WiCrowd: Counting The Directional Crowd with A Single Wireless Link

Lei Zhang, *Member, IEEE,* Yueqiang Zhang, Beibei Wang, *Senior Member, IEEE*, Xiaolong Zheng, *Member, IEEE*, and Liu Yang, *Member, IEEE*

***Abstract*—Wi-Fi based crowd counting is predominant because of its noninvasive and ubiquitous advantages. However, the existing Wi-Fi based crowd counting systems have the constraint that there is always a maximum number of people counted. In order to address this issue, a Wi-Fi based cross-environment crowd counting system, which has the capability of both estimating the walking direction and crowd counting by only one single link, called WiCrowd is proposed. WiCrowd relaxes the restriction of people number counted and demonstrate its extraordinary robustness when environment changes. The signal change trends of the people flow are theoretically analyzed and people flow moving direction is inferred. The unique features using the eigenvalue of convariance matrix of amplitude and phase are derived to effectively detect prominent signal changes led by the crowd movement near LoS. By adopting the augmented feature representations, the robustness of WiCrowd is improved when environment changes. The experimental results in a typical indoor environment demonstrate the superior performance of WiCrowd. This system achieves 87.4%, 85.8%, 79.4% recognition accuracy for the flow movement direction estimation respectively, and 82.4% and 81.6% of overall cross environment accuracy for number of subjects counted in the people flow.**

***Index Terms*—CSI, LoS, crowd counting, direction estimation, CSI power, amplitude, phase.**

I. Introduction

CROWD COUNTING is the process of estimating the number of people in a given area, particularly in a closed environment. It has many potential applications, such as people flow control, guided tour, marketing analysis, train or flight scheduling and boarding, etc [1], [2], [3], [4], [5], [6]. Tracking the people queue could improve the quality of service as well as benefit each individual. However, it also has some basic challenges, including reliable observations and real time processing. There are many techniques for crowd counting, ranging from wearable sensor-based [7], floor sensor-based [8] to computer vision [9] based solutions. However, with the unique advantages of ubiquitous deployment, privacy protection and non-invasive, Wi-Fi based human activity sensing technology appeals lots of attractions. Recently, significant efforts have been made to explore the device-free crowd counting [10], [11], [12], [13], [14] with Wi-Fi. The Wi-Fi signal reflected by human body produces unique variations in Wi-Fi signals due to multipath effect. By taking advantage of the signal processing techniques, people can be counted by Wi-Fi devices.

However, the most existing Wi-Fi based people counting systems have the constraint that there is always a maximum number of people counted in a certain area. If the number of people is larger than maximum, the counting accuracy can not be guaranteed since the same subjects could be counted more than one time. This strong restriction impedes Wi-Fi based crowd counting from being widely used.

In order to tackle this challenge, we present WiCrowd, a Wi-Fi based cross-environment crowd counting system, which has the capability of both estimating the walking direction as well as counting the number of subjects in a given area by only one single link. Particularly, the upper bound of crowd number in a certain area is removed in WiCrowd. WiCrowd consists of one pair of commercial Wi-Fi devices, a transmitter Tx and a receiver Rx, as shown in Fig. 1(a). The device are placed on both sides of the door, both the walking direction and the people flow statistics can be inferred. There are three technical contributions.

The first technical contribution is to employ only one single link to estimate the direction of crowd movement. The CSI dynamics generated by the same activity performed at two symmetrical positions of LoS is similar. Moreover, due to the influence of multiple people movements, many parameters of individual cannot be obtained directly and accurately, such as Angle-of-Arrival (AOA), Time-of-Flight (TOF) and Doppler frequency shift (DFS) [15]. To address this issues, we take advantages of signal attenuation of the walls at both ends of the door and theoretically analyzed the signal change trends of the people flow. The difference between the signal strength of the entry and exit is analyzed and direction estimation of the crowd movement is implemented accordingly.

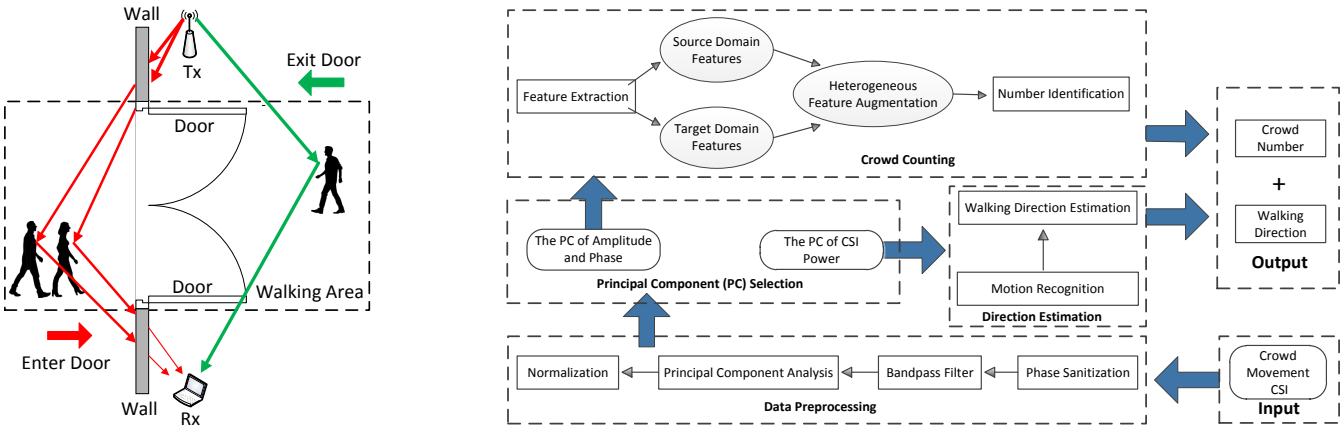
Manuscript received January 23, 2019; revised August 19, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61772364, 62076179, 61732011, 61702358, 61932013 and 61672240, in part by the Beijing Natural Science Foundation under Grant Z180006, in part by the Guangdong Key Research and Development Plan Project under Grant 2019B010153002, and in part by the Tianjin Science and Technology Plan Project under Grant 19ZXZNGX00050. (Corresponding author: Liu Yang.)

L. Zhang and Y. Zhang are with the College of Intelligence and Computing, Tianjin University, Tianjin, China; L. Zhang and Y. Zhang are also with the Tianjin Key Laboratory of Advanced Network Technology and Application, Tianjin, China; L. Zhang is also with the Key Laboratory of Grain Information Processing and Control, Henan University of Technology, Ministry of Education, Henan, China (e-mail: lzhang@tju.edu.cn; zhangyueqiang@tju.edu.cn)

B. Wang is with the Origin Wireless Inc. and also affiliated with the University of Maryland, College Park, USA (e-mail: bebewang@umd.edu)

X. Zheng is with the School of Computer Science & Beijing Key Lab of Intelligent Telecommunication Software and Multimedia Beijing University of Posts and Telecommunications, Beijing, China (e-mail: zhengxiaolong@bupt.edu.cn)

L. Yang is with the College of Intelligence and Computing, Tianjin University, Tianjin, China (e-mail: yangliuyl@tju.edu.cn)



(a) Application scenario

(b) WiCrowd architecture

Fig. 1. WiCrowd system overview

The second technical contribution is to quantify the correlation of the number of walking people and corresponding CSI features. When walking, the torso of human body has much larger surface area than other parts such as arms and legs [16], it reflects more signals and dominates fluctuations in CSI waveform. Different signal states of the torso near LoS are utilized to count the number of people in the crowd flow. WiCrowd is based on two types of crowd flows. One is people pass through the door one by one, and the other is people pass through the door in parallel. The typical waveforms of the crowd flow are classified into two waveform types: different number of people and different types of crowd flow. There are two phenomena. The one is that when the crowd flow type is the same, different number of people flow's waveforms have different number of prominent shapes. The other is that when the number of people in the moving crowd is same, different types of phase prominence in the waveform have different durations. Based on these phenomena, the unique features using the eigenvalue of convariance matrix of amplitude and phase are derived to effectively detect prominent signal changes led by the crowd movement near LoS.

The third technical contribution is to make WiCrowd more robust in various scenarios. The sizes of the sliding windows for crowd counting are adjusted according to the door widths in different scenarios. In this way, the number of features extracted in the sliding window is different in various scenarios, which makes the classification model difficult to build. To address this issues, the power-irrelevant and variation-dependent features are derived and gengerally extensible to various scenarios. Augmented feature representations are employed to tackle the problem of extracting different number of features. By augmenting the transformed samples with their original features and zeros, the dimension of the features are independent of the scenarios [17]. The Support Vector Machine (SVM) algorithm is introduced to seek for a decision function identifying the number of people. Finally, the number of people in the crowd flow can be obtained without specific environment calibration.

Finally, a prototype of the WiCrowd with commercial Wi-Fi devices is built and its superiority in the various indoor environment is demonstrated with extensive experiments. This system achieves 87.4%, 85.8%, 79.4% recognition accuracy for the flow movement direction estimation respectively, and 82.4% and 81.6% of overall cross environment accuracy for number of subjects counted in the people flow.

The remainder of this article is organized as follows. In the next section, we review the related work followed by some preliminary in Section III. The detailed design and implementation of WiCrowd are presented in Section IV. Then, we discuss the experimental results in Section V. Some limitations and discussions are presented in Section VI. Finally, the conclusion is drawn in Section VII.

II. RELATED WORK

Walking Direction Estimation: As an indispensable component of the crowd counting, walking direction estimation is to estimate moving direction of people flow. There are various approaches for the walking direction estimation. Qian *et al.* correlate Doppler shifts with motion directions, and propose a light-weight pipeline to detect, segment, and recognize motions without training [18]. By extracting DFS, AOA and TOF information from CSI, Qian *et al.* enable passive human localization and tracking using a single link of commercial Wi-Fi devices. [15]. Li *et al.* employ Doppler-MUSIC method to extract accurate Doppler velocity information and Doppler-AoA to determine the absolute trajectory of the target by jointly estimating target velocity and location [19]. Wu *et al.* analyze the phase change dynamics from multiple Wi-Fi subcarriers based on Fresnel zone model and the walking direction is inferred with multiple Wi-Fi links [16]. However, these commercial Wi-Fi based direction estimation systems can not work well in the multi-person existing scenarios. In addition, some systems have to recur to multiple Wi-Fi links, which degrades the convenience and practicality.

