Door-Monitor: Counting In-and-Out Visitors With COTS WiFi Devices

Yanni Yang, Jiannong Cao, *Fellow, IEEE*, Xiulong Liu, and Xuefeng Liu

***Abstract*—Visitor counting can be attractive to various applications, like business management and marketing investigation. Recently, many studies have employed wireless signals to achieve visitor counting without people's active participation and privacy intrusion. However, existing systems mainly count the overall visitors inside a certain area, which fails to provide the fine-grained information of the coming and leaving visitor flow. Unlike previous studies, this article proposes to count the in-and-out visitors to monitor visiting frequency and population, which can be applied for many indoor places, such as shops and restaurants. Therefore, we present the first WiFi-based in-and-out visitor counting system, Door-Monitor, which obtains the direction (enter or exit) and the number of visitors passing by the door. The WiFi signals enable us to count the visitors in a low-cost and nonintrusive way, and it can tell the exact number of visitors even when multiple persons pass by the door simultaneously. To detect the visitors' passing direction, we show that the patterns in the phase difference series can indicate the entering and exiting passing directions by analyzing the effects of the passing behavior on the signal's phase information. To count the passing visitors, we perform a short time Fourier transformation on the phase difference series to generate the spectrogram, on which the convolutional neural network is applied for building a counting model. The experimental results show that the average accuracies of passing direction detection and visitor counting are 95.2% and 94.5%, respectively.**

***Index Terms*—Phase difference, spectrogram, visitor counting, WiFi signals.**

I. Introduction

INTERNET-OF-THINGS (IoT)-enabled smart building can improve facility management and occupancy safety, streamline business processes, and expand profits [2], [3]. Counting visitors in the building provides the basic information to achieve better visitor analysis and control. Meanwhile, observing the presence and population of visitors enables a deep and round look into the business of many applications [4], [5]. For example, retailers want to know how many customers show up inside their shops as a reflection of the promotion effects or sales performance. A shop manager would like to adjust the number of shop assistants according to the visitors' population to save the labor cost. In addition to counting the overall visitors, the number and frequency of the dynamic visitors coming in and out of places are also important indicators to show the popularity. More specifically, detecting the incoming and outgoing visitors can provide fine-grained information of the visitor flow, e.g., how many times a place is visited and the number of visitors of each round of presence, while counting the overall people inside the area would miss this kind of information.

Therefore, this article proposes to use the commercial off-the-shelf (COTS) WiFi devices to address the problem of counting the in-and-out visitors passing by a door, which is stated as follows. We assume that the door area is limited so that only visitors with the same direction can pass at one time. The WiFi transmitters and receivers are deployed aside the door of a room, and WiFi signals are collected continuously when visitors pass by the door. Counting the in-and-out visitors is to detect the passing direction (enter or exit) and count the number of visitors passing by the door simultaneously. Then, the frequency of the people visiting the place and the exact number of visitors can be obtained.

To count the visitors dynamically moving in and out of the door of a room, we propose the Door-Monitor. We leverage the WiFi signals which can detect human presence without intruding people's privacy and only rely on the existing indoor WiFi infrastructure. Thus, visitor counting can be achieved in a nonintrusive and low-cost way. The proposed system is different from the previous studies that only count the people inside the room. Our system can be applied to many indoor scenarios, e.g., retail stores and restaurants in the shopping mall, to count in-and-out visitors, and analyze the detailed dynamics of the customer flow.

In our system, we first detect the passing direction of the visitors, i.e., entering or exiting the door, to tell the coming and leaving frequency. By analyzing the effects of the passing behavior on the wireless signals' phase information, we find that the passing direction can be indicated by observing the increasing and decreasing trend in the phase difference series. WiFi transmitters and receivers are deployed around the door to extract the phase information of WiFi signals. Then, the passing direction is inferred by extracting the patterns in the phase difference series. Meanwhile, there can be times when

Manuscript received January 14, 2019; revised August 11, 2019 and October 22, 2019; accepted November 9, 2019. Date of publication November 15, 2019; date of current version March 12, 2020. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB1004801, in part by Hong Kong RGC Research Impact Fund under Grant R5034-18, and in part by the Shenzhen Basic Research Funding Scheme under Grant JCYJ20170818104222072. This work was presented in part at the 27th International Conference on Computer Communications and Networking (ICCCN), Hangzhou, China, July 30–August 2, 2018. *(Corresponding author: Yanni Yang.)*

Y. Yang and J. Cao are with the Department of Computing, Hong Kong Polytechnic University, Hong Kong (e-mail: 17901076r@connect.polyu.hk).

X. Liu is with the College of Intelligence and Computing, Tianjin University, Tianjin 300000, China.

X. Liu is with the Department of Computer Science, Beihang University, Beijing 100083, China.

Digital Object Identifier 10.1109/JIOT.2019.2953713

multiple visitors pass by simultaneously. Through our two-day observation in a shopping mall, we find that two or three persons are more likely to hang out together. Thus, our second target is to accurately count the exact number of visitors passing by the door together. Here, a counting model is built after extracting the features from phase difference series, and then it will classify and tell the number of visitors passing by at a time.

To achieve the visitor counting system, two technical challenges remain to be solved. The first challenge is to accurately detect the passing direction from the noisy phase difference series. The phase information extracted from the commodity WiFi devices contains different sources of noises, including the phase difference ambiguity and random noises. This makes the observed passing pattern unclear and inconsistent for further analysis. Besides, the human body is not a fix-shape reflector while walking. The movement of legs and arms can result in fluctuations in the received phase values, adding difficulties to extract the explicit pattern for detecting the passing direction. To overcome this problem, the calibration on the phase difference series is first performed. To eliminate the phase difference ambiguity, we design an ambiguity removal method based on the clustering algorithm to recover the original phase difference series. Next, the Savitzky–Golay filter is applied to the phase difference series to remove the adverse effects of the random noises and signal fluctuations for passing direction detection.

The second challenge is to extract features from the WiFi signals and build a robust counting model. For counting the passing visitors, we first need to extract features from the WiFi signals to reflect the distinctive patterns of different numbers of passing visitors. Meanwhile, the features should not be strongly affected by the environmental changes. Therefore, the phase difference series, instead of the signal amplitude, are used for feature extraction. This is because the amplitude values are quite sensitive to environmental changes, making the counting model difficult to be adaptive to different environments. Second, by decomposing the effects of the passing behavior on the signals, we find that the walking patterns can be revealed from the phase difference series. To capture the walking patterns, e.g., the stride frequency and torso moving speed for different numbers of passing visitors, short-time Fourier transformation (STFT) is performed on the phase difference series to generate a spectrogram. To fully exploit the spatial information in the spectrogram, we do not extract self-defined features. Instead, the spectrogram is regarded as an image, and the deep convolutional neural network (CNN) is employed on the denoised spectrogram to train the counting model. However, it requires a large number of samples for training the network. To release the labor cost for data collection, we leverage the phase difference series collected from multiple orthogonal frequency-division multiplexing (OFDM) subcarriers and multiple transmitters and receivers as supplementary samples to expand the data set.

The rest of this article is organized as follows. In Section II, the introduction of the related work on people counting is given. Section III gives a brief overview of the visitor counting system. Section IV presents a detailed analysis of the passing behavior on the WiFi phase information. Then, the realization of passing direction detection and visitor counting is illustrated in Sections V and VI, respectively. The evaluation of the system performance is shown in Section VII. At last, we discuss the relevant issues in Section VIII and summarize our work in Section IX.

II. Related Work

Existing people counting systems can be classified into two categories: 1) counting the overall visitors inside the area and 2) counting the visitors passing by a door. Next, we will introduce these two kinds of systems.

A. Count People Inside the Area

To count the people inside certain areas, there are generally three kinds of approaches, including camera-based, device-based, and RF-based approaches.

The camera-based approach uses pattern recognition techniques [6]–[8] to count human objects from images and videos [9]. Much article has been done on utilizing and improving computer vision techniques for human detection. Currently, studies have been dedicated to improving the robustness of human detection algorithms with the help of deep learning algorithms [10], [11]. However, the dead zones and object overlapping problems still lead to the ineffectiveness of the vision-based approach to achieve accurate counting result.

The device-based approach realizes people counting by spreading devices or dedicated sensors in the crowd, e.g., ID tags [12] or using commonly used devices [13]–[15], such as mobile phones as the representative of the human object. On the one hand, the device-based approach raises the cost by allocating a large number of devices. On the other hand, it requires people's active participation to be detected by operating the devices, such as opening the Bluetooth link or running a specific app. There are also approaches that install sensors to sense the environmental temperature [16] or emission of carbon dioxide [17] as a measure for the number of people. However, the estimation result can be easily influenced by the change of the indoor displacement and layout. It also calls for extra deployment of specialized devices in the indoor environment.

The RF-based approach uses different wireless signals, such as WiFi, RFID, and UWB radar, for people counting. Since the presence of human objects can affect the propagation of the wireless signals in the air, human detection can be realized without attaching any sensors on the human body. Existing work mainly leverages the signal's RSS or the amplitude of channel state information (CSI) to learn the relationship between the number of moving people and the variation in the wireless signals. Researchers in [18] deployed wireless sensor networks and estimated the rough crowd density based on the RSS with a clustering algorithm. In [19], the authors counted the people walking between the WiFi transmitter and receiver based on the RSS measurements and performed quantitative analysis on the LoS link and NLoS multipaths of wireless signals. Xi *et al.* [20] found the monotonic relationship between the amplitude of the CSI measurements and the number of moving people in the certain area and count the crowd with the Grey Verhulst model. Guo *et al.* [21] proposed to derive

the number of people through the statistical distribution of CSI amplitude and applied the semi-supervised regression algorithm to obtain the counting result. In [22] and [23], dozens of RFID tags were attached to the wall for counting the moving people around the tags.

