Understanding and Modeling of WiFi Signal Based Human Activity Recognition

Wei Wang[†]   Alex X. Liu[†‡]   Muhammad Shahzad[‡]   Kang Ling[†]   Sanglu Lu[†]

[†]State Key Laboratory for Novel Software Technology, Nanjing University, China
[‡]Dept. of Computer Science and Engineering, Michigan State University, USA
ww@nju.edu.cn, {alexliu,shahzadm}@cse.msu.edu, lingkang@smail.nju.edu.cn, sanglu@nju.edu.cn

Abstract

Some pioneer WiFi signal based human activity recognition systems have been proposed. Their key limitation lies in the lack of a model that can quantitatively correlate CSI dynamics and human activities. In this paper, we propose CARM, a CSI based human Activity Recognition and Monitoring system. CARM has two theoretical underpinnings: a CSI-speed model, which quantifies the correlation between CSI value dynamics and human movement speeds, and a CSI-activity model, which quantifies the correlation between the movement speeds of different human body parts and a specific human activity. By these two models, we quantitatively build the correlation between CSI value dynamics and a specific human activity. CARM uses this correlation as the profiling mechanism and recognizes a given activity by matching it to the best-fit profile. We implemented CARM using commercial WiFi devices and evaluated it in several different environments. Our results show that CARM achieves an average accuracy of greater than 96%.

Categories and Subject Descriptors

C2.1 [**Network Architecture and Design**]: Wireless communication

General Terms

Experimentation,Measurement

Keywords

Channel State Information (CSI);WiFi; Activity Recognition;

1. INTRODUCTION

1.1 Motivation

Human activity recognition is the core technology that enables a wide variety of applications such as health care, smart homes, fitness tracking, and building surveillance. Traditional approaches use cameras [6], radars [2], or wearable sensors [7, 33]. However, camera based approaches have the fundamental limitations of requiring line of sight with enough lighting and breaching human privacy potentially. Low cost 60 GHz radar solutions have limited operation range of just tens of centimeters [2]. Wearable sensors based approaches are inconvenient sometimes because of the sensors that users have to wear. Recently, WiFi signal based human activity recognition systems, such as WiSee [17], E-eyes [27], and WiHear [26], have been proposed based on the observation that different human activities introduce different multi-path distortions in WiFi signals. WiSee uses USRP to capture the OFDM signals and measures the Doppler shift in signals reflected by human bodies to recognize nine gestures. E-eyes uses Channel State Information (CSI) histograms as fingerprints for recognizing daily human activities such as brushing teeth. WiHear uses specialized directional antennas to obtains CSI variations caused by lip movement for recognizing spoken words. Their key advantages over camera and sensor based approaches are that they do not require lighting, provide better coverage as they can operate through walls, preserve user privacy, and do not require users to carry any devices as they rely on the WiFi signals reflected by humans.

1.2 Limitations of Prior Art

The key limitation of these pioneer WiFi based human activity recognition systems is the lack of a model that can quantitatively correlate CSI dynamics and human activities. As such, these systems mostly rely on the statistical characteristics of WiFi signals, such as Doppler movement directions and distributions of signal strength, to distinguish different human activities. The lack of such a model limits the further development of WiFi based human activity recognition technologies. Without such a model, it is difficult to understand the correlation between WiFi signal dynamics and human activities. Furthermore, without such a model, it is difficult to optimize the performance of such systems due to the lack of adjustable parameters, and we have to resort to trial-and-error for performance optimization.

1.3 Proposed Approach

In this paper, we propose CARM, a CSI based human Activity Recognition and Monitoring system. CARM consists of two Commercial Off-The-Shelf (COTS) WiFi devices as shown in Figure 1, one for continuously sending signals, which can be a router, and one for continuously receiving signals, which can be a laptop. When a human activity is performed in the range of these two devices, on the WiFi signal receiver end, CARM recognizes the human activity based on how the CSI value changes. CARM has two theoretical underpinnings that we propose in this paper: a *CSI-speed model* and a *CSI-activity model*. Our CSI-speed model quantifies the correlation between CSI value dynamics and human movement speeds. Our CSI-activity model quantifies the correlation between the movement speeds of different human body parts and a specific human activity. By these two models, we quantitatively build the correlation between CSI value dynamics and a spe-
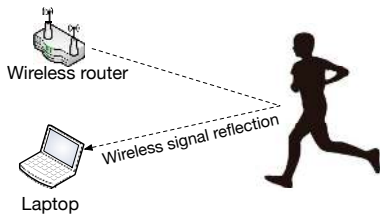
cific human activity. CARM uses this quantitative correlation as the profiling mechanism and recognizes a given activity by matching it to the best-fit profile.

Figure 1: CARM System

Our CSI-speed model and CSI-activity model advance the state-of-the-art on WiFi signal based human activity recognition from two fronts. First, they provide us the theoretical basis to understand, even quantitatively, the relationship between CSI value dynamics and human movement speeds, and the relationship between the movement speeds of different human body parts and human activities. Regarding the relationship between CSI value dynamics and human movement speeds, for example, our model shows that high-speed body part movement generates high-frequency changes in CSI values. Regarding the relationship between the movement speeds of different human body parts and human activities, taking the activity of falling down as an example, our model shows that it can be characterized as a sudden increase in body movement speed in less than one second. Second, these two models provide us the tunable parameters to optimize the performance of WiFi signal based human activity recognition. For example, according to our models, the CSI sampling rate should be chosen as 800 samples per second because the typical human movement speed corresponds to CSI components of lower than 300 Hz.

1.4 Technical Challenges and Our Solutions

The first technical challenge is to estimate human movement speeds from CSI values based on our CSI-speed model. This is challenging because the CSI measurements at the receiver are mixed WiFi signals arrived from multiple paths, which changes as human moves. Furthermore, different human body parts move at different speeds for a given activity and the WiFi signals reflected by different body parts are also mixed at the receiver. Our key observation is that these signals are linearly combined so that their frequencies are preserved when they are mixed together. Therefore, we use Discrete Wavelet Transform (DWT) to separate the frequency components that represent different movement speeds. The advantage of DWT is that it provides a proper tradeoff between time and frequency resolution and enables the measurement of both fast and slow activities.

The second challenge is to build the CSI-activity model that is robust for different humans. For the same activity, to a certain degree, different people perform it differently and even the same person performs it differently at different times. To address this challenge, we propose a Hidden Markov Model (HMM) based human activity recognition approach. We use the patterns of movement speeds for different activities to build their corresponding HMM based models. The features that we extract to infer the speed patterns are only affected by movement speeds of the body and are relatively agnostic to environmental changes. This enables us to recognize activities even when the environment changes. We choose HMM because of its inherent capability to recognize the same activities that are done at different speeds. To recognize a sample of an unknown activity, we evaluate the unknown samples against HMMs of all activities and find the model that gives the highest likelihood.

The third challenge is that CSI values are too noisy to be directly used for human activity recognition. Even in a static environment without any human activity, CSI values fluctuate because WiFi devices are susceptible to surrounding electromagnetic noises. Moreover, the internal state changes in WiFi devices, *e.g.*, transmission rate adaptation and transmission power adaptation often introduce impulse and burst noises in CSI values. General purpose denoising methods, such as low-pass filters or median filters, do not perform well in removing these impulse and bursty noises for two reasons: First, the sampling rates that these methods require are much higher than the frequency of the WiFi signal. Second, the noise density in CSI values is too high for traditional filters, which only work well for low density noise. In this paper, we propose a principal component analysis (PCA) based CSI denoising scheme. This scheme is based on our observation that the signal fluctuations caused by body movements in all subcarriers of the CSI values are correlated.

The fourth challenge is to capture body movements in the presence of carrier frequency offset (CFO). CFO is the dynamically changing difference in carrier frequencies between a pair of WiFi devices, which occurs due to the minor physical differences in hardware and other factors such as temperature changes [8]. CFO causes the phase values of the received signal to change, making it hard to distinguish whether the phase value changed is due to CFO or due to human movement. To address this challenge, we use the CSI signal power to infer the body movement. We show that CSI signal power is not affected by CFO, but retains information about the movement speeds of the body.

The fifth challenge is to automatically detect the start and end of a human activity. To address this challenge, we use the eigenvectors obtained from PCA. The key idea is that in the absence of any activity, the time-series of CSI values contain random noise and consequently, the signal eigenvector varies randomly. During a human activity, the signals in subcarriers become correlated and the signal eigenvector becomes smooth. We capture the smoothness of the eigenvector by calculating its high-frequency energy and compare it to a dynamically adapting threshold to detect start and end.

1.5 Key Technical Novelty and Results

The key technical novelty of this paper is two fold. First, we propose the CSI-speed model and the CSI-activity model to quantify the correlation between CSI value dynamics and a specific human activity. Second, we propose a set of signal processing techniques, such as PCA based denoising and DWT based feature extraction, for human activity recognition based on the CSI-speed model and the CSI-activity model. The key technical depth of this paper lies in the signal processing aspect such as the theoretical analysis of the correlation between CSI values of subcarriers and the relationship between multi-path speeds and CFR power. We implemented CARM on commercial WiFi devices and evaluated it in multiple environments. Our results show that CARM achieves an average activity recognition accuracy of 96%. For a new environment and a new person that the system has never been trained on, CARM can still achieve a recognition accuracy for more than 80%.