Crowd Counting: Crowd counting is to count the number of people in the crowd, which is another indispensable component of the crowd counting. There are a variety of approaches to do the crowd counting. Xi *et al.* apply a Gray Verhulst model to construct a normal profile of the dilated matrix of CSIs for each class [13]. Up to 30 people are counted in both

indoor and outdoor settings using a single AP and 3 or 4 receivers. In [10], deep neural network (DNN) is adopted as the classifier and up to 9 people (10 classes) can be counted in the most typical indoor scenarios. Depatla *et al.* propose a methodology that can extract the total number of people from the inter-event times and show how the probability mass function of the inter-event times carries vital information on the number of people [11]. Up to 20 persons are counted in five different areas. However, without direction estimation, these work has a limitation that there is always an upper bound for people can be counted.

Environment Influence Mitigation: In order to improve the robustness of the crowd counting system, there are various approaches to mitigate the influence of environment changes. Zheng *et al.* model the quantitative relation between complex gestures and CSI dynamics. Extracted velocity profiles of gestures in coordinates are domain-independent and they can act as unique indicators of gestures [20]. Zou *et al.* transform the RSS to a standardized location fingerprint based on a statistical shape analysis method. They also define Signal Tendency Index (STI) to measure the similarity between such standardized location fingerprints, which can address robustness issues with respect to both the device heterogeneity and environmental dynamics [21]. Domenico *et al.* employ differential CSI to classify the number of people [11]. The classifier is trained using 50% of the dataset collected in middle room and tested in large and small room. Zhang *et al.* effectively leverage transfer learning theory to mitigate the impact led by scenario change and maintain satisfied fall detection accuracy [22]. In the work above, the number of features extracted are same for the source domain and target domain. We need to solve the heterogeneous problem here in this work. More specifically, the number of features derived from source domain and target domain have different feature dimensions.

In general, the current systems have the issues of without flow moving direction estimation, having upper bound of counted number and without mitigating the influence of environment changes. We believe that WiCrowd opens up a new direction for the development of commodity Wi-Fi based crowd counting.

III. WiCrowd System

A. WiCrowd Architecture

The designed WiCrowd can estimate people flow moving direction as well as count the number of subjects in the crowd with a single Wi-Fi link. As shown in Fig. 1(b), crowd movement induced CSI measurements are collected in the area near the door, which are used as the input to WiCrowd. The walking direction and the number of subjects in the crowd are the output. WiCrowd, mainly consists of four core modules: 1) data preprocessing; 2) pricipal component (PC) selection; 3) direction estimation; 4) crowd counting. The data preprocessing module including four submodules: 1) phase sanitization; 2) band-pass filter; 3) pricipal component analysis; 4) normalization, works for the direction estimation and crowd counting preparation. The PC selection module includes the PC of amplitude, phase and CSI power submodules. All three submodules help to derive various features for different purposes. The features derived from the amplitude and phase are adopted for crowd counting. The features derived from the CSI power is utilized for direction estimation. The direction estimation module can carry out motion recognition and waking direction estimation. The crowd counting module includes feature extraction and number identification submodules. Due to environment impact mitigation, features are subdivided into source domain features and target domain features, and then they can be applied to number identification after heterogeneous feature augmentation.

B. Data Preprocessing

By adopting Intel 5300 Wi-Fi NIC with slight driver modification, CSI can be obtained from 30 OFDM subcarriers in each transmitting and receiving antenna pair. WiCrowd consists of one antenna at the transmitter and three at the receiver. Therefore, there are totally has $1\times3\times30$=90 subcarriers to record behavior characteristics. As the sampling rate is 1000 samples/s, 1000 CSI values can be obtained in one second for each subcarrier. Due to the frequent changes in internal CSI reference levels, transmit power levels and transmission rates, the CSI values provided by commodity Wi-Fi are inherently noisy. It is necessary to conduct data preprocessing.

1) Phase Sanitization: The measured phase $\hat{\phi_k}$ for the subcarrier k can be expressed as

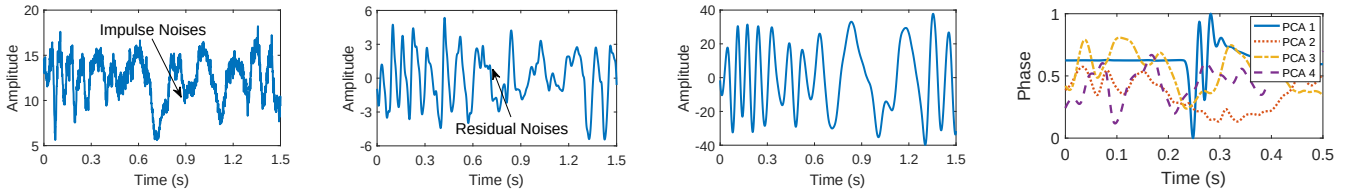
$$\hat{\phi_k} = \phi_k + 2\pi f_k \tau_k + \rho_k + \eta \tag{1}$$

where ϕ_k is the real phase, τ_k is the timing offset at the receiver, ρ_k is phase offset caused by Carrier Frequency Offset (CFO), Sampling Frequency Offset (SFO) and Packet Detection Delay (PDD) [23], [24], and η denotes the minor random noises. Due to the phase errors in listed above, it is infeasible to directly derive the movement induced phase information.

To mitigate the impact of phase errors, the sanitization algorithm in Spotfi [25] is applied. The best linear fit of the unwrapped phase is obtained. And then the unwrapped phase is utilized to subtract the linear fit.

2) Band-Pass Filter: The phase in Fig. 2(a) contains significant low-frequency inferences and burst noises, which are detrimental to movement identification. It is necessary to implement band-pass filter to further denoising CSI. Taking the actual situation consideration that normal human striding velocites are no more than V_s=1m/s, the upper cutoff frequency $f = \frac{2V_s}{\lambda} \approx \frac{2\times1m/s}{0.05m}$ is set to 40HZ. The lower cutoff frequency is set to 5HZ to remove the interference of subtle motions, such as respiration and heartbeat.

3) Principal Component Analysis (PCA): As shown in Fig 2(b), there are still some residual noises that has not been removed. All subcarriers have a strong linear correlation, indicating the redundancy of CSIs [26]. PCA is leveraged to remove the redundancy after the band-pass filter. The correlation between CSI streams can be automatically discovered and CSI streams can be recombined to extract the components, which best represent the variation led by subjects.



(a) The amplitude of raw CSI (b) After band-pass filter (c) After PCA (d) The top 4 phase principal components after normalization

Fig. 2. Data preprocessing.

As demonstrated in Fig. 2(c), amplitude waveform denoised by PCA become smoother.

4) Normalization: The normalization process is as follows:

$$q_i^j = \frac{p_i^j - \min p_i}{\max p_i - \min p_i} \tag{2}$$

where i and j represents the index of principal components and samples, p and q represent the vector before and after normalization, respectively.

Normalization is employed to further reduce the noise and more importantly, solve the problem of model training, and make the decision function converge faster.

C. Principal Component (PC) Selection

1) The PC of CSI Power: CSI power is the square of the amplitude, i.e. $|H(f,t)|^2$. The receiver has three antennas to collect CSI. Because of the various antenna positions, the signals received may vary greatly. For antennas insensitive to signal changes, power changes are difficult to detect. Additionally, different subcarriers have different sensitivity to the movements of walking due to the frequency diversity [27], [28]. However, the CSI streams are correlated. Similar results can be obtained for CSI streams between different antenna pairs because the difference in poisitions of antennas only causes initial phases and attenuations for each multipath to be different. The correlations between CSI streams are analyzed by PCA. Therefore, the signals that are representative of body movements. According to CARM [29], CSI power is not affected by CFO, but is affected by the remain information about the movement speeds of the body. Finally, the second principal component of CSI power is employed which not only has a small influence on the noise component, but also has a significant influence on the walking [30].

2) The PC of Amplitude and Phase: Principle components are in descending order in terms of their variance. The first phase principle component after normalization are selected as the extraction object. In Fig. 2(d), it can be clearly observed that other normalized principal components of phase still have some fluctuations.

D. Direction Estimation

1) Phase Difference: When there are walls in the propagation path, the radio frequency (RF) signal would traverse the wall, reflect off objects and humans, and come back imprinted with a signature of what is inside a closed room. As shown in Fig. 4, when passing through the wall, the signals outside the door attenuate more drastically. In this process, electromagnetic signals are absorbed by conductor in the walls [31]. In Fig. 3, the waveform of signal fluctuations in various conditions is illustrated.

TABLE I
ONE-WAY RF ATTENUATION IN COMMON BUILDING MATERIALS AT 2.4 GHZ [32], [33].

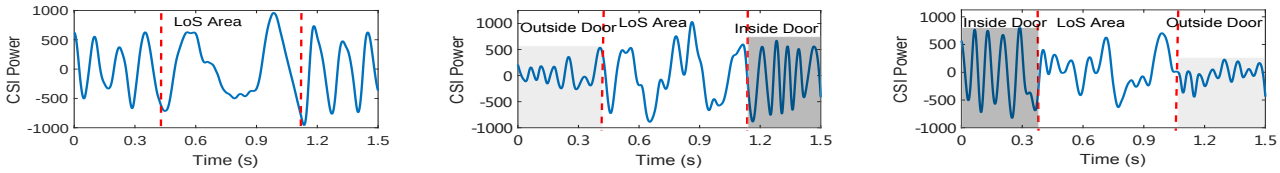
Building Materials	2.4GHz
Solid Wood Door 1.75 inches	6dB
Glass 1 inches	13dB
Steel Fire/Exit Door 1.75 inches	13dB
Interior Solid Wall 5 inches	14dB
Concrete Wall 18 inches	18dB

Observation I: In the scenarios with and without wall, the power variations are obviously different. Specifically, as shown in Fig. 3(a), only the multi-path effect works on the signal propagation in open area and the power variance does not change much. However, as in Fig. 3(b-c), both multi-path and wall blocking work on the signal attenuation. There are distinct power variations going inside and outside the door.