B. Count People Passing By the Door

To count the people passing by the doorway, there are mainly two approaches: one is the camera-based approach; another is the infrared-based approach.

The camera-based approach detects the human presence by mounting the cameras at the entrance or exit. In addition to detecting the number of people, the passing direction of the people is also needed to be detected [24]–[26]. However, it faces the privacy problem even worse, as the camera is quite close to people when they pass by the door.

The infrared-based approach counts people by detecting whether the light beam is blocked by the human object. Generally, multiple sets of infrared sensors are deployed for detecting the entering or exiting people passing by the door [27]–[29]. There are also many other binary sensors for detecting the passing behavior, e.g., ultrasound and laser sensors. They follow the similar detecting principle as the infrared-based approach. However, they cannot count the exact number of visitors when multiple persons are passing by together. Thus, barriers are set around the door to allow only one person passing at a time, which incurs extra deployment cost and inconvenience for the people passing by.

In face of the limitations of existing systems, we propose to solve the problem of counting visitors passing by a door in a nonintrusive and accurate way. To remove the concern of being intruded by the utilization of the camera, we use the COTS WiFi devices to detect the visitors. This also provides us with a low-cost scheme as it only relies on the existing indoor WiFi infrastructure. Besides, to provide more accurate counting result, we also count the exact number of multiple passing visitors, which infrared-based approach fails to do so. Therefore, we achieve dynamic visitor counting in a nonintrusive and accurate manner. The Door-Monitor system is the first WiFi-based solution to count the visiting times and the exact number of visitors.

The work in [30] and [31] is partially similar to this article. The authors were able to achieve decent counting results for observing the passing crowd with wireless signals. Depatla and Mostofi [30] employed the RSS change of the WiFi signals to count the passing people, but it is not for the scenario where visitors pass by the door of a certain place. In [31], IR-UWB radar was leveraged for counting the bi-directional passing people, however, the UWB radar sensor is relatively expensive compared with COTS WiFi devices for dense deployment in every room of the building. This article only relies on the COTS WiFi devices, saving the cost for wide deployment [1].

III. System Overview

As shown in Fig. 1, the Door-Monitor system consists of three modules, i.e., phase collection and calibration, passing direction detection, and passing visitor counting.

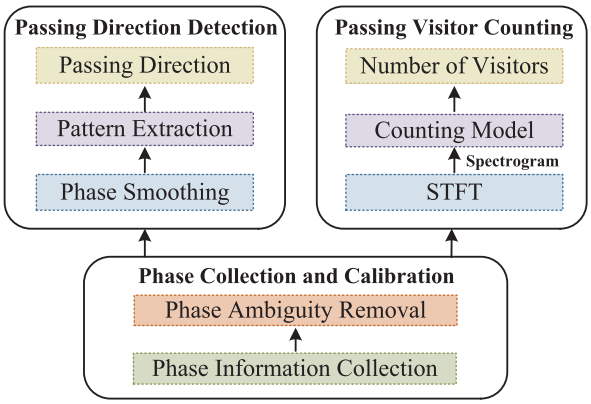


Fig. 1. Sketch of the proposed passing people counting system.

For phase collection and calibration, the phase values are first extracted from the CSI measurements of the WiFi signals, and the raw phase difference series are calculated by subtracting the phase values between two WiFi receiving antennas. Then, the phase difference ambiguity will be eliminated using a clustering-based algorithm for the recovery of the originals phase difference series, which will be articulated in the later section.

For passing direction detection, the aim is to identify the entering and exiting direction of the visitors. By modeling the phase difference values in terms of the passing behavior, we find that under certain deployment of WiFi transmitters and receivers, the increasing and decreasing trend in the phase difference series can indicate the entering and exiting directions, respectively. Because the phase difference series involve random noises and fluctuations caused by the swing of arms and legs, they are first smoothed to get the general trend for recognizing the increasing or decreasing pattern. Then, the derivatives of the smoothed phase difference series are calculated. We draw on the distribution of the sign of the derivatives to detect the passing direction.

For passing visitor counting, we identify how many visitors pass by the door simultaneously. Sometimes, there can be only one person moving in or out of the room at a time. While it is also common in indoor environments, like shopping malls and exhibitions, that people would like to hang out with their friends. Therefore, they would accompany each other when passing by the door, and we need to determine the exact number of passing people. Capturing this information not only helps to obtain an accurate counting result but also provides opportunities for group detection and human dynamics analysis.

To count the number of passing visitors, we perform short-time Fourier transformation on the phase difference series after removing the phase difference ambiguity. The generated spectrogram involves the walking patterns of the visitors, e.g., stride frequency, swing frequency of arms and legs, and other movement information. The distribution and variety of spectrogram can be distinctive for different numbers of passing people. To fully investigate the information in the spectrogram, we employ the deep CNN to exploit the spatial information in the spectrogram and build the counting model.

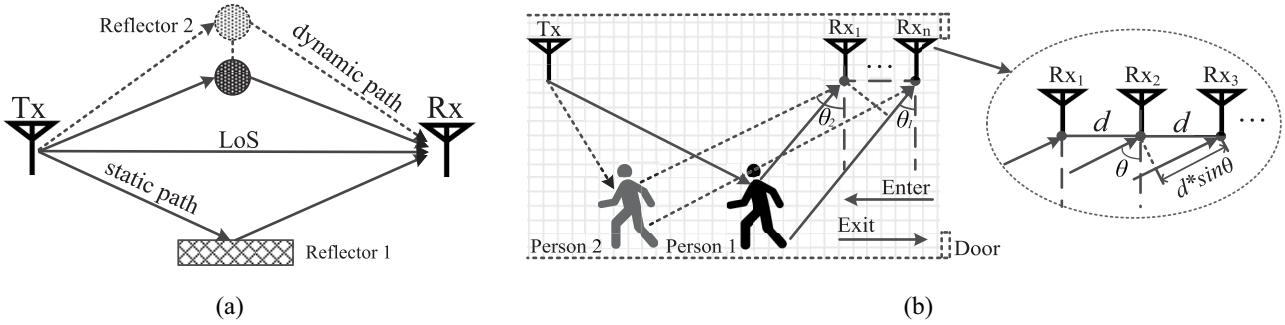


Fig. 2. Illustration of the composition and propagation of wireless signals. (a) Propagation of wireless signals in common indoor environment. (b) Propagation paths of the wireless signals for passing behavior.

To expand the size of the data set for training the CNN model, the phase difference series of different OFDM subcarriers and multiple transmitters and receivers are leveraged as the supplementary training samples as the input to the counting model.

IV. Passing Behavior Analysis Based on WiFi Phase Information

This section starts with the introduction of the preliminary knowledge of the CSI of WiFi signals. Then, we perform a theoretical analysis about the effects of the passing behavior on the propagation of wireless signals. Finally, we show how the passing direction and walking pattern can affect the phase difference series.

A. Preliminaries on CSI

In the modern wireless network, the whole network spectrum is divided into multiple subcarriers using orthogonal frequency division multiplexing (OFDM). The physical (PHY) layer information, CSI, underlying in each subcarrier reflects the linear-combined effects, e.g., reflection, scattering, and diffraction of the wireless signals along different propagation paths. The overall channel response can be represented as follows [32]:

$$H(f,t) = \sum_{i=1}^{n} a_i(f,t) \cdot e^{-j\psi(f,t)} \tag{1}$$

where f denotes the central frequency of each subcarrier and n is the number of propagation paths. $|H(f,t)|$ and $\psi(f,t)$ represent the amplitude and the phase values of the wireless signals, respectively. For m subcarriers, the CSI matrix $\mathbf{H}$ for a given timestamp t is given

$$\mathbf{H} = \big[H(f_1,t), H(f_2,t), \ldots, H(f_m,t)\big]. \tag{2}$$

$|H(f,t)|$ is decided by both $a_i(f,t)$ and $\psi(f,t)$, while the value of $a_i(f,t)$ can be affected by the power amplifier uncertainty and the change of the surrounding environment which is hard to eliminate. So, we turn to use the phase values which are only subject to the length of the propagation paths. However, the phase values $\psi(f,t)$ captured by the commodity network interface card (NIC) also contains timing and phase offset [33], [34]. As a result, the measured $\widehat{\psi_j}$ for the subcarrier j can be expressed as follows:

$$\widehat{\psi_j} = \psi_j + 2\pi \cdot f_j \cdot \alpha_j + \beta_j + Z \tag{3}$$

in which ψ_j is the real phase value; α_j and β_j are the timing and phase offset caused by the packet detection delay, carrier frequency offset, and sampling frequency offset; β_j is a constant value for the same network interface (NIC); and Z denotes the minor random noises in the phase values.

