. In this paper, we are the first work to propose a device-free multi-room human presence detection system based on efficient star network topology. Our proposed conditional recurrent architecture-based multi-room presence detection (C-MuRP) system extracts both spatial and temporal features from the Wi-Fi channel state information (CSI). Associated with a voting scheme, the proposed novel deep-learning architecture classifies the states of multi-room with the condition on current waveform to emphasize present feature states. Real-time experimental results showed that our proposed C-MuRP system can achieve higher accuracy for multi-room presence detection compared to existing methods.**

I. Introduction

In recent years, smart home has brought great attention in diverse fields, such as home care, home appliances intelligence, and family entertainment. Nevertheless, home security is a crucial issue that involves anti-theft, disaster prevention, biometrics, etc. With the growing concerns on home security, presence detection, which detects humans' presence in a region of interest, gains more attention from the public and more research studies on this subject. Until now, vast device-free presence detection technologies are developed, which are featured with different devices placed at fixed locations such as infrared sensors and cameras for detection. However, the infrared sensors require humans to be in line-of-sight (LoS), having difficulty to notice the presence in non-line-of-sight (NLoS). In addition, camera-based resources suffer from the controversy of privacy and have a blind angle problem. Thence, the wireless signal-based techniques are attractive extensively these years.

Currently, the physical layer (PHY) information channel state information (CSI) has been widely mentioned by researchers, which is a fine-grained [1] channel measurement describing how the signal traverses between a transmitter and a receiver. It resolves the problems above and takes advantage of convenience since we can use CSI on commodity Wi-Fi devices ubiquitously. Adopting these radio frequency-based (RF) signals, several applications yield thereby, such as localization [2], [3], gesture detection [4], fall detection [5] and activity recognition [6], [7]. However, they all placed the access points (APs) close to the users. Moreover, intrusion detection [8] or even crowd counting [9] spring up for environmental control; yet, they all demand the transmitter and the receiver placed in the same room.

Nowadays, learning-based algorithms [10], [11] have been widely explored in the field of analyzing CSI data, based on feature extraction from subcarriers. There are massive learning structures proposed, such as support vector machine (SVM) [12] that uses kernel functions for nonlinear classification, multi-layer perceptron (MLP) [13] with a class of feedforward artificial neural network, and long short-term memory (LSTM) algorithm [14] that is adopted for processing sequences of data. Several indoor CSI related systems based on deep learning for both positioning and presence detection [15] have been proposed.

Our system employs a gated recurrent unit (GRU) network to extract temporal features and incorporates it with a convolutional neural network (CNN) providing the condition of spatial features. In this paper, we proposed a conditional recurrent architecture-based multiple room presence detection (C-MuRP) system, where the main contributions are as follows.

1) We proposed the C-MuRP system, in which a novel conditional recurrent network and multi-room probability voting process are presented to enhance the accuracy.
2) To the best of our knowledge, the proposed C-MuRP is the first to detect presence in a multi-room scenario, under the condition that at most a single AP in a room.
3) The experiments are conducted in real-world environments and can be implemented in real-time. The experimental results show that C-MuRP can achieve high accuracy in presence detection for multi-room.

II. System Architecture

A. Channel State Information

Our system aims to achieve a device-free multi-room presence detection of occupants using commercial off-the-shelf

[1]This work was in part funded by Ministry of Science and Technology (MoST) Grants 107-2221-E-009-058-MY3, 109-2218-E-009-002, Higher Education Sprout Project of the National Yang Ming Chiao Tung University and Ministry of Education (MOE), Taiwan.

978-1-7281-8964-2/21/$31.00 ©2021 IEEE
Authorized licensed use limited to: Siksha O Anusandhan University. Downloaded on February 01,2024 at 05:16:33 UTC from IEEE Xplore. Restrictions apply.