Observation II: The CSI fluctuations inside the door are more intense than that of outside the door in Fig. 3(b-c). However, in Fig. 3(b-c), the CSI fluctuations appear in different shifting orders inside and outside the door. In order to describe these differences accurately, the mean absolute deviation (MAD) values of CSI power are calculated. In Fig. 3(b), the MAD values of inside and outside door are 367 and 170, respectively. In Fig. 3(c), the MAD values inside and outside door are 299.5 and 130.5, respectively.

These two observations indicate that the signal fluctuations inside and outside the door are greatly different when entering or exiting. This is an intuitive proof to feasibility of direction estimation with one pair of commercial Wi-Fi devices. TABLE I illustrates a few examples of the one-way attenuation experienced by Wi-Fi signals in common construction materials [32], [33]. It is obvious that the different wall materials produce different attenuation. The more significant attenuation electromagnetic signals suffer, the more serious blocking effect will be. For instance, a one-way traversal of a glass wall and a concrete wall can reduce Wi-Fi signal power by 13dB and 18dB respectively. When the blocking effect is more severe, the power changes inside and outside the door are more obvious, which is helpful for direction estimation.

When the phase changes π, the waveform changes a peak and valley [15]. Therefore, the change of phase difference during the activity is analyzed. An subject that moves by a



(a) People walk in an open area (b) People enter the door of office (c) People exit the door of office

Fig. 3. The waveforms of signal fluctuations in various scenarios.

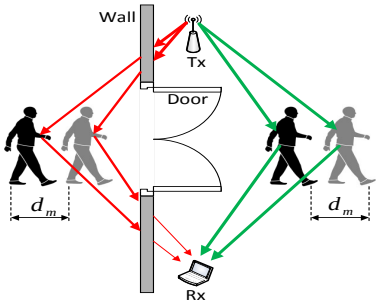


Fig. 4. The paths change when people enter the door.

small distance between time 1 and time 2 is considered. We have the phase difference:

$$\Delta\phi = e^{-j\frac{2\pi d(t_1)}{\lambda}} - e^{-j\frac{2\pi d(t_2)}{\lambda}} \tag{3}$$

where $d(t_1)$ and $d(t_2)$ are two paths, λ is the wavelength. As demonstrated in Fig. 4, the subject walks along the perpendicular bisector of the transceivers and d_m is the distance moved on both sides of the perpendicular bisector. $\Delta\phi_{in}$ and $\Delta\phi_{out}$ are the phase difference between the inside and outside of the door when entering. The phase difference ratio is:

$$\frac{\Delta\phi_{in}}{\Delta\phi_{out}} = \frac{e^{-j\frac{2\pi d_{in}(t_1)}{\lambda}} - e^{-j\frac{2\pi d_{in}(t_2)}{\lambda}}}{e^{-j\frac{2\pi d_{out}(t_1)}{\lambda}} - e^{-j\frac{2\pi d_{out}(t_2)}{\lambda}}} \tag{4}$$

To simplify the equation, $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}$ and $\mathcal{E}$ are employed to represent the terms: $-j\frac{2\pi}{\lambda} = \mathcal{A}$, $d_{in}(t_1) = \mathcal{B}$, $d_{in}(t_2) = \mathcal{C}$, $d_{out}(t_1) = \mathcal{D}$ and $d_{out}(t_2) = \mathcal{E}$. (5) can be simplified as:

$$\frac{\Delta\phi_{in}}{\Delta\phi_{out}} = \frac{e^{\mathcal{AB}} - e^{\mathcal{AC}}}{e^{\mathcal{AD}} - e^{\mathcal{AE}}} \tag{5}$$

The wall blocking can be reflected on the length of the reflection path. This is equivalent to adding a virtual path with distance d_{vir} to the reflection path to achieve the attenuation effect after passing through the wall. i.e. $\mathcal{D} = \mathcal{B} + d_{vir}$ and $\mathcal{E} = \mathcal{C} + d_{vir}$. It is easy to get the phase difference ratio is larger than 1. Therefore, $\Delta\phi_{in} > \Delta\phi_{out}$ relative to moving the same real distance can be concluded. Therefore, the variations of CSI power can be used for direction estimation.

2) Walking Direction Estimation: To detect the flow moving direction, a moving window algorithm is ultilized to detect the changing trend in the second principal component of CSI power. As following, there are four steps to estimate the moving direction.

First, WiCrowd monitors the peaks in the second principal componnet of CSI power waveform to identify the start and end of motion. The peaks are identified by comparing each value in wavaform with the value before and after it. If any value is less than both values before and after it, that value is a local minimum. Similarly, if any value is greater than the values before and after it, it is a local maximum. In order to ensure the peaks found are led by the subject movements rather than noises, a threshold T is used. Only the peaks with local difference $>T$ are considered, where a local difference is equal to a local maximum minus a local minimum.

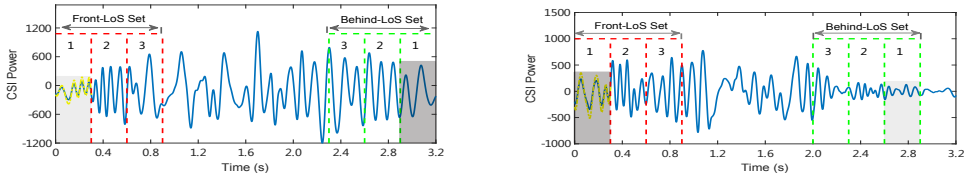
Second, the algorithm checks the number of peaks over a certain sliding window. When the number $\geq N_{peak}$, the window of size W_d is set as a candidate window. The number of candidate window, $N_{candidate}$ are selected from the start and end of the sliding windows, respectively. $2N_{candidate}$ candidate windows are divided into front-LoS and behind-LoS sets.

Third, MAD is calculated for each candidate window at jth iteration to detect the extent of variations of a given time series values. If variance is chosen instead of MAD, the deviations from the mean are squared which gives more weight to extreme values. The MAD is calculated using following equation.

$$\Delta m_j = \frac{1}{W_d} \sum_{i=j}^{j+W_d} |p(i) - \bar{p}(j : j + W_d)| \tag{6}$$

where $\bar{p}(j : j + W_d)$ repersents the vector of means of second PC in jth window. The window with the lowest Δm_j value is selected from each of the two sets.

Fourth, a weight coefficient σ is introduced into the algorithm. In the parameter settings above, the wavebands of front-LoS and behind-LoS are treated equally. This ignores the problem of various walking speeds inside and outside the door. In addition, when the subject walks on a different path or in a different direction with respect to the transceivers, the CSI power changes even if the subject retains the same walking speed. There are two ways to solve this problems. Firstly, according to the previous analysis, the parameters Δm_j and N_{peak} peaks of the front-LoS should be set differently from that of the behind-LoS, which can solve the problems from the source. Unfortunately, for this solution, it is infeasible to find the appropriate parameters due to individual differences. Secondly, a weight coefficient is introduced to the selected window of a set. Specifically, from the empirical study of the data set, we found that when subjects exit the door, the difference between the MAD values of the windows selected from front-LoS and behind-LoS is smaller than when they enter. Therefore, the weight coefficient is added to the selected window of front-LoS. Obviously, the latter way is more



(a) People enter the door

(b) People exit the door

Fig. 5. The CSI power waveforms when people go through the door.

feasible than the former way since it adjusts the parameters based on the empirical data analysis. As shown in Fig. 3(c), when the subject enters, the value of Δm_j inside door is more than twice that of outside door, the σ value is usually in the range of 1 and 2.

In summary, the parameters including T, N_{peak}, $N_{candidate}$, W_d, σ are empirically determined. As illustrated in Fig. 5(a-b), these parameters are set to be 300, 2, 3, 300 and 1.5, respectively. The values of these parameters will vary in different environments.

E. Crowd Counting

1) Amplitude and Phase Analysis: The most difficult and important part of crowd counting is to select the appropriate features representing the movements of people flow. In order to achieve this, it is necessary to quantify the correlation of the number of walking people and corresponding CSI features.

When multiple people walk around the door, one subject's walking induced CSI dynamics is easily overwhelmed by CSI dynamics led by other people walking. However, when walking, the torso of human body has much larger surface area than other parts such as arms and legs, it reflects more signals and dominates fluctuations in CSI waveform. Different signal states of the torso near LoS can always be detected in the waveform. Based on this fact, the correlation between the number of people in the flow and CSI features are built.

As shown in Fig. 6(a), when a subject enters the door, the waveform of the phase is predominant during the time as he passes through the LoS. Then, two subjects with similar walking speeds are selected. When the two walking subjects maintain the distance between them about 0.25m (about half the distance that the normal person take a step), the corresponding phase waveform has two predominance as shown in Fig. 6(b). Each prominent duration is about 0.05s. Fig. 7(a) shows the multipath propagation models when people walk one by one. This explains that this duration is the time consumed for the signals to pass through one subject's torso. When the subjects walk through the door in parallel, the phase waveform has only one clear predominance as shown in Fig. 6(c). However, it is worth noting that this prominence lasts longer than one subject walking through the door. Different from Fig. 7(a), the reflection signals need to pass through two subjects' torsos in Fig. 7(b). The duration of the predominance is about 0.3s. As illustrated in Fig. 6(d), when three subjects walk through the door, the amplitude waveform has three predominance. The signal state of the torso near LoS is more obvious in amplitude. This provides a strong evidence to investigate the relationship between the amplitude and phase to distinguish number of people in the crowd.