2. RELATED WORK

Existing work on device-free human activity recognition and localization can be divided into four categories: Received Signal Strength Indicator (RSSI) based, specialized hardware based, radar based, and CSI based.

RSSI Based: RSSI based human activity recognition systems leverage the signal strength changes caused by human activities [3,22,23]. This approach can only do coarse grained human activity

recognition with low accuracy because the RSSI values provided by the commercial devices have very low resolution [23]. For RSSI based gesture recognition, the accuracy is 56% over 7 different gestures [21]. Sigg *et al.* use software radio to improve the granularity of RSSI values and consequently improve the accuracy of activity recognition to 72% for 4 activities [22]. In comparison, CARM uses CSI values and achieves an accuracy of 96%.

Specialized Hardware Based: Fine-grained radio signal measurements can be collected by software defined radio or specially designed hardware [11, 12, 14, 17]. WiSee uses USRP to capture the WiFi OFDM signals and measures the Doppler shift in signals reflected by human bodies to recognize a set of nine different gestures with an accuracy of 95% [17]. AllSee uses a specially designed analog circuit to extract the amplitude of the received signals and uses their envelopes to recognize gestures within a short distance of 2.5 feet [14]. Wision uses multi-path reflections to build an image for nearby objects [11]. In comparison, CARM requires no specialized hardware and at the same time achieves high activity recognition accuracy at longer distances.

Radar Based: Device-free human activity recognition has also been studied using radar technology [4, 5, 16, 25]. Using the micro Doppler information, radars can measure the movement speeds of different parts of human body [25]. WiTrack uses specially designed Frequency Modulated Carrier Wave (FMCW) signals to track human movements behind the wall with a resolution of approximately 20cm [4, 5]. Compared to the specially designed radar signals such as FMCW or Ultra-wideband (UWB) signals, WiFi signals have much narrower bandwidth. For example, 802.11a/b/g usually use a bandwidth of 20 MHz, while FMCW uses bandwidth of up to 1.79 GHz [4]. Compared to prior work in radar technology, CARM designed a new set of signal processing methods that are suitable for the OFDM signal used in WiFi.

CSI Based: CSI values are available in many commercial devices such as Intel 5300 [9] and Atheros 9390 network interface cards (NICs) [19]. Recently CSI has been used for human activity recognition [10,26,27,30,35] as well as indoor localization [19,32]. Han *et al.* proposed to use CSI to detect a single human activity of falling [10]. Zhou *et al.* proposed to use CSI to detect the presence of a person in an environment [35]. Xi *et al.* proposed to use CSI to count the number of people in a crowd [30]. WiHear uses specialized directional antennas to obtain CSI variations caused by lip movement for recognizing spoken words [26]. E-eyes recognizes a set of nine daily human activities using CSI. Note that WiHear and E-eyes use CSI in quite different ways than CARM. WiHear does not effectively denoise CSI values; thus, it has to use directional antennas to reduce the noise in CSI values to achieve acceptable accuracy. In comparison, we denoise CSI values and use commercial WiFi devices with built-in omnidirectional antennas. E-eyes uses CSI histograms as fingerprints for recognizing human daily activities, such as brushing teeth, taking showers, and washing dishes, which are relatively location dependent. In comparison, CARM uses CSI values based on our CSI-speed and CSI-activity models.

3. UNDERSTANDING WIFI MULTI-PATH

3.1 Overview of CSI

WiFi NICs continuously monitor variations in the wireless channel using CSI, which characterizes the frequency response of the wireless channel [1]. Let $X(f,t)$ and $Y(f,t)$ be the frequency domain representations of transmitted and received signals, respectively, with carrier frequency f. The two signals are related by the expression $Y(f,t) = H(f,t) \times X(f,t)$, where $H(f,t)$ is the complex valued channel frequency response (CFR) for carrier fre-

quency f measured at time t. CSI measurements basically contains these CFR values. Let N_{Tx} and N_{Rx} represent the number of transmitting and receiving antennas, respectively. As CSI is measured on 30 selected OFDM subcarriers for a received 802.11 frame, each CSI measurement contains 30 matrices with dimensions $N_{Tx} \times N_{Rx}$. Each entry in any matrix is a CFR value between an antenna pair at a certain OFDM subcarrier frequency at a particular time. Onwards, we call the time-series of CFR values for a given antenna pair and OFDM subcarrier as *CSI stream*. Thus, there are $30 \times N_{Tx} \times N_{Rx}$ CSI streams in a time-series of CSI values.
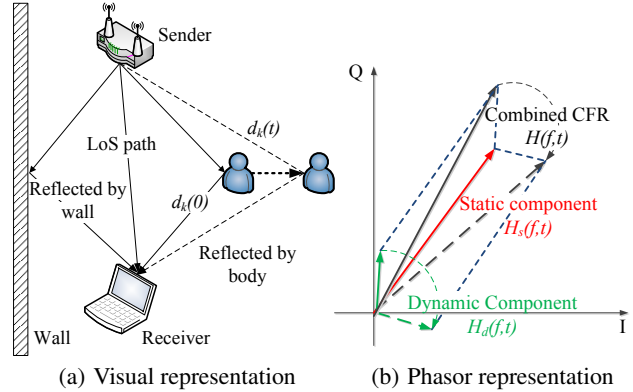

(a) Visual representation   (b) Phasor representation

Figure 2: Multi-paths caused by human movements

3.2 Phase Changes for Paths

Surrounding objects reflect wireless signals due to which a transmitted signal arrives at the receiver through multiple paths. If a radio signal arrives at the receiver through N different paths, then $H(f,t)$ is given by the following equation [24]:

$$H(f,t) = e^{-j2\pi\Delta ft} \sum_{k=1}^{N} a_k(f,t)e^{-j2\pi f\tau_k(t)} \tag{1}$$

where $a_k(f,t)$ is the complex valued representation of attenuation and initial phase offset of the kth path , $e^{-j2\pi f\tau_k(t)}$ is the phase shift on the kth path that has a propagation delay of $\tau_k(t)$, and $e^{-j2\pi\Delta ft}$ is phase shift caused by the carrier frequency difference Δf between the sender and the receiver.

The changes in the length of a path lead to the changes in the phase of the WiFi signal on the corresponding path. Consider the scenario in Figure 2(a), where the WiFi signal is reflected by the human body through the kth path. When the human body moves by a small distance between time 0 and time t, the length of the kth path changes from $d_k(0)$ to $d_k(t)$. Since wireless signals travel at the speed of light, denoted as c, the delay of the kth path, denoted as $\tau_k(t)$ can be written as $\tau_k(t) = d_k(t)/c$. Let f and λ represents the carrier frequency and the wavelength, where $\lambda = c/f$. Thus, the phase shift $e^{-j2\pi f\tau_k(t)}$ on this path can be written as $e^{-j2\pi d_k(t)/\lambda}$, which means that when the path length changes by one wavelength, the receiver experiences a phase shift of 2π in the received subcarrier.

3.3 Practical Limitations

Theoretically, it is possible to precisely measure the phase of the path in systems where sender and receiver are perfectly synchronized, *e.g.*, as in RFID systems [31]. But, unfortunately, commercial WiFi devices have non-negligible carrier frequency offsets (CFO) due to hardware imperfections and environmental variations [8]. IEEE 802.11n standard allows the carrier frequency of a device to

drift by up to 100 kHz from the central frequency of the channel for 5 GHz band [1]. Such frequency drift leads to rapid phase changes in CSI values. Commercial WiFi NICs take one set of CSI measurements per frame. With a transmission rate of 4,000 frames per second, which is around the maximum number of frames that the commercial device can continuously transmit due to the frame aggregation mechanism in 802.11n [1], the phase shift caused by the term $e^{-j2\pi\Delta ft}$ in Equation (1) cloud be as large as 50π between consecutive CSI values.

Our measurements on commercial devices show that phases of CFR are too noisy to be used for activity recognition due to CFO. Figure 3 shows the CSI phase differences for consecutive frames sent through a WiFi link between two commercial devices. Due to the randomness of the packet sending process, the interval Δt between two consecutive frames is randomly distributed in the range of 300~550 microseconds (μs). This gives us a chance to measure the fine grained phase differences for different Δt. Each dot in Figure 3 gives the phase difference for a pair of frame separated by the given Δt, thus we can obtain the relationship between Δt and the phase shift. As shown by Figure 3, the phase difference $2\pi\Delta f\Delta t$ changes by 8π (four vertical strips) when Δt increase from 350 μs to 400 μs. Thus, the CFO can be calculated as $\Delta f = \frac{8\pi}{2\pi(400-350)\mu s} = 80$ kHz. There are two causes that lead to the imprecision of CFR phase. First, from the width of the vertical strips in Figure 3, we observe that CFR phase has measurement error as large as 0.5π. In most cases, the phase changes caused by human reflection are much smaller than 0.5π. Thus, phase changes caused by movements are often buried in phase nosies. Second, our measurements on commercial devices show that the central frequency often drifts by tens of Hz per second, making it hard to predict CFR phase and separate the phase change caused by clock drifts from the small phase shifts caused by body movements. Furthermore, the phase sanitization method introduced in [20] could not work for our case because the phase sanitization process also removes the phase shifts caused by body movements.
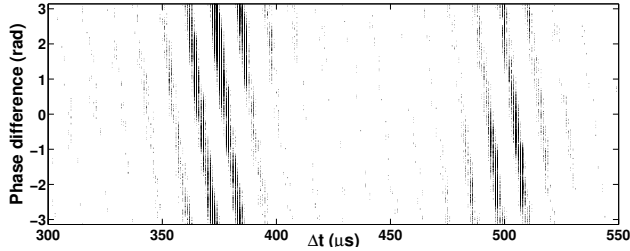

Figure 3: Phase differences for consecutive frames

3.4 CSI-Speed Model

While it is hard to directly measure the phase of a path, it is possible to infer the phase of a path using the CFR power *i.e.*, $|H(f,t)|^2$. The principle behind our method is that when the lengths of multi-paths change, the CFR power varies according to the path length change.

