Understanding WiFi signal frequency features for position-independent gesture sensing

Kai Niu, Fusang Zhang, Xuanzhi Wang, Qin Lv, Haitong Luo, Daqing Zhang

Understanding WiFi Signal Frequency Features for Position-Independent Gesture Sensing

Kai Niu, Fusang Zhang, Xuanzhi Wang, Qin Lv, Haitong Luo, and Daqing Zhang, *Fellow, IEEE*

Abstract—Recent years have witnessed rapid development in the research area of WiFi sensing, which senses human activities in a contactless and non-intrusive manner. One major issue that hinders real-world deployment of these systems is position dependence, i.e., once the human target changes location and orientation, the sensing performance degrades significantly. Existing machine learning based methods aim to solve this problem by either generating high-dimensional features or transfer learning the environment knowledge. However, these methods require significant training effort and yet acquire limited improvement. In this paper, we start by understanding and analyzing the Doppler frequency shift in WiFi sensing. We then develop a WiFi frequency model to quantify the relationship between signal frequency and target position, motion direction and speed for human activities. Based on this theoretical model, we prove that the commonly-used movement speed and motion direction features are position dependent, and further identify movement fragments and relative motion direction changes as two position-independent features. Building upon the frequency model and the position-independent features, we design a suite of position-independent gestures and develop the gesture recognition system accordingly. Evaluation results show that under various conditions (i.e., different locations, orientations, environments, and persons), our system achieves more than 96% recognition accuracy without any training, significantly outperforming state-of-the-art machine learning based solutions.

Index Terms—Time frequency feature, Position-independent, Gesture recognition, Contactless sensing, WiFi sensing.

1 Introduction

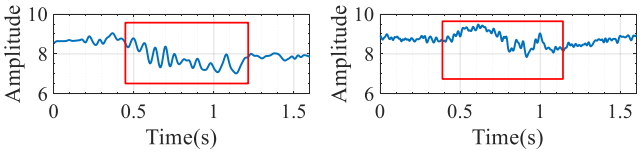
The proliferation of wireless devices in recent years has brought about significant progress in WiFi based contactless sensing, opening a new direction for ubiquitous and non-intrusive sensing of human activities without attaching any device to a target. Various sensing applications have been investigated, including fall detection [1, 2], gesture recognition [3, 4, 5, 6, 7, 8], keystroke detection [9], and vital sign monitoring [10, 11, 12, 13, 14]. Based on the observation that human activities incur changes in the Channel State Information (CSI) of received WiFi signal, existing systems typically extract certain CSI features (e.g., amplitude and phase in the time domain) and employ machine learning methods for training classifiers to recognize human activities. The key assumption here is that there exists a fixed mapping between human activity and CSI signal pattern, i.e., the corresponding CSI signal pattern is consistent for the same activity, yet different for different activities. However, as pointed out by prior works using the Fresnel zone penetration model of wireless signals [15, 16, 17, 18, 19], when the same activity is conducted at different positions (i.e., different location and/or orientation), the CSI signal pattern can vary significantly, resulting in unstable recognition performance. Given the diversity of real-world environments and application scenarios, it is important to develop WiFi based sensing mechanisms that are position-independent [1].

A few different approaches have been explored to address the position dependence problem of WiFi based sensing. A set of solutions [20, 21] resort to advanced machine learning methods (e.g., transfer learning). These methods require significant training and prior knowledge of the different positions. While these methods can handle slight changes in signal patterns, there is no guidance on what knowledge can be transferred and how to ensure the performance when the position of a target changes. More importantly, for large position changes, the same activity can induce very different signal patterns. For example, when a target performs the hand gesture "push" at different positions, the amplitudes of the received signal (shown in Fig. 1) have very different patterns. Without knowing the target's position, machine learning based methods that aim to transfer signal patterns for specific activities do not work well for large position changes. Some recent work aims to tackle this issue by employing frequency-domain CSI features instead of time-domain ones such as amplitude and phase [22, 23, 24]. Since the frequency shift of the signals

- *K. Niu, and X. Wang are with the Key Lab of High Confidence Software Technologies (Peking University), Ministry of Education, Beijing 100871, China, and also with the Department of Computer Science and Technology, School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China (E-mail: xjtunk@pku.edu.cn, xuanzhiwang@stu.pku.edu.cn).*
- *F. Zhang is with the State Key Laboratory of Computer Sciences, Institute of Software, Chinese Academy of Science, Beijing 100190, China (E-mail: zhangfusang@otcaix.iscas.ac.cn).*
- *Q. Lv is with the Department of Computer Science, University of Colorado Boulder, Boulder, CO 80309 USA (E-mail: qin.lv@colorado.edu).*
- *H. Luo is with School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing 102206, China (E-mail: ivanccc@bupt.edu.cn).*
- *D. Zhang is with the Key Lab of High Confidence Software Technologies (Peking University), Ministry of Education, Beijing 100871, China, also with the Department of Computer Science and Technology, School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China, and also with the Telecom SudParis, Institut Polytechnique de Paris, 91011 Evry Cedex, France (E-mail: dqzhang@sei.pku.edu.cn, corresponding author).*

1. In this work, we use position to indicate both the location and orientation of a target.

induced by the static environment (e.g., the line of sight signals and reflections from ambient objects) remains at zero, it is straightforward to relate the non-zero frequency shift in received signals to human motion. This environment-independent property hints that the non-zero frequency features are relatively more robust against environmental changes for human activity recognition. As such, frequency-domain features are explored to address the challenge of position variations in WiFi sensing. However, theoretical studies are still lacking that explore the relationship between human activity and frequency-domain features at different positions.



(a) Position 1 (b) Position 2

Fig. 1: Amplitude for gesture "push" at different positions.

In this work, using WiFi-based gesture recognition as an example, we aim to answer two main research questions:

(1) Can we model the impact of position (location and orientation) changes on frequency-domain CSI features to determine whether they are position-independent?

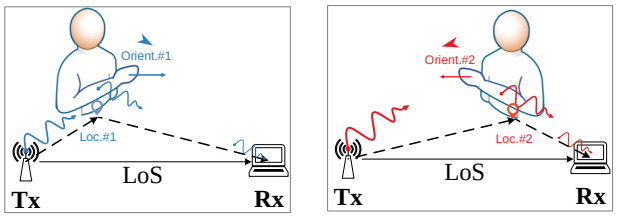
(2) Building upon the model, can we construct frequency-domain CSI features that can support position-independent recognition of gestures?

We start with a validation experiment in which a user performs the "push" gesture at two locations with different orientations, as shown in Fig. 2. The corresponding time-frequency spectrograms are shown in Fig. 3. We observe very different signal frequencies for the same gesture performed at the two positions: about 20Hz at position 1 and about -15Hz at position 2. To understand the reasons behind these differences, we derive a mathematical model quantifying the relationship between signal frequency and sensing target's location, motion direction, and speed of movement, which allows us to explain frequency-domain CSI feature changes at different locations and orientations.

Based on the theoretical modeling of how location and orientation impact CSI in the frequency domain, we prove that the commonly-used features such as movement speed and motion direction in existing gesture recognition systems are position dependent. This is because these features are affected not only by the distinct gestures but also by environmental factors such as the location and orientation of the target with respect to the WiFi transceivers. Guided by the model, we seek to identify and extract position-independent features by leveraging the time-frequency spectrogram from multiple devices. With the position-independent features, we propose a suite of position-independent gestures and build a prototype gesture recognition system that achieves high and robust performance without training. Please see our demo video at: https://youtu.be/o6IbReBig_g.

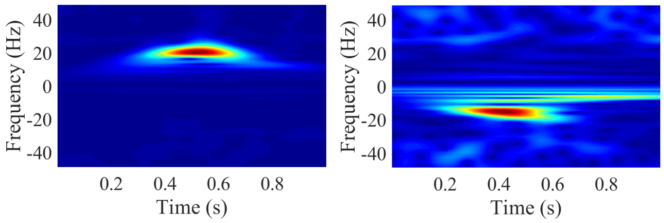
The main contributions of our work are summarized as follows.

- We have developed a mathematical model to quantify the relationship between signal frequency and target position, motion direction, and speed for human activities. It reveals how signal frequency changes with these impact factors.
- Building upon the model, we provide specific guidelines to extract movement fragments and relative direction changes from time-frequency information of multiple WiFi devices as two position-independent features. This enables us to develop a suite of position-independent gestures and the corresponding gesture recognition system, which requires no training and achieves position-independent gesture recognition with high accuracy.
- We have implemented the sensing system on COTS WiFi devices and conducted extensive field experiments (8 gestures, 5 locations and 5 orientations in 3 environments, and comparisons with three state-of-the-art machine learning based approaches). The results demonstrate that our system can recognize the gestures at different locations and orientations with more than 96% accuracy in the 2m×2m sensing area.



(a) Scenario 1 (b) Scenario 2

Fig. 2: A target performs the same "push" activity at two different locations and orientations relative to the WiFi devices.



(a) Scenario 1 (b) Scenario 2

Fig. 3: Time-frequency spectrograms for the two different positions.

The rest of the paper is organized as follows. We summarize related state-of-the-art research in Sec. 2. In Sec. 3, we first model the impact of location and orientation on CSI in the frequency domain, and explain how frequency changes with different locations and orientations. In Sec. 4, we provide guidelines to build position-independent gestures and illustrate how to extract position-independent features from multiple WiFi devices. We present the detailed implementation of the sensing system in Sec. 5. In Sec. 6, we report the evaluation results of our system in diverse settings. Finally, we conclude our work in Sec. 8.

