Learning Human Activities through Wi-Fi Channel State Information with Multiple Access Points

著者	LI He, OTA Kaoru, DONG Mianxiong, GUO Minyi
journal or publication title	IEEE Communications Magazine
volume	56
number	5
page range	124-129
year	2018-05-17
URL	http://hdl.handle.net/10258/00009914

doi: info:doi/10.1109/MCOM.2018.1700083

Learning Human Activities through Wi-Fi Channel State Information with Multiple Access Points

He Li, *Member, IEEE,* Kaoru Ota, *Member, IEEE,* Mianxiong Dong, *Member, IEEE,* Minyi Guo, *Senior Member, IEEE*

***Abstract*—Wi-Fi channel state information (CSI) provides adequate information for recognizing and analyzing human activities. Because of the short distance and low transmit power of Wi-Fi communications, people usually deploy multiple access points (APs) in a small area. Traditional Wi-Fi CSI based human activity recognition methods adopt Wi-Fi CSI from a single AP, which is not so appropriate for a high-density Wi-Fi environment. In this paper, we propose a learning method that analyzes the CSI of multiple APs in a small area to detect and recognize human activities. We introduce a deep learning model to process complex and large CSI information from multiple APs. From extensive experiment results, our method performs better than other solutions in a given environment where multiple Wi-Fi APs exist.**

***Index Terms*—Wi-Fi Channel State Information (CSI), Deep Learning, Human Activity Recognition**

I. Introduction

Human activity recognition is important to many applications, such as crowdsensing, social networks, and recommendation systems [1]. In recent years, recognizing human activities through Wi-Fi channel state information (CSI) has become an emerging technology for activity recognition [2]. As human body may reflect or absorb Wi-Fi signals in a small area, human activities will change the Wi-Fi CSI during Wi-Fi communications. Many works attempt to use the information of Wi-Fi signals to find out various human activities such as walking, sitting and micro activities [3].

Usually, there are several ways to recognize human activities with Wi-Fi CSI. A typical way is to build a signal transmit model to calculate the human movement through the information of Wi-Fi signals [2]. This method is efficient for a simple Wi-Fi environment with a single access point (AP) and given barriers. Unfortunately, as people have to deploy a large number of Wi-Fi APs to cover the communication area, the Wi-Fi environment is very dense and complex. It is very hard to formulate existing Wi-Fi environment through an intuitive mathematical model.

Learning is another way to find out valuable information from complex and large datasets [4]. In Wi-Fi CSI, signal strength is usually visualized into figures, which can be analyzed by many state-of-the-art artificial intelligence techniques. Among all the existing learning methods, deep learning is preferred on the image tasks due to its strong ability of extracting features from a large amount of data. In human activity recognition, deep learning is also considered as a suitable method to find out different human activities from various sensing data. Therefore, we introduce deep learning model for recognizing human activities from Wi-Fi CSI data.

First, we develop a system to capture the Wi-Fi CSI data from all connectable APs. In our work, the data collection is performed in an indoor environment where multiple APs exist. During this process, the corresponding human activities are also recorded to label the CSI data. Then, we design a data structure to encode and storage the collected data. In order to analyze the relationship between CSI and human activity, we design and implement a multi-layer convolutional neural network (CNN) model. In experiments, we train our deep learning network with labeled Wi-Fi CSI datasets and test its actual recognition performance. We also compare our method with a support vector machine (SVM) based learning method.

The main contributions of this paper are summarized as follows.

- We first study the human activity recognition problem with Wi-Fi CSI in environment where multiple APs exist. To the best of our knowledge, this is the first work to recognize human activities with multiple APs.
- We then introduce a deep learning model to recognize human activities from complex and large Wi-Fi CSI datasets. We design an appropriate input data structure and a multi-layer CNN for activity recognition.
- We take the performance evaluation of the deep learning based method with extensive experiments in a given environment. We also compare our recognition method with a traditional SVM based learning method.

The remainder of this paper can be outlined as follows. Section II introduces human activity recognition with Wi-Fi CSI and deep learning technologies in human activity recognition. Section III introduces the scenario of human activity recognition with Wi-Fi CSI in a multi-AP environment. Section IV describes the input data structure and the deep learning network. Section V presents the evaluation results of the deep learning based human activity recognition through extensive experiments, followed by the conclusions drawn in Section VI.