2) Feature Extraction: Different signal states of the torso near LoS are detected by examining walking induced CSI. Motivated by [34], [35], the features of moving crowd are extracted from the covariance eigenvalues of the first normalized principal components of amplitude and phase over the sliding windows, which are absolute power irrelevant and variance dependent. The first normalized principal components of amplitude $||H||_1$ and phase ϕ_1 corresponding covariance vector is:

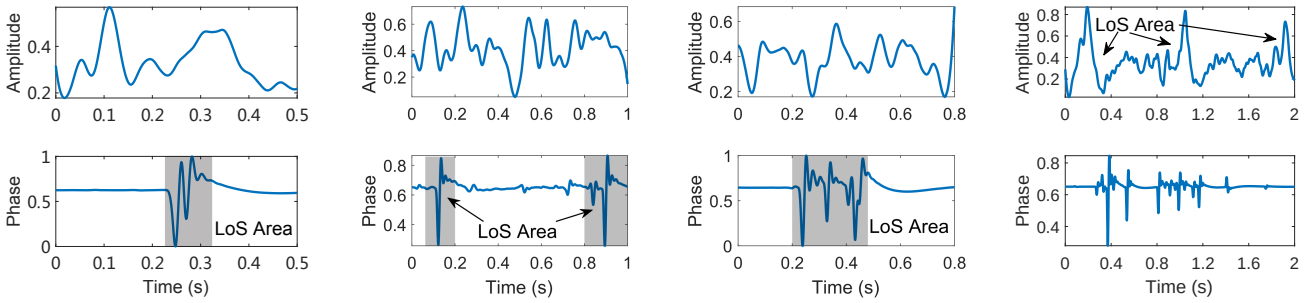
$$cov(||H||_1, \phi_1) = [cov(||H||_1^j, \phi_1^j)]_{1\times N_{slide}} \tag{7}$$

where j and N_{slide} represent the index and number of sliding windows respectively.

As mentioned above, when two subjects pass through a single door in parallel, the duration of walking induced CSI wave predominance is 0.3s, the data window size W_c is set to be 1000×0.3=300 and the step size of each sliding is N_{step}=150. The final number of windows is N_{slide}=28. When the crowd pass through double door, the window size of W_c=300 is no longer suitable. Because in this scenario, the more people pass through the door in parallel, the longer the predominance duration in the walking induced CSI wave lasts. The data window size W_c is empirically set to be 500 and the step size of each sliding N_{step} is is empirically set to be 250. The final number of windows is N_{slide}=16. The larger eigenvalues of sliding window indicate that the torso is closer to LoS area. In contrast, the lower eigenvalues of sliding window would probably indicate that the other parts of the body are close to the LoS area or the torso is far away from the LoS area. The ranges of each sliding window of amplitude and phase are also the final features. They are the maximum value minus the minimum value of the vectors. Range denotes the discrete degree of data, which has a good auxiliary effect on finding the predominance. After obtaining the correlation of amplitude and phase, the training models are built for classification.

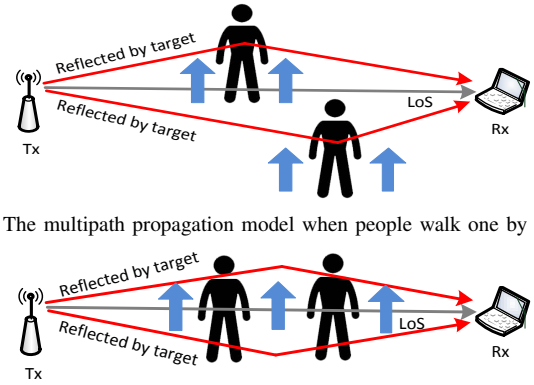
In sum, there are four features extracted on each window. They are the covariance eigenvalues of amplitude and phase, and the full range of them. Then, these features are applied to build the model of multiple classifiers.

3) Number Identification: Since the door width determines the number of subjects passing through the door in parallel, the scenarios are divided into single-door and double-door. The following two issues are cosidered. Firstly, the maximum number of people walking concurrently through the single-door and the double-door are different. This leads to the imbalance



(a) One person enters the door (b) Two persons enter the door one by one (c) Two persons enter the door in parallel (d) Three persons enter the door one by one

Fig. 6. The signal fluctuations when different number of people walking.



(a) The multipath propagation model when people walk one by one

(b) The multipath propagation model when people walk in parallel

Fig. 7. The multipath propagation models of people walking. Both figures omit some static paths, such as ground or furniture reflections.

TABLE II
KEY NOTATIONS.

Symbol	Meaning
n_s/n_t	number of source/target domain samples
$\mathbf{x}_i^s/\mathbf{x}_i^t$	ith source/target domain samples(i=1,...,n_s)
y_i^s/y_i^t	label of $\mathbf{x}_i^s/\mathbf{x}_i^t$
d_s/d_t	dimension of $\mathbf{x}_i^s/\mathbf{x}_i^t$
$\mathbf{I}_n$	$n \times n$ identity matrix
$\mathbf{O}_{n\times m}$	$n \times m$ zero matrix
$\mathbf{0}_n$	$n \times 1$ column vectors of all zeros
$\mathbf{1}_n$	$n \times 1$ column vectors of all ones
$\mathbf{K}_s/\mathbf{K}_t$	kernel of source/target domain samples
$\phi_s(\cdot)/\phi_t(\cdot)$	nolinear feature mapping function induced by $\mathbf{K}_s/\mathbf{K}_t$

in the data distribution of two data sets. Secondly, different scenarios may have different feature dimensions, which leads to the heterogeneous problem. The dimension of feature is determined by the size of the sliding window. Therefore, in different scenarios, the optimal number of features extracted may be different.

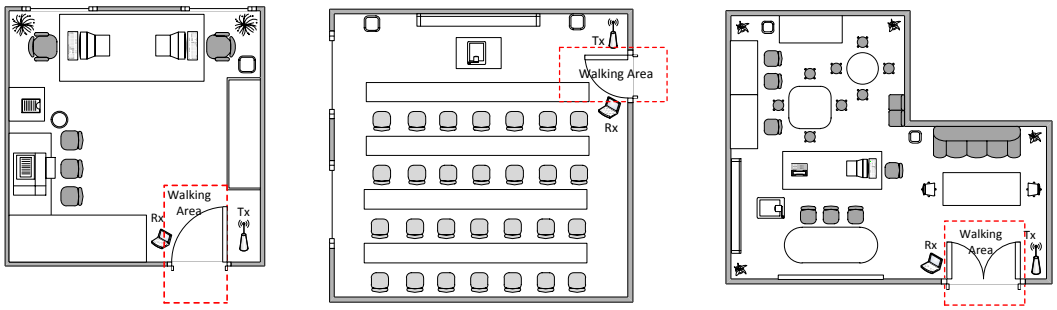
As discussed above, the data distributions of two scenarios are different. The data of double-door scenario are regarded as the source domain and the data of single-door scenario are regarded as the target domain. Because the samples of the source domain are more widely distributed and contain some distributions impossible to exist in the target domain. There are two phases to build the classification models.

In the first phase, the LibSVM tool [36] is appied to build the classification model with the optimal values for the Radial Basis Function (RBF) kernel. Grid search is the method of parameter optimization. The purpose of building a model for each scenario is to verify the extracted features are good enough for crowd counting.

In the second phase, a method Heterogeneous Feature Augmentation (HFA) [17] is employed, which can tackle the two problems above. HFA introduces a common subspace for the source and target data, where the heterogeneous features from two domains can be compared. What needs to be done is to find the corresponding projection matrix of the source domain and the target domain respectively. Some zero elements are needed to build the common subspace. This approach is feasible in terms of its own features. Only the larger eigenvalues are more likely indicate that the torso is near LoS area, while the supplemented zeros have little effect on the final result of model recognition. After feature transformation, the dimensions of the common subspace do not depend on the dimensions of the source domain or the target domain. So, the heterogeneous problem mentioned above is solved. The transformed new features in the common subspace are incorporated into Support Vector Machine (SVM), and the training model can take into account the samples of both the source and target domains, which solves the first problem of imbalanced distribution between the two domains.

4) The Method of HFA: We have a set of labeled training samples $\{(\mathbf{x}_i^s, y_i^s)|_{i=1}^{n_s}\}$ from source domain as well as a limited number of labeled smaples $\{(\mathbf{x}_i^t, y_i^t)|_{i=1}^{n_t}\}$ from the target domain, where y_i^s, $y_i^t \in \{0, 4\}$ are the labels of the sample x_i^s and x_i^t, respectively. The dimensions of $\mathbf{x}_i^s$ and $\mathbf{x}_i^t$ are d_s and d_t respectively ($d_s \neq d_t$). $\mathbf{I}_n$ is defined as the $n \times n$ identity matrix and $\mathbf{O}_{n\times m}$ is defined as a $n \times m$ matrix of all zeros. $\mathbf{0}_n$ and $\mathbf{1}_n$ are defined as the $n \times 1$ column vectors of all zeros and all ones, respectively. The notations used this section are summarized in TABLE II.

In this paper, nonlinear feature transformation is performed to achieve independence from the feature dimension. Feature mapping function can be used to derive a corresponding kernel space for the source and target data. $\mathbf{K}_s = \mathbf{\Phi}_s' \mathbf{\Phi}_s \in \mathbb{R}^{(n_s \times n_s)}$ is used as the kernel on the source domain samples, where $\mathbf{\Phi}_s = [\phi_s(\mathbf{x}_1^s), ..., \phi_s(\mathbf{x}_{n_s}^s)]$ and $\phi_s(\cdot)$ is the nolinear feature



(a) Office (b) Classroom (c) Exhibition Room

Fig. 8. Layouts of three scenarios.

mapping function induced by $\mathbf{K}_s$. Similarly, $\mathbf{K}_t = \mathbf{\Phi}_t'\mathbf{\Phi}_t \in \mathbb{R}^{(n_t \times n_t)}$ is used as the kernel on the target domain samples, where $\mathbf{\Phi}_t = [\phi_t(\mathbf{x}_1^t), ..., \phi_t(\mathbf{x}_{n_t}^t)]$ and $\phi_t(\cdot)$ is the nolinear feature mapping function induced by $\mathbf{K}_t$. Moreover, the dimensions of the nonlinear features $\phi_s(\mathbf{x}_i^s)$ and $\phi_t(\mathbf{x}_i^t)$ are denoted as $\widetilde{d}_s$ and $\widetilde{d}_t$, respectively.