2 Related Work

2.1 Early Practices in WiFi Gesture Sensing

Most early WiFi-based sensing systems employ CSI amplitude information to identify human activities. As an early work in WiFi gesture sensing, WiG [7] extracts four features from CSI amplitude to train an SVM classifier to distinguish four common gestures (i.e., right, left, push, and pull) performed between the transmitter and the receiver. Based on three different change primitives in RSSI/CSI amplitude caused by gesture, WiGest [25] constructs and recognizes mutually-independent gesture families around the receiver through pattern matching. WiFinger [26] extracts patterns in CSI amplitude signals via principal component identification and compares waveform shapes using Dynamic Time Warping (DTW) and kNN (k-Nearest Neighbors) to identify nine different finger gestures in the middle of LoS (line of sight). More recently, researchers have employed temporal changes of activity velocity in the frequency domain to characterize target activities. This is because frequency features are relatively resistant to environment changes while recognizing human activities. WiMU [27] utilizes time-frequency information to classify people's gestures by generating virtual samples. However, this research assumes that time-frequency information is position-independent, which we demonstrate in this work to be false, i.e., performing the same activity at different locations and orientations can result in inconsistent time-frequency spectrum.

2.2 Machine Learning based Methods for Position-Independent WiFi Sensing

Most existing approaches resort to machine learning methods to achieve position-independent sensing. Jiang et al. [20] proposes EI, a deep-learning based contactless activity recognition framework based on Generative Adversarial Network (GAN). However, the adversarial network requires labelled signal data, and the inconsistency of signals makes it difficult for the network to obtain a uniform representation. For activity recognition across different environments, this method's accuracy is less than 80%. CrossSense [21] employs transfer learning with an existing roaming model that generates training samples from one set of measurements for each target environment. It then adopts a mixture-of-experts approach where multiple specialized sensing models are used to capture the mapping from diverse WiFi input to the desired output. However, without knowing the target's exact setting, the transfer learning method has little knowledge to tackle totally different positions that induce very different signal patterns. WiAG [8] and Widar 3.0 [28] propose orientation-independent gesture recognition solutions. WiAG uses a gesture translation function to generate virtual samples of different orientations to increase the training set. However, this method need to be retrained when the location and orientation change, and activities need to be defined ahead of time to estimate the configuration parameters such as location and orientation. Widar 3.0 uses multiple WiFi device pairs to build cross-domain body-coordinate velocity profile from Doppler information for gesture recognition. This method requires a large number of WiFi devices and significant training effort to cover diverse location and orientation settings. Similarly, the location and orientation also need to be given firstly. And the performance can not be guaranteed for incorrect initial position estimations.

Table 1 provides a summary of existing WiFi-based gesture sensing systems. We observe that prior works either ignore the position-dependent nature of WiFi signals or propose partially working solutions to the position-independent human activity sensing problems. We are the first to specifically model the impact of location and orientation on the relationship between frequency domain information and target movements. Furthermore, building upon the in-depth understanding of how frequency changes with location and orientation, we provide specific guidelines and propose novel position-independent features for effective and training-free recognition of gestures.

1. Abbreviations: SVM (Support Vector Machine), PM (Pattern Matching), kNN (k-Nearest Neighbors), SS (Similarity Score), CNN (Convolutional Neural Network), LSTM (Long Short-Term Memory), RSSI (Received Signal Strength Indicator)

3 Modeling the Impact of Target Position on CSI Signal Frequency

In this section, we first review the Doppler effect of Radio Frequency (RF) signal. Building on top of that, we establish a mathematical model to quantify the impact of target position on the relationship between CSI signal frequency and human movement. We then conduct real-world experiments using commodity WiFi devices to validate the proposed

TABLE 1: Summary of Existing WiFi-based Gesture Recognition Systems

	System	Signal & Method	Pos. Independent?	Solution	Training-free	Prerequisite knowledge
Early practices	WiG [7]	Amplitude & SVM	No	×	No	Fixed position
	WiGest [25]	RSSI/Amplitude&PM	No	×	Yes	Certain positions
	WiFinger [26]	Amplitude & kNN	No	×	No	Fixed position
	WiMU [27]	Frequency & SS	No	×	Yes	Assume frequency feature is position-independent
Machine learning based methods	CrossSense [21]	Amplitude	Yes	Transfer learning	No	Initial position
	WiAG [8]	Amplitude	Yes	Translation function	No	Initial position
	Widar3.0 [28]	Velocity profile	Yes	CNN+LSTM	No	Initial position

model by comparing model simulation results with the actual experimental results.

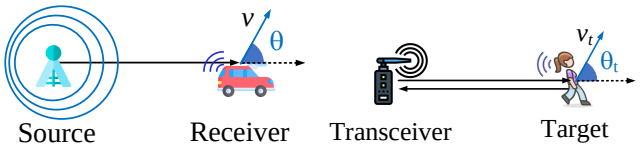
3.1 Doppler Frequency Shift of RF Signal

The Doppler effect is a well-known phenomenon where the signal frequency changes as the wave source or observer moves [29]. This is a foundational concept for human activity sensing. The Doppler effect can be used to quantify the change in the observed signal frequency of a wave (e.g., sound and radio) due to the relative motion of the transmitter and receiver. Different from the sound wave that satisfies the classical Doppler effect, RF signal travels at the speed of light and thus follows the relativistic Doppler effect. As shown in Fig. 4a, when the car moves away from the RF source, the signal received by the RF receiver in the car has lower frequency than the original signal. For motion in an arbitrary direction θ with respect to LoS between the source and the receiver in the source coordinate, let f_s be the frequency of the source signal and f_r be the frequency of the received signal, according to [30],

$$f_r = f_s \frac{c - v cos\theta}{\sqrt{c^2 - v^2}}, \tag{1}$$

where c is the speed of light and v is the relative movement speed between the source and the receiver. We know that the Doppler Frequency Shift (DFS) is the frequency difference between the source signal and the received signal. Based on the assumption that the movement speed is much smaller than the speed of light ($v \ll c$), DFS can be denoted as

$$f_D = f_r - f_s \simeq -\frac{v cos\theta}{c} f_s \tag{2}$$



(a) Doppler effect for RF signal (b) Sensing system with co-located source and receiver

Fig. 4: Doppler effect and its application in sensing systems.

For sensing systems with co-located source and receiver (e.g., radar), the transceiver is stationary while the sensing target moves. When the target moves at speed v_t in motion direction θ_t with respect to the LoS of transceiver and target (Fig. 4b), the DFS in such co-located transceiver sensing system is equivalent to that of the receiver moving at double speed in motion direction θ_t. Thus the DFS induced by moving target in co-located transceiver sensing systems is

$$f_{Dr} \simeq -\frac{2 v_t cos\theta_t}{c} f_s \tag{3}$$

From Equation 3, we observe that the DFS in co-located transceiver sensing system depends on the frequency of RF signal, movement speed and motion direction. In addition, there is a stable $2\times$ relation between the DFS in co-located transceiver sensing systems and the DFS corresponding to real movement speed. However, in WiFi based sensing systems, the source and the receiver are deployed separately, making the DFS more complicated than that in co-located transceiver sensing systems. Next, we will model the DFS in WiFi based sensing systems.

3.2 WiFi Signal Frequency Modeling

Due to the Doppler effect, a moving target will change the frequency of the received WiFi signals. To sense human activity, we extract the frequency feature from received CSI, which characterizes multipath effects in indoor environments for subcarrier with frequency f at arrival time t as

$$H(f,t) = (\sum_{i=1}^{I} a_i(f,t) e^{-j2\pi \frac{d_i(t)}{\lambda}}) e^{-j\theta_o}, \tag{4}$$

where I is the number of paths, a_i and d_i represent the signal attenuation and propagation path length of the i-th path, respectively, λ is the wavelength for subcarrier with frequency f ($\lambda = c/f$), and θ_o is the random phase error caused by central frequency offset, sampling frequency offset and packet boundary detection uncertainty.

In WiFi based contactless sensing, the transceivers are stationary, while the moving target affects the signal frequency at the WiFi receiver. Without loss of generality, we assume there is only one reflection path that corresponds to the target's movement [28]. Thus CSI can be transformed as follows:

$$H(f,t) = (H_s(f) + a(t) e^{-j2\pi \frac{d_0 + v_p t}{\lambda}}) e^{-j\theta_o}, \tag{5}$$

where the constant H_s is the sum of all static signals with zero Doppler frequency shift (e.g., LoS signal), d_0 is the initial path length of dynamic signals reflected by the target (the black dotted line in Fig. 5), and v_p is the speed of reflection path length change that leads to non-zero Doppler frequency shift. Please note that the speed of reflection path length change v_p is not the real velocity of the human target.

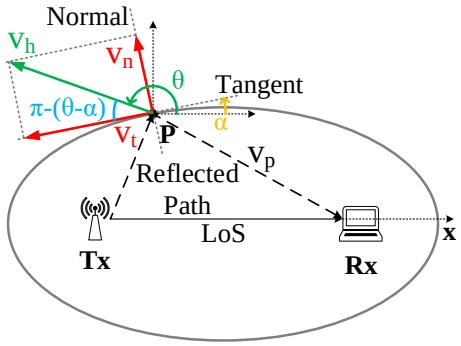


Fig. 5: Projecting target's velocity to both normal and tangent directions of an ellipse.

To establish the relationship between the signal frequency caused by Doppler effect and target movement velocity, we construct an ellipse with foci transmitter Tx and receiver Rx that passes through the target's location P, as illustrated in Fig 5. The target's velocity with magnitude v_h and motion direction θ can be projected to the speed in the normal direction of the ellipse v_n and the one in the tangent direction v_t. Only the normal direction speed v_n can change the path length, while the tangent direction speed v_t does not change the path length. Let α be the angle of

the tangent line for the ellipse curve at position P, whose value is between $-\pi/2$ and $\pi/2$ and is determined by the relative position of the target and the transceivers. Then the angle between the movement velocity and the tangent line is $\pi - (\theta - \alpha)$ (blue arc in Fig. 5). The relationship between the normal direction speed and the movement speed is $v_n = v_h sin(\pi - (\theta - \alpha)) = v_h sin(\theta - \alpha)$. Thus CSI can be denoted as

$$\begin{aligned} H(f,t) &= (H_s(f) + a(t)e^{-j2\pi\frac{d_0 + rv_n t}{\lambda}})e^{-j\theta_o} \\ &= (H_s(f) + a(t)e^{-j2\pi\frac{d_0 + rv_h sin(\theta-\alpha)t}{\lambda}})e^{-j\theta_o}, \end{aligned} \tag{6}$$

where $r = cos\beta_1 + cos\beta_2$ is the ratio coefficient between the speed of reflection path length change v_p and the speed in the normal direction v_n. As the path length change speed is the sum of the speeds in lines TxP and RxP, we project the normal speed v_n to both of them. As shown in Fig. 6, β_1 and β_2 are the angles between the normal direction and the lines TxP, RxP, respectively. Thus, the projections of normal speed v_n in paths TxP and RxP are $v_n cos\beta_1$ and $v_n cos\beta_2$, respectively. Then we can get $v_p = rv_n = rv_h sin(\theta - \alpha)$ and $r = cos\beta_1 + cos\beta_2$. r increases when the target moves away from LoS due to the decrease of β_1 and β_2.