To understand the relationship between CFR power and the length change of a path, we first express CFR as a sum of dynamic CFR and static CFR and then calculate the power. Dynamic CFR, represented by $H_d(f,t)$, is the sum of CFRs for paths whose lengths change with the human movement, and is given by $H_d(f,t) = \sum_{k\in P_d} a_k(f,t)e^{-j2\pi d_k(t)/\lambda}$, where P_d is the set of dynamic paths whose lengths change. Static CFR, represented by

$H_s(f)$, is the sum of CFRs for static paths. Thus, the total CFR is given by the following equation.

$$H(f,t) = e^{-j2\pi\Delta ft}\Big(H_s(f) + \sum_{k\in P_d} a_k(f,t)e^{-j\frac{2\pi d_k(t)}{\lambda}}\Big) \quad (2)$$

The total CFR has time-varying power because in complex plane, the static component $H_s(f)$ is a constant vector while the dynamic component $H_d(f,t)$ is superposition of vectors with time varying phases and amplitudes, as shown in Figure 2(b). When the phase of the dynamic component changes, the magnitude of the combined CFR changes accordingly.

Now, consider how CFR power changes with an object moving around. Let an object move at a constant speed such that the length of the kth path changes at a constant speed v_k for a short time period, *e.g.*, 100 milliseconds. Let $d_k(t)$ be the length of the kth path at time t. Thus, $d_k(t) = d_k(0) + v_k t$. The instantaneous CFR power at time t can be derived as follows (detailed derivations are omitted due to space constraints).

$$|H(f,t)|^2 = \sum_{k\in P_d} 2|H_s(f)a_k(f,t)|\cos\Big(\frac{2\pi v_k t}{\lambda} + \frac{2\pi d_k(0)}{\lambda} + \phi_{sk}\Big)$$
$$+ \sum_{\substack{k,l\in P_d\\k\neq l}} 2|a_k(f,t)a_l(f,t)|\cos\Big(\frac{2\pi(v_k-v_l)t}{\lambda} + \frac{2\pi(d_k(0)-d_l(0))}{\lambda} + \phi_{kl}\Big)$$
$$+ \sum_{k\in P_d} |a_k(f,t)|^2 + |H_s(f)|^2 \qquad (3)$$

where $\frac{2\pi d_k(0)}{\lambda} + \phi_{sk}$ and $\frac{2\pi(d_k(0)-d_l(0))}{\lambda} + \phi_{kl}$ are constant values representing initial phase offsets.

Equation (3) provides a key insight: *the total CFR power is the sum of a constant offset and a set of sinusoids, where the frequencies of the sinusoids are functions of the speeds of path length changes.* By measuring the frequencies of these sinusoids and multiplying them with the carrier wavelength, we can obtain the speeds of path length change. In this way, we can build a CSI-speed model which relates the variations in CSI power to the movement speeds.

3.5 Model Verification

We use a simple moving object to verify our CSI-speed model in Equation (3). We move a steel plate with diameter of 30 cm along the perpendicular bisector of the sender/receiver, similar to the scenario shown in Figure 2(a). Flat steel objects can serve as mirrors for radio waves [34]. Thus, there is only one path dominating the signal reflected by the steel plate and Equation (3) reduces to one sinusoid wave plus a constant offset. The frequency of the sinusoid changes according to the instantaneous moving speed. This can be verified by Figure 4(a), which shows the CSI waveform caused by steel plate movements. When there is only one dominating sinusoid wave, the movement distance can be calculated by measuring the phase change of the signal, which is the integral of the signal frequency over time.

We use Hilbert Transform to calculate the phase change of the waveform as follows. We first remove the DC component that accounts for the static paths. We then use Hilbert Transform to derive the analytic signal from the waveform. The unwrapped instantaneous phase of the analytic signal keeps track of the phase change of the waveform. We can then multiply the phase change with the wavelength to get the path length change. Since the reflected signal goes through a round-trip from the reflector, the path length change is approximately two times of the movement distance of the reflector in this case [29].

The Hilbert Transform based distance measurement has average accuracy of 2.86 cm, as showing in Figure 4(b) and 4(c) . In the experiments, we move the steel plate for a random distance in the
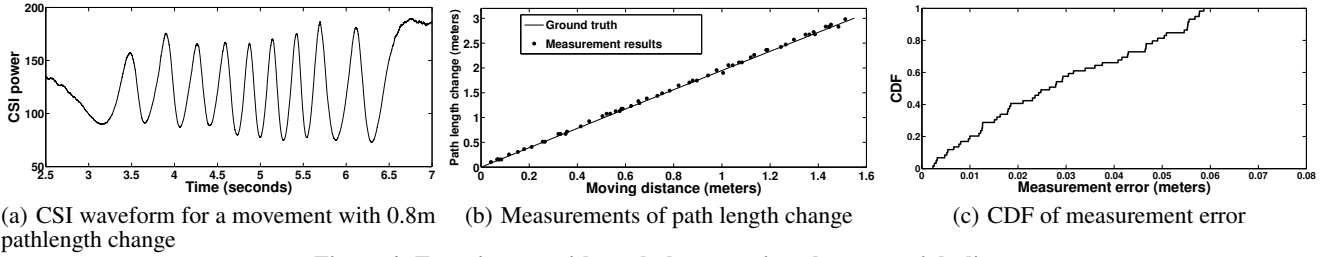
(a) CSI waveform for a movement with 0.8m pathlength change

(b) Measurements of path length change

(c) CDF of measurement error

Figure 4: Experiments with steel plates moving along a straight line.

range of 0∼1.6 m which incurs 0∼3.2 m path length change. The ground truth path length change is measured by a laser rangefinder, which provides distance measurement accuracy of 0.1 cm. Under carrier frequency of 5.825 GHz, which has wavelength of 5.15cm, our path length measurement has maximal error of 5.87 cm and mean error of 2.86 cm. The major error sources are errors in deciding the phase of the starting and ending cycle. Therefore, the measurement error does not increase with the movement distance and is uniformly distributed in the range of 0∼6 cm, see Figure 4(c).

4. MODELING OF HUMAN ACTIVITIES

4.1 Human Activity Characteristics

Modeling CFR power change caused by human activity is challenging. Unlike the simple object used in section 3.5, human bodies have complex shapes and different body parts can move at different speeds. Moreover, the reflections from body parts may go through different paths in complex indoor environments. From Equation (3), we see that the CFR power is a linear combination of all the reflected paths and the speeds of path length change are preserved in the combination process. Therefore, we can use Time-Frequency analysis tools, such as Short-Time Fourier Transform (STFT) or Discrete Wavelet Transform (DWT) to separate these components in the frequency domain. Human activity can be modeled by profiling the energy of each frequency component derived from Time-Frequency analysis tools. As an example, Figure 5 illustrates the waveform and the corresponding spectrogram for three human activities: walking, falling and sitting down. The spectrogram shows how the energy of each frequency component evolves with time, where high-energy components are colored in red. In the spectrogram for the walking activity, there is a high-energy band around 35∼40Hz frequency, as shown in Figure 5(d). With a wavelength of 5.15 cm, these frequency components represent 0.9∼1.0 m/s movement speed after considering the round-trip path length change. This coincides the normal movement speed of human torso while walking [25]. Figure 5(e) shows the spectrogram of falling, which has an energy increase in the frequency range of 40∼80 Hz between 1∼1.5 seconds. This indicates a fast speed-up from below 0.5 m/s speed to 2 m/s, during a short time period of 0.5 seconds, which is a clear sign of falling. The activity of sitting down shown in Figure 5(f) is different from falling, as the speed for sitting down is much slower. Using the energy profile of different frequencies, we can build CSI-activity model, which quantifies the correlation between the movement speeds of different human body parts and a specific human activity.

4.2 Robustness of Activity Speeds

We next study whether the speed based CSI-activity model are robust across different scenarios. It is well known that the path length change is determined by both the position of the sender/receiver and the movement directions [29]. Movements

with the same speed may introduce different path length change speeds when movement directions are different. Furthermore, different people may perform the same activity with different speeds and the multi-path conditions may change under different environments.