The common subspace is defined as $\mathbb{R}^{d_c}$, where $\phi_s(\mathbf{x}_i^s)$ and $\phi_t(\mathbf{x}_i^t)$ can be projected onto it by using two projection matrices $\widetilde{\mathbf{P}} \in \mathbb{R}^{d_c \times \widetilde{d}_s}$ and $\widetilde{\mathbf{Q}} \in \mathbb{R}^{d_c \times \widetilde{d}_t}$ respectively. The feature weight vector $\mathbf{w} = [\mathbf{w_c}', \mathbf{w_s}', \mathbf{w_t}']'$ is defined for the augmented feature space, where $\mathbf{w_c}, \mathbf{w_s}$ and $\mathbf{w_t}$ are also vectors that are defined for the common subspace, the source domain and the target domain respectively. Then, learning $\widetilde{\mathbf{P}}$ and $\widetilde{\mathbf{Q}}$ as well as the weight vector $\mathbf{w}$ by minimizing the structural risk functional of SVM is proposed. Formally, the formulation of HFA method can be present as follows:

$$\min_{\widetilde{\mathbf{P}},\widetilde{\mathbf{Q}}} \min_{\mathbf{w},b,\xi_i^s,\xi_i^t} \quad \frac{1}{2}\|\mathbf{w}\|^2 + C(\sum_{i=1}^{n_s}\xi_i^s + \sum_{i=1}^{n_t}\xi_i^t), \tag{8}$$

$$s.t. \quad y_i^s(\mathbf{w}'\phi_s(x_i^s) + b) \geq 1 - \xi_i^s, \xi_i^s \geq 0; \tag{9}$$

$$y_i^t(\mathbf{w}'\phi_t(x_i^t) + b) \geq 1 - \xi_i^t, \xi_i^t \geq 0; \tag{10}$$

$$\|\widetilde{\mathbf{P}}\|_F^2 \leq \lambda_p, \|\widetilde{\mathbf{Q}}\|_F^2 \leq \lambda_q$$

where $C > 0$ is a regularization parameter, and $\lambda_p, \lambda_q > 0$ is predefined to control the complexities of $\widetilde{\mathbf{P}}$ and $\widetilde{\mathbf{Q}}$, respectively. Dual variables $\{\alpha_i^s\|_{i=1}^{n_s}\}$ and $\{\alpha_i^t\|_{i=1}^{n_t}\}$ are introduced for the constraints in (10) and (11) respectively. By setting the derivatives of the Lagrangian of (9) with respect to $\mathbf{w}, b, \xi_i^s$ and ξ_i^t to zeros, we obtain the Karush-Kuhn-Tucker (KKT) conditions as: $\mathbf{w} = \sum_{i=1}^{n_s}\alpha_i^s y_i^s \phi_s(x_i^s) + \sum_{i=1}^{n_t}\alpha_i^t y_i^t \phi_t(x_i^t)$, $\sum_{i=1}^{n_s}\alpha_i^s y_i^s + \sum_{i=1}^{n_t}\alpha_i^t y_i^t = 0$ and $0 \leq \alpha_i^s, \alpha_i^t \leq C$. With the KKT conditions, the alternative optimization problem is as follows:

$$\min_{\widetilde{\mathbf{P}},\widetilde{\mathbf{Q}}} \max_{\alpha} \quad \mathbf{1'}_{n_s+n_t}\alpha - \frac{1}{2}(\alpha \circ \mathbf{y})'\mathbf{K}_{\widetilde{\mathbf{P}},\widetilde{\mathbf{Q}}}(\alpha \circ \mathbf{y}), \tag{11}$$

$$s.t. \quad \mathbf{y}'\alpha = 0, \mathbf{0_{n_s+n_t}} \leq \alpha \leq C\mathbf{1_{n_s+n_t}}, \tag{12}$$

$$\|\mathbf{P}\|_F^2 \leq \lambda_p, \|\mathbf{Q}\|_F^2 \leq \lambda_q$$

where $\alpha = [\alpha_1^s, ..., \alpha_{n_s}^s, \alpha_1^t, ..., \alpha_{n_t}^t]'$ is a vector of the dual variables, $\mathbf{y} = [y_1^s, ..., y_{n_s}^s, y_1^t, ..., y_{n_t}^t]'$ is a label vector, $\mathbf{K}_{\widetilde{\mathbf{P}},\widetilde{\mathbf{Q}}} = \begin{bmatrix} \mathbf{X_s}'(\mathbf{I_{n_s}} + \widetilde{\mathbf{P}}'\widetilde{\mathbf{P}})\mathbf{X_s} & \mathbf{X_s}'\widetilde{\mathbf{P}}'\widetilde{\mathbf{Q}}\mathbf{X_t} \\ \mathbf{X_t}'\widetilde{\mathbf{Q}}'\widetilde{\mathbf{P}}\mathbf{X_s} & \mathbf{X_t}'(\mathbf{I_{n_t}} + \widetilde{\mathbf{Q}}'\widetilde{\mathbf{Q}})\mathbf{X_t} \end{bmatrix}$ is the derived kernel matrix.

To solve the optimization problem in (11), the dimension of the common subspace d_c must be given beforehand. An intermediate variable $\widetilde{\mathbf{H}} = [\widetilde{\mathbf{P}}, \widetilde{\mathbf{Q}}]'[\widetilde{\mathbf{P}}, \widetilde{\mathbf{Q}}] \in \mathbb{R}^{(\widetilde{d}_s+\widetilde{d}_t)\times(\widetilde{d}_s+\widetilde{d}_t)}$ is defined to replace multiplications in kernel matrix. $\widetilde{\mathbf{H}}$ is positive semidefinite, i.e., $\widetilde{\mathbf{H}} \succeq 0$. After introduced $\widetilde{\mathbf{H}}$, the parameter d_c can be throwed away and do not need to explicitly solve for $\widetilde{\mathbf{P}}$ and $\widetilde{\mathbf{Q}}$.

With the definition of $\widetilde{\mathbf{H}}$, the optimization problem in (12) can be converted as follows:

$$\min_{\widetilde{\mathbf{H}}\succeq 0} \max_{\alpha} \quad \mathbf{1'}_{n_s+n_t}\alpha - \frac{1}{2}(\alpha \circ \mathbf{y})'\mathbf{K}_{\widetilde{\mathbf{H}}}(\alpha \circ \mathbf{y}), \tag{13}$$

$$s.t. \quad \mathbf{y}'\alpha = 0, \mathbf{0_{n_s+n_t}} \leq \alpha \leq C\mathbf{1_{n_s+n_t}}, \tag{14}$$

$$trace(\widetilde{\mathbf{H}}) \leq \lambda,$$

where $\mathbf{K}_{\widetilde{\mathbf{H}}} = \begin{bmatrix} \mathbf{K_s} + \widetilde{\mathbf{L}}_\mathbf{s}'\widetilde{\mathbf{H}}\widetilde{\mathbf{L}}_\mathbf{s}' & \widetilde{\mathbf{L}}_\mathbf{s}'\widetilde{\mathbf{H}}\widetilde{\mathbf{L}}_\mathbf{t}' \\ \widetilde{\mathbf{L}}_\mathbf{t}'\widetilde{\mathbf{H}}\widetilde{\mathbf{L}}_\mathbf{s}' & \mathbf{K_t} + \widetilde{\mathbf{L}}_\mathbf{t}'\widetilde{\mathbf{H}}\widetilde{\mathbf{L}}_\mathbf{t}' \end{bmatrix}, \widetilde{\mathbf{L}}_s' = \begin{bmatrix} \mathbf{I}_{n_s} \\ \mathbf{O}_{n_t \times n_s} \end{bmatrix}\mathbf{K}_s^{1/2}$ and $\widetilde{\mathbf{L}}_t' = \begin{bmatrix} \mathbf{O}_{n_s \times n_t} \\ \mathbf{I}_{n_t} \end{bmatrix}\mathbf{K}_t^{1/2}$. Note the size of $\widetilde{\mathbf{H}}$ is independent from the feature dimensions $\widetilde{d}_s$ and $\widetilde{d}_t$. For a given α, we also arrive at an SDP problem as follows by defining $\beta = \alpha \circ \mathbf{y}$:

$$\min_{\widetilde{\mathbf{H}}\succeq 0} -\frac{1}{2}\beta'\mathbf{K}_{\widetilde{\mathbf{H}}}\beta, \quad s.t. \quad trace(\widetilde{\mathbf{H}}) \leq \lambda. \tag{15}$$

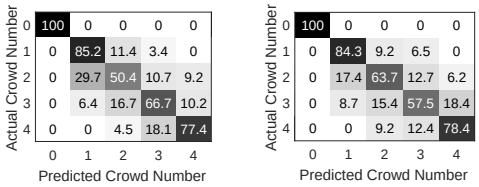
After obtaining the optimal solution α and $\widetilde{\mathbf{H}}$ to (13), sample $\mathbf{x}$ can be predicted from the target domain by using the following target decision function:

$$f(x) = \mathbf{w}'\phi_t(\mathbf{x}) + b$$

$$= ((\beta_s'\widetilde{\mathbf{L}}_s' + \beta_t'\widetilde{\mathbf{L}}_t')\widetilde{\mathbf{H}}\begin{bmatrix} \mathbf{O}_{n_s \times n_t} \\ \mathbf{I}_{n_t} \end{bmatrix})\mathbf{k}_t + b, \tag{16}$$

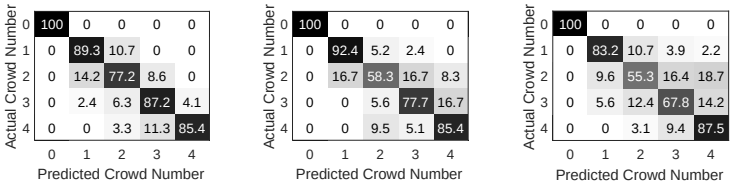
where $\mathbf{k}_t = [k(\mathbf{x}_1^t, \mathbf{x}), ..., k(\mathbf{x}_{n_t}^t, \mathbf{x})]$ and $k(\mathbf{x}_i, \mathbf{x}_j) = \phi_t(\mathbf{x}_i)'\phi_t(\mathbf{x}_j)$ is a predefined kernel function for two data samples $\mathbf{x}_i$ and $\mathbf{x}_j$ with the same feature dimension.