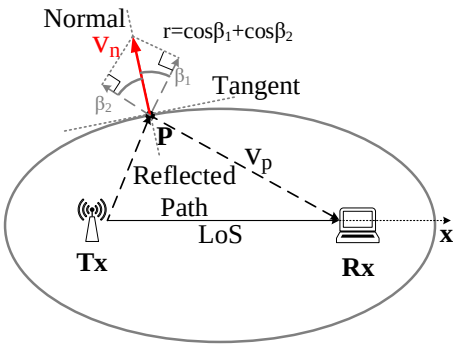


Fig. 6: Ratio coefficient r between the speed of reflection path length change and the normal speed.

We further rewrite Equation 6 as

$$H(f,t) = (H_s(f) + a(t)e^{j(2\pi\frac{-rv_h sin(\theta-\alpha)}{\lambda}t - \delta)})e^{-j\theta_o}, \tag{7}$$

where $\delta = 2\pi d_0/\lambda$ is the initial phase caused by the initial path length d_0 of dynamic signals at position P. From Equation 7, we obtain the relationship between signal frequency and target's location, motion direction, and speed as follows:

$$f_d = \frac{-rv_h sin(\theta - \alpha)}{\lambda} = \frac{-rv_h sin(\theta - \alpha)}{c} f \tag{8}$$

Note that the sign of f_d depicts whether the target moves toward the transceivers (positive) or moves away from the transceivers (negative). Unless otherwise specified, the value of f_d in the following text does not contain the sign.

From Equation 8, we can observe that signal frequency is directly related to four parameters that correspond to two factors in the physical world: (i) target location, which determines r and α in the equation; and (ii) target motion velocity, including both movement speed v_h and motion direction θ. We discuss the two factors in detail below.

- **Target location.** (1) *We first analyze how target location affects parameter r.* Consider two different target locations P_1 and P_2, the corresponding ellipse curves are shown in Fig. 7 and the angle of tangent lines are the same (i.e., $\alpha_1 = \alpha_2$). Location P_1 corresponds to a larger r value than location P_2, leading to higher frequency in the received signal. When the location is far enough from LoS, the value of r approaches 2. (2) *Next, we analyze how target location also affects parameter α.* For example, P_1 and P_3 on the same ellipse curve correspond to different angles of the tangent lines: α_1 has a small positive value (in counter-clockwise direction from the x axis), while α_3 has a large negative value (in clockwise direction from the x axis). Suppose the target moves away perpendicularly from LoS ($\theta = 90°$), location P_1 has a smaller α value thus larger $sin(\theta - \alpha)$ value than that of location P_3, inducing higher frequency. From the analysis, we know that target location has a strong influence on parameters r and α in Equation 8, which in turn impacts signal frequency.

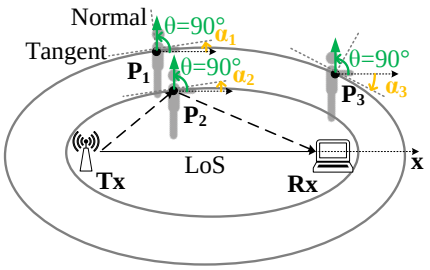


Fig. 7: Impact of target location.

- **Target movement velocity.** Target movement velocity consists of movement speed v_h and motion direction θ. When the parameters r and α are fixed, the larger the speed v_h, the higher the signal frequency. This property can be leveraged to recognize activities with large speed differences at certain positions. However, the range of speed is usually limited for most activities (e.g., hand gestures). In contrast, motion direction can change sharply when target orientation changes. As illustrated in Fig. 8, the target performs the same "push" gesture at the perpendicular bisector of LoS ($\alpha = 0°$). For orientation 1, the motion direction $\theta_1 = 0°$, corresponding to $sin(\theta_1 - \alpha) = 0$ and the frequency is also 0. In comparison, for orientation 2, the motion direction $\theta_2 = 90°$, corresponding to $sin(\theta_2 - \alpha) = 1$ and the frequency is at its maximum. In other words, when orientation changes, motion direction also changes and can significantly affect signal frequency. And the influence of motion direction is much bigger than that of movement speed.

Comparing Equation 3 and Equation 8, we identify two main differences between co-located transceiver sensing systems and WiFi based contactless sensing systems: (i) The ratio coefficient r is fixed at $2\times$ for co-located transceiver systems, but r changes by position in WiFi based sensing systems; and (ii) The DFS is independent of target position in co-located transceiver systems, but is position dependent in WiFi based sensing systems. Therefore, the DFS in WiFi based sensing systems is more complicated than that in

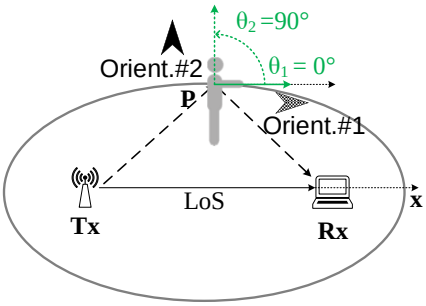


Fig. 8: Impact of target orientation.

co-located transceiver sensing systems. As such, position-independent sensing is a challenging problem for WiFi based sensing systems.

3.3 WiFi Signal Frequency Model Validation

Based on the modeling process above, we know that target location and motion direction (thus target orientation) are two key factors that can affect signal frequency. Next, using model simulations and benchmark experiments, we further study the impact of these two factors and validate the proposed frequency model.

3.3.1 *Experiment 1: Verifying the effect of target location*

Experimental setup: We first verify the effect of target location with real-world experiments. The experimental setup is shown in Fig. 9. The two commodity WiFi transceivers (MiniPCs equipped with Intel 5300 NIC adapter) are placed 2m apart. To precisely control target movement speed, we employ a sliding track to move a metal cylinder to simulate moving human target. The target and the transceivers' antennas are placed at the same height of 25cm. We set 20 uniformly spaced locations (green labels in Fig. 9) in both horizontal and vertical directions as the starting locations of target movement. The distance between adjacent locations is 50cm. For all the locations, we set the speed of the sliding track as a constant of $11cm/s$ to move 1m distance. The sliding track is perpendicular to the transceiver pair to ensure that the target's motion direction is $90°$.

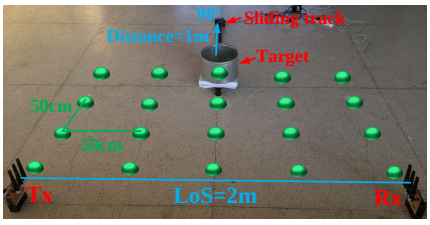
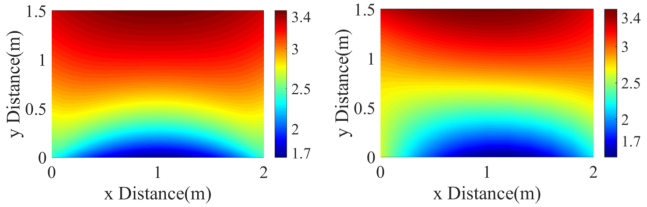


Fig. 9: Experimental setup to verify the effect of target location.

Result analysis: We first use our model to calculate the frequency value at all locations, spaced uniformly with a step size of 0.1m in both horizontal and vertical directions. The distance range is $0m$ to $2m$ for the horizontal direction, and $0m$ to $1.5m$ for the vertical direction. The simulation results are shown in Fig. 10a. For the real-world experiments, we first extract signal frequency for the movements. We adopt Continuous 1-D Wavelet Transform (CWT) to get the time-frequency spectrogram from the received signal, then calculate the mean value of the frequency with maximum power in each sample as the frequency of received signal. Then quadratic interpolation is adopted to adjust the space size from $0.5m$ to $0.1m$. The experimental results are shown in Fig. 10b. Comparing these two figures, we can observe that the simulation results and the experimental results match very well with each other. Furthermore, for the motion with direction $90°$, we note that the frequency value increases as the vertical distance to LoS increases, which explains how target location affects parameter r and further impacts received signal frequency. The change of r dominates frequency change in locations that are close to LoS. For locations that are far from LoS, the change of r is small, frequency is maximal at the middle locations and decreases on both sides, which reveals the effect of parameter α. In summary, we find that frequency is highly dependent on target location.



(a) Simulation results (b) Experiment results

Fig. 10: Verifying the effect of target location on frequency.

3.3.2 *Experiment 2: Verifying the effect of motion direction*

Experimental setup: The experimental setup for verifying the effect of motion direction (and target orientation) is shown in Fig. 11. Two locations (green labels in Fig. 11) with the same vertical distance to LoS are chosen to verify the effect of motion direction. In each location, the sliding track is positioned at different directions relative to LoS, from $0°$ to $180°$ with a step size of $15°$ (blue arrow line in Fig. 11). In total, 13 different directions are considered. For each direction, the target moves 1m from the starting point of the arrow line to the end point with the help of the sliding track. Other configurations are the same as that in Experiment 1.