B. Analysis of Passing Behavior on Phase Information

To remove the phase errors from the raw phase values and analyze the effects of the passing behavior on the phase information, we first illustrate how the phase values are formed and affected. For the propagation of wireless signals with the presence of multiple reflectors in the environment, there are multipath signals traveling from the WiFi transmitter (Tx) to the receiver (Rx) except for the line-of-sight (LOS) path. As illustrated in Fig. 2(a), there are LOS path and multipaths reflected by the Reflector 1 and 2. For the ith path, the phase value can be calculated as follows:

$$\psi_i = \{l_i/\lambda\} \mod 2\pi \tag{4}$$

where l_i is the length of the ith propagation path and λ is the wavelength of the wireless signals. Suppose that reflector 2 is a moving object, then the received wireless signals consist of the static propagation paths (P_s) and dynamic propagation paths (P_d). Therefore, the overall phase change ψ in the received signals is the combination of the static and dynamic shift of phase values, which can be represented with the following equation:

$$\psi = \left\{ \frac{\sum_{i\in P_s} l_s + \sum_{i\in P_d} l_d}{\lambda} \right\} \mod 2\pi. \tag{5}$$

Since the passing behavior only contributes to the change of the dynamic paths, we employ the phase difference between two receiving antennas as the indicator of the passing behavior. As shown in Fig. 2(b), the transmitting and receiving antennas are horizontally displayed. The distance between two receiving antennas Rx_1 and Rx_2 is d, and θ is the angle of arrival of the wireless signals reflected by the moving human object.

Then the phase difference ψ_{21} between Rx_1 and Rx_2 can be formulated as

$$\begin{aligned}\psi_{21} &= \psi_2 - \psi_1 = \left\{\frac{L_{s_2}+L_{d_2}}{\lambda} - \frac{L_{s_1}+L_{d_1}}{\lambda}\right\} \mod 2\pi \\ &= \left\{\frac{\left(L_{s_2}-L_{s_1}\right)+\left(L_{d_2}-L_{d_1}\right)}{\lambda}\right\} \mod 2\pi \\ &= \left\{\frac{L_0 + d\cdot\sin\theta}{\lambda}\right\} \mod 2\pi \end{aligned} \tag{6}$$

where $L_s = \sum_{i\in P_s} l_s$, $L_d = \sum_{i\in P_d} l_d$. L_0 is the difference between L_{s_1} and L_{s_2}, representing the static phase difference caused by the static propagation paths, and the phase difference of the dynamic paths is induced by the moving targets, which is $d\cdot\sin\theta$, as shown in Fig. 2(b). However, the measured phase values include several sources of errors, so the reported phase difference $\widehat{\psi_{21}}$ is formulated as follows:

$$\begin{aligned}\widehat{\psi_{21}} &= \widehat{\psi_2} - \widehat{\psi_1} \\ &= \psi_{21} + 2\pi f\cdot\Delta\alpha + \Delta\beta + \Delta Z. \end{aligned} \tag{7}$$

For the above equation, the difference of timing offset $\Delta\alpha$ equals to zero, since the two receiving antennas on the same NIC card use the same clock and down-converter frequency. Thereby, $2\pi f\cdot\Delta\alpha$ can be removed, and $\widehat{\psi_{21}}$ can be shortened in the following equation:

$$\begin{aligned}\widehat{\psi_{21}} &= \psi_{21} + \Delta\beta + \Delta Z \\ &= \left\{\frac{L_0 + d\cdot\sin\theta}{\lambda} + \Delta\beta + \Delta Z\right\} \mod 2\pi. \end{aligned} \tag{8}$$

In the above equation, $\Delta\beta$ is a constant value for the same NIC and ΔZ is the random noise with minor influence on the phase values. Hence, the change in the phase difference mainly resides in the term $d\cdot\sin\theta$. When the human object moves toward the left, i.e., entering the room in Fig. 2(b), θ will increase within the range $(0, \pi/2)$. If we make d to be the wavelength of the wireless signals, then the phase difference between Rx_1 and Rx_2 will monotonically increase as $d\cdot\sin\theta$ goes up. Conversely, θ will decrease from $\pi/2$ to 0 when the human object moves to the right, making the phase difference to decline.

Therefore, we can detect the passing direction via the increasing and decreasing trend in the phase difference. For instance, we deploy the WiFi transceivers according to the deployment setting in Fig. 11(a). We plot the raw phase difference series when one person enters and exits the room in Fig. 3(a) and (b), respectively. When the person enters the room, there is an increasing trend; and it shows a decreasing pattern when exiting the room. The presence of the four separate time series is caused by the four-way phase ambiguity, which will be removed via phase calibration.

In (8), the human object is treated as a general reflector with fixed shape when moving. While in practice, the swinging of arms and legs could incur periodic changes in the phase difference series. Then, the dynamic paths reflected by the human object can be represented as

$$L_d(t) = L_{\mathrm{torso}}(t) + L_{\mathrm{arm}}(t) + L_{\mathrm{leg}}(t) + L_n(t) \tag{9}$$

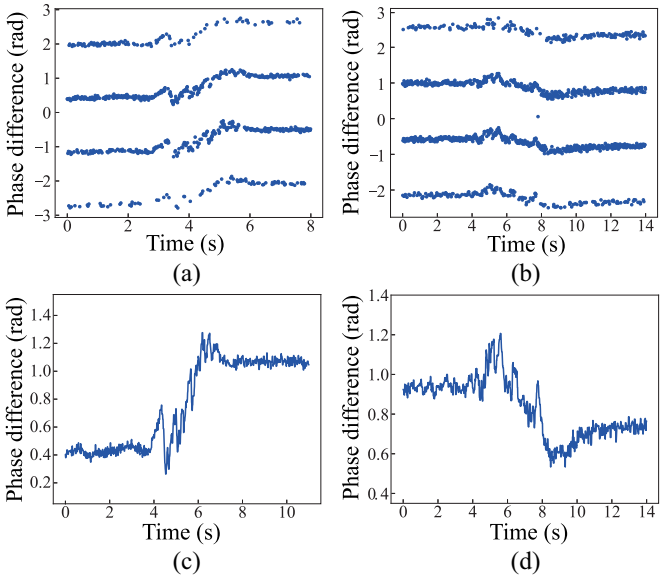


Fig. 3. Raw phase difference series for a person: (a) entering the door and (b) exiting the door; phase difference after phase ambiguity removal for (c) entering the door and (d) exiting the door.

where $L_{\mathrm{torso}}(t)$ is the path reflected by the torso, $L_{\mathrm{arm}}(t)$ and $L_{\mathrm{leg}}(t)$ denotes the paths reflected by the arms and legs, and $L_n(t)$ are the paths affected by other minute body movements. Since the movements of arms and legs during walking are mainly sinusoidal and periodic, there will be some minor peaks and valleys in the real phase difference series, as illustrated in Fig. 3(c) and (d).

V. Passing Direction Detection

In this section, we calibrate and preprocess the raw phase difference series and then detect the moving direction by extracting the increasing and decreasing pattern from the recovered phase difference series.

A. Phase Ambiguity Removal

For the commodity NIC, the four-way phase ambiguity causes the real difference to be θ, $\theta+\pi/2$, $\theta-\pi/2$ or $\theta-\pi$ for 2.4-GHz wireless signals (5-GHz wireless signals have two-way phase ambiguity) [35]. In Fig. 3(a) and (b), the real phase difference is separated into four groups with $\pi/2$ spacing. One way to retrieve the real phase difference is to compare the difference between the two consecutive phase values. If the difference is around $\pi/2$ and π, then it is expected to experience a phase change and we can add the corresponding change to the current phase to recover the original phase value. However, there are some outliers and random noises in the phase difference, making the above approach ineffective in dealing with the phase ambiguity.

Here, we design a phase ambiguity removal method based on the clustering algorithm. The raw phase difference series are segmented into small windows with a length of w and the k-means clustering is applied. For the four-way phase ambiguity, there would be four clusters with the spacing of each cluster's centroid to be $\pi/2$. Then, we sort the four centroids

Algorithm 1: Phase Ambiguity Removal

Input: $\Delta\psi, w, num, \sigma$; Output: $\Delta\psi'$;
for $i = 1{:}w{:}length(\Delta\psi) - w$ **do**
  $[index, center] = kmeans(\Delta\psi[i, i + w - 1], num)$;
  $baseline = max(sort(center))$;
  if $baseline - \Delta\psi(i) \in [3/2\pi - \sigma, 3/2\pi + \sigma]$ **then**
    $\Delta\psi'(i) = \Delta\psi(i) + 3/2\pi$; break;
  end
  if $baseline - \Delta\psi(i) \in [\pi - \sigma, \pi + \sigma]$ **then**
    $\Delta\psi'(i) = \Delta\psi(i) + \pi$; break;
  end
  if $baseline - \Delta\psi(i) \in [\pi/2 - \sigma, \pi/2 + \sigma]$ **then**
    $\Delta\psi'(i) = \Delta\psi(i) + \pi/2$; break;
  end
end

and choose the highest one as the baseline so that the sample points in other clusters can add $\pi/2$, π, and $3\pi/2$ respectively. Thereby, they can be integrated into a single time series. The pseudo-code for phase ambiguity removal is illustrated in Algorithm 1. In our experiments, w is set as 100, num equals to 4. The examples of the recovered phase difference series are shown in Fig. 3(c) and (d). The phase ambiguity removal is performed for all the subcarriers of one sample for later use.

B. Passing Direction Extraction

As discussed in Section IV-B, the passing direction can be inferred from the increasing and decreasing trend of the phase difference series. To extract the signal trend from the recovered phase difference series, the subcarrier with the highest standard deviation is selected. Then, we need to wipe out the random noises and the signal fluctuations in the selected phase difference series caused by the arm and leg movements. Here, we apply the Savitzky–Golay filter [36] on the phase difference series. The Savitzky–Golay filter is based on the least-square polynomial approximation, which can smooth the details and maintain the contour of the time series. For the filter, we set the polynomial order as 3 and the length of the frame as 50. The filtered phase difference series are illustrated in Fig. 4(a) and (b).