(COTS) Wi-Fi devices. The multiple-input multiple-output (MIMO) system combined with the orthogonal frequency division multiplexing (OFDM) technique is supported from 802.11n. OFDM technique can divide wideband signals into multiple orthogonal subcarriers in the range of system bandwidth. Each subcarrier travels through a narrow-band flat fading channel to combat frequency selective fading in multipath propagation environments. In a typical indoor environment, the Wi-Fi signal propagates through multiple paths from a transmitter (TX) to receiver (RX), experiencing reflection, refraction, scattering, and diffraction caused by occupants, walls, and furniture. CSI provides channel characteristics, containing multipath amplitude attenuation and phase shift for each subcarrier. OFDM technique enables simple least-squares channel estimation for CSI per subcarrier efficiently in the frequency domain. MIMO system allows more pairs of CSI measurements with multiple transmitting and receiving antennas. In the frequency domain, the channel model can be expressed as

$$y_{k,l} = h_{k,l}x_{k,l} + n_{k,l}, \tag{1}$$

where $x_{k,l}$ and $y_{k,l}$ are the transmitted and received signal at k-th subcarrier of l-th antenna pair, respectively. $h_{k,l}$ and $n_{k,l}$ are the channel response and additive white Gaussian noise (AWGN). In the OFDM system, the CSI of the k-th subcarrier and the l-th antenna pair can be estimated as

$$\hat{h}_{k,l} = \frac{y_{k,l}}{x_{k,l}}. \tag{2}$$

The CSI at the receiver for a total of K subcarriers and L antenna pairs can be expressed as

$$\hat{H} = \begin{bmatrix} \hat{h}_{1,1} & \hat{h}_{1,2} & \dots & \hat{h}_{1,L} \\ \hat{h}_{2,1} & \hat{h}_{2,2} & \dots & \hat{h}_{2,L} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{h}_{K,1} & \hat{h}_{K,2} & \dots & \hat{h}_{K,L} \end{bmatrix}. \tag{3}$$

Each element of CSI matrix is modeled as

$$\hat{h}_{k,l} = |\hat{h}_{k,l}|e^{j(\angle \hat{h}_{k,l})}, \tag{4}$$

where $|\hat{h}_{k,l}|$ and $\angle \hat{h}_{k,l}$ denote the amplitude response and phase response of the k-th subcarrier and l-th antenna pair. In this paper, the proposed method is based on the amplitude response of CSI.

B. System Component

Our system architecture is as Fig. 1, to detect the presence in multi-room in an indoor environment. A transmitter and several receivers are deployed in the whole area, and the amount of receivers is determined by the number of rooms. Note that each room is deployed with one Wi-Fi AP. The transmitter sends out RF signals, and the receivers estimate the CSI from the TX-RX channels and then send it to the central server. Based on the raw CSI $\hat{H}$, we obtain the CSI amplitude by taking its absolute values. We conduct normalization for each antenna pair using the maximum and minimum values

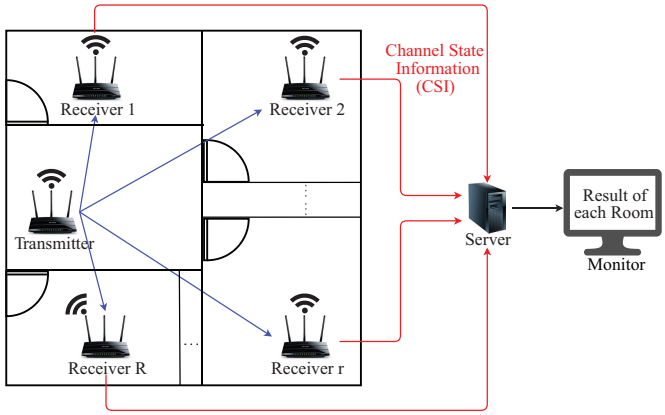


Fig. 1: System architecture for presence detection.

of CSI amplitude from all $K = 56$ subcarriers. The normalized CSI amplitude of the k-th subcarrier of the l-th antenna is

$$|h'_{k,l}| = \frac{|\hat{h}_{k,l}| - \min\limits_{q}|\hat{h}_{q,l}|}{\max\limits_{q}|\hat{h}_{q,l}| - \min\limits_{q}|\hat{h}_{q,l}|}, q = 1, \dots, K, \tag{5}$$

We conduct normalization because distinct power variation of APs results in changes of CSI amplitude in different times, which causes data not unified. Since it is not related to the environment change, we eliminate this variation. Besides, due to our observation that there are specific CSI shape patterns when the environment or the presence is in different situations, we aim to exploit the shape of amplitude. Moreover, the MIMO system provides spatial diversity, so we perform normalization on each antenna pair independently. The normalized CSI amplitude for a total of K subcarriers and P antenna pairs can be expressed as H', where $H' = \left\{ |h'_{k,l}| \middle| k = 1, \dots, K \text{ and } l = 1, \dots, P \right\}$.