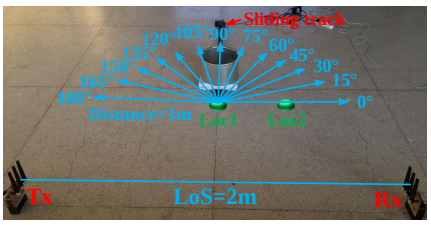
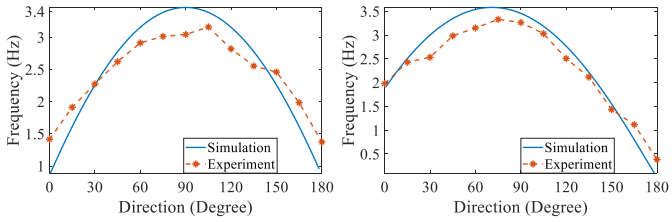


Fig. 11: Experimental setup to verify the effect of motion direction.

Result analysis: The results for the two locations are shown in Fig. 12a and Fig. 12b, respectively. In Fig. 12, the blue line shows the simulated frequency results using our model, and the red line shows the experimental results. For



(a) Location 1 (b) Location 2

Fig. 12: Verifying the effect of motion direction on frequency.

both locations, we can observe that the experimental results match the simulation results very well. Fig. 12a also shows that the maximum frequency for location 1 is in the direction of $90°$. And the frequency decreases gradually when the direction changes to $0°$ and $180°$. The changing process is symmetrical. In comparison, the direction corresponding to maximum frequency of location 2 is less than $90°$. The reason is that the α for location 2 is less than 0, so the maximum value involves a left shift. The frequency also decreases gradually with the direction changing for both sides and the changing process is asymmetrical. For the case $\alpha > 0$, the maximum value will shift to the right, i.e., the corresponding direction will be greater than $90°$ and the changing process is asymmetrical. The results show that frequency is dependent on motion direction (and thus target orientation).

4 Guiding the Position-Independent Gesture Recognition

In this section, guided by the WiFi signal frequency model, we demonstrate that movement speed and motion direction are position dependent, and propose two position-independent features: gesture fragments and relative motion direction changes. Then we show how to extract these position-independent features from the time-frequency spectrograms of multiple WiFi devices. Finally, we propose a suite of position-independent gestures and develop the corresponding system to achieve position independent gesture recognition.

4.1 Guidelines for Gesture Recognition

4.1.1 Commonly-used features for WiFi sensing

Movement speed and motion direction are two commonly-used features in previous WiFi sensing systems [4, 22]. Using the frequency model we have developed, we first analyze these two features and demonstrate that they are position dependent.

Movement speed. Based on the observation that large movement speed induces high frequency value, existing work employs the frequency value feature to represent movement speed, so as to distinguish activities with large speed differences [22]. The basic assumption is that the frequency value for higher speed is always larger than that of lower speed. However, our model reveals that this assumption is true only at certain positions and not for all positions. For instance, Fig. 13 shows a target performing the "Push" gesture with different speeds at two positions in the perpendicular bisector of LoS. Suppose the lower speed at position 1 with orientation $90°$ is v and the higher speed at position 2 with orientation $0°$ is $2v$. According to Equation 8, the frequency value for position 1 is $f_{p1} = rvsin90°f/c = rvf/c$, while the frequency value for position 2 is $f_{p2} = r2vsin0°f/c = 0$. Obviously, the frequency value for the lower speed gesture at position 1 is larger than that of the higher speed gesture at position 2, indicating that movement speed is a position dependent feature.

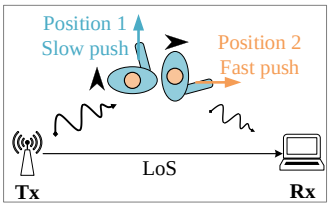


Fig. 13: "Push" gesture with different speeds at different positions.

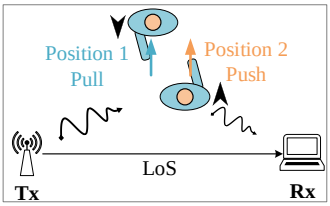


Fig. 14: "Pull" and "Push" gestures at different positions.

Motion direction. Existing work WiSee [4] employs the sequence of the sign of frequency to recognize nine gestures, such as "push" and "pull". This is because the sign of frequency represents the relative motion direction between the target and the transceivers, i.e., positive for moving towards the transceivers and negative for moving away from the transceivers. However, this sign only works for a certain range of motion directions. As shown in Fig. 14, the target performs "push" and "pull" at two positions with opposite orientations. We observe that the two gestures induce the same signal frequency, thus the same extracted motion directions (i.e., moving away from the transceivers), indicating that motion direction is also not a position-independent feature.

4.1.2 Position-independent features for gesture recognition

In this work, we have identified the following two features as position-independent features for gesture recognition.

Gesture fragments. A gesture typically consists of a number of basic strokes, which can be used to divide the gesture into a series of fragments. For example, the "Zigzag" gesture consists of three fragments: '¯', '/' and '_'. The number of fragments in a gesture is an intrinsic feature and is resistant to position changes. As shown in Fig. 15, the "Zigzag" gesture has three fragments, which is true for both positions and does not vary with target position. In other words, gesture fragments is a position-independent feature. Furthermore, our model shows that when the motion direction changes, the corresponding frequency also changes. This characteristic provides an opportunity for us to segment a gesture into multiple fragments and get the fragment number. Next section will discuss how to extract gesture fragment number from the time frequency spectrum with multiple WiFi devices.

Series of motion direction changes. As revealed by our model in Section 4.1.1, the absolute motion information depends on the target position. Thus a position-independent feature need to be independent (i.e., remains the same) of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2021.3063135, IEEE Transactions on Mobile Computing

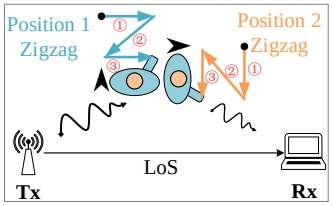


Fig. 15: Gesture fragment number for the"Zigzag" gesture at different positions.

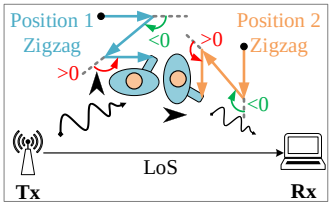


Fig. 16: Motion direction changes for the"Zigzag" gesture at different positions.

the absolute motion information. We propose the series of motion direction changes, which satisfies this condition and can be a position-independent feature. The motion direction change depicts the relative change between the motion directions of adjacent fragments of a gesture. As shown in Fig. 16, we also take the "Zigzag" gesture as an example. The motion direction change between the first fragment and second fragment is clockwise, defined as a negative angle change. As a comparison, the motion direction change between the second fragment and third fragment is anti-clockwise, defined as a positive angle change. Thus the series of motion direction changes for the "Zigzag" gesture can be expressed as '¯', negative angle change, '/', positive angle change, and '_'. Fig. 16 illustrates that this feature is also position-independent.

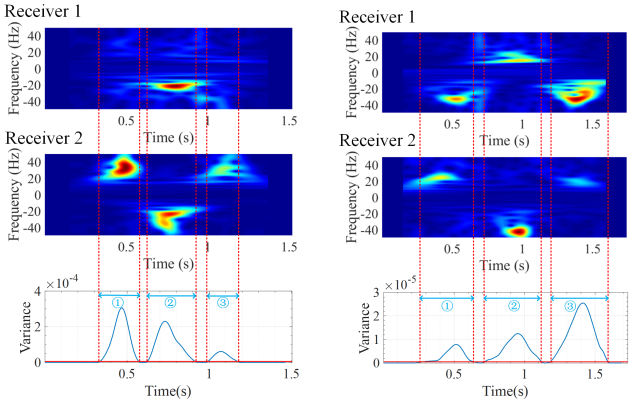
In summary, movement speed and motion direction are position-dependent features, while gesture fragment number and series of motion direction changes are position-independent features.

4.2 Extracting Position-Independent Features from Multiple WiFi Devices

From the analysis above, we know that the gesture fragments and the series of motion direction changes are position-independent features for gesture recognition. In this section, we present how to extract these position-independent features from time-frequency spectrogram of multiple WiFi devices (one transmitter and two receivers). Our method has two main steps: (1) segmenting gesture fragments and (2) extracting direction changes. The reason for employing multiple devices is two folds. First, when using a single pair of transceivers, the frequency may be zero for a specific movement, and that movement fragment may not be detected. Second, multiple devices are needed to construct the frequency profile and extract the series of direction changes, which can not be achieved using a single pair of transceivers.

4.2.1 Segmenting gesture fragments from time-frequency spectrogram of multiple devices

To obtain the number of motion fragments, we leverage the time-frequency spectrogram of multiple devices. We adopt the CWT algorithm [31] to extract time-frequency spectrogram from the received signal for both receivers. The spectrogram shows how the energy of each frequency component varies over time, where high-energy components are colored in red. In Fig. 17a, the top two figures show the time-frequency spectrogram from two receivers when performing the "Zigzag" gesture. For receiver 1, the frequencies for the first and third fragment movements are nearly zero, while the frequency for the second fragment movement is negative (moving away from the WiFi link). For receiver 2, the frequencies for the first and third fragment movements are positive (moving toward the WiFi link), while the frequency for the second fragment movement is negative. We employ the variance of the time-frequency spectrogram to segment the gesture fragments. First, we get the variance for each time-frequency spectrogram by using a 0.1s sliding window with 0.025s step size. Then we calculate the sum of variance from the two WiFi links (shown in the bottom figure of Fig. 17a). Finally, a dynamic threshold (red line in the bottom figure) is set to identify the start point and end point for each fragment. For "Zigzag", we can clearly observe three fragments from the time-frequency spectrogram of multiple devices, which can be segmented accurately using the sum variance of spectrogram. Fig. 17b shows the time-frequency spectrogram and variance for the "Zigzag" gesture in another position. While the time-frequency spectrograms are very different, we can accurately segment the gesture fragments and identify the correct number of movement fragments in a gesture.



(a) Position 1 (b) Position 2

Fig. 17: Fragment segmentation of gesture "Zigzag" for different target positions.