Next, a threshold is set to segment out the phase difference series affected by the passing behavior. We compare the distribution of the standard variance of the phase difference series when there is no person and one or two persons passing by, as shown in Fig. 5. We set the threshold as 0.01 as it can cut the two distributions without much overlapping. The segment between the dashed lines in Fig. 4(a) and (b) is the phase difference series affected by the passing behavior. To detect the increasing and decreasing trend from the recovered phase difference series, we first calculate the derivative of each sample point and smooth the derivatives with a median filter. Fig. 4(c) and (d) shows the smoothed derivatives of the affected series. The gradients of the increasing segment are mostly positive and those of the decreasing segment show more negative numbers. Thus, we count the presence of the negative and positive numbers in the derivatives. If the positive numbers appear more than the negative ones, then there

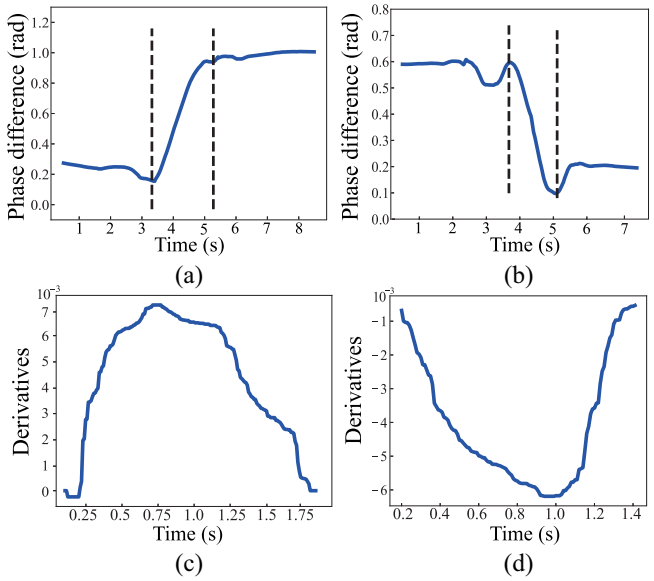


Fig. 4. Phase difference series after denoising for a person: (a) entering behavior and (b) exiting behavior; derivatives on the phase difference series of the (c) entering segment and (d) exiting segment.

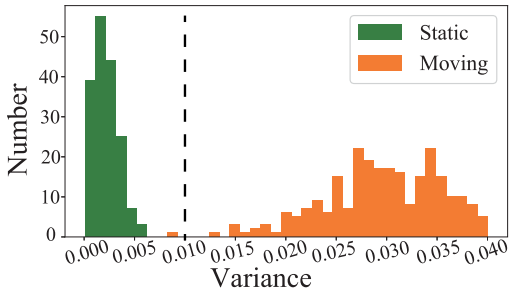


Fig. 5. Standard variance of the phase difference series with no person passing (static) and with one or two persons passing (moving).

is an increasing trend. Otherwise, there would be a decreasing trend. After finding out the pattern in the phase difference series, we can determine the passing direction. When multiple persons enter or exit the doorway together, there also appears an increasing or decreasing trend, respectively.

VI. Passing Visitor Counting

In this section, the focus is mainly put on counting the number of visitors passing by a door. There can be one person entering or exiting the room alone, while it is also common in shopping malls and other indoor places where people would hang out with their friends. Therefore, we need to count the exact number of visitors passing by a door simultaneously to provide an accurate counting result. To count the passing people, we first apply time-frequency analysis on the recovered phase difference series and then build the model for passing visitor counting with CNN.

A. Spectrogram Extraction

In (9), it shows that the phase difference series involve the movements of different body parts for the passing behavior. With the increasing number of passing people, the phase difference series will contain more movement

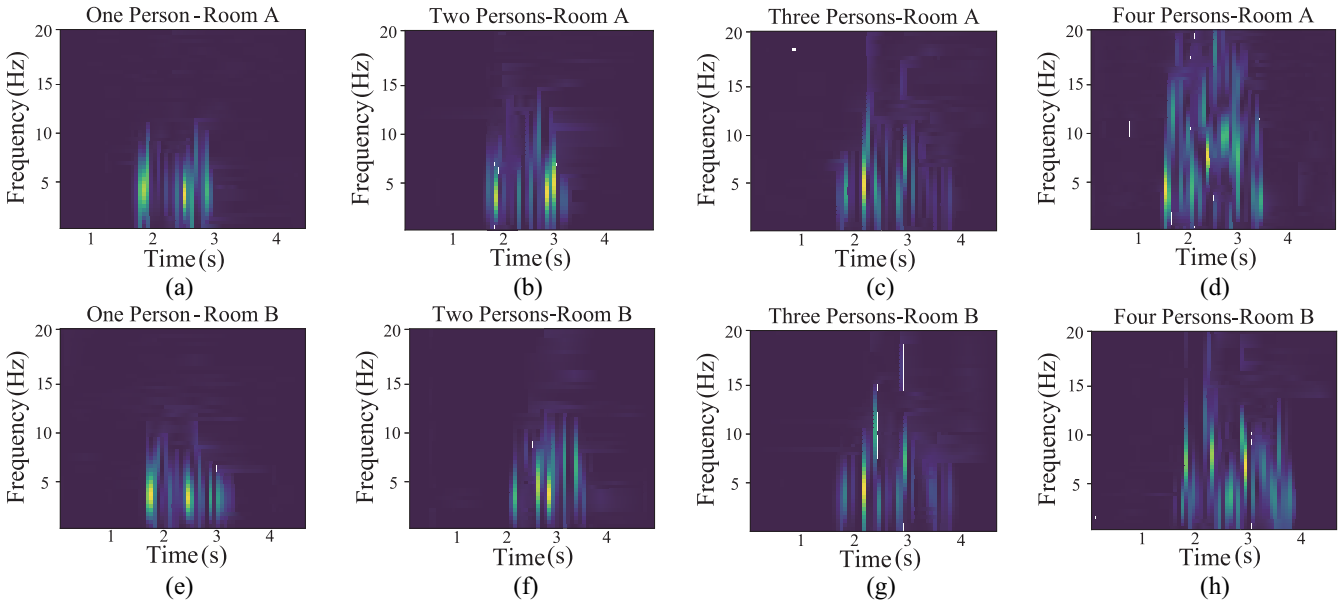


Fig. 6. Examples of spectrogram for different numbers of passing people in different environments: (a)–(d) 1, 2, 3, and 4 persons pass by the door of room A; (e)–(h) 1, 2, 3, and 4 persons pass by the door of room B.

information of different persons. To extract the passing movement information from the phase difference series, we perform time-frequency analysis using the short-time Fourier transformation and transform the recovered phase difference series into the spectrogram. The spectrogram shows the change of the frequency components for different body parts while passing by the door.

STFT is done by segmenting the phase difference series with a sliding window and applying fast Fourier transformation (FFT) on each window. Since the length of the window for FFT decides the frequency and time resolution of STFT, we need to choose a reasonable window size. Practically, the walking speed is in the range of 0.8–1.5 m/s. The frequency of leg and arm swing during walking is in the range of 2–6 Hz [37]. The window size for FFT is selected to be 256 samples and the width of the sliding window is 28 samples. The overlapping of the sliding window is 14 samples. To avoid frequency leakage, we add a hamming window for performing FFT. Then, the frequency resolution is 0.39 Hz and the time resolution is 0.14 s for the sampling rate of 100 samples per second, which are proper for tracking the passing behavior.

Examples of spectrograms with different numbers of passing people in two different environments are depicted in Fig. 6. Fig. 6(a)–(d) shows the spectrograms of one to four persons passing by the door of room A, and Fig. 6(e)–(h) illustrates the spectrograms when people pass by the door of room B. The lighter parts in the spectrograms indicate larger values along the frequency and time axes. It can be seen that the spectrograms of more passing visitors involve more variations for the distribution of the frequency components, and there would be higher frequency values. The time span of the light area for more passing people is longer.

To explain this observation, one reason is that different people can have different walking patterns. Thus, the spectrogram presents more variations when more people show up. Another reason is that wireless signals can be consecutively reflected by different persons. Therefore, when more people get involved, the more disperse the frequency components would be and higher frequency components appear. Meanwhile, we also compare the spectrograms in different environments. The frequency components in the spectrogram show a similar distribution for the same number of passing people in different environments, as the background effects have been removed from the phase difference series. Therefore, the spectrogram can be leveraged as an effective feature for building the counting model to classify the number of passing visitors.

B. Counting Model Building

Previous studies build the classification model with handcrafted features, like mean, standard deviation or other statistical values from the spectrogram [38], [39]. For this article, we can also extract self-defined features from the spectrogram. But we will lose the valuable information during the process of compressing the spectrogram into several feature points. Another way is to stretch the matrix of the spectrogram into a one-dimension feature vector [40]. However, we will lose the spatial relationship among different pixels in the spectrogram. To fully exploit the spectrogram, we propose to use CNN for building the counting model. The convolutional layers in CNN can help to automatically learn the features in the spectrogram. Here, CNN is exploited for multiclass classification. In Fig. 7, it shows that the number of concurrent passing visitors is mainly below six for many popular places. Therefore, we set the maximum number of people passing by the door as six. Accordingly, we have six classes corresponding to different numbers of passing people. In Section VII-D5, the performance of our method and other feature extraction approaches will be compared to show the effectiveness of exploiting spectrogram with CNN.