II. Related Work

In this section, we first introduce some previous works in CSI-based human activity recognition, then discuss deep learning in human activity recognition.

He Li, Kaoru Ota and Mianxiong Dong are with Muroran Insitute of Technology, Japan.

Minyi Guo is with Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China.

A. Human activity recognition with Wi-Fi CSI

Radio based human activity recognition is a method to detect the changing of radio information as human body will reflex or absorb radio waves. As there are many methods for radio based human activity recognition, Wi-Fi radio based activity recognition is an emerging solution that uses existing network infrastructure. In most recent research works, people can use pre-installed wireless network adapters to retrieve enough information from Wi-Fi APs for activity recognition.

In Wi-Fi radio based activity recognition, earlier works usually use received signal strength indication (RSSI), which is considered as a single amplitude measurement of signal information [5]. Comparatively, Wi-Fi CSI provides phase information as well as amplitude measurement for each sub-carrier, which provides higher accuracy than RSSI based methods. Wang et al. [2] first introduced fine-grained Wi-Fi CSI to recognize human activities in home environments. For recognizing home activities, they collected Wi-Fi CSI data from a given AP and modeled the radio transferring with human activities.

Wi-Fi CSI based recognition is also able to detect human activities behind the walls. Adib et al. [6] proposed a recognition solution through the interference nulling in multiple input multiple output (MIMO) Wi-Fi system to find out the human activity behind walls. The recognition method adopts a single AP with two antennas and measures nulled signals to find out the reflections off in transmitting. An interesting idea in the recognition method is a technique named inverse synthetic aperture radar is introduced to track moving objects.

Signal preprocessing can improve the accuracy of activity recognition through Wi-Fi CSI. Pu et al. [7] presented a signal preprocessing that transforms received signals into narrowband pulse to detect a Doppler shift in small-scale human activities such as hand gestures. The authors repeated the orthogonal frequency division multiplexing (OFDM) symbol in the received signal than performed fast Fourier transform (FFT).

These works focused on the simple Wi-Fi environment in which there is only one AP transferring signals without references. In a high-density Wi-Fi environment, as Wi-Fi CSI is interfered by complex wireless radios, the accuracy of these methods will also be influenced.

B. Deep learning in human activity recognition

Human activity recognition is an emerging application of deep learning technologies. Many works focus on activity recognition with different deep learning models. In human activity recognition, there are two categories of methods including unimodal methods and multimodal methods. Unimodal methods recognize human activities from single modality data while multimodal recognize activities from multiple modality data. Deep learning is adopted in both unimodal and multimodal methods for human activity recognition. As one of the unimodal methods, Rahmani et al. [8] proposed a deep learning based action recognition from multiple camera views. The authors introduced learning model to find out semantic relationships from images of multiple camera views and map images into a single view.

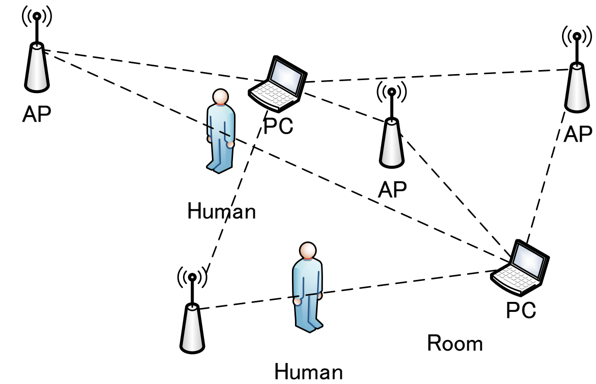


Fig. 1. Wi-Fi CSI based human activity recognition with multiple APs

As one type of unimodal methods, shape-based methods first model the motion of human body parts then efficiently recognize human activities. In a shape-based method, deep learning is usually used for estimating human poses from multisource data. Toshev et al. [9] adopted deep neural networks (DNN) regression for pose estimates in a holistic manner.

Deep learning plays an important role in multimodal methods. Martinez et al. [10] adopted deep learning to explore the most informative features for recognition of human affective states. Deep learning performs efficiently for activity recognition based on social networking.