4.2.2 Extracting series of direction changes from time-frequency spectrogram of multiple devices

After segmenting the gesture fragments, we can get the time-frequency spectrogram for every fragment in the gesture. Then for each fragment of the gesture, we compute the mean of the frequency with maximum energy from the time-frequency spectrogram for both receivers (left subfigures of Fig. 18) as the frequency feature. As shown in the right subfigure of Fig. 18, f_{d1i} and f_{d2i} are the frequency features extracted from the two receivers for the i-th fragment, respectively. Thus the frequency feature for the i-th fragment is $F_i = (f_{d2i}, f_{d1i})$. The frequency feature of each fragment in the gesture can be further constructed as the frequency profile for the gesture $F = \{F_1, F_2, ..., F_i, ..., F_L\}$, where L is the number of movement fragments.

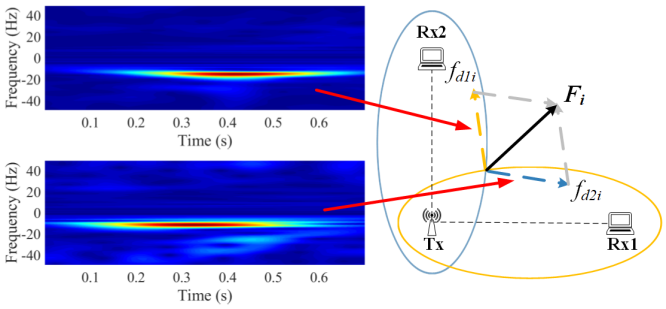
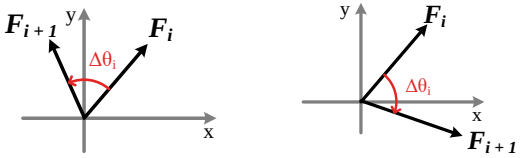


Fig. 18: Constructing frequency feature from two receivers.

Then we extract the series of direction changes [2] from the frequency profile of a gesture. Specifically, the angle of F_i is approximately equal to the motion direction of i-th fragment. Thus the direction change $\Delta\theta_i$ between the i-th fragment and $(i+1)$-th fragment can be calculated as $angle(F_{i+1}) - angle(F_i)$. There are two situations for the direction change between the adjacent fragments. If F_{i+1} rotates counter-clockwise relative to F_i, the direction change $\Delta\theta_i$ is positive (Fig. 19a), otherwise is negative (Fig. 19b). For every two adjacent fragments, we can get the direction change, then construct the direction change feature vector $\{\Delta\theta_1, \Delta\theta_2, ..., \Delta\theta_i, ..., \Delta\theta_N\}$, where $N = L - 1$.



(a) Positive direction (b) Negative direction

Fig. 19: Two different cases of motion direction change.

4.3 Achieving Position-Independent Gesture Recognition

Note that not all gestures can be recognized position independently. Table 2 summarizes the common gestures that are used in prior work [4, 7, 8, 25, 28]. According to the guidelines in Section 4.1, gestures 1-8 (Group 1) include only one fragment without motion direction change. For certain positions, movement speed and motion direction can help identify them, such as "Push" vs. "Pull", "Right" vs. "Left", and "Sweep" vs. "Slide". However, these gestures are not position-independent and cannot be distinguished for all possible positions. Similarly, gestures 9-14 (Group 2) can be divided into two fragments with one motion direction change, gestures 15-17 (Group 3) have three fragments with two motion direction changes, and gesture 18 (Group 4) corresponds to four fragments with three motion direction changes. Although gestures in the same group cannot be classified position independently, we can achieve position-independent recognition for gestures in different groups. Using this guiding principle, we extend these gestures into eight gestures, as shown in Figure 20, which form a suite of position-independent gestures and can be further extended to include gestures with more fragments and direction changes. The eight gestures' corresponding position-independent features are listed in Table 3. Please note that these features can be defined in advance and extracted accordingly. As such, no training is needed to recognize such gestures.

2. The gestures are performed in the plane constructed by Tx and two Rxs. Thus the direction changes are the same for both Rxs.

TABLE 2: Eighteen Common Gestures in Prior Work

ID	Gesture	ID	Gesture	ID	Gesture
1	Push	7	Flick	13	Up-down
2	Pull	8	Strike	14	Infinity
3	Right	9	Circle	15	Zigzag
4	Left	10	Drag	16	Down-Pause-Up
5	Sweep	11	Throw	17	Up-Pause-Down
6	Slide	12	Down-up	18	Punch×2

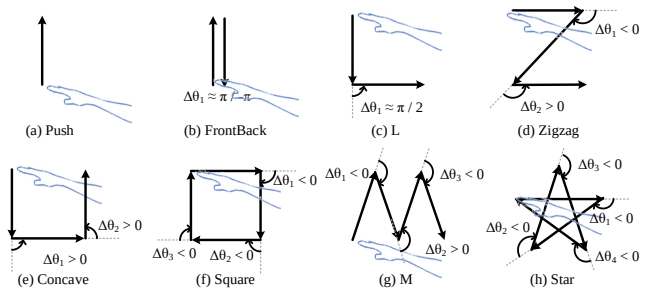


Fig. 20: A suite of position-independent gestures.

TABLE 3: Position-Independent Features for the Eight Gestures: #Fragments and Series of Direction Changes

Gesture	#Fragments	$\Delta\theta_1$	$\Delta\theta_2$	$\Delta\theta_3$	$\Delta\theta_4$
Push	1	-	-	-	-
Frontback (FB)	2	$\approx \pi / -\pi$	-	-	-
L	2	$\approx \pi/2$	-	-	-
Zigzag (Zig.)	3	< 0	> 0	-	-
Concave (Con.)	3	> 0	> 0	-	-
Square (Sq.)	4	< 0	< 0	< 0	-
M	4	< 0	> 0	< 0	-
Star	5	< 0	< 0	< 0	< 0

5 System Implementation

In this section, we design and implement a prototype gesture recognition system. Fig. 21 gives an overview of our system, which consists of four core modules: (1) data acquisition, (2) data preprocessing, (3) feature extraction, and (4) gesture recognition. In the data acquisition stage, raw CSI readings from two commodity WiFi receivers are collected as input. To remove phase offset and amplitude noise, we apply a series of algorithms to smooth out the random noise in both phase and amplitude in the data preprocessing stage. Then we extract position-independent features from the smoothed signal. Finally, the features are matched with predefined rules to recognize gestures.

(1) Data acquisition: Our system employs one transmitter (Tx) and two receivers (Rx). The Tx is a wireless AP (access point) such as a router or a laptop. The Rx can be any smart device with WiFi capabilities, such as smart TV

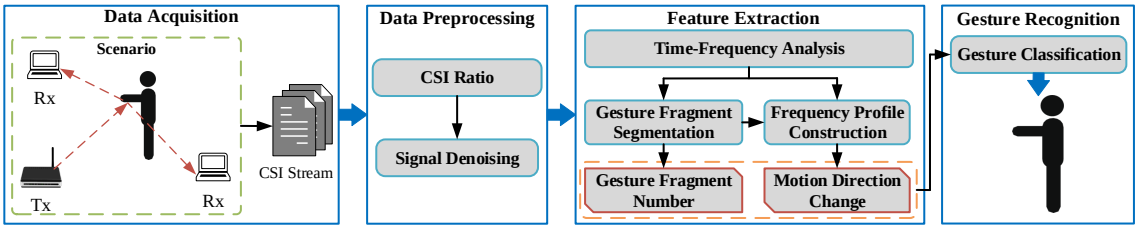


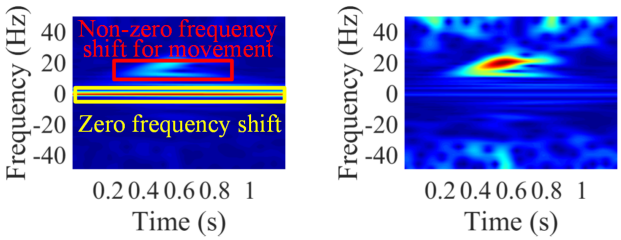
Fig. 21: System overview: Position-independent gesture recognition.

or refrigerator. In our system implementation, we configure Gigabyte MiniPCs equipped with Intel 5300-NIC adapters as the transceivers. Each Rx is equipped with three omni-directional antennas (3cm apart), corresponding to three CSI streams. When using the system, the two Rxs continuously receive wireless packets from Tx and extract three CSI streams from three antennas. A Dell laptop (Precision 5520) with Xeon CPU and 16G RAM is connected to the two receivers to collect CSI samples and process the data in real time.

(2) Data preprocessing: *(i) CSI ratio.* From each sampled packet, we can obtain three CSI streams simultaneously from the antennas. To get the detailed frequency feature, we first use CSI ratio between CSIs from two antennas in the same WiFi adapter to eliminate phase offset in the original CSIs [32, 33]. The first two CSI streams with the highest power are chosen to calculate CSI ratio. *(ii) Signal denoising.* Furthermore, we denoise the signal using a least-square smoothing filter named Savitzky-Golay filter [34], which fits successive subset of adjacent data points with a polynomial using the linear least square method.

(3) Feature extraction: After the denoising process, we employ the CWT algorithm to extract the time-frequency spectrogram from the smoothed signal. However, we notice that the signal still contains static component with zero frequency shift (yellow box in Fig. 22a), which has high energy and makes it difficult to extract the non-zero frequency of movement (red box in Fig. 22a). To eliminate the zero frequency shift so that the time-frequency spectrogram only reflects the target movement, we differentiate the smoothed signal with respect to time. This differentiation also eliminates the contribution of any static object in the environment, making our system resilient to static changes in the environment, e.g., changing the layout of a room. Then we leverage CWT to get time-frequency spectrogram without zero frequency shift. As shown in Fig. 22b, the zero frequency shift in the time-frequency spectrogram is removed effectively. Now, based on the time-frequency spectrogram from multiple receivers, we use the methods proposed in Section 4.2 to extract the number of gesture fragments and direction changes as position-independent features.