In conclusion, when a data sample from office or classroom (target domain) is obtained, the original features need to be extracted. After augmented feature representations are performed, by feeding them to the trained model, the output of the model is the number of crowd.



(a) Room A (b) Room B

Fig. 9. Confusion matrices of cross environment models.



(a) Room A (b) Room B (c) Room C

Fig. 10. Confusion matrices of single environment models.

IV. Experiments and Evaluation

A. Implementation

1) Experimental Environment: WiCrowd is implemented with two ThinkPad T-series laptops equipped with Intel 5300 Wi-Fi cards. One with an antenna works as a transmitter, while the other with three antennas works as a receiver. The both sender and receiver are 1.2m high. All laptops are installed with the Linux 802.11n CSI Tool [37] on Ubuntu desktop 11.04 LTS OS and work in monitor mode on Channel 161 at 5.825 GHz. The transmission rate is set to be 1000 Hz and the CSI measurements are processed by MATLAB.

2) Experimental Methodology: We prototype WiCrowd with commodity Wi-Fi devices and evaluate its performance in some typical scenarios in academic buildings, such as classroom, office and exhibition room, as shown in Fig. 8. For convenience, they are called room A, room B and room C respectively. Room A or B has a single wooden door, whose width is 0.9m. In both scenarios, the length of LoS is d_L=1.2m. Room C has a double wooden door, whose width is 1.4m. In this scenario, the length of LoS is d_L=1.8m. The transceiver placements are illustrated in Fig. 8.

To fully understand the variations in user diversity, 35 volunteers, including 22 males and 13 females, have been selected to test the effectiveness of WiCrowd over six months. They varied in age from 18 to 36 years old, in height from 158 to 190cm and in weight from 45 to 75kg. Up to 4 subjects were tested each time. The subjects were asked to walk freely through the door within the range of $0.75d_L$ outside the door and $1.5d_L$ inside the door.

B. Overall Performance

The experimental results from room A, room B and room C are summarized. As illustrated in Fig. 11, WiCrowd achieves average direction estimation accuracy of 87.4%, 85.8% and 79.4% respectively. For the crowd counting, the overall accuracies are demonstrated in Fig 9. It is obvious that WiCrowd performs well in different scenarios, with the mean accuracy of 82.4% and 81.6% in room A and room B in the target domain, respectively. As presented in Fig. 10, the mean accuracy of the model trained in room A, B, anc C is 86.4%, 84.2%, 80.2%, respectively. Compared with single-door scenario, WiCrowd performs worse in double-door scenario. The reason is the wider door allows more people go through, which induces more mobile multipath. This superimposed multipath deteriorates the performace of WiCrowd [38].

C. Direction Estimation

1) Impact of Walking Speed Diversity: According to the literature [39], [40], the human comfortable walking speed v is 1$\sim$1.5m/s. In order to test the impact of walking speed, the subjects were asked to walk at three different speed levels: slow (below $(v-0.3)$m/s), normal ($(v-0.3)$m/s$\sim(v+0.3)$m/s) and fast (above $(v+0.3)$m/s). As illustrated in Fig. 12, for all scenarios, with decrease of the walking speed, the direction estimation accuracy declines. Generally, fast walking speed is more favorable for direction estimation in all the scenarios. This is because the faster speed going through the door, the larger MAD difference it is. MAD difference refers to the sliding window MAD difference the crowd is inside and outside the door. The larger MAD difference is more beneficial to the direction estimation.

2) Impact of Window Size: The core approach for direction estimation is to find two sliding windows that best represent the movement inside and outside the door and compare the extent of variations within the window. It is necessary to select an appropriate window size. If the selected window size is too small, the increased computation load leads to the prolonged response time. If the selected window size is too large, the direction estimation accuracy can not be guaranteed. The accuracy results are shown in Fig. 13. It is worth noting that when the window size exceeds 500, the direction estimation accuracies of room B and C significantly reduce. The main reason is that the states of the people passing through the door change dramatically and the instant state changes can not be accurately captured. The optimal window size of room A, B and C is 200, 300 and 300 respectively. The optimal window size varies with the environment.

3) Impact of the Distance between the Transceivers: To evaluate the impact of the distance between the two transceivers on direction estimation, some experiments were conducted. As shown in Fig. 14, the distance between devices changes have insignificant impact on the direction estimation accuracy in room A and B, each of which has a single door with 0.9m width. When the distance between devices is greater than 0.9m, there will be a wall blocking effect [14], [41], which is beneficial to the direction estimation. Therefore, for each room (room A or B), direction estimation accuracy increases when the distance between the two transceivers approaches to 0.9m. After the distance reaches 1.2m, the accuracies begin to decline. In an another word, 1.2m is the best distance for the direction estimation. For room C, the overall tendency is similar to the ones in room A and B. Since the width of double door is 1.2m, the best estimation

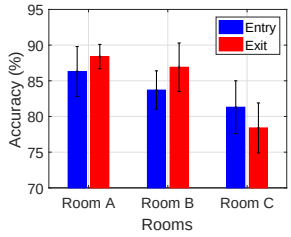


Fig. 11. Mean accuracy of direction estimation.

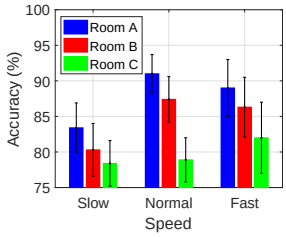


Fig. 12. Impact of different speed diversity on direction estimation.

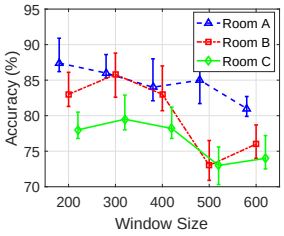


Fig. 13. Impact of different window size on direction estimation.

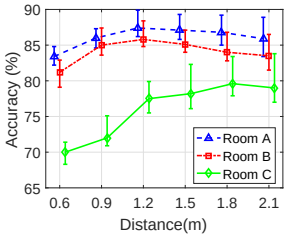


Fig. 14. Impact of transceivers' distance on direction estimation.

accuracy is obtained when the distance is 1.8m. After the transceivers' distance reaches 1.2m, the estimation accuracy slightly improves. When the transceivers' distance reaches 1.8m, WiCrowd performances the best in Room C. The reason of estimation accuracy declines afterward for all three rooms is with the increase of distance, signal power becomes weak. This offsets the benefit of wall blocking effect, reversing the rising trend of estimation accuracy.

4) Impact of Direction Inconsistency: Direction inconsistency refers to the situation that two crowd groups enter and exit the door at the same time. The impact of inconsistency on the direction estimation is shown in Fig. 15. In all three rooms, the accuracy in the consistent direction is much higher than that of the inconsistent direction. The average accuracy of room A, B and C drops from 89.4%, 87.7%, 82.3% to 78.3%, 77.2% 74.3% respectively. This is because the walking induced CSI dynamics from different walking directions are mixed. It is difficult to derive the direction related variations. This is the fundamental issue of the direction estimation with one single Wi-Fi link.

D. Crowd Counting

1) Impact of Window Size: This evaluation is to test the impact of window size on crowd counting in single environment. For crowd counting, the larger the size of sliding window can guarantee the better the performance since if window size is too small, the entire movement induced the signal states can not be completely captured. However, when sliding window size exceeds some threshold, the intrinsic variance of CSI will become comparable to variance caused by human motion, which offsets the benefit of increasing window size, reversing the rising trend of recognition accuracy [34]. Therefore, it is necessary to experientially select an appropriate window size. The impact of different window size on crowd counting is shown in Fig. 16. The optimal window size of room A, B and C is 200, 300 and 400 respectively.

2) Impact of the Amount of Training Data: To clarify the relationship between the recognition accuracy and the amount of training data in the target domain, the ratio is applied. This ratio represents the proportion the number of target domain training samples to the number of source domain training samples. As shown in Fig. 17, when the ratio is equal to 0, there is no sample data from target domain to train classifier. This returns to the case of the single environment motion identification. It is obvious that as the ratio increases, the performance gets better. The accuracy reaches its maximum when the ratio is 0.15 and almost gets constant afterward. This means more training samples can not improve the performance after the thresholds.

3) Robustness: To evaluate robustness, WiCrowd is tested with both single environment changes and cross environment changes. The single environment changes refer to moving the furniture around in the same room. And cross environment changes refer to the tests done in totally different rooms. Specifically, for the single environment changes, the bookcase in room A was moved to another side of the room. In room B, the desks in the first row were moved to the back of the room. The table with two chairs in room C was relocated to the other room. For single environment evaluation, WiCrowd was trained in original scenarios, and tested in the changed scenarios. For cross environment evaluation, the data collected in room C (double-door scenario) were used as the training dataset, and the data collected in other rooms (single-door scenarios) were used as testing datatests. These environmental changes are regarded as interference. The results with and without changes are shown in Fig. 19. It is obverious that all the models are environment dependent and optimal performance can be achieved only by fine-tuning the model in a specific environment [42], [43]. In Fig. 19(a), the environment changes lead to similar tendency in room A, B and C. The average accuracy of room A, B and C decreases from 86.4%, 84.2%, 80.2% to 84.8%, 82.1%, 78.2%, respectively. The movement of furnitures has insignificant impact on the crowd counting accuracy since only the movements passing through LoS are considered. In Fig. 19(b), the average accuracy of room A, B declines from 82.4%, 81.5% to 81%, 79.3%, respectively. It is obvious that the average accuracy only drops a little bit even if the environment changes totally. This validates the effectiveness of augmented feature representations in the environment dependence mitigation. In summary, WiCrowd is robust to various environment changes.

4) Impact of Walking Speed Diversity: According to the literature [39], [40], the human comfortable walking speed v is 1$\sim$1.5m/s. In order to check the impact of walking speed, the subjects were asked to walk at three different speed levels: slow (below $(v-0.3)$m/s), normal ($(v-0.3)$m/s$\sim(v+0.3)$m/s) and fast (above $(v+0.3)$m/s). The results are shown in Fig. 20. For single environment evaluation, WiCrowd was trained and tested in the same room. For cross environment evaluation, the data collected in room C were used as the training dataset, and the data collected in room A and B were used as testing datatests. The changing trends of the three speed levels under

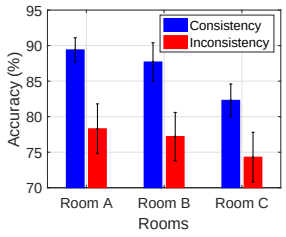


Fig. 15. Impact of direction inconsistency.

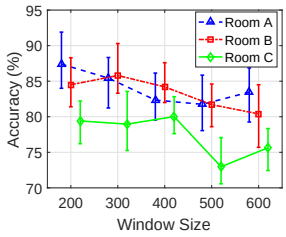


Fig. 16. Impact of different window size on crowd counting.

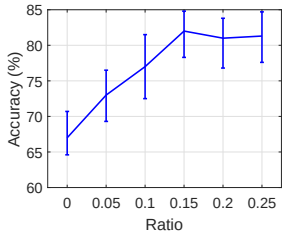


Fig. 17. Impact of the amount of training data on crowd counting.

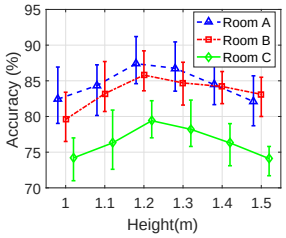
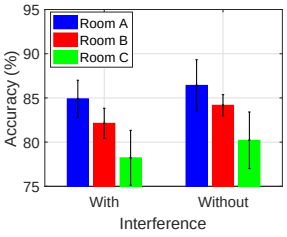
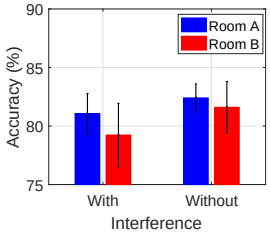


Fig. 18. Impact of transceiver height on crowd counting.

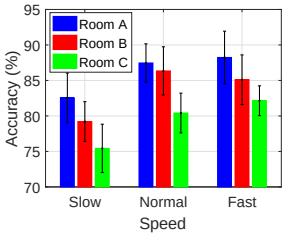


(a) Single environment

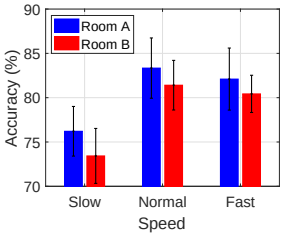


(b) Cross environment

Fig. 19. Impact of robustness on crowd counting.



(a) Single environment



(b) Cross environment

Fig. 20. Impact of different speed diversity on crowd counting.

the single environment and cross environment are similar. In Fig. 20(a), the average accuracy of slow speed level is 82.5%, 79.2%, 75.4%, normal speed level is 87.5%, 86.3% 80.4%, the fast speed level is 88.2%, 85.1%, 82.1%, respectively. In Fig. 20(b), the average accuracy of slow speed level is 76.2% 73.4%, normal speed level is 83.3%, 81.4%, fast speed level is 82.1%, 80.4%, respectively. It is worth noting that the slow speed has a negative effect on the crowd counting accuracy either in single environment or cross environment. For slow movement, the signal fluctuations caused by the torso are more difficult to be distinguished from the signal fluctuations led by the movements of the body's other limbs, such as arms and legs. Compared with single environment, the average accuracy of cross environment is only slightly lower. This verifies the effectiveness of augmented feature representations in reducing environment dependency from another aspect.

V. LIMATATIONS AND DISCUSSION

The WiCrowd system shows the feasibility of signle link commodity Wi-Fi based crowd counting and direction estimation. However, the current implementation of the WiCrowd has several limitations.

A. Direction Estimation

In this section, we will discuss some typical limitations in the direction estimation.

1) Direction Inconsistency: The basic principle of WiCrowd is to estimate the direction by using the signal fluctuation generated before and after passing through the door and finding their difference. When people pass through the door from the opposite directions, with one pair of transceivers, it is challenging to distinguish the fluctuation variance of each direction from the mixed signals. Adding an extra receiver might be a possible solution to the problem. The positions of the current two transceivers do not change. The added receiver needs to be in the same line with the existing receiver and there must be a certain distance between the two receivers. When walking from opposite directions, there will be variations in CSI power at two receivers. Therefore, direction estimation might be achieved in this case.

2) Parameters Setting: The parameters set by the direction estimation algorithm are scenario dependent. This becomes a big issue when environment changes and has a great change. For example, the range of weight coefficient σ should be related to the variation of CSI power inside and outside the door. In other words, the thickness and formation of the wall will affect the setting of σ. To solve this problem, an approach similar to grid search might be applied to find the optimal parameter values by limiting the scope and varying the step size.

3) Interval Between Sequential People: When the crowd go through the door sequentially, the signals reflected by multiple persons are more likely to cover each other, there will be an additive effect at the receiver, which weakens the received signal strength. When the interval between sequential people is too large, the signals of multiple people will be mixed together one after another, resulting in the failure to detect the variation trend of signal power.

B. Crowd Counting

In this section, the key issues of crowd counting are discussed.

1) Window Size: Firstly, as demonstrated before, when people walk through the door with slow speed, WiCrowd's performance gets deteriorated in crowd counting. The main reason is that the sliding window size is a fixed value. With the fixed window size, signal fluctuations led by multiple people sometimes are very similar to the ones led by a single person,

specially when walking speed is low. This makes it difficult to distinguish the exactly number of people.

2) Transceiver Height: There is a constraint that a pair of Wi-Fi devices should be deployed at a height similar to the human height. The heights of transceivers vary from 1 to 1.5m and the corresponding accuracies are plotted in Fig. 18. The best crowd counting accuracy can be achieved when the Wi-Fi devices are deployed at 1.2m. From the tendency in Fig. 18, the conclusion that the height of people is a key point influencing the counting accuracy can be drawn. If the heights of people vary greatly, it will be very difficult to set the heights of Wi-Fi devices. If people are in the similar heights, a fixed and appropriate height of the Wi-Fi devices must be set for the accurate crowd counting. Otherwise, even the height adjustment can not guarantee the crowd counting accuracy. If the counted people are too hight or too short, the reflective surface will decrease. This is the main reason for this constrain. How to tackle this issue will be an interesting future work.

VI. CONCLUSION

In this paper, WiCrowd, the first commercial Wi-Fi based crowd counting system without upper bound of people counted has been proposed. The key technical novelty of WiCrowd lies in the removal the maximum number of people counted as well as the robustness is improved by the mitigation of environment influence on flow counting. In order to achieve this, a series of signal processing techniques have been adopted to analyze signal change trends of the people flow, quantify the correlation of the number of walking people and corresponding CSI features. The augmented feature representations are employed to mitigate the environment change impact on crowd counting. Extensive experiments under different typical indoor environment have been conducted to validate the superiority of WiCrowd from different aspects. This system achieves 87.4%, 85.8%, 79.4% recognition accuracy for the flow movement direction estimation respectively, 82.4% and 81.6% of overall cross environment accuracy for number of subjects counted in the people flow. We firmly believe that WiCrowd is a promising step toward a practical wireless human-computer prototype, which underpins new insights for future wireless sensing applications.

REFERENCES

[1] Yuvraj Agarwal, Bharathan Balaji, Rajesh Gupta, Jacob Lyles, Michael Wei, and Thomas Weng. Occupancy-driven energy management for smart building automation. *BuildSys'10 - Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, pages 1–6, 11 2010.

[2] Omid Ardakanian, Arka Bhattacharya, and David Culler. Non-intrusive techniques for establishing occupancy related energy savings in commercial buildings. In *the 3rd ACM International Conference*, 2016.

[3] Olga Perdikaki, Saravanan Kesavan, and Jayashankar M. Swaminathan. Effect of traffic on sales and conversion rates of retail stores. *Manufacturing & Service Operations Management*, 14(1):145–162, 2012.

[4] Bluescan. www.bluescan.org/english/counting/traffic-counting/.

[5] Anuj Dimri, Harsimran Singh, Naveen Aggarwal, Bhaskaran Raman, K. K. Ramakrishnan, and Divya Bansal. Barosense: Using barometer for road traffic congestion detection and path estimation with crowdsourcing. *ACM Trans. Sen. Netw.*, 16(1), November 2019.

[6] Xu Wang, Zimu Zhou, Yi Zhao, Xinglin Zhang, Kai Xing, Fu Xiao, Zheng Yang, and Yunhao Liu. Improving urban crowd flow prediction on flexible region partition. *IEEE Transactions on Mobile Computing*, 19(12):2804–2817, 2020.

[7] Lan Zhang, Kebin Liu, Yonghang Jiang, Xiang Yang Li, Yunhao Liu, Panlong Yang, and Zhenhua Li. Montage: Combine frames with movement continuity for realtime multi-user tracking. *IEEE Transactions on Mobile Computing*, 16(4):1019–1031, 2017.

[8] Han Ding, Jinsong Han, Alex X Liu, Wei Xi, Jizhong Zhao, Panlong Yang, and Zhiping Jiang. Counting human objects using backscattered radio frequency signals. *IEEE Transactions on Mobile Computing*, 18(5):1054–1067, 2019.

[9] Min Li, Zhaoxiang Zhang, Kaiqi Huang, and Tieniu Tan. Estimating the number of people in crowded scenes by mid based foreground segmentation and head-shoulder detection. In *2008 19th International Conference on Pattern Recognition*, pages 1–4, 2008.

[10] Yen Kai Cheng and Ronald Y. Chang. Device-free indoor people counting using wi-fi channel state information for internet of things. In *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, pages 1–6, 2017.

[11] Simone Di Domenico, Mauro De Sanctis, Ernestina Cianca, and Giuseppe Bianchi. A trained-once crowd counting method using differential wifi channel state information. In *Proceedings of the 3rd International on Workshop on Physical Analytics*, WPA '16, page 37–42, New York, NY, USA, 2016. Association for Computing Machinery.

[12] Simone Di Domenico, Giovanni Pecoraro, Ernestina Cianca, and Mauro De Sanctis. Trained-once device-free crowd counting and occupancy estimation using wifi: A doppler spectrum based approach. In *IEEE International Conference on Wireless & Mobile Computing*, 2016.

[13] Wei Xi, Jizhong Zhao, Xiang Yang Li, Kun Zhao, Shaojie Tang, Xue Liu, and Zhiping Jiang. Electronic frog eye: Counting crowd using wifi. In *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, pages 361–369, 2014.

[14] Chenren Xu, Bernhard Firner, Robert S. Moore, Yanyong Zhang, Wade Trappe, Richard Howard, Feixiong Zhang, and Ning An. Scpl: Indoor device-free multi-subject counting and localization using radio signal strength. In *2013 ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, pages 79–90, 2013.

[15] Kun Qian, Chenshu Wu, Yi Zhang, Guidong Zhang, and Yunhao Liu. Widar2.0: Passive human tracking with a single wi-fi link. In *the 16th Annual International Conference*, 2018.

[16] Dan Wu, Daqing Zhang, Chenren Xu, Yasha Wang, and Hao Wang. Widir: walking direction estimation using wireless signals. In *Acm International Joint Conference*, 2016.

[17] Li, W., Duan, L., Xu, D., Tsang, and W. I. Learning with augmented features for supervised and semi-supervised heterogeneous domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(6):1134–1148, 2014.

[18] Kun Qian, Chenshu Wu, Zimu Zhou, Yue Zheng, Zheng Yang, and Yunhao Liu. Inferring motion direction using commodity wi-fi for interactive exergames. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, page 1961–1972, New York, NY, USA, 2017. Association for Computing Machinery.

[19] Xiang Li, Daqing Zhang, Qin Lv, Jie Xiong, and Hong Mei. Indotrack: Device-free indoor human tracking with commodity wi-fi. *Proceedings of the ACM on Interactive Mobile Wearable and Ubiquitous Technologies*, 1(3):1–22, 2017.

[20] Yue Zheng, Yi Zhang, Kun Qian, Guidong Zhang, Yunhao Liu, Chenshu Wu, and Zheng Yang. Zero-effort cross-domain gesture recognition with wi-fi. In *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '19, page 313–325, New York, NY, USA, 2019. Association for Computing Machinery.

[21] Han Zou, Baoqi Huang, Xiaoxuan Lu, Hao Jiang, and Lihua Xie. A robust indoor positioning system based on the procrustes analysis and weighted extreme learning machine. *IEEE Transactions on Wireless Communications*, 15(2):1252–1266, 2016.

[22] Lei Zhang, Zhirui Wang, and Liu Yang. Commercial wi-fi based fall detection with environment influence mitigation. In *2019 16th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, pages 1–9, 2019.

[23] Nan Yu, Wei Wang, Alex X. Liu, and Lingtao Kong. Qgesture: Quantifying gesture distance and direction with wifi signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2:1–23, 03 2018.

[24] Youwei Zeng, Dan Wu, Jie Xiong, Enze Yi, Ruiyang Gao, and Daqing Zhang. Farsense: Pushing the range limit of wifi-based respiration sensing with csi ratio of two antennas. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 3(3), September 2019.

[25] Manikanta Kotaru, Kiran Joshi, Dinesh Bharadia, and Sachin Katti. Spotfi: Decimeter level localization using wifi. *Acm Sigcomm Computer Communication Review*, 45(4):269–282, 2015.

[26] 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*, MobiCom '15, page 90–102, New York, NY, USA, 2015. Association for Computing Machinery.

[27] Sheng Tan and Jie Yang. Wifinger: Leveraging commodity wifi for fine-grained finger gesture recognition. MobiHoc '16, page 201–210, New York, NY, USA, 2016. Association for Computing Machinery.

[28] Wei Wang, Alex Liu, and Muhammad Shahzad. Gait recognition using wifi signals. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '16, page 363–373, New York, NY, USA, 2016. Association for Computing Machinery.

[29] Wei Wang, Alex X. Liu, Muhammad Shahzad, Kang Ling, and Sanglu Lu. Understanding and modeling of wifi signal based human activity recognition. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, MobiCom '15, page 65–76, New York, NY, USA, 2015. Association for Computing Machinery.

[30] Fusang Zhang, Daqing Zhang, Jie Xiong, Hao Wang, Kai Niu, Beihong Jin, and Yuxiang Wang. From fresnel diffraction model to fine-grained human respiration sensing with commodity wi-fi devices. 2(1), March 2018.

[31] Andre K. T. Assis. Michael eckert: Arnold sommerfeld: Science, life and turbulent times 1868- 1951, translated by tom artin. *Science & Education*, 23(3):707–710, 2014.

[32] City of Cumberland Report. How signal is affected.

[33] Fadel Adib and Dina Katabi. See through walls with wifi! *SIGCOMM Comput. Commun. Rev.*, 43(4):75–86, August 2013.

[34] Kun Qian, Chenshu Wu, Zheng Yang, Yunhao Liu, and Zimu Zhou. Pads: Passive detection of moving targets with dynamic speed using phy layer information. In *2014 20th IEEE International Conference on Parallel and Distributed Systems (ICPADS)*, pages 1–8, 2014.

[35] Youwei Zeng, Dan Wu, Ruiyang Gao, Tao Gu, and Daqing Zhang. Full-breathe: Full human respiration detection exploiting complementarity of csi phase and amplitude of wifi signals. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(3), September 2018.

[36] Chih-Chung Chang and Chih-Jen Lin. Libsvm: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, 2(3):1–27, 2011.

[37] Daniel Halperin, Wenjun Hu, Anmol Sheth, and David Wetherall. Predictable 802.11 packet delivery from wireless channel measurements. *SIGCOMM Comput. Commun. Rev.*, 40(4):159–170, August 2010.

[38] Kai Niu, Fusang Zhang, Jie Xiong, Xiang Li, and Daqing Zhang. Boosting fine-grained activity sensing by embracing wireless multipath effects. In *Proceedings of the 14th International Conference on Emerging Networking EXperiments and Technologies*, CoNEXT '18, page 139–151, New York, NY, USA, 2018. Association for Computing Machinery.

[39] S. Fritz and M. Lusardi. Walking speed: The sixth vital sign. *Journal of Geriatric Physical Therapy*, 32(2):46–49, 2009.

[40] Bohannon and W. Richard. Comfortable and maximum walking speed of adults aged 20-79 years: Reference values and determinants. *Age and ageing*, 26:15–9, 01 1997.

[41] Hai Zhu, Fu Xiao, Lijuan Sun, Ruchuan Wang, and Panlong Yang. R-ttwd: Robust device-free through-the-wall detection of moving human with wifi. *IEEE Journal on Selected Areas in Communications*, 35(5):1090–1103, 2017.

[42] Wenjun Jiang, Chenglin Miao, Fenglong Ma, Shuochao Yao, Yaqing Wang, Ye Yuan, Hongfei Xue, Chen Song, Xin Ma, Dimitrios Koutsonikolas, Wenyao Xu, and Lu Su. Towards environment independent device free human activity recognition. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, MobiCom '18, page 289–304, New York, NY, USA, 2018. Association for Computing Machinery.

[43] Hao Wang, Daqing Zhang, Yasha Wang, Junyi Ma, Yuxiang Wang, and Shengjie Li. Rt-fall: A real-time and contactless fall detection system with commodity wifi devices. *IEEE Transactions on Mobile Computing*, 16(2):511–526, 2017.

Lei Zhang received the Ph.D. degree in Computer Science from Auburn University (Auburn, AL, USA) in 2008. She worked as an assistant professor from 2008-2011 in the Computer Science Dept. at Frostburg State University (Frostburg, MD, USA); She is now an associate professor in College of Intelligence and Computing at Tianjin University (Tianjin, P.R. China). Her research interests include mobile computing, computer networks, data mining. She is a member of the ACM and IEEE.

Yueqiang Zhang received the B.E. degree from Central South University, Changsha, China, in 2018. He is currently pursuing master degree at Tianjin University, Tianjin, China. His research interests include wireless communication and data mining.

Xiaolong Zheng received the Ph.D. degree from the Hong Kong University of Science and Technology, China, in 2015. He is currently a research associate professor with the School of Computer Science and Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia, Beijing University of Posts and Telecommunications, China. His research interests include Internet of Things, wireless networks, and ubiquitous computing.

Beibei Wang received the Ph.D. degree in electrical engineering from the University of Maryland, College Park in 2009. Since 2015, she has been with Origin Wireless Inc., where she is currently the Vice President of Research and Director of Intellectual Properties. She is also affiliated with the University of Maryland, College Park. Her research interests include Internet of Things, mobile computing, wireless sensing and positioning, and communications and networking. She is a senior member of IEEE.

Liu Yang received the Ph.D. degree in computer science from the School of Computer and Information Technology, Beijing Jiaotong University, China, in 2016. She is currently an Associate Professor with the College of Intelligence and Computing, Tianjin University. Her research interests include data mining and machine learning.