Kim et al. [11] introduced deep belief networks (DBM) to recognize human emotions from audio datasets. They adopted DBM in both supervised and unsupervised models to acknowledge non-linear relationships. Shao et al. [12] applied deep learning to combine different modalities such as appearance and motion features collected from social networks. Gan et al. [13] proposed an approach for recognizing human events in videos by introducing deep learning to identify the most important frames in video sequences.

Deep learning is an efficient way to recognize human activities in complex information, which is also appropriate for exploring complex Wi-Fi information data from multiple APs.

III. Wi-Fi CSI based Human Activity Recognition with Multiple APs

In this section, we will discuss a typical scenario of the human activity recognition with multi-AP CSI. We first use an example to introduce the scenario, then address the problems for human activity recognition.

People often use many Wi-Fi APs for better signal quality and network performance. Thus, in multiple Wi-Fi environments, there are more than one APs near to the computers or devices. As shown in Fig.1, there are five APs and two personal computers. All computers can receive signals from all five APs. We assume all computers can record the Wi-Fi CSI from each AP, and the interferences from human activities can be sensed.

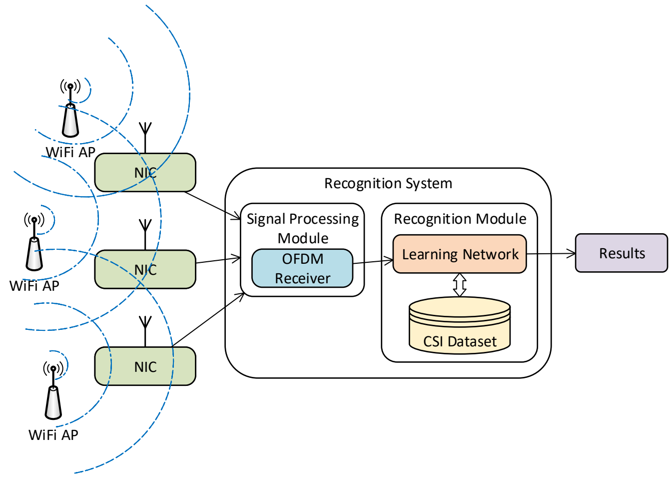


Fig. 2. Structure of Wi-Fi CSI based human activity recognition system with multiple APs

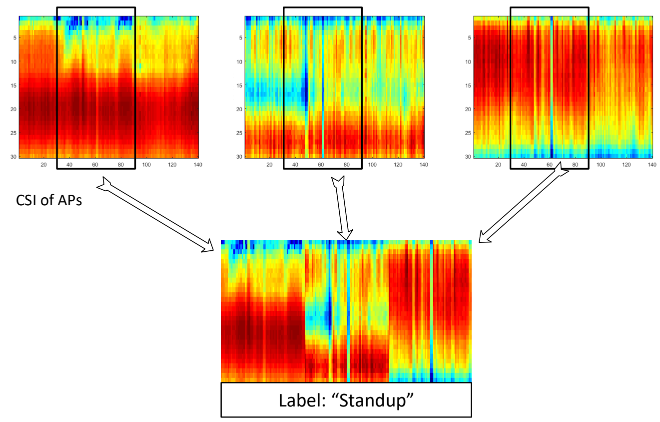


Fig. 3. Labeled Wi-Fi CSI data with multiple APs during a people stands up

In the scenario, two people absorb and reflect a part of Wi-Fi radio from APs to computers. When a people moves its body, Wi-Fi CSI will be influenced since the absorbed or reflected radio is changed. Wi-Fi CSI based activity recognition means recognizing different Wi-Fi CSI changing during human activities. As there have been several tools for recording CSI data in general Wi-Fi hardware, it is possible to sense Wi-Fi CSI data for activity recognition. For a single AP environment, the problem is to find an efficient way for recognizing activities in Wi-Fi CSI information. In a multiple APs environment, sensing Wi-Fi CSI data from a single AP will be interfered by other Wi-Fi communications. Fortunately, network adapters can also sense additional information from other APs. Therefore, to improve the recognition accuracy in the scenario, we design an activity recognition system that obtains Wi-Fi CSI from multiple APs, and based on that, recognizes the corresponding human activities. As shown in Fig. 2, we add an OFDM receiver in the signal processing module for receiving Wi-Fi CSI from multiple active APs. In our system, we combine the multiple Wi-Fi CSI data into the input data structure for training our deep learning model. We will introduce the input data structure in next section.

There are two phases in deep learning based recognition, namely, the training phase and inference phase. For training phase, we collect the Wi-Fi CSI data and label them with the corresponding human activities. For example, if we record the Wi-Fi CSI data during a person stands up, the label of this combined dataset will be set as "Standup". We maintain a knowledge dataset for all labeled data. In the training phase, the system will load the combined data from the Wi-Fi CSI dataset and train the learning network. In the inference phase, the system will recognize different human activities from the input Wi-Fi CSI information using the learning network.

IV. Learning Model

In this section, we first present the data structure in the training phase, then introduce the proposed learning model.

There are different Wi-Fi CSI data from multiple APs. We have to combine them into one data structure for network training. A Wi-Fi CSI record includes tens of subcarriers with amplitudes in thousands of time slots. For example, if we can receive signals from multiple APs, we can choose a multiple-tuples data structure to combine these multiple records. We use an example to show a typical data structure shown in Fig. 3, in which we record Wi-Fi CSI data from multiple different APs.

We first transform the values in each record into heatmaps. For the Wi-Fi CSI record, the x-axis is set to the packet index, the y-axis is set as subcarrier index, and the colors mean the amplitude values. We use a time window to clip the Wi-Fi CSI data during human activities. In the collection of our training data, we use a camera to recode the human activities in the same timeline of Wi-Fi CSI records. Then, from the video record, we label all Wi-Fi CSI data in the time windows. Then, we combine multiple heatmaps and the label as the input data structure for training the deep learning network.

With the input data from multiple APs, we design a learning network with multiple inputs and one output. As shown in 4, we design a CNN-based learning network. There are six layers in the learning network and multiple branches for multiple input data. The Wi-Fi CSI records provide fine-grain amplitude values of all subcarriers, which contain many features. Therefore, we use three convolutional layers for extracting features in Wi-Fi CSI data. After feature extraction in convolutional layers, we use two fully connected layers as training layers. Finally, we use a softmax loss function as the classifier for the output of the neural network.

In the training phase, training records will be sliced into multiple parts including all Wi-Fi CSI Input. We have set several different activities to label the CSI data, including standing up, sitting down, hand raising, hand moving, and the default label. The default label is used for any other activities which we have not defined in our work, including no activity of course. Related labels will be put into the softmax classifier.

In the inference phase, we use the sliding window to clip a part of the input Wi-Fi CSI data. The CNN-based network will learn the features of the input data and recognize the label. If the target perform any static activities or some activities not considered in the training stage, the proposed CNN model

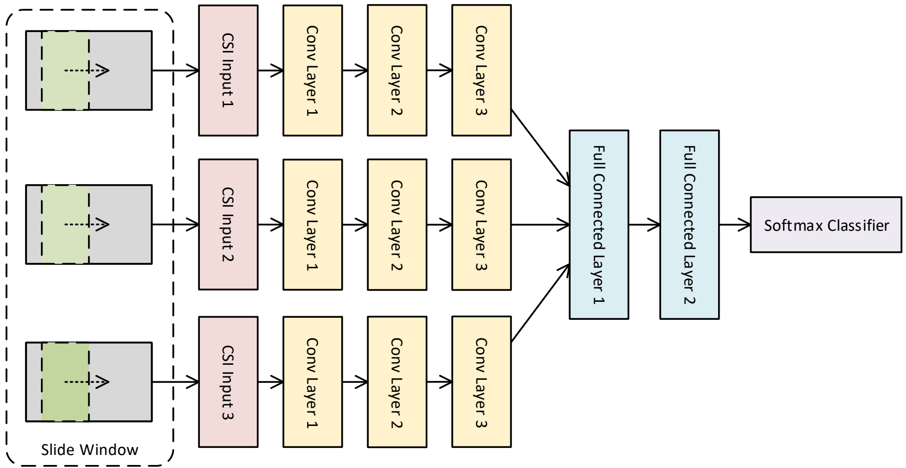
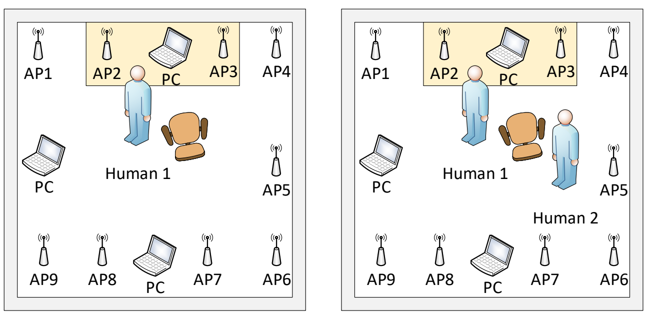