(4) Gesture recognition: From the analysis in Section 4.3, we can guarantee that: (1) the features for the same gesture at different locations and orientations are consistent; (2) the features for different gestures are unique for each gesture. According to the position-independent features identified in Table 3, we predefine the features as rules to recognize gestures. Given the features obtained in the previous step, we match them with the predefined rules, and then output



(a) Time-frequency spectrogram with zero frequency shift (b) Time-frequency spectrogram without zero frequency shift

Fig. 22: Remove zero frequency shift by differentiation.

the recognized type of the gesture performed. As a result, we are able to achieve training-free gesture recognition that are independent of location and orientation changes.

6 Evaluation

In this section, we evaluate the overall performance of our system, and compare it with state-of-the-art machine learning based approaches. We also evaluate our system under various conditions in terms of location and orientation changes, environment diversity, user variety and real-life deployments.

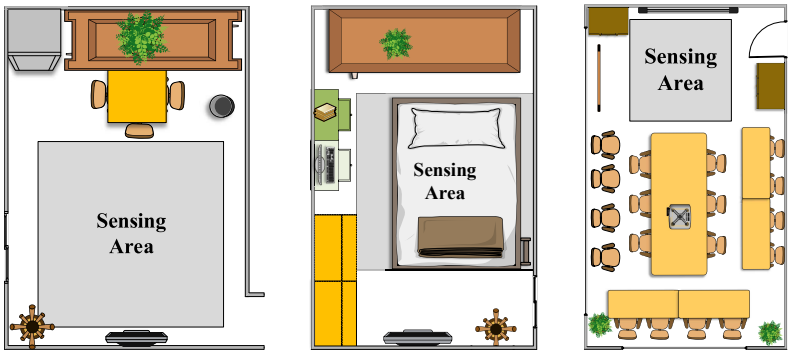
6.1 Experimental Setup

6.1.1 Devices

In all the experiments, the receivers are configured to work under the monitor mode and capture packets from the transmitter simultaneously. We mount one transmitter and two receivers on tripods. The antennas of all devices are placed vertically to the ground at the height of 1m, in order to better capture a user's gesture motions in the sensing area. The open-source Linux CSITool [35] is installed in the receiver to collect CSI data. The frequency of WiFi signal is set to 5.32GHz and the bandwidth is set to 40MHz. The transmitter is set to send 1000 packets per second at a transmitting power of 15dBm.

6.1.2 Environments

To comprehensively evaluate the performance of our system, we conduct experiments in three typical indoor environments shown in Fig. 23: (a) a living room ($2.4m \times 3.6m$) furnished with a dining table, chairs, and cabinet; (b) a bedroom ($2.5m \times 3.85m$) furnished with a bed and desks; and (c) a large meeting room ($5.1m \times 6.9m$) furnished with desks and chairs. The sensing area in each environment is



(a) Living room (b) Bedroom (c) Meeting room

Fig. 23: Layouts of three evaluation environments.

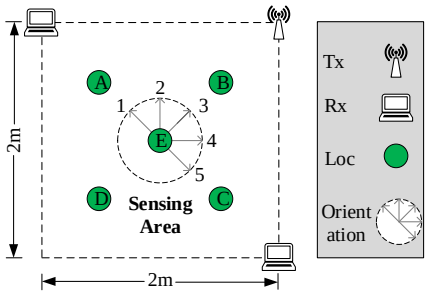


Fig. 24: A typical setup of WiFi devices for gesture recognition.

marked with gray color. Fig. 24 shows a typical deployment set up of the devices in the sensing area, which is a $2m \times 2m$ square. Note that the $2m \times 2m$ square is a typical setting to perform interactive gestures for recognition, especially in the scenario of smart homes where WiFi modules are usually embedded in smart devices such as smart TV and refrigerator [28].

6.1.3 Participants

We have recruited 10 participants (7 male and 3 female), aging from 20 to 40. The height of the participants varies from 155cm to 190cm. The participants include college students and staff members. Among the participants, two are authors of this paper, the other eight participants know nothing about our system. All the participants follow the same instructions to conduct experiments. When we collect data samples, each participant is asked to perform the 8 gestures shown in Fig. 20 in the sensing area. All the gestures are performed in the plane constructed by Tx and two Rxs, which is in parallel with the ground.

6.1.4 Comparison methods

We compare our system with three state-of-the-art machine learning based methods: SVM, CNN, and CNN-LSTM. Following the typical processing procedures in the learning based methods, we normalize the signals, segment the gestures and label the sensing signals with corresponding gestures. Then the sensing signals are used as input to train the model to recognize gestures. Specifically, we use the following methods and setups for comparison:

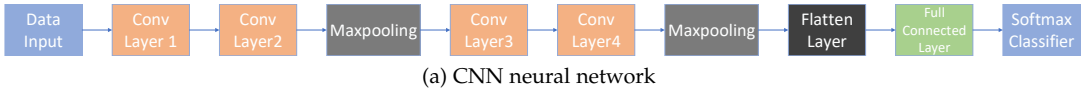
SVM based method: Support Vector Machine (SVM) is widely adopted as a supervised machine learning model. A number of studies have leveraged SVM for activity recognition [1, 2, 23]. To train the SVM model, we extract the following features from the CSI amplitude: mean and standard deviation of CSI amplitude, maximum and minimum value, skewness, kurtosis, root sum square, and q-quantiles (q=0.25,0.5,0.75). We extract totally 20 features from the received CSI signals of the two Rxs and feed them into an SVM classifier with the radial basis function kernel.

CNN based method: Some works have employed Convolutional Neural Network (CNN) to automatically extract signal features to recognize human activities [36, 37, 38]. A typical CNN-based neural network framework is shown in Fig. 25a. During the training process, the input spectrogram data from two Rxs are normalized and resized to a $6 \times 40 \times 40$ tensor. The convolutional layers with 3×3 convolution kernel are used to extract features. We use max pooling and set the pooling size to be 2. All the activation functions are ReLU (ReLU(x)=max(0,x)). The flatten layer is employed to flatten the data into a vector, then fed into a fully connected layer. At last, the softmax Loss function is used as the classifier to generate the output of the neural network. The method is implemented using PyTorch [39].

CNN-LSTM based method: Since WiFi CSI signals are time series with temporal dependency, the combined CNN and LSTM based method [40] can extract features in both spatial and temporal domains. The framework of CNN-LSTM neural network is shown in Fig. 25b. The additional LSTM layer is employed to capture temporal dependencies in sequential data for feature learning. This model is also implemented using PyTorch [39].

6.2 Overall Performance

We first compare the overall performance of our method with the three state-of-the-art machine learning based methods. Then we present the recognition accuracy of each gesture using our system.



(a) CNN neural network

(b) CNN-LSTM neural network

Fig. 25: The CNN and CNN-LSTM neutral networks for performance comparison.

6.2.1 Comparison with state-of-the-art machine learning based methods

In this experiment, we have collected 2400 gesture samples (6 users×5 positions×5 orientations×8 gestures×2 instances) as the dataset. For the machine learning based methods, we use 75% and 25% of the dataset for training and testing, respectively. Fig. 26 shows the recognition accuracy of the four methods for each gesture. Among all the methods, SVM has the lowest average accuracy of 45.4%. The results demonstrate that the amplitude features are easily affected by target location and orientation changes. In contrast, CNN extracts frequency features from the spectrogram and achieves 69.6% average accuracy, whereas CNN-LSTM is able to achieve 65.8% accuracy with the spatial-temporal features extracted from the time-frequency spectrogram. We note that, to some extent, the time-frequency feature is more robust against location and orientation changes than time-domain features. However, the time-frequency feature still has a serious position-dependent issue, even with the help of advanced machine learning methods. In comparison, our method achieves the highest recognition accuracy for all gestures without any training. The average accuracy of our method is 96.8% for all eight gestures, a significant increase of more than 25% in accuracy as compared with other methods. Furthermore, the proposed features can be fed into machine-learning methods. For instance, using our features can increase the accuracy of SVM from 45.4% to 98.1%, indicating the effectiveness of the proposed features.

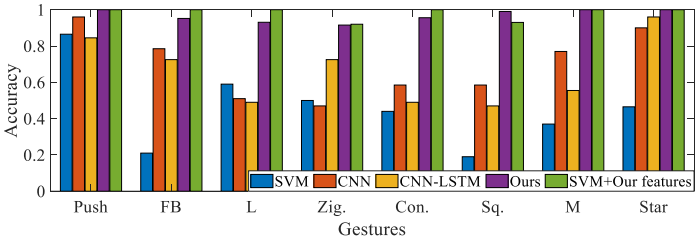


Fig. 26: Comparison of different gesture recognition methods.

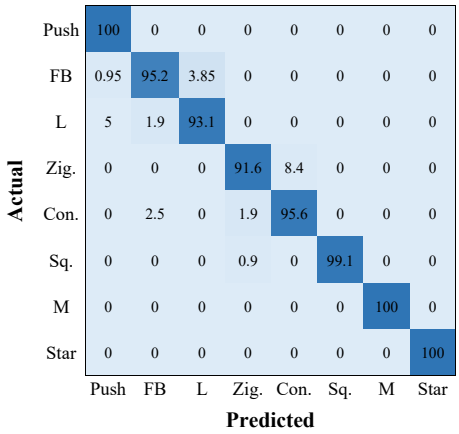


Fig. 27: Confusion matrices of different gestures.

6.2.2 Recognition accuracy

Fig. 27 shows the confusion matrices of recognizing the 8 gestures using our system. We observe that the gestures "Push", "M", and "Star" can be recognized with 100% accuracy. This indicates that using frequency spectrogram analysis from multiple devices, gesture movements can be clearly segmented with high accuracy. Another observation is that the "Zigzag" gesture has the lowest recognition accuracy (still more than 90%) among all gestures. The reason is that this gesture has two successive changes in moving direction, which have a similar impact as the direction changes of gesture "Concave". Nevertheless, the overall recognition accuracy of our system is more than 96.8% on average, this is because we capture the relative change of motion direction rather than the absolute value of CSI signal features.