Our experiments show that different human activities actually incur path length change speed with significant difference, so that the minor measurement differences caused by movement direction and the different ways to perform the same activity can be safely ignored. To study the robustness of the movement speeds, we collect more than 780 activity samples for three activities, walking, running and sitting down, performed by 25 volunteers with different ages and genders. The activities are performed at different locations with different directions, *e.g.*, we ask the volunteer to walk around a large table so that four different walking directions are captured. Figure 6 shows the estimated torso speed distribution for the three different activities. Note that we estimate the torso speed by dividing the speed of path length change by two. This usually gives a smaller estimation than the actual speed because depending on the movement direction, moving by 1 cm usually cause less than 2 cm path length change [29]. Even with different movement directions, we observe that the three activities have different speeds in Figure 6. Such speed difference can be used for activity classification. As an example, we can achieve a classification accuracy of 88% for all three activities, when we divide the samples to three types with estimated speed of 0∼0.61 m/s, 0.61∼1.0 m/s and above 1.0 m/s. By looking at various different activities, we found that most human activities contains speed components ranging from 0∼2.5 m/s and the frequency components for a given activity are stable across different scenarios, including apartments, offices, and large open area, see our evaluations in Section 8. Therefore, the strength of the frequency components can serve as a robust feature for human activities.
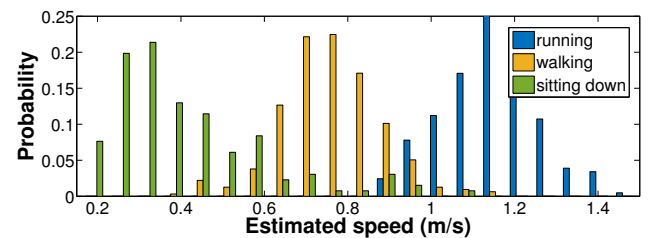

Figure 6: Histogram of speeds for different activities

4.3 CSI-Activity Model

We propose to use Hidden Markov Model (HMM) to build CSI-activity models that consist of mutiple movement states. As an example, we observe that the action of falling comprises several states from Figure 5(e). The person first moves slowly, with most CSI energy on the low frequency (slow movement) components. Then, there is a fast transition to very high speed movement where sub-
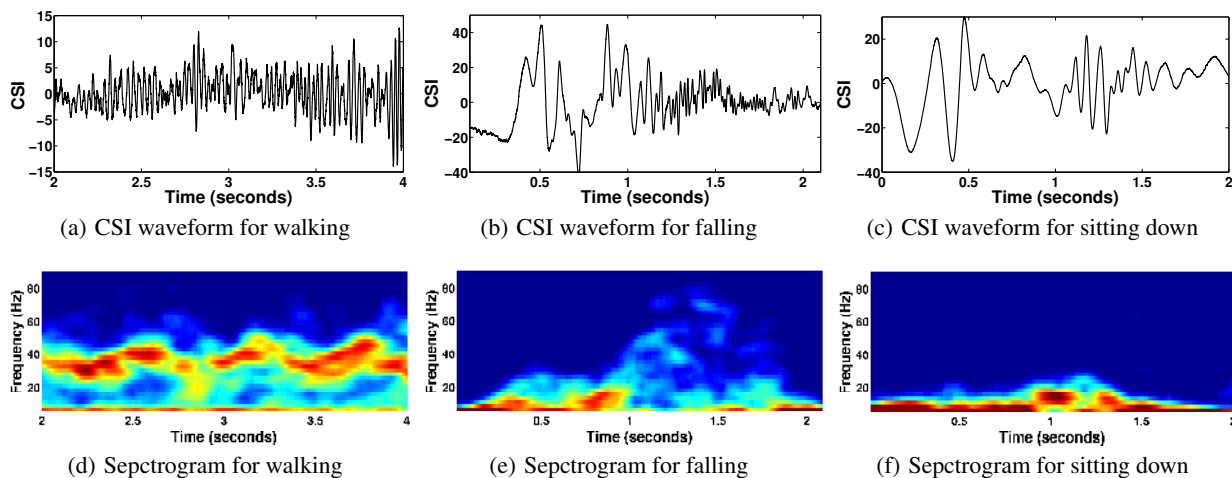
(a) CSI waveform for walking   (b) CSI waveform for falling   (c) CSI waveform for sitting down

(d) Sepctrogram for walking   (e) Sepctrogram for falling   (f) Sepctrogram for sitting down

Figure 5: Waveforms and spectrograms for different activities.

stantial energy are on high frequency components. After that, there is a quick transition to the silent state, where the movement energy reduces to nearly zero. By looking at these transitions between different states, we can infer that the person is possibly falling. Similarly, other human activities also contain states with can be characterized by their movement speeds.

Hidden Markov Model (HMM) is a suitable tool to build state transition models using time-dependent features. It has been extensively used in several recognition applications such as speech recognition [18], handwriting recognition, and gesture recognition in videos [6]. Use of HMMs for activity recognition is based on the assumption that the sequence of observed feature vectors corresponding to an activity is generated by a Markov model, which is a finite state machine that changes state once every time unit. Each time a state is entered, a feature vector is generated from a probability density called *output probability density*. Furthermore, the transition from one state to another or back to itself is also probabilistic and is governed by a discrete probability called *transition probability*. Hidden Markov Models are called hidden because in practice, the sequence of feature vectors is known but the underlying sequence of states that generated those feature vectors is hidden.

HMM can capture information from all training samples and thus works very well even when there is high within-class variance. Provided that a sufficient number of representative training samples of an activity are available, an HMM can be constructed that implicitly models all of the many sources of variability inherent in the activity. Compared to existing works which uses statistical features along a long period [10, 27], HMM based models utilizes the transitions within the activity that provide more details about the activity. For details related to HMM model training and classification, please refer to Section 7.

4.4 Discussion

Detection of high-speed and low-speed movements: CARM can reliably detect both high-speed movement and low-speed movements. Commercial WiFi devices provide CSI values with sampling rates high enough to accurately obtain the values of these frequencies. From our extensive activity dataset, we have observed that indoor human movements introduce frequency components of no more than 300Hz in the CFR power, which corresponds to a top human movement speed of about $300 \times 0.0515/2 = 7.7$m/s, when

the wavelength is 5.15 cm. Note that 7.7 m/s is already too fast a speed for a human to move with. Commercial WiFi NICs can easily sample CSI values at a rate of up to 2500 samples/second, which is far greater than the sampling rate required by the Nyquist criteria for a 300Hz signal. Thus, we can apply signal processing techniques on the denoised CSI values and get the frequency components in CFR power and infer the speed. For slow movements that only move a few centimeters per second, CARM utilizes wavelet transforms to extract low frequency components below 1 Hz to capture slow movements such as brushing teeth, see Section 6.

Movement of different body parts: Time-frequency analysis tools can separate the movement of different body parts when they move at different speeds. For example, the weak energy bands in frequency components between 50∼70 Hz in Figure 5(d) are actually caused by swing of legs when walking [25]. In general, CFR power changes caused by the movements of arms/legs have smaller in energy compared to torso movements as the reflection areas for arms/legs are smaller. Our feature extraction process captures both the body movement and arms/legs movements. Therefore, CARM can distinguish whether the activity involves the whole body or just arms/legs. For example, CARM can recognize falling, running and boxing, which are all high speed movements but involve different body parts.

Scenarios with multiple persons: When there are multiple persons within the same room, CARM can recognize the activity when only one person is moving, see details in Section 8. When both persons are actively moving, we need multiple sender/receivers to capture the actions. Activities that are closer to the given sender/receiver introduce higher distortions in CFR power. Therefore, we can use blind signal separation methods [15] to extract CFR power distortions caused by different person. However, this is out of the scope of this paper and will be studied in our future works.

5. PCA BASED CSI DENOISING SCHEME

CARM builds the HMM model in following three steps as described in Sections 5, 6, and 7, respectively. First, CARM collects CSI values and removes the noises in the measurements. Second, CARM extracts human movement features from the denoised CSI values using DWT. Third, CARM trains an HMM model for each activity and uses the CSI-activity models to recognize activities in real time.

5.1 Sources of Noise in CSI

The CSI streams provided by commercial WiFi devices are extremely noisy. Figure 7 plots a noisy CSI stream that we collected from an Intel 5300 NIC at a sampling rate of 2.5 kHz for a period of 1.75 seconds. One major sources of noise in CSI streams are the internal state transitions in sender and receiver WiFi NICs such as transmission power changes, transmission rate adaptation, and internal CSI reference level changes. These internal state transitions result in high amplitude impulse and burst noises in CSI streams. An interesting feature of these impulse and burst noises is that their effect is highly correlated across all CSI streams, *i.e.*, they affect samples in all streams at the same time. For example, if sender WiFi NIC increases the transmission power by 0.5dB, all streams see a power increase of 0.5dB.

5.2 Traditional Filter based Denoising

Traditional filters such as low-pass filters or median filters do not perform well in removing the impulse and burst noises. Theoretically, a low-pass filter, such as a Butterworth filter, should be able to remove such noise. However, due to the high energy and high bandwidth of impulse noises in CSI, the pass band for the low-pass filters usually need to be less than one-twentieth of the sampling rate so that the energy of residual noise in the pass band becomes negligible compared to the signal energy [26]. When the sampling rate is not high enough, the residual noises can still distort the filtered stream. Figure 7(b) shows the output of a low-pass filter with a cutoff frequency of 100 Hz when applied to the CSI stream in Figure 7(a), which has sampling rate of 2,500 samples per second. We observe that the filtered stream is still severely distorted and the low-pass filter could not effectively denoise it. Another type of filters, called median filters, is specifically designed to remove impulse noise, but they do not work well on CSI streams because the density of the noises is very high. Figure 7(c) shows the output of a 5-point median filter when applied to the CSI stream in Figure 7(a). We again observe that the filtered stream is still severely distorted.