Fig. 4. Learning network in Wi-Fi CSI based human activity recognition system with multiple APs



(a) One people (b) Two people

Fig. 5. Recognition accuracy with different number of people in the same room

will output the default label. Then, the recognition system continues moving the sliding window with a time interval and recognizes the rest of Wi-Fi CSI data. With the sliding window, the recognition system also supports recognize human activities from the streaming Wi-Fi CSI data. By this means, this CNN model can correctly recognize the pre-defined activities without a detection model.

V. Evaluation

In this section, we present the evaluation results of our activity recognition system. We first introduce setting of the experiments and the data collection in our laboratory. Then, we discuss the accuracy of our system and compare it with a traditional SVM based activity recognition method.

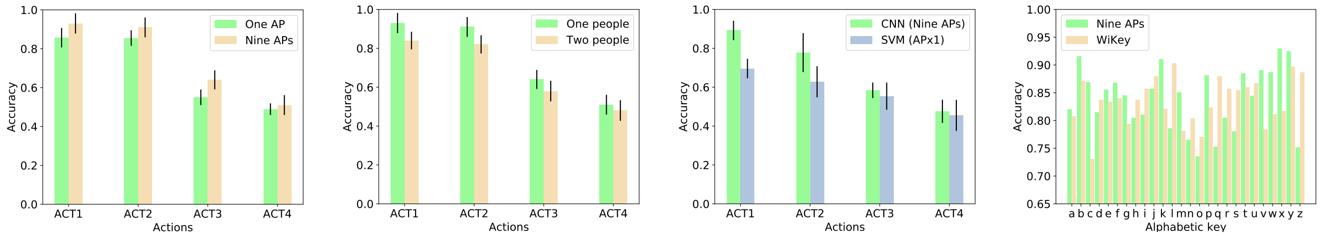
We use the Linux 802.11n CSI Tool [14] to record the Wi-Fi CSI data from Wi-Fi network adapters. We use three Dell E6230 laptop computers which have three Mini PCI Express slots as the receiver and install three Intel Wi-Fi Link 5300 AGN network adapters in the receiver. We put nine Wi-Fi routers denoted by "AP1", "AP2", ..., and "AP9" in the same room and connect each network adapter to one router. We use Ubuntu 14.04 with Linux Kernel 4.10 as the operating system of the computer.

For our activity recognition system, we use a workstation computer from G-Tune, a high-performance computer manufacturer in Japan. The computer equips a Nvidia GeForce GTX 1080 graphics card, an Intel Core i7-7700K processor, 32 GB memory and a 8 TB hard disk. We install Ubuntu 16.04.2 LTS as the operating system and adopt Keras and TensorFlow as the development framework. As shown in Fig. 5, for testing the recognition performance with different number of people, we use two different settings in our experiments, which are denoted by "One people" and "Two people", shown in 5(b) and 5(a), respectively.

Data collection is very important to our experiments. We use a camera to record the activity of the people in the room and recode the CSI with the same timeline. Then, we choose several activity types for testing, and find out the time periods and activity types according to the video. We choose standing up, sitting down, hand raising, and hand moving as the activity types, which are denoted by "ACT1", "ACT2", "ACT3", and "ACT4", respectively.

We first record the training data, which consists of 180 times of standing up and sitting down, 300 times of hand raising, and 900 times of hand moving. We set the period of the time window as 1 second. For evaluation, we first choose three-fourths of the entire records as training data and the rest one-fourth data as the testing data. We also test the accuracy of our recognition system with only AP1. Meanwhile, for performance comparison, we also test a SVM-based recognition method [15] in the experiments. All experiments are executed 20 times, and we record the mean value and errors.