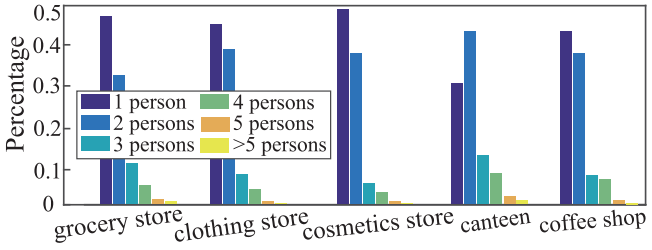


Fig. 7. Percentages of different numbers of people enter or exit different indoor places after our two-day observation in a shopping mall.

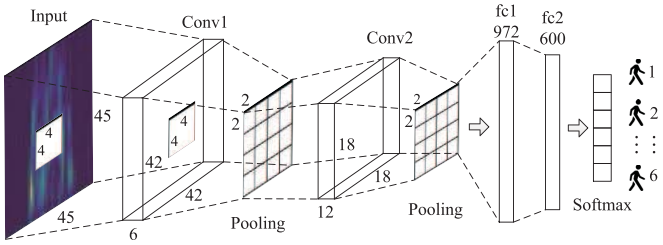


Fig. 8. CNN architecture for training the counting model.

We adopt the CNN architecture illustrated in Fig. 8, for classifying the number of passing visitors. The input is the spectrogram which is reshaped into a 45×45 matrices. Two convolutional layers come next. The kernel size is 4×4 and the stride is 1 without padding. The rectified linear unit (ReLU) is chosen as the activation function. Then, the pooling layer is inserted between the convolutional layers to reduce the spatial size and computation for the network. Here, we choose the max-pooling which down-samples the input representation with the kernel size of 2×2. The performance of counting visitor with different pooling layers will be discussed in the evaluation part. The convolutional layer, the ReLU layer, and the max-pooling layer are integrated as one unit, on which batch normalization is performed. Batch normalization helps to speed up the training process of CNN, and enables easier initialization for the start of the network [41]. Next, two fully-connected layers are added so that the neurons will have full connections to all the neurons in the previous layer. To avoid over-fitting, we perform dropout between the two fully-connected layers. The dropout operation throws out a certain proportion of the neurons in the network to reduce the interdependency among the neurons [42]. At last, the softmax layer is added in the final layer for multiclass classification.

To build the CNN counting model, a large amount of training data is required, which could result in much effort on data collection. To release the labor of collecting massive training samples, we leverage the multiple OFDM subcarriers to expand the training set in a lightweight way. For the NIC we use, it can report 30 sets of phase time series measured from 30 subcarriers. The central frequency of each subcarrier is different, leading to a difference in the wavelength λ in (8). While the difference of λ is quite small, which is around 0.208 cm for all the 30 subcarriers. Thereby, different subcarriers show generally similar passing pattern, but slightly differ from each other. Examples of the recovered phase difference series on different subcarriers are shown in Fig. 9. In this way, there

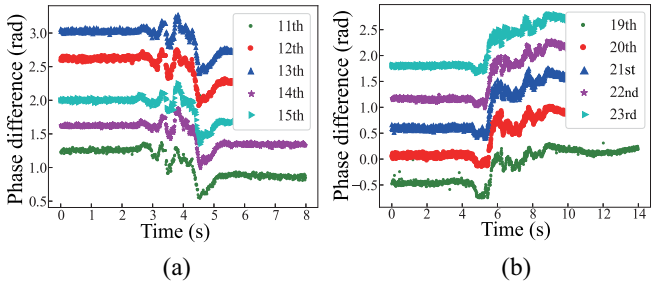


Fig. 9. Phase difference series of different subcarriers for (a) one person exits the room and (b) two persons enter the room.

will be 30 phase difference series with one transmitter and two receivers.

If we use the WiFi transmitter with two antennas (Tx_1, Tx_2) and the receiver with two antennas (Rx_1, Rx_2), two groups of phase difference series from the antenna pairs of Tx_1-Rx_1-Rx_2 and Tx_2-Rx_1-Rx_2 can be obtained, respectively. Then, one sample of the phase difference series is expanded into $30 \times 2 = 60$ training samples. In practice, the initial counting model can be trained with a certain number of samples first. Then, the training set can be incremented by adding the new coming testing samples of which the counting results are accurate, then the counting model can be retrained constantly.

To get the counting result with testing samples, we select ten sets of the recovered phase difference series from all the subcarriers of all the antenna pairs. The selected ten phase difference series have the highest variance of the phase values. The ten sets of phase difference series will be separately transferred to the counting model. Then, the final result on the number of passing people can be obtained by calculating the highest voting of all the results from the ten set of phase difference series.

VII. Evaluation

In this section, we first introduce the experimental setup of the system. Then, the evaluation metrics are given for evaluating the performance of passing direction detection and passing visitor counting. Finally, the experimental results on passing direction detection and passing visitor counting are discussed under different settings and scenarios.

A. Experiment Setup

For the deployment of our system, we leverage a TP-Link WR641N WiFi router with two antennas as transmitters (Tx). A Lenovo B460 laptop equipped with the Intel 5300 network interface is used to collect the transmitter WiFi signals, where the first two antennas on the interface are used as the receivers (Rx). The receiving antennas are directional, which are put 0.5 m aside the door. The distance between the two receiving antennas is 12 cm, which is the wavelength of the 2.4-GHz WiFi signals. Rx and Tx are put in a line with around 1-m distance to each other and 1.5-m height above the floor. The door is around 1 m wide in our experiments, and the door area is about 2 m^2. The distance between the transmitters and receivers can be adjusted according to the size of the door area.

TABLE I
DATA COLLECTION OF DIFFERENT GROUPS IN DIFFERENT ROOMS

No. of people		1	2	3	4	5	6
Room 1	Group A	65(in), 65(out)	65(in), 65(out)	65(in), 65(out)	65(in), 65(out)	65(in), 65(out)	65(in), 65(out)
	Group B	55(in), 55(out)	55(in), 55(out)	55(in), 55(out)	55(in), 55(out)	55(in), 55(out)	55(in), 50(out)
Room 2	Group A	42(in), 42(out)	42(in), 42(out)	42(in), 42(out)	42(in), 42(out)	42(in), 42(out)	42(in), 42(out)
	Group B	42(in), 42(out)	42(in), 42(out)	42(in), 42(out)	42(in), 42(out)	42(in), 42(out)	42(in), 42(out)

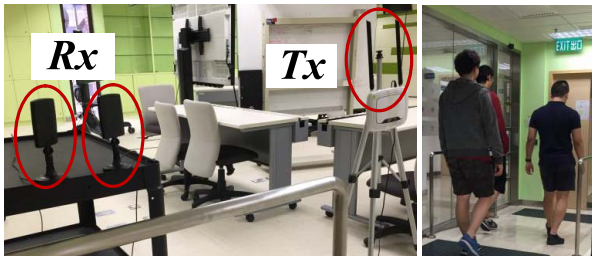


Fig. 10. System deployment with two WiFi transceivers and two WiFi receivers, and three persons passing by the door.

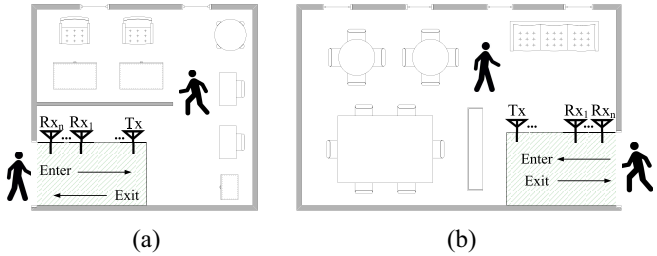


Fig. 11. Experiment environment layout: (a) room A and (b) room B; the WiFi transmitters (Tx) and receivers (Rx) are placed horizontally beside the doorway.

The CSI information is extracted through the CSI tool which modifies the firmware under the Linux system for exposing the CSI amplitude and phase values [43], which are processed with MATLAB2016a. Fig. 10 shows an example of the system deployment.

We recruit 12 volunteers, including eight males and four females, and they are randomly divided into two groups: group A and group B. They are asked to enter and exit the door of the two rooms alone or with other volunteers. The layout of the two rooms is shown in Fig. 11. The number of samples collected with the two groups of volunteers are given in Table I. In the table, the "in" and "out" refer to the entering and exiting direction, respectively. When multiple persons pass by the door, they can walk one by one or side by side along with each other. The extra samples in room 1 of group A and group B are collected with different walking speeds and sampling rates to investigate their effects on passing direction detection. The remaining samples are divided into two sets, i.e., training and testing data sets for evaluating the performance of counting passing visitors. While collecting the data, volunteers are allowed to walk and pass by the door as they get used to.

B. Evaluation Metrics

To evaluate the performance of passing direction detection and passing visitor counting, the following evaluation metrics are defined accordingly.

Evaluation on Passing Direction Detection: There are two possible results for detecting the passing direction, which is either entering or exiting the room. We define the accuracy of passing direction detection as the ratio between all the correct detected instances and the total number of instances under each case.

Evaluation on Passing Visitor Counting: To evaluate the performance of the counting model on obtaining the number of people passing by the doorway, we define the accuracy of people counting as the percentage of the correctly identified instances over the total number of instances for different numbers of passing people. Besides, the error of people counting is defined as the difference between the estimated number of people and the real number of people.

C. Performance of Passing Direction Detection

In the following sections, the accuracies on passing direction detection will be discussed in detail concerning the impacts of different walking speeds, sampling rates, numbers of visitors, and the surrounding moving people.

1) Impact of Walking Speed: Volunteers are asked to enter or exit the doorway of the two rooms with three different walking speeds, including the slow (about 0.6−0.7 m/s), normal (about 1 m/s), and fast (about 1.3−1.4 m/s) walking speeds. The accuracies of passing direction detection with different walking speeds are shown in Fig. 12(a). It can be seen that the average accuracy among the three kinds of walking speed is around 94%. The detection accuracy of slow and normal walking speeds is quite similar to each other, while walking with faster speed can lead to more detection failure (around 91%). This is because the effective time series will be less affected by the moving human objects and suffer from more noises when the person is moving too fast. The corresponding period will be shorter which makes the extraction of the signal trend suffer from more difficulties.

2) Impact of Sampling Rate: We also investigate the impact of different sampling rates on the accuracy of passing direction detection. In fact, the sampling rate is relevant to the walking speed. If the sampling rate is too low, the effective time series of the passing behavior will be too short to be detected. By contrast, if the sampling rate is too high, then it can lead to more noises in the wireless signals. Fig. 12(b) depicts the accuracy on passing direction detection under different sampling rates, which are 100 packets per second (p/s), 200 p/s, and 400 p/s. It shows that the sampling rate of

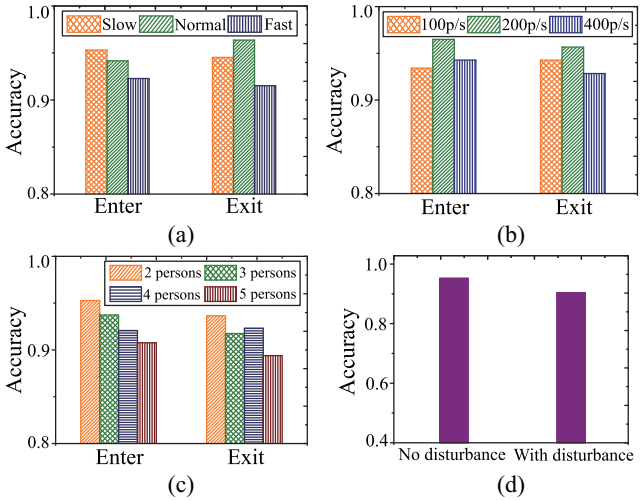


Fig. 12. Accuracy of passing direction detection with (a) different walking speeds, (b) different sampling rates, (c) different numbers of people, and (d) surrounding moving people as disturbance.

200 p/s can achieve the best performance with 95.5% detection accuracy, and the accuracies of another two sampling rates are also above 90%. In our case, the distance between the transmitter and two receivers is about 1 m. The sampling rate can be adaptively adjusted with the change on the distance, like applying a higher sampling rate for longer distance.

3) Impact of the Number of Passing People: The above experiments are only done with a single person. Other than the single-person scenario, we also explore the effects of the number of people on the performance of passing direction detection. The accuracies of direction detection are shown in Fig. 12(c) of which 2, 3, 4 and 5 human objects move into or go out of the door together. Although the performance on passing direction detection drops with the rising of the number of people, the accuracy is still around 90%. More people passing by the door induces more variations in the WiFi phase values owing to the superposition of multiple human objects as the reflectors and various walking patterns, leading to more misinterpretation on the pattern of the phase difference series.

4) Impact of Surrounding Moving People: The presence of surrounding moving human objects inside the room may also influence the propagation of the wireless signals. Therefore, we discuss the effects of the surrounding human activities on the passing direction detection. As introduced in the experimental setup, directional antennas are employed, and the displacement of the transmitter and receivers are only targeted at the passing human object. Hence, the effects of the surrounding changes will be lowered to the least. We allow several volunteers to walk around in the room freely to see its effects on the performance of the direction detection. As depicted in Fig. 12(d), the average accuracy can still reach around 90% for passing direction detection when people are moving inside the room. This is due to the usage of directional antennas, which only receive the signals reflected by the passing human objects.

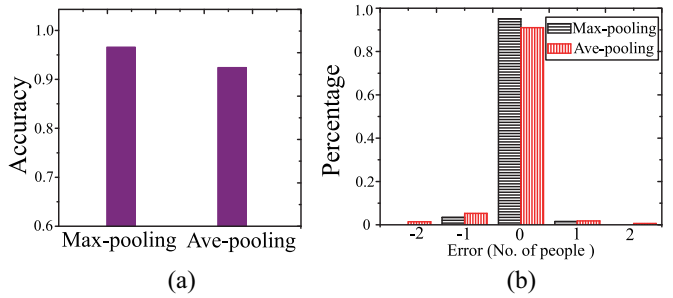


Fig. 13. Passing visitor counting performance with respect to pooling layers: (a) counting accuracy with different pooling layers and (b) error on counting people with different pooling layers.

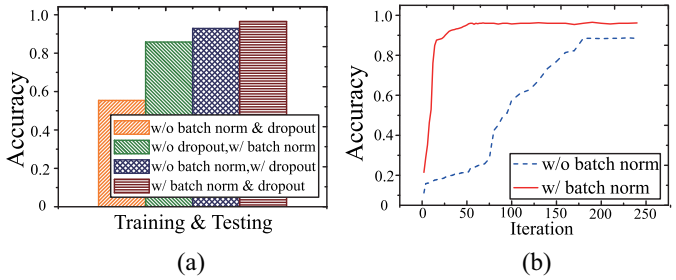


Fig. 14. Passing visitor counting performance: (a) counting accuracy with or without batch normalization and dropout and (b) counting accuracy different numbers of iteration.

D. Performance of Passing Visitor Counting

To evaluate the performance of passing people counting, both the estimation error on the number of people and the overall accuracy are calculated. We investigate the impact of different pooling layers, batch normalization and dropout in the CNN model on the counting performance. Finally, we discuss the effects of different environments and humans objects and compare different feature extraction and training algorithms on the accuracy of the counting passing people.

1) Impact of Pooling Layer: First, we compare the influence of the average-pooling and max-pooling for the pooling layer in terms of the accuracy and estimation error on counting people. Fig. 13 shows that detailed results, which tell that the counting accuracy with max-pooling is around 5% higher than that of average-pooling. For the distribution of the estimation error in Fig. 13(b), the error on the real number of people and the estimated number of people is mainly 1 person with max-pooling. While the average-pooling could incur 2 persons' error. Hence, we apply max-pooling in the pooling layer.

2) Impact of Batch Normalization and Dropout: The results of the accuracy on counting people without both batch normalization and dropout, with only batch normalization, with only dropout, and with both batch normalization and dropout are illustrated in Fig. 14(a). The counting accuracy is only about 55% when neither batch normalization nor dropout is applied. The operation of dropout makes the accuracy increase to 85% and the use of batch normalization also improves the counting accuracy. Furthermore, to demonstrate the influence of the batch normalization on accelerating the training process, the accuracy on people counting of different times of iteration

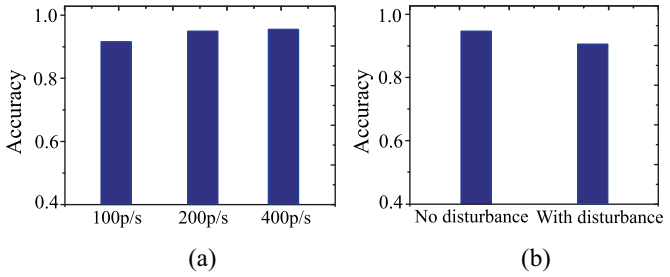


Fig. 15. Passing visitor counting performance: (a) counting accuracy of different sampling rates and (b) counting accuracy with and without surrounding moving people as disturbance.

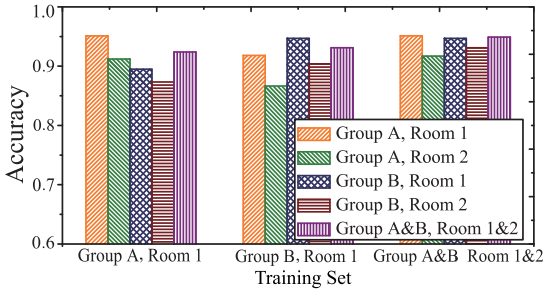


Fig. 16. Counting accuracy in different environments and with different groups of human objects.

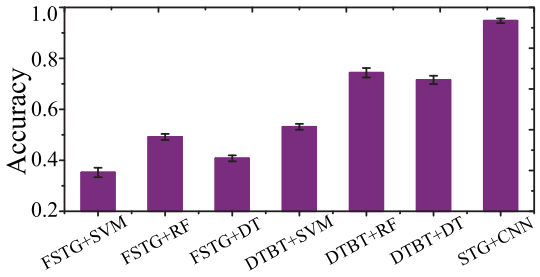


Fig. 17. Comparison of the counting accuracy of different feature extraction and training algorithms.

is shown in Fig. 14(b). It cost less number of iterations for the training phase to reach the best performance with batch normalization, at the same time, the accuracy is higher than that without batch normalization.

3) Impact of Sampling Rate and Surrounding People: Since the sampling rate can affect the number of samples in the window of STFT, we build three counting models with the sampling rates of 100, 200, and 400p/s and get the counting accuracy with the testing samples whose sampling rate is the same as the training data. The results are shown in Fig. 15(a). The counting accuracy increases with a higher sampling rate. The promotion of accuracy from 100p/s (91.3%) to 200p/s (94.4%) is more obvious compared with from 200p/s to 400p/s (95.2%). Therefore, the sampling rate of 200p/s is enough.

We also show the influence of the surrounding moving people on the accuracy of visitor counting. Several persons are asked to move randomly in the room to create the disturbance. The results with and without moving people as the disturbance are shown in Fig. 15(b). While the accuracy is still above 90% because we use the directional antennas that only orient to the door area.

4) Impact of the Environment and Human Objects: In the training data, the samples that are collected when people pass by the door one by one and side by side are combined together to build the counting model. To test the performance of counting people under different scenarios, including different environments and groups of human objects, we devise separate evaluation: 1) three training sets: one collected from Group A in Room 1; one collected from Group B in Room 1; and one collected from both Group A and Group B in Room 1 and Room 2 and 2) five testing sets: the first set collected from Group A in Room 1; the second set collected from Group A in Room 2; the third set collected from Group B in Room 1; the fourth set collected from Group B in Room 2; and the last one collected from both Group A and Group B in Room 1 and Room 2.

The accuracies of counting people with the above divided data sets are shown in Fig. 16. The counting accuracy can reach around 95% in the same environment with the same group. When the counting model is built under different environments, the average counting accuracy is 91.8%, which is only approximately 5% lower than the model built under the same environment. The counting accuracy will drop to 87.4% when the model is built with different groups of human objects under different environments. While the involvement of more human objects in the training phase can help to keep the counting accuracy above 90%. This is due to the reason that the human objects in different groups could have some variations in the passing patterns. Therefore, we can gradually increment the training set dynamically with new samples of more people to enhance the counting performance.

5) Impact of Feature Extraction and Training Algorithms: We compare the counting accuracy of different counting models to show the effectiveness of CNN with the spectrogram as the input. In [40], the 2-D spectrogram is flattened into a 1-D array (shortened as FSTG) as the input to the classification model for gait identification. In [38], the distribution of the energy spectrum on different frequency points in the spectrogram (shortened as DTST) is extracted as the features for gait recognition. We apply the above two methods for extracting features from the spectrogram and leverage different classification algorithms for building the counting model, including support vector machine (SVM), random forest (RF), and decision tree (DT). Then, the counting accuracy is calculated with the above approaches and our method (STG+CNN).

The results are shown in Fig. 17. It shows that the accuracies of using the flattened spectrogram (FSTG) and the distribution of spectrogram (DTBT) for passing visitor counting are all below 80%, which are all lower than our proposed method (STG+CNN). This is mainly due to the reason that they discard the potential features and ignore the spatial information in the spectrogram.

6) Computation Time for Training and Testing: Finally, the computation time for training the counting model and detecting the passing direction and estimating the number of passing visitors with one testing sample is given. The counting model is trained on a desktop installed with the Ubuntu 16.04 system

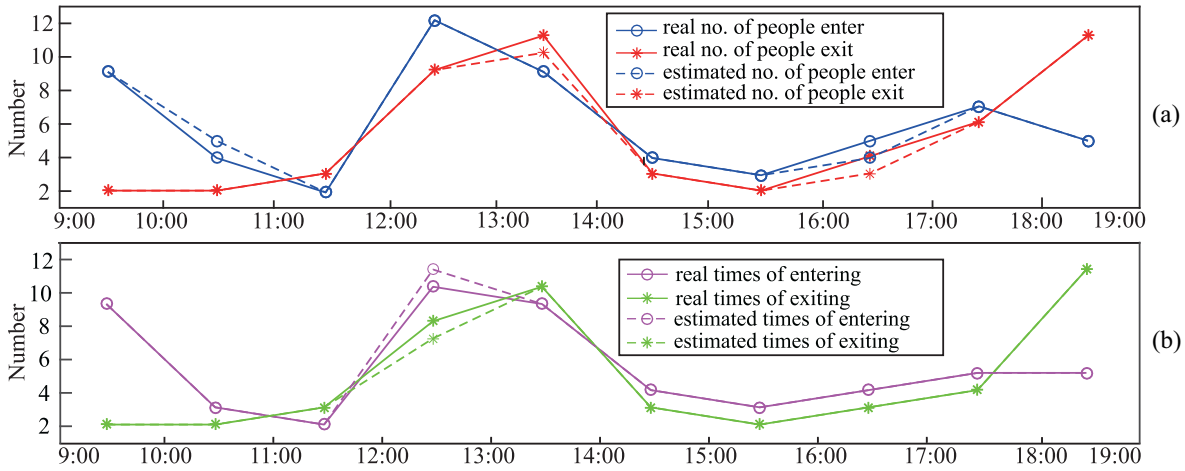


Fig. 18. Real-time visitor counting from 9:00 to 18:00: (a) number of people enter and exit a room and (b) number of times people enter and exit a room.

and equipped with 16G RAM and GPU GeForce GTX 1080Ti. The average training time for 11 520 training samples with 100 times of iteration is about 17.4 min. The overall testing time for passing direction detection and counting is around 0.0268 s for each testing sample. It shows that the results on direction detection and visitor counting can be obtained immediately.

E. Real-Time Visitor Counting

We deploy the system in a student lab to illustrate the real-time results of the students coming in and out, including the number of people and times when they enter and exit the lab in an hourly base, as shown in Fig. 18. One camera is put aside the door to obtain the ground truth. The solid lines are the real results, while the dashed lines are the estimated results. In Fig. 18(a), during 10:00–11:00, 13:00–14:00, and 16:00–17:00, the number of people is wrongly estimated with the error of one person. In Fig. 18(b), the error only appears between 12:00–13:00. The students in the lab enter and exit the lab more frequently during 9:00–10:00, 12:00–13:00 and 18:00–19:00, which are the times when they arrive at the lab, go for lunch and dinner. While during the period of 10:00–12:00 and 14:00–17:00, fewer students are moving in and out because most of them are working at the table.

VIII. Discussion

Door-Monitor aims to accurately count the passing visitor using WiFi signals. However, there are still two main issues that we want to illustrate and discuss.

The first concern is the security issue of the WiFi systems. If the WiFi router is hacked by attackers, the information of the visitor can be leaked, and the availability of the visitor counting service can even be destroyed. Therefore, it is important to avoid breakdown and make the system more secure to use. Researchers have made great effort to enhance the security of IoT systems [44]. We believe that the security issue would be well recognized and addressed to guarantee better service.

Another problem we want to demonstrate is about our assumption of the system. In practice, common shops or restaurants are visited intermittently, and the door area is limited. Thereby, we assume that visitors would only enter or exit the door with the same direction, otherwise, they will crash into each other. However, there can be cases when people enter and exit at the same time if the door area is large and wide enough. To make our system more scalable, we can separate the door area into two lanes to let people with opposite directions to pass. We do not need to assign the passing direction for each lane, so people can still choose any of them to pass. The two sets of WiFi transceivers will be deployed for each lane to detect the direction and count the passing visitors. The design of two lanes is not as convenient as the one lane. However, it is a challenging task to count passing visitor if they pass by in opposite directions. This is because we need to estimate the exact number of visitors for each direction, i.e., counting how many people enter and how many exit, respectively. We will continue to solve this issue with a more flexible approach.

IX. Conclusion

In this article, we proposed Door-Monitor, a visitor counting system which counts the visitors dynamically moving in and out of the door using WiFi signals. The system can be deployed under the existing indoor WiFi infrastructure and visitors can be counted without active participation and carrying any devices. The passing behavior is modeled by analyzing the phase information of the WiFi signals. Then, we identify the pattern in the passing difference series for detecting the visitors' passing direction, i.e., entering or exiting, and generate the spectrogram which represents the variety of different numbers of passing visitor. The spectrograms from different subcarriers are input into a 7-layer CNN to build the model for counting passing visitors. Extensive experiments show the effectiveness of the visitor's counting system.

References

[1] Y. Yang, J. Cao, X. Liu, and X. Liu, "Wi-count: Passing people counting with COTS WiFi devices," in *Proc. 27th Int. Conf. Comput. Commun. Netw. (ICCCN)*, Hangzhou, China, 2018, pp. 1–9.

[2] K. Lin, M. Chen, J. Deng, M. M. Hassan, and G. Fortino, "Enhanced fingerprinting and trajectory prediction for IoT localization in smart buildings," *IEEE Trans. Autom. Sci. Eng.*, vol. 13, no. 3, pp. 1294–1307, Jul. 2016.

[3] G. Fortino, A. Guerrieri, G. M. P. O'Hare, and A. Ruzzelli, "A flexible building management framework based on wireless sensor and actuator networks," *J. Netw. Comput. Appl.*, vol. 35, no. 6, pp. 1934–1952, 2012.

[4] *Intelligent Counting: People Counting and Visitor Counting*. Accessed: Aug. 2018. [Online]. Available: https://www.intelligentcounting.com/

[5] *Countwise: People Counting Solution—Visitor Counting System*. Accessed: Aug. 2018. [Online]. Available: http://www.countwise.com/

[6] M. Li, Z. Zhang, K. Huang, and T. Tan, "Estimating the number of people in crowded scenes by MID based foreground segmentation and head-shoulder detection," in *Proc. IEEE 19th Int. Conf. Pattern Recognit. (ICPR)*, Tampa, FL, USA, 2008, pp. 1–4.

[7] J. D. Nichols *et al.*, "Multi-scale occupancy estimation and modelling using multiple detection methods," *J. Appl. Ecol.*, vol. 45, no. 5, pp. 1321–1329, 2008.

[8] X. Zhao, E. Delleandrea, and L. Chen, "A people counting system based on face detection and tracking in a video," in *Proc. 6th IEEE Int. Conf. Adv. Video Signal Based Surveillance (AVSS)*, Genoa, Italy, 2009, pp. 67–72.

[9] S. A. M. Saleh, S. A. Suandi, and H. Ibrahim, "Recent survey on crowd density estimation and counting for visual surveillance," *Eng. Appl. Artif. Intell.*, vol. 41, pp. 103–114, May 2015.

[10] C. Zhang, H. Li, X. Wang, and X. Yang, "Cross-scene crowd counting via deep convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, 2015, pp. 833–841.

[11] D. B. Sam, S. Surya, and R. V. Babu, "Switching convolutional neural network for crowd counting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, 2017, pp. 4031–4039.

[12] T. Germa, F. Lerasle, N. Ouadah, and V. Cadenat, "Vision and RFID data fusion for tracking people in crowds by a mobile robot," *Comput. Vis. Image Understand.*, vol. 114, no. 6, pp. 641–651, 2010.

[13] J. Weppner and P. Lukowicz, "Bluetooth based collaborative crowd density estimation with mobile phones," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. (PerCom)*, San Diego, CA, USA, 2013, pp. 193–200.

[14] L. Schauer, M. Werner, and P. Marcus, "Estimating crowd densities and pedestrian flows using Wi-Fi and Bluetooth," in *Proc. 11th Int. Conf. Mobile Ubiquitous Syst. Comput. Netw. Services (MOBIQUITOUS)*, 2014, pp. 171–177.

[15] P. G. Kannan, S. P. Venkatagiri, M. C. Chan, A. L. Ananda, and L.-S. Peh, "Low cost crowd counting using audio tones," in *Proc. 10th ACM Conf. Embedded Netw. Sensor Syst. (SenSys)*, 2012, pp. 155–168.

[16] K. P. Lam *et al.*, "Occupancy detection through an extensive environmental sensor network in an open-plan office building," *IBPSA Build. Simulat.*, vol. 145, pp. 1452–1459, Jan. 2009.

[17] C. Jiang, M. K. Masood, Y. C. Soh, and H. Li, "Indoor occupancy estimation from carbon dioxide concentration," *Energy Build.*, vol. 131, pp. 132–141, Nov. 2016.

[18] Y. Yuan, C. Qiu, W. Xi, and J. Zhao, "Crowd density estimation using wireless sensor networks," in *Proc. IEEE 7th Int. Conf. Mobile Ad Hoc Sensor Netw. (MSN)*, Beijing, China, 2011, pp. 138–145.

[19] S. Depatla, A. Muralidharan, and Y. Mostofi, "Occupancy estimation using only WiFi power measurements," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 7, pp. 1381–1393, Jul. 2015.

[20] W. Xi *et al.*, "Electronic frog eye: Counting crowd using WiFi," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Toronto, ON, Canada, 2014, pp. 361–369.

[21] X. Guo, B. Liu, C. Shi, H. Liu, Y. Chen, and M. C. Chuah, "WiFi-enabled smart human dynamics monitoring," in *Proc. 15th ACM Conf. Embedded Netw. Sensor Syst. (SenSys)*, 2017, pp. 1–13.

[22] H. Ding *et al.*, "Human object estimation via backscattered radio frequency signal," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, 2015, pp. 1652–1660.

[23] H. Ding *et al.*, "Counting human objects using backscattered radio frequency signals," *IEEE Trans. Mobile Comput.*, vol. 18, no. 5, pp. 1054–1067, May 2019.

[24] A. M. Ashkanani, A. S. M. Roza, and H. Naghavipour, "A design approach of automatic visitor counting system using video camera," *IOSR J. Elect. Electron. Eng. (IOSR-JEEE)*, vol. 10, no. 2, pp. 62–67, 2015.

[25] T.-H. Chen, T.-Y. Chen, and Z.-X. Chen, "An intelligent people-flow counting method for passing through a gate," in *Proc. IEEE Conf. Robot. Autom. Mechatronics*, 2006, pp. 1–6.

[26] J. García, A. Gardel, I. Bravo, J. L. Lázaro, M. Martínez, and D. Rodríguez, "Directional people counter based on head tracking," *IEEE Trans. Ind. Electron.*, vol. 60, no. 9, pp. 3991–4000, Sep. 2013.

[27] F. Wahl, M. Milenkovic, and O. Amft, "A distributed PIR-based approach for estimating people count in office environments," in *Proc. IEEE 15th Int. Conf. Comput. Sci. Eng. (CSE)*, 2012, pp. 640–647.

[28] A. Sikdar, Y. F. Zheng, and D. Xuan, "An iterative clustering algorithm for classification of object motion direction using infrared sensor array," in *Proc. IEEE Int. Conf. Technol. Practical Robot Appl. (TePRA)*, Woburn, MA, USA, 2015, pp. 1–6.

[29] *Counting System for a Library or a Museum*. Accessed: Sep. 2018. [Online]. Available: https://sensmax.eu/solutions/details/id//visitor-counting-system-for-a-library-or-a-museum

[30] S. Depatla and Y. Mostofi, "Passive crowd speed estimation and head counting using WiFi," in *Proc. 15th Annu. IEEE Int. Conf. Sens. Commun. Netw. (SECON)*, 2018, pp. 208–216.

[31] J. W. Choi, X. Quan, and S. H. Cho, "Bi-directional passing people counting system based on IR-UWB radar sensors," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 512–522, Apr. 2018.

[32] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of WiFi signal based human activity recognition," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, Paris, France, 2015, pp. 65–76.

[33] H. Wang, D. Zhang, Y. Wang, J. Ma, Y. Wang, and S. Li, "RT-fall: A real-time and contactless fall detection system with commodity WiFi devices," *IEEE Trans. Mobile Comput.*, vol. 16, no. 2, pp. 511–526, Feb. 2017.

[34] X. Wang, C. Yang, and S. Mao, "PhaseBeat: Exploiting CSI phase data for vital sign monitoring with commodity WiFi devices," in *Proc. IEEE 37th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, Atlanta, GA, USA, 2017, pp. 1230–1239.

[35] Y. Zhuo, H. Zhu, H. Xue, and S. Chang, "Perceiving accurate CSI phases with commodity WiFi devices," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Atlanta, GA, USA, 2017, pp. 1–9.

[36] R. W. Schafer, "What is a Savitzky–Golay filter? [Lecture notes]," *IEEE Signal Process. Mag.*, vol. 28, no. 4, pp. 111–117, Jul. 2011.

[37] T. Ji and A. Pachi, "Frequency and velocity of people walking," *Struct. Eng.*, vol. 84, no. 3, pp. 36–40, 2005.

[38] W. Wang, A. X. Liu, and M. Shahzad, "Gait recognition using WiFi signals," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput. (UbiComp)*, 2016, pp. 363–373.

[39] Y. Zeng, P. H. Pathak, and P. Mohapatra, "WiWho: WiFi-based person identification in smart spaces," in *Proc. 15th ACM/IEEE Int. Conf. Inf. Process. Sensor Netw. (IPSN)*, 2016, pp. 1–12.

[40] F. Hong, X. Wang, Y. Yang, Y. Zong, Y. Zhang, and Z. Guo, "WFID: Passive device-free human identification using WiFi signal," in *Proc. 13th Int. Conf. Mobile Ubiquitous Syst. Comput. Netw. Services (MOBIQUITOUS)*, 2016, pp. 47–56.

[41] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach Learn.*, 2015, pp. 448–456.

[42] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[43] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11n traces with channel state information," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 1, p. 53, 2011.

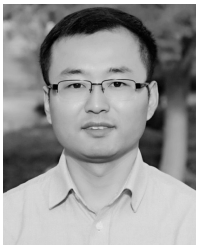
[44] M. Frustaci, P. Pace, G. Aloi, and G. Fortino, "Evaluating critical security issues of the IoT world: Present and future challenges," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 2483–2495, Aug. 2018.

Yanni Yang received the B.E. and M.Sc. degrees from the Ocean University of China, Qingdao, China, in 2014 and 2017, respectively. She is currently pursuing the Ph.D. degree with the Department of Computing, Hong Kong Polytechnic University, Hong Kong.

Her research interests include wireless human sensing, pervasive and mobile computing, and Internet of Things.

Jiannong Cao (M'93–SM'05–F'15) received the M.Sc. and Ph.D. degrees in computer science from Washington State University, Washington, DC, USA.

He is currently a Chair Professor with the Department of Computing, Hong Kong Polytechnic University, where he is also the Director of the Internet and Mobile Computing Laboratory and the Director of the University's Research Facility in Big Data Analytics. His research interests include parallel and distributed computing, mobile computing, and big data analytics.



Xiulong Liu received the B.E. and Ph.D. degrees from the Dalian University of Technology, Dalian, China, in 2010 and 2016, respectively.

He is currently a Professor with the College of Intelligence and Computing, Tianjin University, Tianjin, China. He also worked as a Visiting Researcher with Aizu University, Aizuwakamatsu, Japan, a Postdoctoral Fellow with Hong Kong Polytechnic University, Hong Kong, and a Postdoctoral Fellow with the School of Computing Science, Simon Fraser University, Burnaby, BC, Canada. His research interests include wireless sensing and communication, indoor localization, and networking.

Xuefeng Liu received the M.S. degree from the Beijing Institute of Technology, Beijing, China, in 2003, and the Ph.D. degree from the University of Bristol, Bristol, U.K., in 2008.

He is currently an Associate Professor with the Department of Computer Science, Beihang University, Beijing, China. His research interests include wireless sensor networks, and distributed computing and in-network processing.

Dr. Liu has served as a reviewer for several international journals/conference proceedings.