5.3 Correlation in CSI Streams

Our denoising method leverages the fact that the changes introduced in all CSI streams by body movement are correlated. CSI streams of different subcarriers are linear combinations of the same set of time-varying signals and thus they are highly correlated. To show this, consider an object that moves by a small distance between time 0 and time t. Let the length of the path changes by $\Delta_k(t)$ between time 0 and time t when the object moves. Thus, $d_k(t) = \Delta_k(t) + d_k(0)$, where $d_k(0)$ is the initial length of the path. When the initial phase offset of the subcarrier is ϕ_k, the phase of the subcarrier s in Equation (3) seen by the receiver at time t is given by the following equation.

$$
\begin{aligned}
\cos\left(\frac{2\pi d_k(t)}{\lambda_s} + \phi_k\right) &= \cos\left(\frac{2\pi\Delta_k(t)}{\lambda_s} + \frac{2\pi d_k(0)}{\lambda_s} + \phi_k\right) \\
&= \cos\left(\frac{2\pi d_k(0)}{\lambda_s} + \phi_k\right)\cos\left(\frac{2\pi\Delta_k(t)}{\lambda_s}\right) \\
&\quad - \sin\left(\frac{2\pi d_k(0)}{\lambda_s} + \phi_k\right)\sin\left(\frac{2\pi\Delta_k(t)}{\lambda_s}\right) \quad (4)
\end{aligned}
$$

Consider two subcarriers with wavelengths λ_1 and λ_2 that traverse the kth multipath. The difference between wavelengths of subcarriers in a WiFi channel is small. For example, in a 20MHz WiFi channel, the lowest and highest frequency subcarriers are separated by about 17MHz. Thus, in 5GHz band, wavelengths λ of subcarriers differ by at most 0.34%. This slight difference in wavelengths usually does not change the number of mul-

tipaths across subcarriers. We make two observations from Equation (4). First, the time-varying terms in the equation above are approximately equal, *i.e.*, $\cos\left(\frac{2\pi\Delta_k(t)}{\lambda_1}\right) \approx \cos\left(\frac{2\pi\Delta_k(t)}{\lambda_2}\right)$ and $\sin\left(\frac{2\pi\Delta_k(t)}{\lambda_1}\right) \approx \sin\left(\frac{2\pi\Delta_k(t)}{\lambda_2}\right)$ because $\Delta_k(t)$ is small and λ_1 and λ_2 differ only slightly. Second, the constant terms in Equation (4) are unequal for the two subcarriers because initial path length $d_k(0)$ is much greater than $\Delta_k(t)$ and thus, results in a non-negligible initial phase difference between the two subcarriers even though the wavelengths differ only slightly. For example, for a path length of 10 meters, a radio signal with wavelength of 5.150 cm traverses distance equal to 194.1 full wavelengths on this path, where as a radio signal with wavelength of 5.168 cm ($= 5.15 \times 1.0034$) traverses distance equal to 193.5 wavelengths. Thus, there is an initial phase difference of $(194.1 - 193.5) \times 2\pi = 1.2\pi$ between these two signals at the receiver.

These two observations show that CFR for different subcarriers is a linear combination of the same set of time-varying waveforms with different initial phases, *i.e.*, $\cos\left(\frac{2\pi\Delta_k(t)}{\lambda_s}\right)$ and $\sin\left(\frac{2\pi\Delta_k(t)}{\lambda_s}\right)$. Therefore, the CSI streams are correlated. Similar results can be obtained for CSI streams between different antenna pairs because the difference in positions of antennas only causes initial phases and attenuations for each multipath to be different.

Our measurements confirm the observation that CSI streams are correlated. Figure 8 plots the 180 CSI streams for a link with $N_{Tx} = 2$ and $N_{Rx} = 3$ when a human is walking around. We group the CSI streams in their transmission/receiving antenna pairs, *e.g.*, streams 1∼30 are the 30 subcarriers for transmitting antenna 1 and receiving antenna 1. Each CSI stream is a curve that is similar to the one in Figure 5(a), where the amplitudes for CSI values are represented by the color, *i.e.*, red colors are "peaks" and blue colors are "valleys" in the curve. We have following observation in the CSI streams. First, CSI streams are correlated. The "peaks" and "valleys" have similar shapes in all CSI streams across different antenna pairs and different subcarriers. Moreover, the phases of CSI streams changes smoothly across different subcarriers in the same antenna pair, *e.g.*, streams 1∼30 and 151∼180, because the subcarriers on the same antenna pair only differs slightly in their frequencies. Second, there is no single "good" CSI stream. Although we can see clear "peaks" and "valleys" in stream 151∼180 at time between 2.5∼2.6 and 2.8∼2.9 seconds, the changes in measurements are vague during the time 2.9∼3 seconds for the same set of streams. However, we observe streams 1∼30 give clear CSI fluctuations during time 2.9∼3 seconds. This implies that we need to combine different streams to get optimal observations in the movements. Third, simply using weighted average over CSI streams [13] cannot provide good results. We see that the phase of different CSI streams are different so that if we simply add them up, they can cancel each other as the time point for the "peak" of a stream may be the "valleys" of other streams. Therefore, it is important to find a good way to combine CSI streams.
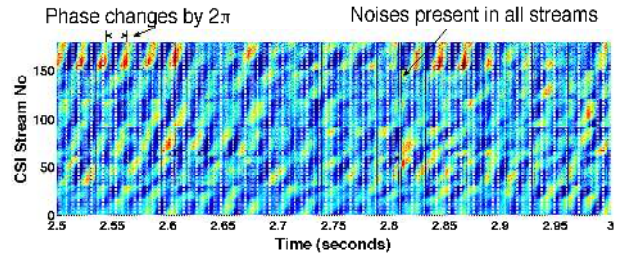

Figure 8: Correlation in CSI streams

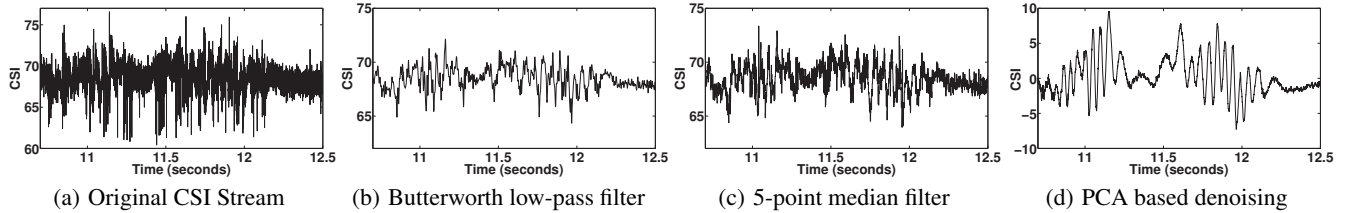| (a) Original CSI Stream | (b) Butterworth low-pass filter | (c) 5-point median filter | (d) PCA based denoising |

Figure 7: Denoising the time-series of CSI values

5.4 Principal Component Analysis

To address the challenges in combining CSI streams, we apply PCA to discover the correlations between CSI streams. With PCA, we can track the time-varying correlations between CSI streams, and optimally combine them to extract principal components of CSI streams. CARM applies PCA to CSI streams using the following four steps.

(1) *Preprocessing:* In this step, CARM first removes the static path components from each CSI stream by subtracting the corresponding constant offsets from the streams. It calculates the constant offset for each stream through long-term averaging over that stream, *i.e.*, average CSI amplitude for 4 seconds. After that, it cut CSI streams into chunks that contain samples obtained in 1-second interval and arrange chunks of different CSI streams in columns to form a matrix of **H**. We choose interval size to be 1 second so that the distance moved by the object is short and at the same time the number of samples is large enough to ensure accurate correlation estimation, which is the next step.

(2) *Correlation estimation:* CARM calculates the correlation matrix as $\mathbf{H}^T \times \mathbf{H}$. The correlation matrix has dimension of $N \times N$, where N is the number of CSI streams. For the example in Figure 8, we have $N = 180$.

(3) *Eigendecomposition:* CARM performs Eigendecomposition of the correlation matrix to calculate the eigenvectors.

(4) *Movement Signal Reconstruction:* In this step, CARM constructs the principal components using the equation $\mathbf{h}_i = \mathbf{H} \times \mathbf{q}_i$, where $\mathbf{q}_i$ and $\mathbf{h}_i$ are the ith eigenvector and the ith principal components, respectively.

CARM discards the first principal component $\mathbf{h}_1$ and retains the next five principal components to be used for feature extraction. As discussed in 5.1, noises caused by internal state changes present in all CSI streams, which are the vertical lines appear in Figure 8. Due to the high correlation, these noises are captured in $\mathbf{h}_1$ along with the human movement signal. However, an interesting result is that all the information about the human movement signal captured in $\mathbf{h}_1$ is also captured in other principal components, because by Equation (4), the phase of a subcarrier is a linear combination of two orthogonal components: $\cos\left(\frac{2\pi\Delta_k(t)}{\lambda}\right)$ and $\sin\left(\frac{2\pi\Delta_k(t)}{\lambda}\right)$. Since the PCA components are uncorrelated, the first principal component only contains one of these orthogonal components and the other component is retained in the rest PCA components. Therefore, we can safely discard the first principal component without losing any information. The number of PCA components used for feature extraction is empirically selected to achieve a good tradeoff between classification performance and computational complexity. Figure 7(d) shows the second PCA component of our denoising scheme. We observe that our proposed method outperforms traditional filtering methods and does not contain the high frequency noise.

6. FEATURE EXTRACTION

6.1 Extracting Features from CSI

To obtain activity features from CSI, CARM needs to extract frequency components from different activities at different time scales. This is because human activities have two aspects associated with them, duration and frequency. Duration represents the time a person takes to perform an activity and frequency represents the speed of multi-paths due to body movements during the activity. Different activities may have similar durations but different frequencies. For example, sitting down and falling both have short durations but the speeds of paths are significantly higher in falling than in sitting down. Consequently, the frequencies in CFR power for falling are greater than the frequencies for sitting down. Similarly, different activities may have similar frequencies but different durations. For example, running and falling both have similar frequencies but the duration of falling is shorter than running. Thus, to analyze CFR power for human activities, we need to extract frequencies from it at multiple resolutions on multiple time scales.

The most relevant signal processing tool that can enable us to extract frequencies at multiple resolutions on multiple time scales is discrete wavelet transform (DWT). DWT provides high time resolution for activities with high frequencies in CFR signals and high frequency resolution for activities with slow speeds. DWT calculates the energies in different levels at any given time in the CFR signals, where each level corresponds to a frequency range. The frequency ranges of adjacent DWT levels decrease exponentially. For example, if level 1 DWT represents a frequency range of 150∼300Hz, which corresponds to 3.85∼7.7 m/s movement speed in 5GHz band, then level 2 DWT represents a frequency range that is half of the frequency range for level 1, *i.e.*, 75∼150Hz, which corresponds to 1.925∼3.85 m/s. The higher the energy in a DWT level is, the more likely it is that the speed of the path is in a range associated with the frequency range of that level. Figure 9(a) shows the wavelet transform for a falling action, where higher brightness represents higher energy level. Although DWT has lower resolution compared to spectrogram in Figure 5(e), we can see the high energy region moves from level 6 to level 2 from 1 to 1.5 seconds. The advantage of DWT compared to STFT is as follows: First, DWT has nice tradeoffs in time and frequency resolutions. DWT naturally groups frequencies that differ by several orders of magnitude into a few levels so that both high speed movements and low speed movements can be captured. Second, DWT reduces the size of data so that the classification algorithm can run in real time.

To extract features for classification from a sample of an activity, CARM applies DWT to decompose the PCA components into 12 levels that span the frequency range from 0.15Hz to 300Hz. The DWT results of the five PCA components are averaged to capture the movement information present in different PCA components. From the output of DWT on each 200ms interval, CARM extracts a 27 dimensional feature vector that includes three types of features. 1). The energy in each level, which represents the intensity

of movement in each speed range. 2). Difference in the energy of each level between consecutive 200ms intervals, which represents the rate of change of speed of a multi-path for the activity. 3). Estimated torso and leg speeds using the percentile method introduced in Doppler radar [25].

6.2 Resilience to Environmental Changes

Environmental changes such as adding an extra chair in a room change the number of multi-paths arriving at the receiver. However, even when the number of multi-paths changes due to environmental changes, the speed of change in lengths of multi-paths does not change because it depends on the movement of human body and not on the number of multi-paths in an environment. Consequently, the frequency components in the CFR power stay the same as long as the person performs the same activity, and DWT gives higher energy in the same levels regardless of how many multi-paths have appeared or disappeared. Figures 10(a) and 10(b) plot the time-series of denoised CSI values for the activity of falling. To emulate the change in environment, our volunteer performed the action at two different locations. Figures 9(a) and 9(b) plot the DWTs of these two waveforms. We observe from these figures that even though the time-series of denoised CSI values look very different for the same activity in different environments, the features that CARM extracts look very similar. Compared to directly using the CSI waveform, the features that CARM uses are resilient to environmental changes, see detailed evaluations in Section 8.
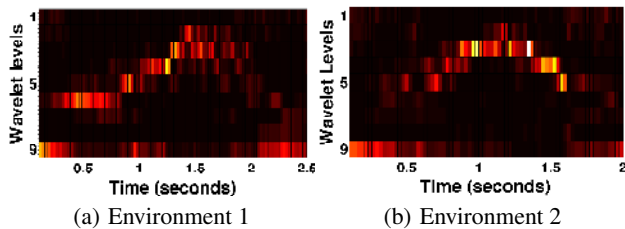

(a) Environment 1   (b) Environment 2

Figure 9: DWT of time-series for falling
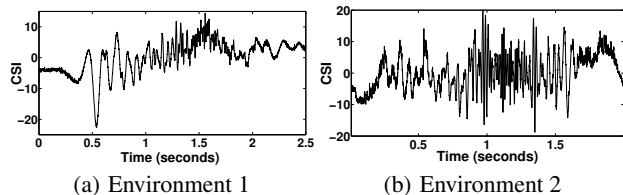

(a) Environment 1   (b) Environment 2

Figure 10: Denoised CSI time-series for falling

7. CLASSIFICATION & RECOGNITION

7.1 Building Activity Models

CARM constructs an HMM for each activity using the training samples of that activity. It also constructs an activity model for the situation when there is no activity in the room. To estimate mean vector and covariance matrix corresponding to each state and the transition probabilities for the HMM, CARM uses the well-known Baum-Welch algorithm [28]. Baum-Welch algorithm needs a rough guess of these probabilities to start with. To guess the initial values, CARM first divides the sequence of feature vectors from each training sample equally amongst the states and then calculates the initial values for the mean vector and covariance matrix of each state using the feature vectors assigned to that state. CARM

also calculates the initial transition probabilities by first counting the number of transitions between every pair of states from the sequence of feature vectors of all training samples divided equally amongst states and then dividing the counts by total number of transitions in all training samples. To decide the number of states, CARM iterates through various number of states and selects the number that provides highest cross validation accuracy. To avoid overfitting of Baum-Welch algorithm to a particular person or moving direction, in generating the model of an activity, we include samples of that activity from different people and different movement directions. Furthermore, we evaluate the models using both 10-fold cross validation and separated testing samples collected in different environments to ensure that the models do not overfit on samples from specific scenarios.

7.2 Real Time Activity Recognition

Once CARM generates HMMs for all activities, it can recognize activities in real-time. We will first explain how CARM detects the start and end of the activity and then explain how it recognizes the unknown activity.

Activity Detection: To detect the start and end of an activity, CARM monitors the second eigenvector $\mathbf{q}_2$ and the corresponding principal component $\mathbf{h}_2$. Our activity detection method is based on two key observations. First, in the absence of an activity, the eigenvector $\mathbf{q}_2$ varies randomly over neighboring subcarriers because CSI streams contain uncorrelated values; whereas, in the presence of an activity, the CSI streams become correlated and $\mathbf{q}_2$ varies smoothly over neighboring subcarriers. Second, in the absence of an activity, the principal component $\mathbf{h}_2$ has smaller variance; whereas in the presence of an activity, it has higher variance.

CARM empirically calculates the variance, $\mathbb{E}\{\mathbf{h}_2^2\}$, of the time series $\mathbf{h}_2$ and the mean of first difference of the eigenvector $\mathbf{q}_2$ given by $\delta_{\mathbf{q}_2} = \frac{1}{S-1} \sum_{l=2}^{S} |\mathbf{q}_2(l) - \mathbf{q}_2(l-1)|$, where S is the number of CSI streams and $|\mathbf{q}_2(l) - \mathbf{q}_2(l-1)|$ is the difference in coefficients for neighboring subcarriers. In the presence of an activity, $\mathbb{E}\{\mathbf{h}_2^2\}$ has a higher value because the time-series for human movement signal varies more frequently, whereas $\delta_{\mathbf{q}_2}$ has a smaller value because the eigenvector becomes more smooth. Therefore, we define the *activity indicator* as $\mathbb{E}\{\mathbf{h}_2^2\}/\delta_{\mathbf{q}_2}$. Figure 11 plots $\mathbb{E}\{\mathbf{h}_2^2\}$, $\delta_{\mathbf{q}_2}$, and the activity indicator over a period of about 10 seconds. We observe that activity indicator increase at 1.4 second and decreases at 9 seconds, which is the start and end times of the action, respectively. The activity indicator $\mathbb{E}\{\mathbf{h}_2^2\}/\delta_{\mathbf{q}_2}$ has better detection performance because it has sharper edges than the metrics of $\mathbb{E}\{\mathbf{h}_2^2\}$ and $\delta_{\mathbf{q}_2}$.
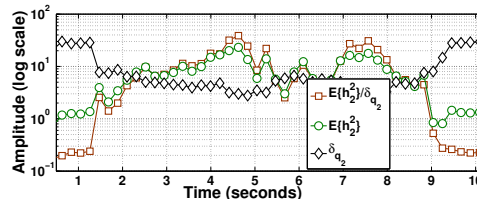

Figure 11: Activity detection indicators

To automatically detect the start or end of an activity, CARM compares the activity indicator with a threshold that it adjusts dynamically based on the background noise level. To dynamically adjust the threshold, we use an Exponential Moving Average (EMA) algorithm to update the detection threshold. In case of a sudden increase in noise level, CARM can incorrectly detect the start of an activity. To handle this, when CARM builds activity models, it also

collects some samples for the situation when there is no activity and builds a model for "no activity" using these samples. CARM classifies a detected activity using all activity models and decides whether a detected activity indeed had an activity in it or not. If it finds that there was no activity, it adjusts the detection threshold accordingly.

Activity Recognition: CARM identifies the activity in following four steps. First, it takes all CSI values between the start and end times and denoises them using the PCA based denoising method described in Section 5.4. Second, from every 200ms interval between the start and end times, it extracts a 27 dimensional feature vector as described in Section 6.1. Third, it uses dynamic programming to calculate the likelihood of each HMM generating this sequence of feature vectors [18]. Finally, the model with the highest likelihood identifies the activity.

8. IMPLEMENTATION & EVALUATION

8.1 Implementation

CARM consists of two components: a laptop with a WiFi card and a commercially available WiFi access point (AP). We implemented CARM on a Think-pad X200 laptop equipped with Intel 5300 WiFi card and tested it using two commercially available 802.11ac APs: NETGEAR JR6100 and TP-Link TL-WDR7500. To obtain CSI values from regular data frames transmitted by the access point, we installed the CSI tool developed by Halperin *et al.* on the laptop [9].

All the experiments that we report in this paper were performed in the 5GHz frequency band with 20MHz bandwidth channels. We choose 5GHz band for shorter wavelength, which lead to better movement speed resolution. We also tested CARM in 2.4GHz. Due to the longer wavelength, frequencies of the CSI waveforms in 2.4GHz are lower for the same activity compared to those in 5GHz. CARM acquires CSI measurements from the CSI tool and processes it in real-time using MATLAB. CARM is computationally efficient. For example, on a Think-pad laptop with Intel i5-3320 CPU and 4GB RAM, CARM takes 85.6ms to process 200ms CSI values sampled at a rate of 2,500 samples/second. Note that it is possible to further improve the computational efficiency of CARM by implementing it in C or using DSP to accelerate the signal processing.

8.2 Evaluation Setup

We collected training samples for eight different activities in our lab, which is 7.7m in length and 6.5m in width, as shown in Figure 12(a). We collected a total of 1,400 samples for the testing activities from 25 volunteers. The volunteers included 20 male and 5 female graduate/undergraduate students with ages in the range of 19–22.

We evaluated the recognition accuracy of CAMR through two sets of experiments, one is in the trained environments and the other is in the untrained environments. We use the lab where we collected the training dataset as the trained environments. For the untrained environments, we evaluate on three typical indoor scenarios, including: 1) a large open lobby area, which has a length, width, and height of 45m, 5.3m, and 4m, respectively; 2) a small apartment, which has area of $70m^2$ as shown in Figure 12(b); 3) a small office, which has a length, width, and height of 5.6m, 3.4m, and 2.7m, respectively.

The activities for which we collected training samples are listed in Table 1, along with their abbreviations and number of training samples for each activity. We collected training samples for each activity except walking and running at the location marked with star in Figure 12(a). For walking and running, our volunteers followed
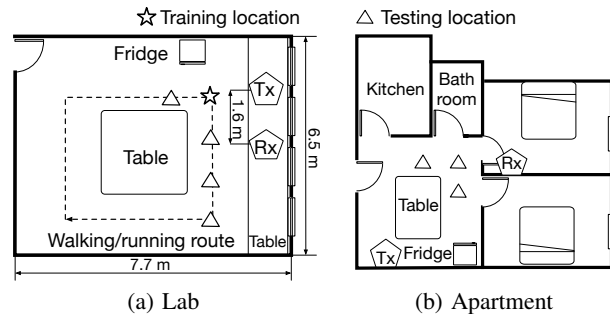

(a) Lab   (b) Apartment

Figure 12: Floor plans

the path marked with a dashed line around the table in the center of the lab. Tx and Rx represent the locations of the transmitter AP and receiver laptop. While collecting the training data, we requested the volunteers to change their orientation to ensure the generality of the collected data. Triangles in Figures 12(a) and 12(b) represent the locations where our volunteers performed activities when evaluating accuracy of CARM. The total training time for our activity data set with 1,400 samples was 100.55 seconds on a laptop with Intel i5-4285 CPU, as shown in Table 1. Note that our activities have no location dependency. Furthermore, once CARM is trained on the given training set, it can be directly applied to environments and persons that have not been included in the training set. Thus, CARM does not need the on-site training data collection as for E-eyes [27]. Consequently, the one-time training can be done on a large training set using a data center.

Activity	Samples	Training Time
(R) Running	205	16.38s
(W) Walking	315	26.84s
(S) Sitting down	266	14.49s
(O) Opening refrigerator	213	13.49s
(F) Falling	98	5.02s
(B) Boxing	75	4.88s
(P) Pushing one hand	72	7.00s
(T) Brushing teeth	96	7.35s
(E) Empty (*i.e.*, no activity)	60	5.10s

Table 1: Summary of activity dataset

8.3 Activity Detection

Now we present the accuracy CARM achieves in detecting the presence of an activity. We evaluate the accuracy using two metrics: true positive rate (TPR) and false alarm rate (FAR). TPR is the ratio of the number of times CARM correctly detects the presence of an activity to the total number of times the activity is performed. FAR is the ratio of the number of times that CARM incorrectly detects the presence of an activity when actually there is no activity.
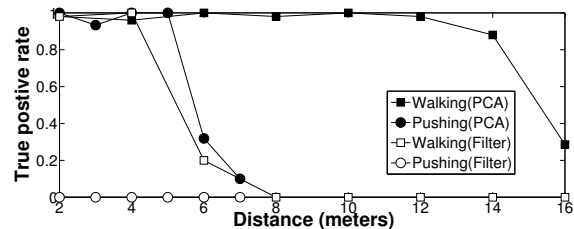

Figure 13: Detection range of CARM

CARM detects small movements such as pushing one hand and large movements such as walking with a TPR larger than 98% at distances of up to 5 meters and 12 meters, respectively. Figure 13 plots CARM's TPR for two different activities in the open lobby area as described in section 8.2, where the sender and receiver were separated by 3.5 meters. Each TPR for an activity was calculated from 20 samples of that activity collected at three different locations while ensuring that the distance of each point from the receiver was the same. Our PCA denoising method has a much longer detection range for walking compared to the range of 4 meters achieved by the low-pass filtering approach. Furthermore, without PCA, low-pass filtering cannot reliably detect small movements such as pushing at a distance of 2 meters. This shows that compared to the conventional low-pass filtering approach, PCA is more efficient in extracting small changes in CSI values caused by human movements. Using PCA denoising, CARM achieves a large coverage area: with a single Tx/Rx pair, it reliably detects the presence of activities in a $450m^2$ open region.

CARM has a low FAR of 1.4 false alarms per hour. To measure CARM's FAR, we record the activity log for 14 hours from 6:00pm to 8:00am on a certain day when no one was around the sender and receiver. During these 14 hours, we saw only 20 false alarms, which were primarily caused by sudden increase in noise levels.

8.4 Activity Recognition

CARM achieves an average cross validation accuracy of 96.5% across all activities. Figure 14 shows the recognition accuracy of 10-fold cross validation using the training data set collected in the "lab". Except for "sitting down", "opening fridge" and "falling", CARM achieves close to 100% accuracy, when the sampling rate is 2,500 samples per second. The accuracy for "empty" is 100%, which means that if CARM incorrectly detects the presence of an activity when in reality there was no activity, it can correctly find out that there was no activity as soon as it evaluates the detected sample against the activity models.
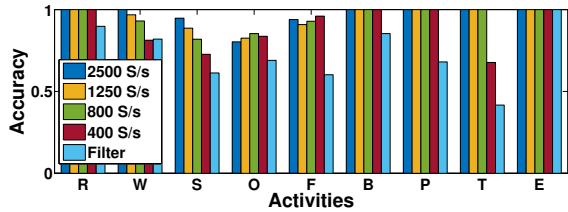

Figure 14: Impact of sampling rates on accuracy

Accuracy of CARM improves when CSI values are sampled at a higher sampling rate, but the increase is not significant beyond the sampling rate of 800 samples/second. Figure 14 plots CARM's recognition accuracy for each activity under four different sampling rates. From this figure, we observe that with sampling rates of 800 samples/second and higher, the average cross validation accuracy is over 94.8%, which is only a slight drop compared to 96.5% achieved for a sampling rate of 2,500 samples/second. Sampling rate of 800 samples/second is very close to the Nyquist sampling rate because the frequency components in CFR power go up to 300Hz due to human movements, as discussed in Section 5.2. When sampling rate further reduces to 400 samples/second, the average accuracy reduces significantly to 87%. In comparison, the average recognition accuracy for low-pass filtering approach is only 73% with the sampling rate of 2,500 samples/second. In the low-pass filtering approach, the accuracies for activities of "falling" and small activities such as "brushing teeth" drop significantly because filtering cannot reliably preserve the high frequency com-

ponents and small fluctuations in CSI values. In our real-time experiments, CARM tolerates wireless interference caused by other devices that are using the same channel. When facing a high packet loss rate, CARM interpolates the missing CSI values and uses a lower sampling rate to achieve an acceptable recognition accuracy.
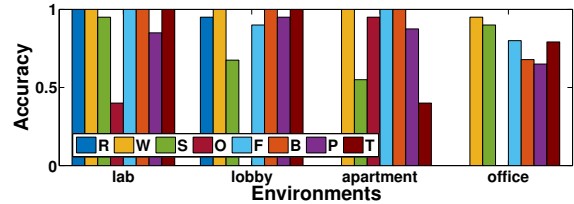

Figure 15: Accuracy in different environments

CARM achieves an accuracy of more than 80% for environments and persons that it has not been trained on. Figure 15 shows the accuracy CARM achieves in different environments for each activity. In each environment, we randomly picked 3 to 6 locations and performed each activity for three times in each location to calculate the recognition accuracy. Two volunteers who were not in the training set performed these activities. Note that this figure does not have results for "opening refrigerator" in the lobby and office, "running" in the apartment and office due to equipment and space constraints. In the lab, lobby, office and apartment, CARM achieves average accuracy of 90%, 93%, 83% and 80%, respectively. For the environment that is the same as the training environment, *i.e.*, "lab", the average accuracy reduces moderately to 90% with lower accuracy for a single activity, "opening refrigerator", due to the different ways of performing the activity. For the more challenging case where CARM is applied to a new environment that it have no training data, the average accuracy drops to around 80%. Most of the recognition errors are in a few activities, *e.g.*, "sitting down" for the lobby and apartment case and "brushing teeth" for the apartment case. The major causes for the errors are due to the drastic change in multi-path conditions. For example, there are no LOS path between the sender and the receiver in the apartment case as shown in Figure 12(b), while our training data are all collected in the lab with LOS path. The performance of CARM can be improved by including more scenarios in the training set. During the experiments, there are other persons sitting or using computers in the same room. CARM is robust to multiple persons coexisting in the same room given that they do not move at the same time. Recognizing simultaneous activities of multiple persons is left as future work of this paper. Currently, CARM uses a single Tx/Rx pair. The performance of CARM can be improved by simultaneously using multiple Tx/Rx pairs located at different positions in the environment. We will explore this direction in our future work.

9. CONCLUSION

In this paper, we make the following three key contributions. First, we propose the CSI-speed model, which quantifies the correlation between CSI value dynamics and the movement speeds of different human body parts, and the CSI-activity model, which quantifies the correlation between the movement speeds of different human body parts and a specific human activity. Second, we propose several signal processing techniques, such as PCA based denoising and DWT based feature extraction, for human activity recognition based on the two models. Third, we implemented CARM using commercial WiFi devices and evaluated it in several different environments. Our results show that CARM achieves an average accuracy of greater than 96%.

Acknowledgments

10. REFERENCES

[1] Enhancements for higher throughput. IEEE Standard 802.11n, 2009.

[2] Google project Soli. https://www.youtube.com/watch?v=0QNiZfSsPc0.

[3] H. Abdelnasser, M. Youssef, and K. A. Harras. WiGest: A ubiquitous wifi-based gesture recognition system. In *Proceedings of IEEE INFOCOM*, 2015.

[4] F. Adib, Z. Kabelac, and D. Katabi. Multi-person motion tracking via RF body reflections. In *Proceedings of Usenix NSDI*, 2015.

[5] F. Adib, Z. Kabelac, D. Katabi, and R. C. Miller. 3D tracking via body radio reflections. In *Proceedings of Usenix NSDI*, 2013.

[6] J. K. Aggarwal and S. R. Michael. Human activity analysis: A review. *ACM Computing Surveys*, 43(3), 2011.

[7] E. Ertin, N. Stohs, S. Kumar, A. Raij, M. al'Absi, and S. Shah. AutoSense: unobtrusively wearable sensor suite for inferring the onset, causality, and consequences of stress in the field. In *Proceedings of ACM Sensys*, 2011.

[8] J. Gjengset, J. Xiong, G. McPhillips, and K. Jamieson. Phaser: Enabling phased array signal processing on commodity WiFi access points. In *Proceedings of ACM MobiCom*, 2014.

[9] D. Halperin, W. Hu, A. Sheth, and D. Wetherall. Tool release: Gathering 802.11n traces with channel state information. *ACM SIGCOMM CCR*, 41(1):53, 2011.

[10] C. Han, K. Wu, Y. Wang, and L. M. Ni. Wifall: Device-free fall detection by wireless networks. In *Proceedings of IEEE INFOCOM*, pages 271–279, 2014.

[11] D. Huang, R. Nandakumar, and S. Gollakota. Feasibility and limits of Wi-Fi imaging. In *Proceedings of ACM SenSys*, pages 266–279, 2014.

[12] K. R. Joshi, S. S. Hong, and S. Katti. Pinpoint: Localizing interfering radios. In *Proceedings of Usenix NSDI*, pages 241–253, 2013.

[13] B. Kellogg, A. Parks, S. Gollakota, J. R. Smith, and D. Wetherall. Wi-Fi backscatter: Internet connectivity for RF-powered devices. In *ACM SIGCOMM*, 2014.

[14] B. Kellogg, V. Talla, and S. Gollakota. Bringing gesture recognition to all devices. In *Proceedings of Usenix NSDI*, 2014.

[15] T. Kim, H. T. Attias, S.-Y. Lee, and T.-W. Lee. Blind source separation exploiting higher-order frequency dependencies. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(1):70–79, 2007.

[16] B. Lyonnet, C. Ioana, and M. G. Amin. Human gait classification using microdoppler time-frequency signal representations. In *Proceedings of IEEE Radar Conference*, pages 915–919, 2010.

[17] Q. Pu, S. Gupta, S. Gollakota, and S. Patel. Whole-home gesture recognition using wireless signals. In *Proceedings of ACM MobiCom*, 2013.

[18] L. R. Rabiner and B.-H. Juang. *Fundamentals of speech recognition*. Prentice Hall, 1993.

[19] S. Sen, J. Lee, K.-H. Kim, and P. Congdon. Avoiding multipath to revive inbuilding WiFi localization. In *Proceeding of ACM MobiSys*, pages 249–262, 2013.

[20] S. Sen, B. Radunovic, R. R. Choudhury, and T. Minka. You are facing the mona lisa: spot localization using phy layer information. In *Proceedings of ACM MobiSys*, pages 183–196, 2012.

[21] S. Sigg, U. Blanke, and G. Troster. The telepathic phone: Frictionless activity recognition from WiFi-RSSI. In *IEEE PerCom*, pages 148–155, 2014.

[22] S. Sigg, M. Scholz, S. Shi, Y. Ji, and M. Beigl. RF-sensing of activities from non-cooperative subjects in device-free recognition systems using ambient and local signals. *IEEE Transactions on Mobile Computing*, 13(4):907–920, 2014.

[23] S. Sigg, S. Shi, F. Buesching, Y. Ji, and L. Wolf. Leveraging RF-channel fluctuation for activity recognition: Active and passive systems, continuous and rssi-based signal features. In *Proceedings of ACM MoMM*, 2013.

[24] D. Tse and P. Viswanath. *Fundamentals of wireless communication*. Cambridge university press, 2005.

[25] P. Van Dorp and F. Groen. Feature-based human motion parameter estimation with radar. *IET Radar, Sonar & Navigation*, 2(2):135–145, 2008.

[26] G. Wang, Y. Zou, Z. Zhou, K. Wu, and L. M. Ni. We can hear you with Wi-Fi! In *Proceedings of ACM MobiCom*, 2014.

[27] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, and H. Liu. E-eyes: In-home device-free activity identification using fine-grained WiFi signatures. In *Proceedings of ACM MobiCom*, 2014.

[28] L. R. Welch. Hidden markov models and the baum-welch algorithm. *IEEE Information Theory Society Newsletter*, 53(4):10–13, 2003.

[29] N. J. Willis. *Bistatic Radar*. SciTech Publishing Inc., 2005.

[30] W. Xi, J. Zhao, X.-Y. Li, K. Zhao, S. Tang, X. Liu, and Z. Jiang. Electronic frog eye: Counting crowd using WiFi. In *Proceedings of IEEE INFOCOM*, 2014.

[31] L. Yang, Y. Chen, X.-Y. Li, C. Xiao, M. Li, and Y. Liu. Tagoram: real-time tracking of mobile RFID tags to high precision using cots devices. In *ACM MobiCom*, pages 237–248, 2014.

[32] Z. Yang, Z. Zhou, and Y. Liu. From RSSI to CSI: Indoor localization via channel response. *ACM Computing Surveys*, 46(2):25, 2013.

[33] K. Yatani and K. N. Truong. Bodyscope: a wearable acoustic sensor for activity recognition. In *Proceedings of ACM UbiComp*, pages 341–350, 2012.

[34] X. Zhou, Z. Zhang, Y. Zhu, Y. Li, S. Kumar, A. Vahdat, B. Y. Zhao, and H. Zheng. Mirror mirror on the ceiling: flexible wireless links for data centers. In *Proceedings of ACM SIGCOMM*, pages 443–454, 2012.

[35] Z. Zhou, Z. Yang, C. Wu, L. Shangguan, and Y. Liu. Towards omnidirectional passive human detection. In *Proceedings of IEEE INFOCOM*, pages 3057–3065, 2013.