6.3 Evaluation of System Robustness

Next, we evaluate the robustness of our system against location and orientation changes, environment diversity, and user variety. When evaluating a specific factor, we keep the other impact factors constant.

6.3.1 Location variation

To validate the location independence capability, we evaluate our system at 5 different locations with various distances in each environment, as shown in Fig. 24. The recognition accuracy at location A, C and E are relatively high with 99.8%, 99.5% and 96.1% accuracy, respectively. The accuracy decreases to 90% when targets perform gestures at location D. This is because location D is at the corner of the sensing area and is far away from the transmitter. The wireless signal reflected by the human-body at this location becomes weaker after a longer distance of propagation, resulting in inadequate time-frequency features. Location B also has slightly lower accuracy than other locations. According to our model, movements near LoS induce lower energy in the frequency spectrum. Moreover, if a target's arm happens to pass through LoS, the signal diffraction dominates other signals in this case, which also indicates a low frequency energy. As such, location B has slightly lower accuracy.

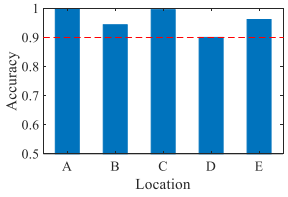


Fig. 28: Accuracy for different locations.

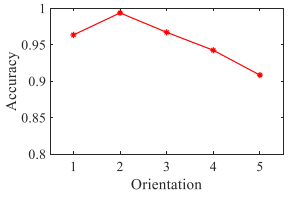


Fig. 29: Accuracy for different orientations.

6.3.2 Orientation variation

In this experiment, we select 5 different orientations facing the WiFi transmitter from orientation 1 to 5 with 45^o angle difference for each. As shown in Fig. 29, the average accuracy for all orientations are all above 90%. The recognition accuracy is highest at orientation 2 (100%), while orientations 5 have slightly lower accuracy. The reason is that the target's arm may be shadowed by his/her body in this orientations. If the target faces the opposite direction of the transmitter, the shadowing effect becomes even worse.

Generally, it is reasonable to assume that the user would face towards the smart device (e.g., TV or screen) in common gesture recognition applications.

6.3.3 Environment diversity

In this experiment, we collect gesture samples in the three multipath-rich environments: living room, bedroom, and meeting room. As shown in Fig. 30, our system achieves an average accuracy of 96% and 97% in the living room and meeting room, respectively. The accuracy is lower (90%) in the bedroom, since we let the users perform the gestures in bed, and users may not perform the gestures properly, thus affecting the sensing performance. Overall, our system is robust against different environments.

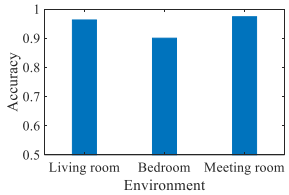


Fig. 30: Accuracy for different environments.

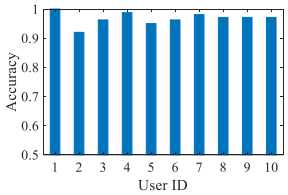


Fig. 31: Accuracy for different users.

6.3.4 User variety

Different users may perform the same gestures differently, such as the movement speed, which can impact the frequency features. Instead of using the original absolute frequency, we extract the motion direction change information to alleviate this problem. To evaluate the performance of system for different users, we collect gesture samples from 6 participants. Fig. 31 shows that the accuracy is above 92% across all 10 users. For user 2, the accuracy is slightly low, since the user has a larger body size, which shadows the signals in some orientations. Overall, our system achieves high accuracy for different users.

6.4 Performance in Challenging Real-life Scenarios

To test whether the proposed features can work in more complicated scenarios in the real world, we conduct experiments under the through-wall scenario and the vertical transceiver deployment.

6.4.1 Through-wall scenario

We first evaluate through-wall performance of our system. The experiment is conducted in the environment of Fig. 23a. The WiFi receivers are placed on the top right and bottom left corners of the sensing area, while the WiFi transmitter (e.g., a router in smart home) is placed in a different room behind the right-side wall near the entrance. The wall between the two rooms is about 30cm thick and is made of solid concrete. Note that concrete can cause more RF attenuation than other common building materials [3], thus our setting is representative of typical indoor environments in modern buildings. Fig. 32 shows the accuracy for eight gestures performed in different positions and the average accuracy is 93.87%. The results indicate that the proposed features are position-independent and can recognize the suite of position-independent gestures in the challenging through-wall scenario. We also observe that the accuracy in the through-wall scenario is slightly lower than that of the unobstructed scenario in Section 6.2. The reason is that the signal reflected off the human target becomes even weaker after penetrating the wall. Note that the state-of-the-art methods (e.g., Widar 3.0) have not evaluated their system performance under the through-wall scenario.

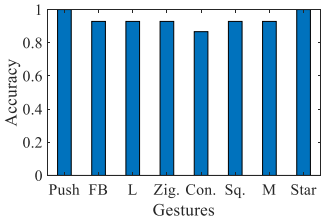


Fig. 32: Accuracy for the through-wall scenario.

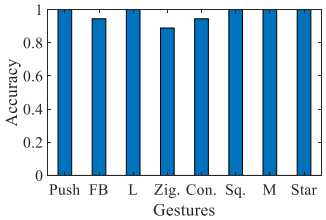


Fig. 33: Accuracy for the vertical deployment.

6.4.2 Vertical transceiver deployment

To further evaluate the capability of the proposed features, we change the deployment of the transceivers. We conduct experiments in the same setting of Fig. 23a. But the plane of the WiFi transceivers is changed to be vertical to the ground. The WiFi transmitter and one receiver are placed at the height of 1.55m (maximum height of the tripods) with 2m LoS, while another receiver is placed under the transmitter at the height of 0.5m. The target performs gestures in the plane composed of the WiFi transceivers. This type of deployment can be applied to human interaction with smart screens and so on. The accuracy is shown in Fig. 33 and the average accuracy is 97.22%, indicating that our position-independent features are robust against different deployment of the transceivers. Note that the accuracy in the vertical deployment is slightly higher than that of the deployment in Section 6.2. This is because there is little occlusion for signal reflected off the human arm in the vertical deployment.

7 Discussion and Future Work

Resolution of motion direction changes. As a relative feature, motion direction changes can be effectively utilized to position-independently differentiate gestures in a training-free fashion. The resolution of motion direction changes is affected by target's position, frequency algorithm, and noises in CSI readings. In our experiments, the resolution is empirically shown to be about $30°$, which is sufficient for recognizing the devised gestures. In our future work, we will further improve the resolution to refine the features with multiple antennas and receivers so that more gestures can be identified position-independently without training.

Generalizability of gestures. The most important premise for generalizability is that the performance can be guaranteed at various conditions. As a step forward, the proposed features can guarantee the recognition performance of gestures at different positions without training. Furthermore, our theoretical model and the position-independent features provide guidance and design criteria to effectively generalize to other gestures. As long as the

criteria are satisfied, the performance can be guaranteed for the generalized gestures. Once if the resolution of features can be further improved, it is possible to generalize to more gestures that can be differentiated position-independently with performance guarantee.

Impact of bandwidth. The proposed model is established on the single-tone frequency, which means that one frequency is sufficient to apply our model to extract position-independent features from multiple receivers without specific requirement of bandwidth. Giving that the current WiFi protocol can provide a maximum bandwidth of 160MHz, it can be used to estimate target's location, movement speed and motion direction by various frequencies. Thus there is an opportunity to further increase the resolution of position-independent features with higher bandwidth for gesture/activity recognition. We take this as our future work.

8 Conclusion

In this work, we have developed a WiFi-based signal frequency model that quantifies the relationship between signal frequency and target position, motion direction and speed, thus providing a theoretical foundation for explaining how frequency features change with target location and orientation. Building upon this model, we prove that the commonly-used movement speed and motion direction features are position dependent, and propose gesture fragments and relative direction changes as two position-independent features, which can be extracted from the time-frequency spectrogram of multiple devices. Then we successfully use these features for position-independent gesture recognition. Compared with three state-of-the-art machine learning based methods, we can improve the gesture recognition accuracy by more than 25% without any training. We further demonstrate that our system is robust against the changes of location, orientation, environment, and user. The model we have developed allows for an in-depth understanding of the relationship between frequency features and human activities, and provides valuable guidelines for selecting frequency features to develop activity recognition applications. Furthermore, we believe that the proposed signal frequency model also provides insights to address the position-dependent problem in RF-based (e.g., WiFi, 4G/5G) sensing systems.

Acknowledgments

This work was supported by the NSFC A3 Project 62061146001, the Project 2019BD005 supported by PKU-Baidu Fund, the EU CHIST-ERA RadioSense Project, the NSFC Grant No.61802373, the Youth Innovation Promotion Association, Chinese Academy of Sciences (No. 2020109).

References

[1] Y. Wang, K. Wu, and L. M. Ni, "Wifall: Device-free fall detection by wireless networks," *IEEE Transactions on Mobile Computing*, vol. 16, no. 2, pp. 581–594, Feb 2017.

[2] 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 Transactions on Mobile Computing*, vol. 16, no. 2, pp. 511–526, Feb 2017.

[3] F. Adib and D. Katabi, "See through walls with wifi!" *SIGCOMM Comput. Commun. Rev.*, vol. 43, no. 4, pp. 75–86, Aug 2013.

[4] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *Proceedings of the 19th Annual International Conference on Mobile Computing and Networking (MobiCom '13)*. New York, NY, USA: ACM, 2013, pp. 27–38.

[5] P. Melgarejo, X. Zhang, P. Ramanathan, and D. Chu, "Leveraging directional antenna capabilities for fine-grained gesture recognition," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14)*. New York, NY, USA: ACM, 2014, pp. 541–551.

[6] H. Abdelnasser, M. Youssef, and K. A. Harras, "Wigest: A ubiquitous wifi-based gesture recognition system," in *2015 IEEE Conference on Computer Communications (INFOCOM)*, April 2015, pp. 1472–1480.

[7] W. He, K. Wu, Y. Zou, and Z. Ming, "Wig: Wifi-based gesture recognition system," in *2015 24th International Conference on Computer Communication and Networks (ICCCN)*, Aug 2015, pp. 1–7.