To provide the CSI characteristic of the time, by representing the normalized CSI amplitude at the current timestep t as H'_t, we consider a normalized CSI time window at timestep t as

$$\mathbf{H}'_t = [H'_{t-\tau}, H'_{t-\tau+1}, \dots, H'_t], \tag{6}$$

where τ is the length of the considered time window size, which in our system is 50. The merged normalized CSI window can be expressed as $\mathbf{H}'$. Utilizing $\mathbf{H}'$, the central server would then conduct a training process and analyze the signal from different rooms since the received RF signals vary due to distinct furnishings and positions of different rooms. Finally, the result would be displayed on the monitor for the users.

The proposed system block diagram of C-Murp is shown in Fig. 2, which includes the offline and online stages. In the offline stage, there are two inputs in our system, the main input and a condition input. We take the normalized CSI window $\mathbf{H}'$ and the element of $\mathbf{H}'_t$, that is H'_t, to be the inputs and extract their features. In the main part, we extract the *dynamic moving feature* and *dynamic spatial feature*, and in the condition part, we extract the *current spatial feature* to be the model's input.

Authorized licensed use limited to: Siksha O Anusandhan University. Downloaded on February 01,2024 at 05:16:33 UTC from IEEE Xplore. Restrictions apply.

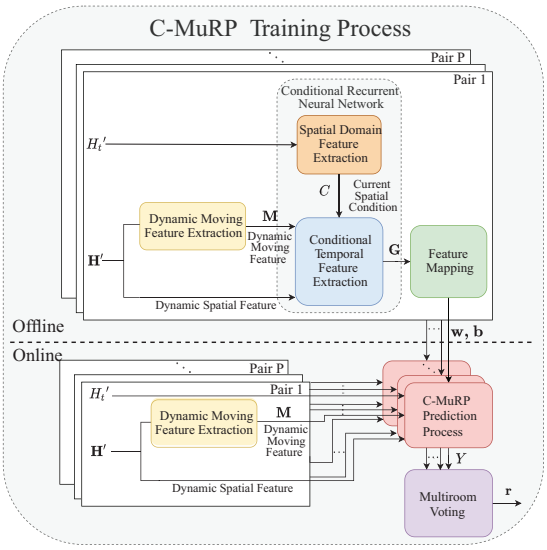


Fig. 2: Block diagram for proposed C-MuRP system.

Next, letting the current spatial feature be the condition while dynamic moving feature and dynamic spatial feature be the main input, we collect the conditional temporal feature and execute feature mapping. Finally, we can get the parameters of the prediction model. In the online stage, the inputs are the same as in the offline stage. After receiving the real-time normalized CSI window, we conduct feature extraction for each input. Again, the dynamic moving feature and dynamic spatial feature be the main input, and the current signal be the condition. We import them into the model trained in the offline stage. We can get the prediction vector containing each antenna pair. Last, exploiting each antenna pair's prediction, we apply the majority vote to get the result whether there are occupants or not for each room.

III. Proposed C-MuRP System

In this section, we will introduce the proposed C-MuRP training process and discuss the significance and benefit of adding conditions to our system.

A. Dynamic Feature Extraction

As Fig. 2 shows, there are several feature extraction blocks in our model, which contains both spatial and temporal features of the input. We extract *dynamic moving feature* by exploiting amplitude shape trend (AST), which can be expressed as

$$\mathbf{M}_t = \mathbf{H}'_t - \mathbf{H}'_{t-1}, \tag{7}$$

where $\mathbf{M}_t = [M_{t-\tau}, \ldots, M_t]$. This feature is obtained by subtracting the previous normalized CSI window from the current normalized CSI window. After this extraction, we can obtain the shape variety of each signal currently, that is, the amplitude change in time. When people are moving in the room, the absolute value of the element of $\mathbf{M}_t$ might be greater. On the contrary, when there are no people in the room, the waveform would be smooth, and the absolute value of the element of $\mathbf{M}_t$ would be smaller. $\mathbf{M}_t$, which is the output signal of AST, can be merged into $\mathbf{M}$ as the dynamic moving feature. Furthermore, we also use the normalized CSI window $\mathbf{H}'$ as the input. The amplitude variation in a duration provides the spatial information with the shape transform whether there is the presence or not in the space. We denote this feature as the *dynamic spatial feature*.

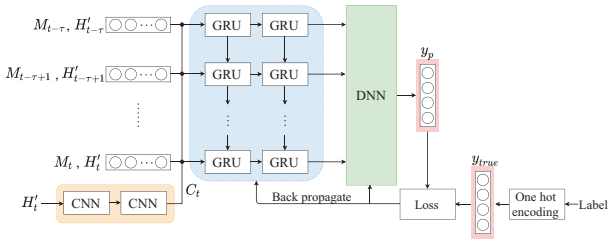


Fig. 3: Proposed conditional recurrent neural network.

B. Conditional Recurrent Neural Network Design

To enhance the accuracy and the stability of our system, we adopt a conditional recurrent neural network. As shown in Fig. 3, for the conditional feature, we extract by two layers of CNN, since CNN conducts convolution on graphs and specific feature detector, also called filter, and outputs a feature map. In this condition part, we first input the current normalized CSI H'_t, which is the shape of the current waveform. After the CNN calculation, the output can be regarded as the spatial feature of the present time. We denote the output of CNN as the current spatial condition C. To input C to GRU, we flatten the feature map and duplicate it τ times to align with the size of the CSI time window, then concatenate it with $\mathbf{M}$ and $\mathbf{H}'$. In this way, the feature at each timestep of a CSI window would have the condition, which is the current spatial feature C. Having the above spatial features, we next extract temporal features. For the conditional temporal feature extraction block, we adopted GRU. Similar to LSTM, it can extract the feature of time and prevent the vanishing gradient problem. Further, GRU has less gate than LSTM, which leads to fewer parameters, and this gives the advantage of lower complexity and reduced training time. Finally, we apply a fully connected layer (FCN) for feature mapping, then we can get the parameters $\mathbf{w}$ and $\mathbf{b}$ of our C-MuRP process, which can be used for online prediction.

In our system, we added a spatial domain feature extraction block, and it helps improve the performance. Since the output of the original model with only a recurrent network is the feature of the comprehensive time duration, it comes out that GRU extracts important features with adjusting weights of the model. However, the block length limited results in that some of the features will be ignored after the signal passes through massive neurons in the hidden layer. For example, assuming that there are two signals with different shapes, it may appear that the output features of the two signals from the recurrent

Authorized licensed use limited to: Siksha O Anusandhan University. Downloaded on February 01,2024 at 05:16:33 UTC from IEEE Xplore. Restrictions apply.

network GRU are the same. In this circumstance, the added information/condition forces the entire historical signals refer to the current state, which can be regarded as the conditional probability of the input. $\mathbf{G}$ is denoted the output of conditional temporal feature extraction block as

$$\mathbf{G} = f(\mathbf{M}, \mathbf{H}'|C), \tag{8}$$

where the function f represents the model of GRU block. The intuition of this condition can be thought of as increasing features since GRU's output with the same cluster may split into several subclusters after adding the condition. Taking an example of two scenarios, one with presence in only the room of RX and the other with presence in both rooms of RX and TX. Assuming that for the two scenarios, the person in the room with RX moves in the same pattern, and with the scenario that there is a presence in the room with TX, the person stands still in the room. For these two scenarios, the feature of the time domain may be the same. However, observing only the amplitude, the shape would significantly differ from that with a person standing there, which is our input of condition. This gives prominence to the importance of the condition. Hence, adding conditions improves the precision of classifying multiple rooms due to the growth in each room's number of features and distinguishes with better results.

C. Multi-Room Voting

According to Fig. 1, we can see that the signal from a pair of transmitter and receiver passes through two rooms; therefore, we could judge the presence of the two rooms through the signal with two APs. However, the room with TX would have a poorer performance since there is a wall between TX and RX. When the CSI signal passes through the wall, a wide CSI variation happens. If the signal varies in the TX room then passes through the wall, the feature would relatively become small comparing to the signal varying after passing the wall. Hence, the feature would become less significant, which affects the prediction. The signal after being influenced by people in the room with TX propagates with a longer distance, which attenuated the feature generated by people. The result probabilities of presence from several TX-RX pairs of our model can be represented as

$$Y = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_p, \ldots, \mathbf{y}_P], \tag{9}$$

where $\mathbf{y}_p$ represents the probability vector of the different cases of the p-th TX-RX pair. With two rooms, there are a total of four cases for $\mathbf{y}_p = [y_{p,1}, y_{p,2}, y_{p,3}, y_{p,4}]^T$, where the subscript number 1 to 4 corresponds to

- case 1: both empty,
- case 2: presence in TX room,
- case 3: presence in RX room,
- case 4: both presence.

For each room, we want to estimate the probability of the binary case, i.e., probability of [empty, presence]. Thus, we conduct a voting scheme by merging probabilities from the above four cases to whether there are people in the room with TX and to that with RX. For the room with RX, the probability becomes

$$\mathbf{V}_p^{RX} = [y_{p,1} + y_{p,2}, y_{p,3} + y_{p,4}]^T, \tag{10}$$

and for the room with TX, the probability is

$$\mathbf{V}_p^{TX} = [y_{p,1} + y_{p,3}, y_{p,2} + y_{p,4}]^T. \tag{11}$$

In this way, the number of cases reduces to two, we can judge the presence of the room with TX and the room with RX, respectively. Since several receivers are sharing the same TX, the room with TX has plenty of predictions from each match, we can exploit voting by averaging those output probabilities to yield the ultimate outcome, that is,

$$\mathbf{V}^{TX} = \frac{1}{P} \sum_{p=1}^{P} \mathbf{V}_p^{TX}. \tag{12}$$

Finding the element (empty or presence) with higher probability in vectors $\mathbf{V}^{TX}$ and $\mathbf{V}_p^{RX}$ for $p = 1, \ldots, P$, respectively, we can get the final prediction result $\mathbf{r}$ of the presence in each room.

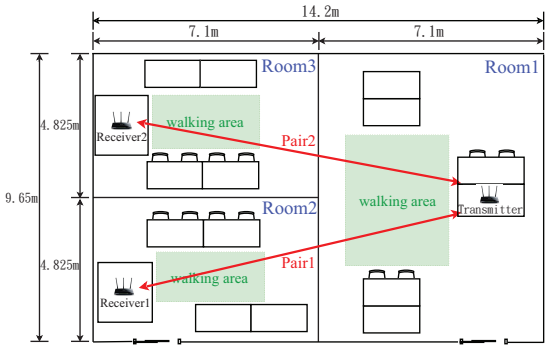
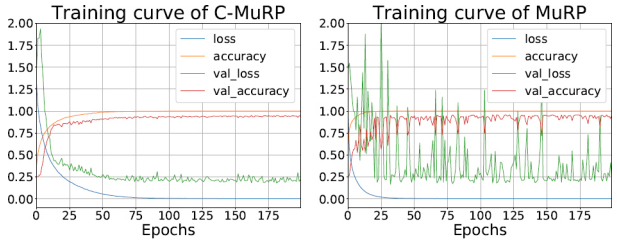


Fig. 4: Experimental setting for the three-room scenario.

IV. Performance Evaluation

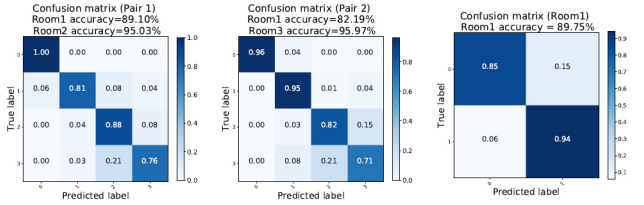
In our C-MuRP system, TL-WDR4300 wireless routers are used as TX and several RXs for implementation. All of the routers are operated in IEEE 802.11n, and each has two antennas, which yields four propagation paths. CSI data is transmitted and collected with central carrier frequency 2.447 GHz and carrier bandwidth 20 MHz. Each TX-RX pair has 56 subcarriers. The experiment environment of a 14.2m x 9.65m meeting room, which contains three compartments, is illustrated in Fig. 4, and the routers were placed on the table of each room. As shown in Fig. 4, there are three rooms in our experiment, which yields eight cases of presence and generates two antenna pairs to predict. The data collection frequency is 10 Hz, and we collected 5000 data each case for training data, and after two hours, we collected 2000 data each case for testing data, then a total of 56000 data was collected, so it took around one and a half hour for collecting data. For each case, the testers walk arbitrarily in the green area shown in Fig. 4, and randomly halt for about eight to ten seconds at random spots.

Authorized licensed use limited to: Siksha O Anusandhan University. Downloaded on February 01,2024 at 05:16:33 UTC from IEEE Xplore. Restrictions apply.



(a) Training curve of C-MuRP. (b) Training curve of MuRP.

Fig. 5: Training curve comparison between C-MuRP and MuRP.



(a) Confusion matrix of antenna pair 1. (b) Confusion matrix of antenna pair 2. (c) Confusion matrix of room 1 after voting.

Fig. 6: Benefit of voting scheme.

To demonstrate the proposed C-MuRP system, we compare our system with the system without adding condition, termed as MuRP, with different factors. Fig. 5 shows the training curves of C-MuRP and MuRP, which expressed the advantage of condition scheme, since the process without condition Fig. 5(b) is unsteady and the model may vary by a large margin between different rounds of training, on the contrary, Fig. 5(a) illustrates a smooth training curve.

Fig. 6 shows the benefits of voting by the results of SVM, which we replace the conditional recurrent neural network block into the SVM algorithm, and it would be a baseline in our final performance. Begin with Figs. 6(a) and 6(b), the confusion matrix prove that when there are people in the room with RX, the system cannot distinguish the cases of whether there are people in the room with TX or not, which confirms the statement above. The accuracy for each antenna pair after merging the probabilities into two cases is shown in Figs. 6(a) and 6(b), and the final voting for room 1 is illustrated in Fig. 6(c). It reveals that despite the poor results for certain rooms, with our voting scheme, the accuracy would be improved. Fig. 7 summarizes the F1-score comparison between C-MuRP and the other algorithms at different experimental rooms. It can be seen that the proposed C-MuRP scheme can provide the highest F1-score compared to MuRP, FCN, and SVM methods.

V. Conclusion

This paper proposes the first multi-room presence detection system, a conditional recurrent architecture-based multiple room presence detection (C-MuRP) cooperating with a voting scheme. Exploiting the amplitude of CSI for all subcarriers in 2x2 MIMO OFDM systems, we utilize both spatial and temporal information. In the spatial stage, we extract the dynamic moving feature and dynamic spatial feature. For

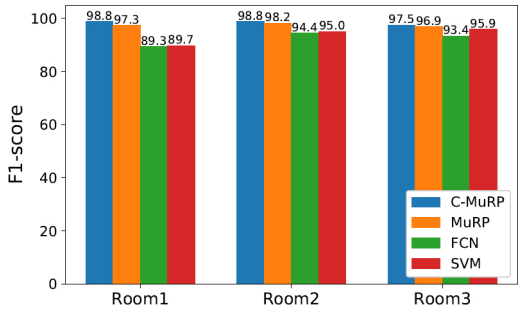


Fig. 7: F1-score of C-MuRP, MuRP, FCN and SVM.

the condition input, we extract the spatial domain feature. Then, in the temporal stage, we extract conditional temporal features in our training process. To further enhance accuracy for multi-room, we added a voting scheme that averages the probabilities. It is shown in the experiments that our proposed C-MuRP can achieve the best performance in multi-room presence detection compared to other existing schemes.

References

[1] K. Wu, J. Xiao, Y. Yi, Min Gao, and L. M. Ni, "FILA: Fine-Grained Indoor Localization," in *Proc. IEEE International Conference on Computer Communications*, 2012, pp. 2210–2218.

[2] H. Tsai, C. Chiu, P. Tseng, and K. Feng, "Refined Autoencoder-Based CSI Hidden Feature Extraction for Indoor Spot Localization," in *Proc. IEEE Vehicular Technology Conference*, 2018, pp. 1–5.

[3] K. Lu, C. Chiu, K. Feng, and P. Tseng, "Device-Free CSI-Based Wireless Localization for High Precision Drone Landing Applications," in *Proc. IEEE Vehicular Technology Conference*, 2019, pp. 1–5.

[4] J. Yang, H. Zou, Y. Zhou, and L. Xie, "Learning Gestures From WiFi: A Siamese Recurrent Convolutional Architecture," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10 763–10 772, 2019.

[5] Y. Wang, K. Wu, and L. M. Ni, "WiFall: Device-Free Fall Detection by Wireless Networks," *IEEE Transactions on Mobile Computing*, vol. 16, no. 2, pp. 581–594, 2017.

[6] J. Yang, H. Zou, H. Jiang, and L. Xie, "Device-Free Occupant Activity Sensing Using WiFi-Enabled IoT Devices for Smart Homes," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3991–4002, 2018.

[7] J. Ding and Y. Wang, "WiFi CSI-Based Human Activity Recognition Using Deep Recurrent Neural Network," *IEEE Access*, vol. 7, pp. 174 257–174 269, 2019.

[8] Z. Tian, Y. Li, M. Zhou, and Z. Li, "WiFi-Based Adaptive Indoor Passive Intrusion Detection," in *Proc. IEEE International Conference on Digital Signal Processing (DSP)*, 2018, pp. 1–5.

[9] H. Zou, Y. Zhou, J. Yang, W. Gu, L. Xie, and C. Spanos, "FreeCount: Device-Free Crowd Counting with Commodity WiFi," in *Proc. IEEE Global Communications Conference*, 2017, pp. 1–6.

[10] X. Wang, L. Gao, S. Mao, and S. Pandey, "CSI-Based Fingerprinting for Indoor Localization: A Deep Learning Approach," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 1, pp. 763–776, 2017.

[11] Y. Zhang, D. Li, and Y. Wang, "An Indoor Passive Positioning Method Using CSI Fingerprint Based on Adaboost," *IEEE Sensors Journal*, vol. 19, no. 14, pp. 5792–5800, 2019.

[12] R. Zhou, X. Lu, P. Zhao, and J. Chen, "Device-Free Presence Detection and Localization With SVM and CSI Fingerprinting," *IEEE Sensors Journal*, vol. 17, no. 23, pp. 7990–7999, 2017.

[13] C. Hsieh, J. Chen, and B. Nien, "Deep Learning-Based Indoor Localization Using Received Signal Strength and Channel State Information," *IEEE Access*, vol. 7, pp. 33 256–33 267, 2019.

[14] Y. Zhang, C. Qu, and Y. Wang, "An Indoor Positioning Method Based on CSI by Using Features Optimization Mechanism With LSTM," *IEEE Sensors Journal*, vol. 20, no. 9, pp. 4868–4878, 2020.

[15] Y. Huang, A. Hsiao, C. Chiu, K. Feng, and P. Tseng, "Device-Free Multiple Presence Detection Using CSI with Machine Learning Methods," in *Proc. IEEE Vehicular Technology Conference*, 2019, pp. 1–5.

Authorized licensed use limited to: Siksha O Anusandhan University. Downloaded on February 01,2024 at 05:16:33 UTC from IEEE Xplore. Restrictions apply.