We conduct another test for keystroke recognition, also as a comparison with an existing method [3]. The difference is, in this test, we use an automatic key record program to label the captured CSI data, instead of the former adopted video analysis approach. Therefore, the dataset size is significantly increased due to the high labeling efficiency. We recorded 4000 times per each alphabetic key and three-fourths of the entire records as training data and the rest one-fourth data as the testing data.



(a) Different number of APs (b) Different number of people (c) Different number of people (d) Keystroke

Fig. 6. Accuracy with different receiver positions and methods

As shown in Fig. 6(a), we first test the recognition accuracy with different number of APs. It can be seen that the recognition system achieves different accuracy when recognizing different activities. Since the Wi-Fi CSI varies obviously with body movements, the recognition system performs much better in these scenarios than recognizing the hand movements. Meanwhile, changing the number of APs will influence the accuracy of activity recognition. From the experiment results, we can see that the recognition accuracy with 9 APs is better than the accuracy with a single AP.

We also test the recognition accuracy with different number of people in the room. As shown in Fig. 6(b), the recognition system achieves different accuracy with different number of people in the room. An increasing number of people will decrease the accuracy of activity recognition because another people will reflect or absorb Wi-Fi signals. From the experiment results, we can see that the recognition accuracy with two people is worse than the accuracy w1ith one people.

We compare the performance of our method with a SVM-based recognition method. We choose the settings of one people in the comparison experiment. We use “CSI (Nine APs)” and “SVM” to denote our solution with multiple APs and the SVM based recognition, respectively. As shown in Fig.6(c), two solutions achieve different accuracy of recognizing four activity types. Our solution outperforms the SVM-based method in recognizing large scale body movements. In recognizing hand movements, our solution has a similar performance with the SVM-based recognition. Meanwhile, with more Wi-Fi CSI data from multiple APs, the recognition accuracy of our solution improves obviously. The results demonstrate that, a well-designed CNN model, which is based on the CSI data of multiple APs, can be an efficient approach to recognize human activities in complicated wireless environments.

We compare the performance of our method with WiKey, a previous work on keystroke recognition. We choose the settings of one people in this experiment. As shown in Fig. 6(d), two solutions have different recognition accuracy for each alphabetic key. Our solution achieves an average recognition accuracy of 83.98% with 26 alphabetic keys, which is a little better than WiKey’s with an average accuracy of 83.46%. Meanwhile, the accuracy of keystroke recognition is higher than small-scale human activity recognition, which means our solution performs better with bigger training dataset.

VI. Conclusion

In the paper, we propose a human activity recognition method using multiple APs’ CSI data. Instead of complex mathematical models and analysis approaches, we adopt the state-of-the-art deep learning method in our research. We transform the Wi-Fi CSI data into heatmaps, and regard them as the input of our specially-designed deep network. Due to its strong abilities of feature extraction, the proposed CNN model is very suited to the CSI-based human activity recognition task. In addition, GPU acceleration also gives a huge boost to the learning and inference process, which is of great help for the actual application of the proposed system. The experiment results show that our solution is accurate and efficient in recognizing human activities, especially the large-scale body movements. Future work includes implementing a streaming based method to recognize real-time human activities. We also plan to introduce our system in a cloud environment to provide cloud service for human activity recognition.

Acknowledgments

This work is supported by JSPS KAKENHI Grant Number JP16K00117, JP15K15976, JP17K12669, KDDI Foundation, and Research Fund for Postdoctoral Program of Muroran Institute of Technology.

References

[1] B. Guo, H. Chen, Q. Han, Z. Yu, D. Zhang, and Y. Wang, “Worker-contributed data utility measurement for visual crowdsensing systems,” *IEEE Transactions on Mobile Computing*, vol. 16, no. 8, pp. 2379–2391, 2016.

[2] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, “Understanding and modeling of wifi signal based human activity recognition,” in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, ser. MobiCom ’15. New York, NY, USA: ACM, 2015, pp. 65–76.

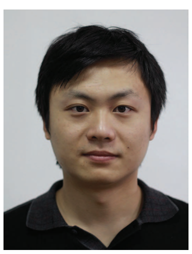
[3] K. Ali, A. X. Liu, W. Wang, and M. Shahzad, “Keystroke recognition using wifi signals,” in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, ser. MobiCom ’15. New York, NY, USA: ACM, 2015, pp. 90–102.

[4] 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, Feb 2017.

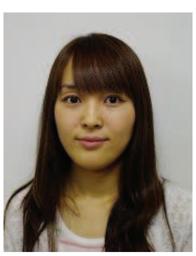
[5] S. Sigg, U. Blanke, and G. Trster, “The telepathic phone: Frictionless activity recognition from wifi-rssi,” in *2014 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, March 2014, pp. 148–155.

[6] F. Adib and D. Katabi, *See through walls with WiFi!* ACM, 2013, vol. 43, no. 4.

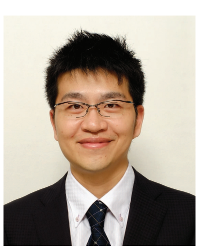
[7] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *Proceedings of the 19th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '13. New York, NY, USA: ACM, 2013, pp. 27–38.
[8] H. Rahmani and A. Mian, "Learning a non-linear knowledge transfer model for cross-view action recognition," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.
[9] A. Toshev and C. Szegedy, "Deeppose: Human pose estimation via deep neural networks," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, ser. CVPR '14. Washington, DC, USA: IEEE Computer Society, 2014, pp. 1653–1660.
[10] H. P. Martnez, G. N. Yannakakis, and J. Hallam, "Don't classify ratings of affect; rank them!" *IEEE Transactions on Affective Computing*, vol. 5, no. 3, pp. 314–326, July 2014.
[11] Y. Kim, H. Lee, and E. M. Provost, "Deep learning for robust feature generation in audiovisual emotion recognition," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 3687–3691.
[12] J. Shao, K. Kang, C. C. Loy, and X. Wang, "Deeply learned attributes for crowded scene understanding," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 4657–4666.
[13] C. Gan, N. Wang, Y. Yang, D.-Y. Yeung, and A. G. Hauptmann, "Devnet: A deep event network for multimedia event detection and evidence recounting," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 2568–2577.
[14] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11n traces with channel state information," *ACM SIGCOMM CCR*, vol. 41, no. 1, p. 53, Jan. 2011.
[15] D. Zhang, H. Wang, Y. Wang, and J. Ma, *Anti-fall: A Non-intrusive and Real-Time Fall Detector Leveraging CSI from Commodity WiFi Devices*. Cham: Springer International Publishing, 2015, pp. 181–193.



He Li received the B.S., M.S. degrees in Computer Science and Engineering from Huazhong University of Science and Technology in 2007 and 2009, respectively, and Ph.D. degree in Computer Science and Engineering from The University of Aizu in 2015. He is currently a Postdoctoral Fellow with Department of Information and Electronic Engineering, Muroran Institute of Technology, Japan. His research interests include cloud computing and software defined networking.



Kaoru Ota was born in Aizu Wakamatsu, Japan. She received M.S. degree in Computer Science from Oklahoma State University, USA in 2008, B.S. and Ph.D. degrees in Computer Science and Engineering from The University of Aizu, Japan in 2006, 2012, respectively. She is currently an Assistant Professor with Department of Information and Electronic Engineering, Muroran Institute of Technology, Japan. She serves as an editor for IEEE Communications Letter.



Mianxiong Dong received B.S., M.S. and Ph.D. in Computer Science and Engineering from The University of Aizu, Japan. He is currently an Associate Professor in the Department of Information and Electronic Engineering at Muroran Institute of Technology, Japan. Dr. Dong serves as an Editor for IEEE Communications Surveys and Tutorials, IEEE Network, IEEE Wireless Communications Letters, IEEE Cloud Computing, and IEEE Access.

Minyi Guo received the BSc and ME degrees in computer science from Nanjing University, China, and the PhD degree in computer science from the University of Tsukuba, Japan. He is currently a Zhiyuan chair professor and a chair of the Department of Computer Science and Engineering, Shanghai Jiao Tong University (SJTU), China. He received the national science fund for distinguished young scholars from NSFC in 2007.