[8] A. Virmani and M. Shahzad, "Position and orientation agnostic gesture recognition using wifi," in *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '17)*. New York, NY, USA: ACM, 2017, pp. 252–264.

[9] 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)*. New York, NY, USA: ACM, 2015, pp. 90–102.

[10] X. Liu, J. Cao, S. Tang, J. Wen, and P. Guo, "Contactless respiration monitoring via off-the-shelf wifi devices," *IEEE Transactions on Mobile Computing*, vol. 15, no. 10, pp. 2466–2479, 2016.

[11] S. Shi, Y. Xie, M. Li, A. X. Liu, and J. Zhao, "Synthesizing wider wifi bandwidth for respiration rate monitoring in dynamic environments," in *2019 IEEE Conference on Computer Communications (INFOCOM)*, 2019, pp. 181–189.

[12] X. Wang, C. Yang, and S. Mao, "Phasebeat: Exploiting csi phase data for vital sign monitoring with commodity wifi devices," in *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*, 2017, pp. 1230–1239.

[13] J. Liu, Y. Chen, Y. Wang, X. Chen, J. Cheng, and J. Yang, "Monitoring vital signs and postures during sleep using wifi signals," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 2071–2084, 2018.

[14] K. Niu, F. Zhang, J. Xiong, X. Li, E. Yi, and D. 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)*. New York, NY, USA: ACM, 2018, pp. 139–151.

[15] H. Wang, D. Zhang, J. Ma, Y. Wang, Y. Wang, D. Wu, T. Gu, and B. Xie, "Human respiration detection with commodity wifi devices: Do user location and body

orientation matter?" in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. New York, NY, USA: ACM, 2016, pp. 25–36.

[16] D. Wu, D. Zhang, C. Xu, Y. Wang, and H. Wang, "Widir: Walking direction estimation using wireless signals," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. New York, NY, USA: ACM, 2016, pp. 351–362.

[17] D. Zhang, H. Wang, and D. Wu, "Toward centimeter-scale human activity sensing with wi-fi signals," *Computer*, vol. 50, no. 1, pp. 48–57, Jan 2017.

[18] F. Zhang, D. Zhang, J. Xiong, H. Wang, K. Niu, B. Jin, and Y. Wang, "From fresnel diffraction model to fine-grained human respiration sensing with commodity wi-fi devices," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, no. 1, pp. 53:1–53:23, Mar 2018.

[19] F. Zhang, K. Niu, J. Xiong, B. Jin, T. Gu, Y. Jiang, and D. Zhang, "Towards a diffraction-based sensing approach on human activity recognition," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 3, no. 1, 2019.

[20] W. Jiang, C. Miao, F. Ma, S. Yao, Y. Wang, Y. Yuan, H. Xue, C. Song, X. Ma, D. Koutsonikolas, and et al., "Towards environment independent device free human activity recognition," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking (MobiCom '18)*. New York, NY, USA: ACM, 2018, pp. 289–304.

[21] J. Zhang, Z. Tang, M. Li, D. Fang, P. Nurmi, and Z. Wang, "Crosssense: Towards cross-site and large-scale wifi sensing," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking (MobiCom '18)*. New York, NY, USA: ACM, 2018, pp. 305–320.

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

[23] W. Wang, A. X. Liu, and M. Shahzad, "Gait recognition using wifi signals," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. New York, NY, USA: ACM, 2016, pp. 363–373.

[24] Y. Xu, W. Yang, J. Wang, X. Zhou, H. Li, and L. Huang, "Wistep: Device-free step counting with wifi signals," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, no. 4, Jan 2018.

[25] H. Abdelnasser, K. Harras, and M. Youssef, "A ubiquitous wifi-based fine-grained gesture recognition system," *IEEE Transactions on Mobile Computing*, vol. 18, no. 11, pp. 2474–2487, 2019.

[26] H. Li, W. Y. Wei, J. Wang, Y. X. Yang, and L. Huang, "Wifinger: Talk to your smart devices with finger-grained gesture," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. New York, NY, USA: ACM, 2016, pp. 250–261.

[27] R. H. Venkatnarayan, G. Page, and M. Shahzad, "Multi-user gesture recognition using wifi," in *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '18)*. New York, NY, USA: ACM, 2018, pp. 401–413.

[28] Y. Zheng, Y. Zhang, K. Qian, G. Zhang, Y. Liu, C. Wu, and Z. 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)*. New York, NY, USA: ACM, 2019, pp. 313–325.

[29] K. S. Suslick, "Encyclopedia of physical science and technology," *Sonoluminescence and sonochemistry, 3rd edn. Elsevier Science Ltd, Massachusetts*, pp. 1–20, 2001.

[30] A. Einstein *et al.*, "On the electrodynamics of moving bodies," *Annalen der physik*, vol. 17, no. 10, pp. 891–921, 1905.

[31] C. Torrence and G. P. Compo, "A practical guide to wavelet analysis," *Bulletin of the American Meteorological society*, vol. 79, no. 1, pp. 61–78, 1998.

[32] K. Niu, F. Zhang, Y. Jiang, J. Xiong, Q. Lv, Y. Zeng, and D. Zhang, "Wimorse: A contactless morse code text input system using ambient wifi signals," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 9993–10 008, Dec 2019.

[33] Y. Zeng, D. Wu, J. Xiong, E. Yi, R. Gao, and D. Zhang, "Farsense: Pushing the range limit of wifi-based respiration sensing with csi ratio of two antennas," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 3, no. 3, 2019.

[34] R. W. Schafer, "What is a savitzky-golay filter?[lecture notes]," *IEEE Signal processing magazine*, vol. 28, no. 4, pp. 111–117, 2011.

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

[36] H. Chen, Y. Zhang, W. Li, X. Tao, and P. Zhang, "Confi: Convolutional neural networks based indoor wi-fi localization using channel state information," *IEEE Access*, vol. 5, pp. 18 066–18 074, 2017.

[37] Y. Ma, G. Zhou, S. Wang, H. Zhao, and W. Jung, "Signfi: Sign language recognition using wifi," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, no. 1, Mar 2018.

[38] Y. Ma, G. Zhou, and S. Wang, "Wifi sensing with channel state information: A survey," *ACM Comput. Surv.*, vol. 52, no. 3, Jun 2019.

[39] "Pytorch." https://pytorch.org/, 2019, online, accessed Aug 2019.

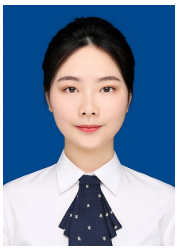
[40] M. T. Islam and S. Nirjon, "Wi-fringe: Leveraging text semantics in wifi csi-based device-free named gesture recognition," *arXiv preprint arXiv:1908.06803*, 2019.

Kai Niu received the M.E. degree in computer technology from the School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, China, in 2016. He is currently pursuing the Ph.D. degree in computer science with the School of Electronics Engineering and Computer Science, Peking University, Beijing, China.

His current research interests include ubiquitous computing, context-aware computing and wireless sensing.

Fusang Zhang received the M.S. and Ph.D. degrees in computer science from the Institute of Software, Chinese Academy of Sciences, Beijing, China, in 2013 and 2017, respectively. He is currently an Associate Professor with the Institute of Software, Chinese Academy of Sciences.

His current research interests include mobile and pervasive computing, ad hoc network, and wireless contactless sensing.



Xuanzhi Wang received the bachelor degree in software engineering from the School of Software, Northwestern Polytechnical University, Xi'an, China, in 2020. She is currently pursuing the Ph.D. degree in computer science with the School of Electronics Engineering and Computer Science, Peking University, Beijing, China.

Her current research interests include ubiquitous computing and mobile computing.

Qin Lv received her Ph.D. degree in Computer Science from Princeton University, Princeton, NJ, USA.

She is an Associate Professor with the Department of Computer Science, University of Colorado Boulder, Boulder, CO, USA. She has published over 100 peer-reviewed papers with over 7000 citations. Her research focuses on full-stack data analytics, which integrates systems, algorithms, and applications for effective and efficient data analytics in ubiquitous computing and scientific discovery. Her research is interdisciplinary in nature and interacts closely with a variety of application domains, including environmental research, Earth sciences, renewable and sustainable energy, transportation electrification, as well as the information needs in people's daily lives, such as mobile environmental sensing, indoor localization, driving behavior analysis, user profiling, and cybersafety. Her current research interests include mobile/wearable/Internet of Things computing, online social networks, spatial-temporal data, anomaly detection, recommender systems, and multimodal data fusion.

Dr. Lv was a recipient of the SenSys 2018 Best Paper Runner-Up Award, the 2017 Google Faculty Research Award, the VLDB 2017 Ten Year Best Paper Award, the ICTAI 2017 Best Student Paper Award, two Best Paper Award nominations, and the Pervasive 2012 Computational Sustainability Award. She is an Associate Editor of PACM IMWUT and has served on the technical program committee and organizing committee of many international conferences.

Haitong Luo is pursuing the bachelor's degree in electronic engineering from the School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing, China.

His research interests include mobile computing, machine learning and signal processing.

Daqing Zhang (Fellow, IEEE) received the Ph.D. degree from the University of Rome "La Sapienza", Italy, in 1996.

He is a Chair Professor with the Department of Computer Science and Technology, Peking University, China, and Telecom SudParis, France. His current research interests include context-aware computing, urban computing, mobile computing, big data analytics, and pervasive elderly care. He has published over 280 technical papers in leading conferences and journals.

Prof. Zhang was a recipient of the Ten-Years CoMoRea Impact Paper Award at IEEE PerCom 2013, the Honorable Mention Award at ACM UbiComp 2015 and 2016, the Best Paper Award at IEEE UIC 2012 and 2015. He served as the General or Program Chair for over 17 international conferences, giving keynote talks at more than 20 international conferences. He is an Associate Editor for IEEE Pervasive Computing, ACM Transactions on Intelligent Systems and Technology, and Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies.