Harmony: Exploiting Coarse-Grained Received Signal Strength from IoT Devices for Human Activity Recognition

Zicheng Chi, Yao Yao, Tiantian Xie, Zhichuan Huang, Michael Hammond, and Ting Zhu
Department of Computer Science and Electrical Engineering
University of Maryland, Baltimore County
{zicheng1, yaoyaoumbc, xtiant1, zhihu1, mhammon1, zt}@umbc.edu

Abstract*—The emerging smart health and smart home applications require pervasive and non-intrusive human activity recognition and monitoring. Traditional technologies (e.g., using cameras or accelerometers and gyroscopes) may introduce privacy issues or require people to wear sensors. To address these issues, recent approaches exploit fine-grained wireless signals for activity recognition. However, these approaches require devices that are costly or need to provide unique wireless features (e.g., Doppler shifts or phase information). With the increasingly available Internet of Things (IoT) devices, in this paper, we propose *Harmony*, a human activity recognition and monitoring middleware which can utilize the coarse-grained (but pervasively available) received signal strength (RSS) measurements from the radios of IoT devices. We implement a complete evaluation platform (from data collection to data analysis) of the middleware on top of low cost ZigBee compliant MICAz nodes and a laptop. We also conducted extensive experiments. Our results show that our design can achieve similar accuracy as fine-grained WiFi channel state information (CSI) measurement-based approaches. Specifically, our overall human activities recognition accuracy is up to 74% and 90% for RSS readings from a single pair and 3 pairs of IoT devices, respectively.

I. Introduction

Recognizing and monitoring human activities have a lot of potential applications (e.g., home energy management, personalized health, and elder assistance). For example, the recognition and monitoring of activities such as *running, jumping, and sit-up* can be used for fitness tracking and calorie burned calculation. More importantly, by detecting the *falling* activity, we are able to help more than 2.5 million patients per year in U.S. [1]. Especially, for the elderly (65+), the unintentional fall death rate is linearly increasing since 2004 and higher than 56 per 100,000 in 2013, which is mainly because of no immediate detection after a fall.

Traditional approaches use cameras or wearable sensors (e.g., accelerometers and gyroscopes) to recognize human activities, which may reveal private information through cameras or introduce uncomfortableness while wearing sensors. To address these issues, researchers propose to recognize human activities by measuring the fine-grained wireless signals, such as Doppler shifts [2], [3], [4], [5], [6], time of arrival [7], angle of arrival [8], and channel state information (CSI) [9], [2]. However, these approaches require either customize [8], [2], [3] or unique [9], [2] hardware platforms that can measure the fine-grained wireless signals. Therefore, the adoption of these approaches are limited.

On the other hand, the number of internet of things (IoT) devices has reached 14.4 billion in 2014 and will grow exponentially to reach 50.1 billion by 2020 [10]. Most of the wireless radios (e.g., WiFi, ZigBee, Bluetooth) on the IoT devices are able to measure the received signal strength (RSS). It is highly possible that in the near future we will witness a paradigm shift from fine-grained (but limited available) wireless signal-based to coarse-grained (but pervasively available) RSS-based design for all kinds of human activity/gesture recognition and monitoring. Therefore, in this paper, we take the first attempt to explore the feasibility of human activity recognition by measuring RSS. A simplified example is shown in Fig. 1. Within the transmission range of a pair of IoT devices (i.e., a sender and a receiver), a person performs activities, which will affect the RSS values at the receiver side.

In order to utilize the coarse-grained RSS values for human activity recognition, we encountered five main challenges: *(i)* the resolution of RSS is extremely low (i.e., 8 bits for most of the IoT devices), while the reflected signal from human body is much weaker than the line-of-sight (LoS) signal at the receiver side; *(ii)* different IoT devices have different sampling rates (e.g., 1,000 samples for ZigBee, while 2,500 samples for WiFi); *(iii)* RSS values are extremely coarse-grained (e.g., 8 bits for a MICAz node) when compared with the fine-grained channel state information (e.g., Intel 5300 [11] has 3 antennas and each antenna has 30 groups of subcarriers. Therefore, its channel state information contains 30 matrices with 3×3 dimension and each element in the matrix has 64 bits); *(iv)* with the extremely coarse-grained RSS values, we need to build a robust model that can recognize the same activity from different people. This is the most challenge part, because different people may perform differently for the same activity. Moreover, for the same activity, the same person may perform differently over time; and *(v)* we also need to incorporate the RSS values from multiple IoT sender-receiver pairs from different dimensions to further improve the accuracy.

To address these challenges, we design *Harmony*, a human activity recognition and monitoring middleware, which can utilize existing IoT devices' radios (e.g., WiFi, ZigBee, and Bluetooth) to provide the add-on service for various applica-

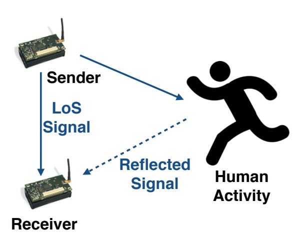
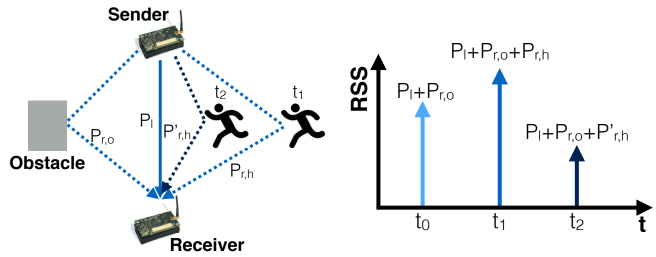


Fig. 1: Human Activity Recognition Scenario



(a) Multipath (b) RSS

Fig. 2: An example to prove the human activity could be captured by RSS variation though multipath effect

tions, such as home energy management, personalized health monitoring, and fall detection. Specifically, we use oversampling technique to compensate the extremely low resolution of RSS. To accommodate different sampling rates of RSS, we develop data pre-processing technique. Then, we conduct the time-frequency analysis to address the coarse-grained RSS issue. After that, we apply the machine learning technique to classify and recognize the human activities. In summary, our main contributions are as follows:

- To the best of our knowledge, this is the first work that exploits the coarse-grained RSS values from the existing IoT devices for human activity recognition.
- We have designed and implemented *Harmony*, a middleware that is i) generic enough to accommodate different RSS sampling rates from multiple IoT devices simultaneously; and ii) robust enough to recognize the same activities from different people by using extremely low resolution RSS values.
- We have conducted extensive real-world evaluations with 9 different types of human activities that can be categorized into three categories: i) daily life; ii) accident; and iii) fitness. The overall accuracy is 74% for a single stream of RSS readings and 90% for RSS readings from 3 pairs of IoT devices.
- We have also extensively evaluated the factors (i.e., different RSS sampling rates, ZigBee channels, and number of clusters) that can potentially affect the recognition accuracy. Results reveal that our approach is reliable and has relatively high accuracy under different scenarios. We also find that the non-line-of-sight human activities do not have obvious impact to the data packet receptions among IoT devices.

II. Background

In this section, we describe the relationship between human activities and their corresponding RSS measurements at the IoT receiver. In indoor environment, the received signal strength $P_R(t)$ at the receiver side is a composite signal that contains the line-of-sight (LoS) signal and multiple reflected signals from the environment that can be expressed by using the following equation:

$$P_R(t) = |H(t)|^2 \times P_t \tag{1}$$

Where P_t is the transmission power, $H(t)$ is the channel response which can be calculated by using:

$$H(t) = \sum_{k=1}^{N} a_k(t) e^{-j2\pi f \tau_k(t)} \tag{2}$$

Where $a_k(t)$ and $2\pi f \tau_k(t)$ are the magnitude and phase of the received signal from the kth path, respectively. N is the total number of paths. For the sake of clarity, let us consider the simplified example shown in Fig. 2. At time t_0, there is no human present, the RSS value $P_R(t_0)$ is the sum of P_l and $P_{r,o}$ which are the signal strength from the LoS path and the signal strength reflected from the obstacle (as shown in Fig. 2(b)). At time t_1, a person is in the radio transmission range (as shown in Fig. 2(a)), so the RSS value $P_R(t_1)$ needs to include another element $P_{r,h}$ which is received signal strength reflected from the human body. At time t_2, the person walked to a new location, so the element changes to $P'_{r,h}$. Therefore, the new RSS value $P_R(t_2) = P_l + P_{r,o} + P'_{r,h}$.

III. Technical Challenges and Our Solutions

In this section, we firstly discuss the technical challenges and then briefly introduce our solutions.

A. Challenge 1: Low Resolution of RSS

The first challenge is that the resolution of RSS is extremely low (usually 8 bits) for most of the IoT devices. On the other hand, the reflected signal strength from the human body is much weaker than the direct LoS signal strength. In order to extract the weaker signal out of the stronger signal, we need to improve the resolution of RSS. In our experiments, we observed that different human activities can be reflected in the coarse-grained RSS values if we increase the sampling rates (i.e., conducting oversampling).

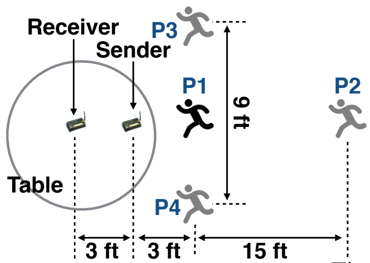
In our experiment, a ZigBee compliant MICAz sender transmits 1,000 packets per second on channel 26 to another MICAz receiver which is three feet away from the sender (shown in Fig. 3). The receiver measures the RSS values of each received packet and saves the RSS values to its on-board memory. During the packet transmissions, a person performed nine different activities (illustrated in Fig. 3). As shown in Fig. 4, we can visually identify different human activities by using oversampling. The reason is as follows:

Based on the Nyquist-Shannon sampling theorem, to characterize the signal $x(t)$ with maximum frequency component f, the sampling rate $f_s = 2f$ is sufficient. This is true in analog sampling schemes (i.e., discrete time and continuous amplitude). However, when converting the sampled analog values to digital values, each analog value is quantitized into a limited set of numbers (i.e., discrete time and discrete amplitude) which will lose some information during the quantitization stage. the SNR for quantizing noise can be calculated below [12]:

$$SNR_{dB} = 6.02n + 1.76dB \tag{3}$$

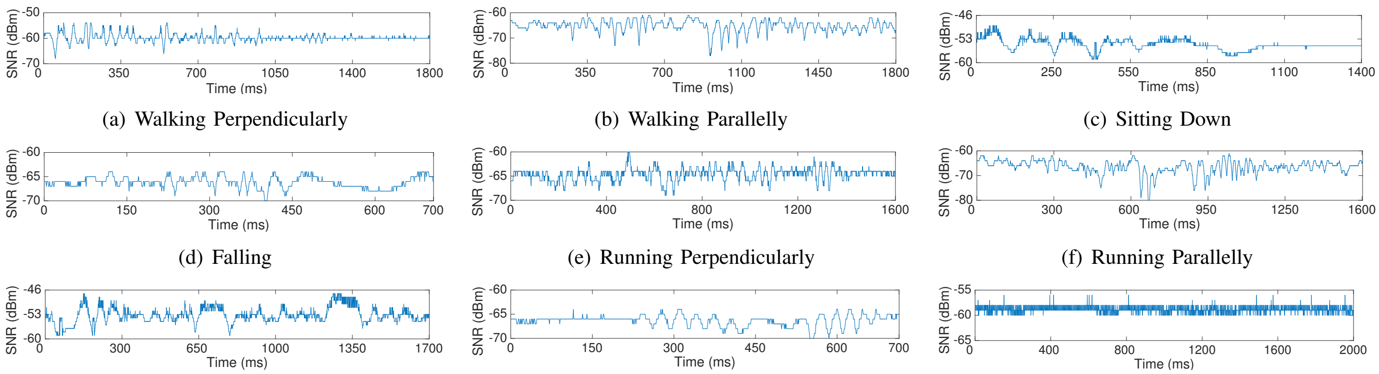
Where n is the number of quantitization bits (e.g., an 8-bit ADC has a maximum signal-to-noise ratio of $6.02 * 8 + 1.76dB = 49.78dB$). Therefore, the larger the number of quantitization bits n, the higher the SNR value. The higher the SNR value, the higher the probability to recover the original signal.

From Eq. 3, it is very hard to recognize human activities based on IoT devices' low resolution RSS, which is usually



Category	Activities	Activity Description
Daily	Walking Perpendicularly Walking Parallelly Sitting Down	Walking from $P3$ to $P4$ (perpendicularly to S and R) Walking from $P1$ to $P2$ (parallelly to S and R) sitting down at position $P1$
Accident	Falling	Falling at position $P1$
Fitness	Running Perpendicularly Running Parallelly Jumping Sit-up	Running from $P3$ to $P4$ (perpendicularly to S and R) Running from $P1$ to $P2$ (parallelly to S and R) Jumping at position $P1$ Sit-up at position $P1$
Steady	No Activity	Standing still with no activities

Fig. 3: Experiment Setup and Activity Descriptions



(a) Walking Perpendicularly (b) Walking Parallelly (c) Sitting Down

(d) Falling (e) Running Perpendicularly (f) Running Parallelly

(g) Jumping (h) Sit-up (i) No Activity

Fig. 4: An example to illustrate that we can use oversampling to identify human activities

8 bits. Therefore, we need to conduct oversampling which spreads the quantitized noise from $1 \sim f_s/2$ to $1 \sim R \times f_s/2$ (where f_s is the original sampling rate, R is the oversampling ratio, and $R \times f_s$ is the oversampling rate) because the number of quantitized bits does not change. By using the oversampled RSS as inputs to our middleware, we apply re-constructor and low-pass filter to extract the signal below frequency $f_s/2$ with lower noise. So the SNR could be expressed as Eq. 4:

$$SNR_{dB} = 6.02n + 1.76dB + 10lg\frac{R \times f_s}{2f_m} \quad (4)$$

Where f_m is the maximum frequency component in the original signal. For each $R = \{1, 4, 16, \cdots\}$, the SNR could be increased by $\{0dB, 6dB, 12dB, \cdots\}$, and the corresponding number of quantitized bits can be increased by $\{0bit, 1bits, 2bits, \cdots\}$.

B. Challenge 2: Different RSS Sampling Rates

The second challenge is that different IoT devices's radios have different RSS sampling rates. For example, ZigBee compliant MICAz node can sample RSS at 1,000 samples/second, while WiFi devices can sample RSS at 2,500 samples/second. As a generic middleware, *Harmony* needs to conduct the data pre-processing (detailed in Section IV-B) that dynamically adjusts the sampling rate of RSS input to form a normalized series for extracting the features of different human activities. Therefore, another inherited challenge is to preserve as many features of human activities as possible while reducing the number of RSS samples.

C. Challenge 3: RSS is a Coarse Mixture of Multiple Signals

Channel state information (CSI) from WiFi has been utilized for human activity recognition. This is because CSI is fine-grained information that is measured on 30 selected OFDM subcarrier groups for a received 802.11 frame. Each CSI measurement contains 30 matrices with $N_T \times N_R$ dimensions, where N_T and N_R represent the number of transmitting and receiving antennas. Moreover, each element in the matrix is 64 bits.

Different from the fine-grained CSI from WiFi, the 8-bit RSS value is a coarse mixture of multiple signals (i.e., the LoS signal and multiple reflected signals) from the *sole* antenna that works in the *sole* channel. Therefore, depending on the differences in the path lengths of the reflected signals, the multipath effect has constructive or destructive contribution to the RSS value. For example, the RSS value increases when the LoS signal has the same phase with the reflected signals, and vice verse. Therefore, it is extremely difficult to identify the human activities base on RSS values at the time domain. To address this issue, we propose to apply Short-time Fourier transform (STFT) to obtain frequency components for human activity recognition (detailed in Section IV-C).

D. Challenge 4: Need to Recognize Activities from Different People with One Training Set

The fourth challenge is how to build a robust middleware that can use one training set to recognize activities from different people. This is challenging because for the same activity, different people has different body size and shape or conducts differently which have different impact to the RSS values. Even for the same person, he/she may conduct the same activity differently over time. To address this challenge, we propose to build a Markov Model to describe the transition among different states of an activity (detailed in Section IV-D).

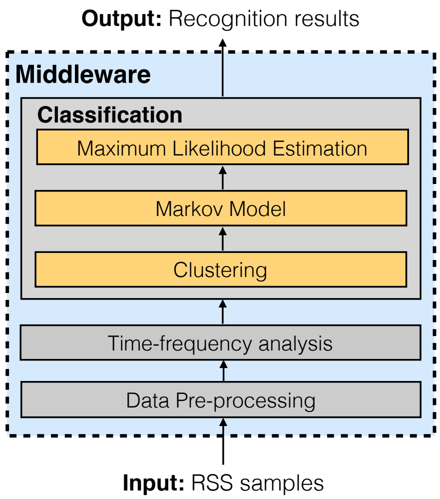


Fig. 5: Middleware Overview

E. Challenge 5: Incorporate RSS from Multiple Sources

Since IoT devices are predicted to be pervasive, it is highly possible that multiple pairs of IoT devices are conducting the packet transmission and getting the RSS readings when a human is conducting some activities. Therefore, our middleware needs to consider how to leverage RSS readings from multiple pairs of IoT sender and receivers for more accurate human activity recognition. The challenge is how to incorporate RSS from multiple sources to improve the recognition accuracy while minimizing the noise/interference introduced by different sources. To overcome this challenge, we mathematically formulate the problem as an optimization problem and provide the optimal solution (detailed in Section IV-E).

IV. Middleware Design

To address the challenges described in the previous section, we propose *Harmony*, a middleware that utilizes RSS values from IoT devices for human activity recognition. In this section, we first introduce the overview of the design, and then describe each component of *Harmony* in detail.

A. Design Overview

As shown in Fig. 5, *Harmony* contains three components:
i) Data pre-processing module, which reconstructs the RSS time series from the input into a normalized sampling stream. The main design goals are to i) improve the resolution of amplitude; and ii) reserve the low-frequency features and trends that contains the information of human activities (detailed in Section IV-B).
ii) Time-frequency analysis module, which uses sliding Short-time Fourier Transform (STFT) to transform the reconstructed and smoothed RSS time series into frequency component vectors (detailed in Section IV-C.
iii) Classification module, which classifies the frequency component vectors into k states. Based on the Markov state transition matrix, the maximum likelihood is calculated and the activities are estimated (detailed in Section IV-D).

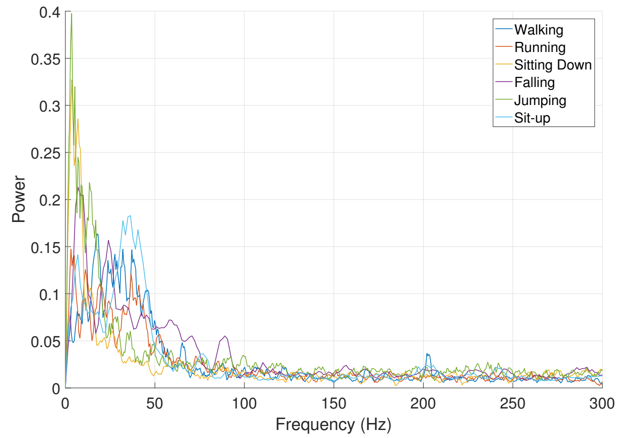


Fig. 6: Frequency Components of different Activities

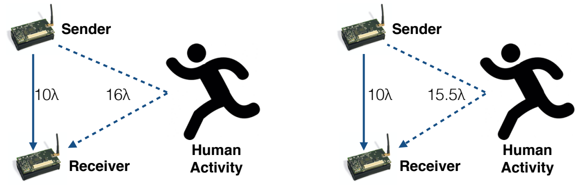
B. Data Pre-processing

The design goal of our middleware is to make it generic so that it can be applied on top of different IoT systems. However, due to various physical layers coexisted in IoT communication systems, the RSS sampling rates are completely different (e.g., up to 1,000 samples/second for ZigBee radio and 2,500 samples/second for WiFi radio [9]). Moreover, for the same IoT system, the sampling rate can change over time due to the changes of the data traffic demand. In this section, we describe how to normalize the received RSS samples to support human activity recognition.

Since different human activities introduce different frequency components in the RSS values. To identify a good RSS sampling rate, which can capture human activities, we conducted extensive experiments with 9 different activities. Fig. 6 shows the magnitudes of different frequency components for activities such as *walking, running, siting down, falling, jumping, and sit-up*. Generally, the frequency components in human activities we tested do not exceed $f(t) = 150Hz$, which is far less than the highest sampling rate of the IoT devices' radios, which is normally more than 1 kHz.

We can count the time stamps of the input RSS time series to obtain the sampling period $T_o(t)$ and convert to sampling rate $f_o(t) = 1/T_o(t)$. Then, based on the above maximum frequency component $f(t) = 150Hz$, we can get the time-varying oversampling ratio $R(t) = f_o(t)/2f(t)$. As mentioned in Section III-A, whenever $R(t)$ is increasing four times, it is equivalent to increase one bit of quantitization. With the oversampling ratio $R(t)$, we can reconstruct the RSS time series by applying the averaging factor of $R(t)$. After the reconstruction, the RSS time series decreases $R(t)$ times but the precision of each value increases $log_4 R(t)$ bits.

To decrease the input time series $R(t)$ times for reconstruction, normally, equal weight is used. For example, if the oversampling ratio $R(t) = 4$, the weight set is $\{1/4, 1/4, 1/4, 1/4\}$. Instead of applying the equal weight, we choose binomial expansion as the weight to prevent the sharp edges introduced by equal weight which can cause spurious oscillation and leakage into the smoothed output. The weight can be calculated by using the convolution of $(1/2, 1/2)^{R(t)-1}/(R(t)-1)$. For example, if the oversampling



(a) Integer Times of λ (b) Non-integer Times of λ

Fig. 7: Constructive and destructive contribution on RSS by different human position

ratio $R(t) = 4$, the weight set is $\{1/8, 3/8, 3/8, 1/8\}$.

C. Time-frequency Analysis

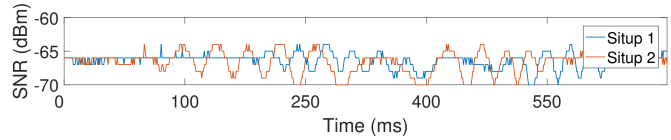
In this section, we discuss what signatures to use for characterizing human activities. We propose to perform the time-frequency analysis of human activity to obtain the frequency component vector because the magnitude of RSS depends on not only the LoS path but also the other paths that reflected signals (because of the multipath effect).

Though different activities have different patterns in the time domain (as shown in Fig. 4), it is not reliable to use these patterns as signatures for classifying activities, because the multipath effect may have constructive or destructive contribution to the magnitude of RSS depending on the differences in the path lengths of the reflected signals from human body. This means even the same people performs exactly the same activity at slightly different positions, the multipath effect will be different. Therefore, the RSS values will also be affected.

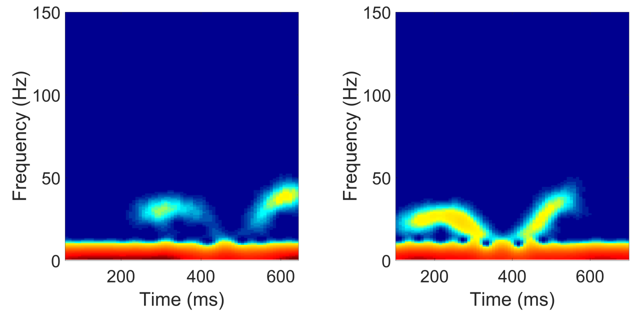
For example, in Fig. 7, we assume that the distance between sender and receiver is 10 times of wavelength (λ), if the total distance of the reflected path (i.e., sender-human-receiver) is dividable by λ (e.g., 16λ as shown in Fig. 7(a)), the RSS is strengthened. If the reflected path is not dividable by λ (e.g., 15.5λ as shown in Fig. 7(a))the RSS is weakened. Because the λ value for 2.4 GHz (i.e., the frequency band for our daily used WiFi, Bluetooth, and ZigBee devices) is around 3cm, a slight change of a human's position will lead to different readings of RSS.

Another choice is to extract the signature in frequency domain since the frequency components of a certain activity are relatively stable (as shown in Fig. 6). However, a long time Fourier transform drops so much information in time domain and the frequency components are overlapped with each other over activities. Actually, a human activity could be resolved into multiple states. In a single state, one or more main parts of human body (arm, leg, or torso) moves with different speeds which result in the magnitude change of frequency components. So a time-frequency analysis method is promising for human activity recognition. Here, we use STFT to transform the time domain series to frequency component vector $\mathbf{v_i}(\mathbf{t})$.

We take the activity *sit-up* as an example. The comparison of two *sit-up* samples in time domain are shown in Fig. 8(a). It is hard to conclude the patterns from this figure. However, in the spectrogram (Fig. 8(b)), we can clearly observe two states: i) torso curling up from 200ms to 420ms and ii) torso stretching

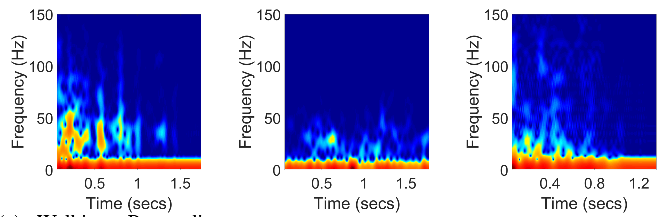


(a) Time Domain Signal of Two *sit-up* Activities

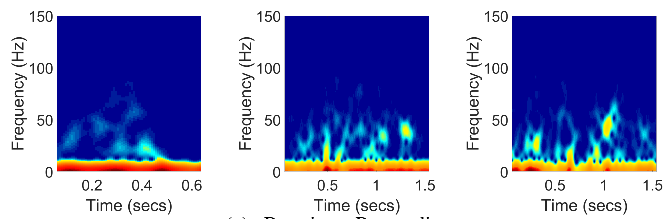


(b) Spectrogram of *sit-up* One (c) Spectrogram of *sit-up* Two

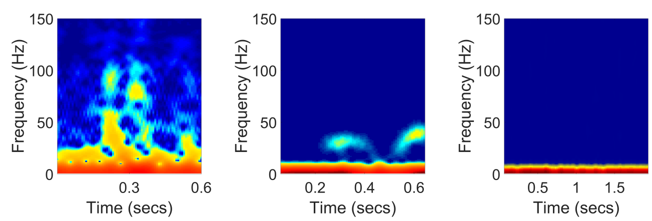
Fig. 8: Time Domain Signal and Spectrogram of Sit-up from Two Samples



(a) Walking Perpendicularly (b) Walking Parallelly (c) Sitting Down



(d) Falling (e) Running Perpendicularly (f) Running Parallelly



(g) Jumping (h) Sit-up (i) No Activity

Fig. 9: Spectrograms for Different Activities

out from 450ms to 620ms. Also, in the second sample shown in Fig. 8(c), we can observe almost the same two-state *sit-up* patterns. Therefore, the frequency component vectors $\mathbf{v_i}(\mathbf{t})$ are stable within the same human activity.

We also plot the spectrogram of one sample for all nine activities: *no activity, walking perpendicularly/parallelly, running perpendicularly/parallelly, sitting down, falling, jumping and sit-up* in Fig. 9. The no activity spectrogram (Fig. 9(i)) shows that the RSS stays stable during the 2 seconds period. The composite activities such as walking and running in

Fig. 9(a), 9(b), 9(e), and 9(f) show periodical patterns. The single-action activities such as sitting down, falling, jumping, and sit-up in Fig. 9(c), 9(d), 9(g), and 9(h), respectively, show different clear multi-state patterns. Thus, the frequency component vector $\mathbf{v_i(t)}$ can be applied to classify the human activities.

D. Classification

With extracted frequency component vectors of different activities, we present how to recognize the activities of different people in this section. Specifically, we propose to apply k-means (because its simplicity and relative good performance) to classify the state of frequency component vector at each time t and Markov model to describe the transitions among different states of an activity. Finally, we use maximum likelihood estimation to recognize the activities based on the transition matrix of learned Markov model.

Before classification, we first reconstruct and filter the RSS time series as described in Section IV-B; and calculate the frequency component vectors by STFT as described in Section IV-C. At each time t, $\mathbf{v}_i(t)$ represents the frequency component vector. The frequency component vectors of a sample i can be represented as a matrix $\mathbf{V}_i = [\mathbf{v}_i(1), \cdots, \mathbf{v}_i(T)] \in \mathbb{Z}^{F\times T}$, where F is maximum frequency and T is the length of the sample. Each sample i is labelled as activity θ_i.

The classification process is split into training stage and testing stage and dataset is also split into training dataset $\mathbf{D_r}$ and testing dataset $\mathbf{D_t}$. The training stage can be executed as follows:

i) **Clustering**. k-means is applied to cluster frequency component vectors of all the samples in the training dataset into k clusters and each frequency component vector $\mathbf{v}_i(t)$ is clustered into one of the k clusters, which is represented as $x_i(t)$.

ii) **Transition Matrix Training**. The Markov state transition matrix $\mathbf{M}_i$ of each sample i is trained based on the clustered state transitions from $x_i(t)$ to $x_i(t+1)$ for $1 \leq t \leq T$ with all the samples labelled as activity θ_i in the training dataset.

The k cluster centroids and trained Markov model will be applied for classification for testing dataset. In testing stage, the classification process of each sample in testing dataset can be executed as follows:

i) **Cluster Labeling**. Based on the trained k cluster centroids, the frequency component vector $\mathbf{v}_i(t)$ over time of the sample is labelled based on the minimum Euclidean distance between vector $\mathbf{v}_i(t)$ and k cluster centroids as $x_i(t)$, which is one of the k states.

ii) **Likelihood Probability Calculation**. The log-likelihood probabilities of each sample under the trained transition probability matrice of all activities are calculated. Specifically, the log-likelihood probability of sample i under transition matrix θ_j can be calculated as follows:

$$ln\mathcal{L}(\mathbf{M_j}; x_i(1), ..., x_i(T)) = \sum_{i=1}^{T-1} lnP(x_i(t), x_i(t+1)|\mathbf{M}_j) \tag{5}$$

where $P(x_i(t), x_i(t+1)|\mathbf{M}_j)$ is the transition probability from state $x_i(t)$ to $x_i(t+1)$ under the transition probability matrix $\mathbf{M}_j$.

iii) **Maximum Likelihood Estimation**. With the likelihood probability of the activity i under different transition matrix, the sample i can be classified as activity θ_i as:

$$\hat{\theta}_i = j = arg \max_{j\in\mathbf{A}} ln\mathcal{L}(\mathbf{M}_j; x_i(1), ..., x_i(T)) \tag{6}$$

where $\hat{\theta}_i$ is the estimated activity of the input RSS time series sample impacted by a human activity j and $\mathbf{A}$ is the set of all the possible activities.

E. Multi Pairs Harmony

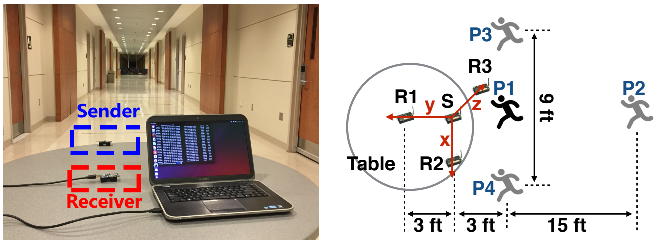
In previous sections, we introduced the design of single-source RSS middleware with the assumption that there is only one pair of nodes in the environment. However, in reality, there may exist multiple pairs of nodes deployed in the environment. In this section, we discuss how we can utilize the time series RSS from different pairs of nodes to further improve the human activity recognition accuracy.

Suppose there are n pairs of nodes in the environment and the time series RSS from n nodes are collected. Before classification, we first conduct data-preprocessing and time-frequency analysis for n time series RSS of each activity i to get the n frequency component vectors $\{\mathbf{v}_i^{(1)}(t), \cdots, \mathbf{v}_i^{(n)}(t)\}$ over time. With the frequency component vectors, one naive solution is to append all the vectors to form a new $\mathbf{v}'_i(t)$ and then apply the same classification process. However, the length of vector $\mathbf{v}'_i(t)$ will be n times compared to $\mathbf{v}_i(t)$, which makes the clustering process much harder to converge and may even get worse clustering performance. Meanwhile, the different activities may have different impact on the time series RSS for n different nodes. Therefore, we extend maximum likelihood estimation with one set of frequency component vectors over time to maximum likelihood estimation with n sets of frequency component vectors over time. The maximum likelihood function can be rewritten as:

$$\hat{\theta}_i = j = arg \max_{j\in\mathbf{A}} \sum_{k=1}^{n} ln\mathcal{L}(\mathbf{M}_j^{(k)}; x_i^{(k)}(1), ..., x_i^{(k)}(T)) \cdot w_k \tag{7}$$

where $\{x_i^{(k)}(1), \cdots, x_i^{(k)}(T)\}$ are the clustered states with the time series RSS from node k. $\mathbf{M}_j^{(k)}$ is the Markov state transition matrix for activity j based on the time series RSS from node k. w_k is the weight for likelihood probability of frequency component vectors with different time series RSS. w_k is trained in training stage by solving a minimization problem as:

$$\begin{aligned} \min \quad & \sum_{i\in\mathbf{D_r}} \frac{\mathbb{1}_{\theta_i}(\hat{\theta}_i)}{|\mathbf{D_r}|} \\ \textit{s.t.} \quad & 0 \leq w_k \leq 1 \;\; \forall k \in [1, n] \quad (a) \\ & \sum_{k=1}^{n} w_k = 1 \qquad\qquad (b) \end{aligned}$$



(a) Real World Deployment (b) Three Pairs Setting

Fig. 10: Experimental Setup

where $\mathbb{1}()$ is the indicator function. If $\hat{\theta}_i = \theta_i$, $\mathbb{1}_{\theta_i}(\hat{\theta}_i) = 1$; otherwise $\mathbb{1}_{\theta_i}(\hat{\theta}_i) = 0$. Because the two constraints and the object function are all linear function, therefore, the problem can be solved with linear programming. With the trained w_k, classification can be conducted with maximum likelihood estimation based on Equation (7).

V. Evaluation Setup

In this section, we describe how we implemented our middleware *Harmony* in real world.

A. Implementation

We evaluate our middleware on top of the low cost ZigBee wireless sensor network. The evaluation platform is consisted of two parts: ZigBee compliant MICAz nodes and a laptop. We collect RSS data from MICAz nodes which are fully compatible with IEEE 802.15.4. We use two MICAz nodes to form a pair in order to record the RSS values during continuously transmitting. Within the pair, one node works as a sender by sending packets continuously. The payload only contained packet ID in order to i) increase packet sending rate and ii) count packet loss rate. The other node works as a receiver who accepts packets from the sender. For each received packet, the receiver measures the RSS and forwards the packet to the laptop through serial port. The pair can be set to work on ZigBee channels 15, 20, or 26 in order to evaluate the impact of different ZigBee channels. The data processing and analyzing approaches (as described in Section IV), which takes the RSS data as input and human activities recognition as output, are deployed on the laptop.

B. Experiment Setup

We evaluated our middleware in a hallway on the third floor of an engineering building which is covered by multiple 2.4 GHz WiFi Access Points (APs) working on different WiFi channels (e.g., channel 1,6, and 11). The hallway is measured as 14 feet in width by 150 feet in length. We put a round table (4 feet of diameter and 2.3 feet in height) in the middle of the hallway (as shown in Figure 10(a)). The data collection client and all the analysis programs are written in Python running on the laptop. Totally, we collected 6360 samples (as shown in Table I) of 9 human activities from 6 volunteers. The volunteers include 5 male and 1 female. To obtain the ground truth, we used a camera to record the activities of volunteers during all the experiments.

Category	Activities	Number of Samples	Notation
Daily	Walking Perpendicularly	1080	WP1
	Walking Parallelly	1080	WP2
	Sitting Down	660	SD
Accident	Falling	540	F
Fitness	Running Perpendicularly	720	RP1
	Running Parallelly	720	RP2
	Jumping	420	J
	Sit-up	540	SU
Steady	No Activity	600	NA

TABLE I: Activities Dataset

We evaluated two settings as follows: i) the one pair (deployment is the same as shown in Fig. 3) and ii) the three pairs (deployment shown in Fig. 10(b)). We adopt recognition accuracy as the evaluation metric and evaluate the middleware using 5-fold cross validation.

VI. Experimental Results

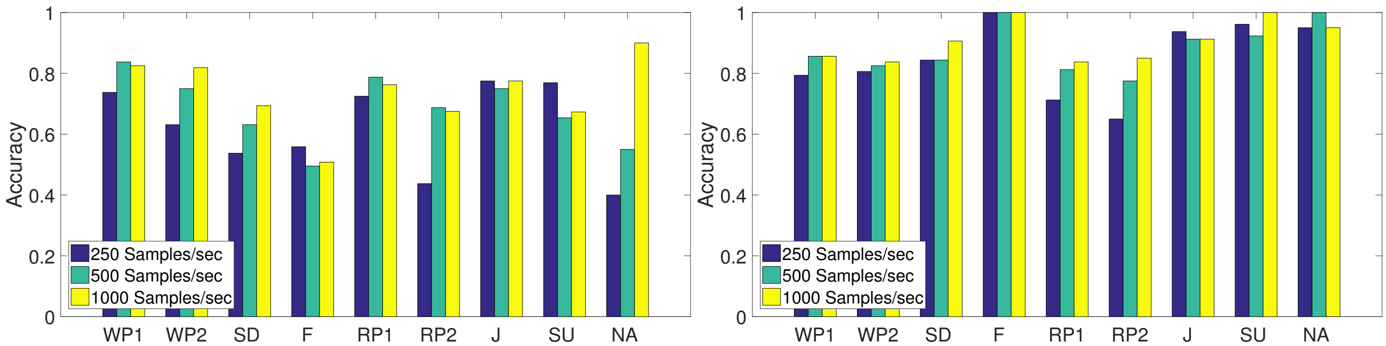
In this section, we show the results from our comprehensive experiments which explore the factors (including RSS sampling rate, ZigBee channel, and number of clusters) that have impact on the human activity recognition accuracy for both two settings stated in the previous section. Then, we study the impact of human activity on Packet Reception Ratio (PRR). Finally, we compare our result with the result from state-of-the-art CSI approach.

A. Impact of RSS Sampling Rate

As described in Section IV-B, our middleware has to work with different wireless radio systems under different sampling rates. Thus we evaluated three different sampling rates 250 (without reconstructing), 500, and 1,000 samples/second on MICAz nodes. Fig. 11 presents the detailed accuracy under different sampling rates with the parameter (channel = 20 and number of cluster = 17). The average accuracy is 61%, 68%, and 73% of sampling rates 250, 500, and 1,000 samples/second, respectively, in one pair setting. The average accuracy is 85%, 88%, and 90% of sampling rates 250, 500, and 1,000 samples/second, respectively, in three pairs setting. For one pair setting, the accuracy drops sharply when sampling rate decreases since the details of high frequency components are dropped. In three pairs setting, the accuracy does not change so much which may because the information on different axes can compensate each other.

B. Impact of Channel

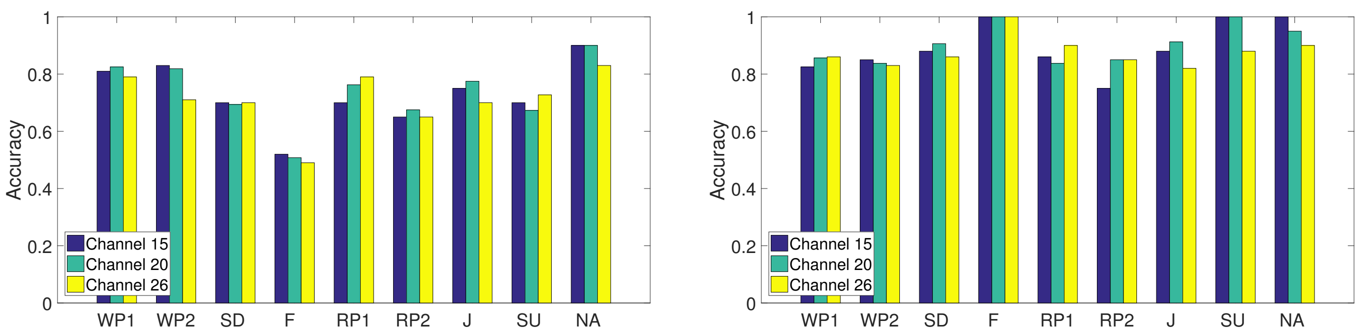
The detailed recognition accuracy with the parameters (sampling rate = 1,000 samples/second and number of cluster = 17) across three ZigBee channels (15, 20, and 26) are shown in Fig. 12. The average accuracy over all activities are 72.9%, 73.7%, and 71% on channel 15, 20, 26 respectively, for one pair setting. The average accuracy over all activities are 89.4%, 90%, and 87.8% on channel 15, 20, 26 respectively, for three pair setting. Since different channels have different center frequencies which correspond to different wavelengths, the multipath effect varies across the channels. However, as we discussed in Section IV-C, the frequency component vectors



(a) One Pair

(b) Three Pairs

Fig. 11: **Accuracy under Different Sampling Rates:** We can observe the trend in both of the two settings that the accuracy increases along with the sampling rate increases. The average accuracy is 61%, 68%, and 73% of sampling rate 250, 500, and 1,000 samples/second, respectively, for one pair. The average accuracy is 85%, 88%, and 90% of sampling rate 250, 500, and 1,000 samples/second, respectively, for three pairs.



(a) One Pair

(b) Three Pairs

Fig. 12: **Accuracy under Different Channels:** The average accuracy over all activities are 72.9%, 73.7%, and 71% on channel 15, 20, 26 respectively, for one pair setting. The average accuracy over all activities are 89.4%, 90%, and 87.8% on channel 15, 20, 26 respectively, for three pairs setting. The differences are not significant across channels.

can well characterize the human activity, so the differences within the results are not significant.

C. Impact of Number of Clusters

Fig. 13 shows the accuracy with the parameters (channel = 20 and sampling rate = 1,000 samples/second) under different number of clusters. In Fig. 13, it does not show an obvious trend since different activities have different time duration (e.g., walking is much longer than falling). So the states' numbers of Markov Model are also different and the results do not have a certain trend. Overall, the accuracy for both one pair setting and three pair setting reaches the highest 73% and 90%, respectively, when the number of clusters is 17.

D. Confusion Matrix

Fig. 14 shows the synthesized confusion matrix of one pair setting with the parameters (sampling rate = 1,000 samples/second, channel = 20, dimension = x axis, and number of clusters = 17). Our middleware has a overall accuracy 73.68% to detect the different activities.

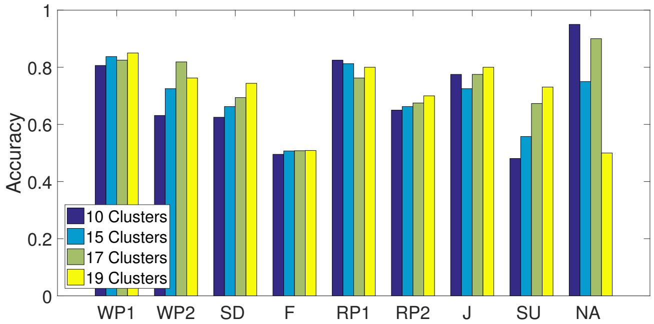
Fig. 15 shows the result in three pairs setting with the parameters (sampling rate = 1,000 samples/second, channel = 20, and number of cluster = 17). Our middleware has a high accuracy up to 90%. We note that the four activities with lowest accuracy are composite activities, such as *walking perpendicularly/parallelly* and *running perpendicularly/parallelly*. It is reasonable that the four periodical activities have similar body movements which leads similar states transition matrix in the Markov Model.

We note that the recognition accuracy of activities with lowest accuracy (*falling* and *sit-up*) in one pair setting increases dramatically in three pairs setting because the information on different dimensions can compensate each other.

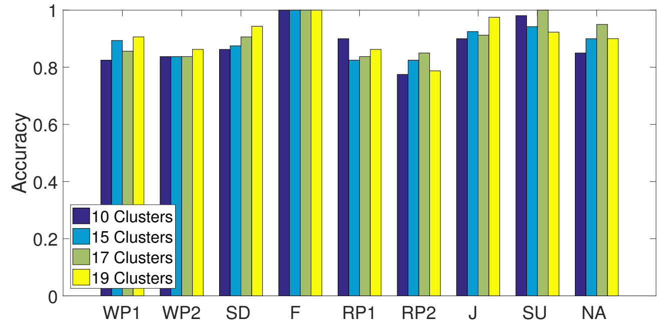
Since the similarity of *walking perpendicularly/parallelly* and *running perpendicularly/parallelly* is high, the accuracy does not improve too much from one pair setting to three pairs setting.

E. Packet Reception Ratio under Human Activities

We evaluate the packet reception ratio (PRR) impacted by different human activities in this section. The PRR is the percentage of packets that have been received out of all sent packets. As shown in Fig. 16, the minimum PRR is 99.38% which occurs during the activity *walking parallelly* (labeled as "WP2") and the PRR is 99.8% under the same setting but without human activity (labeled as "NA"). There is no obvious difference between human activity presence and



(a) One Pair



(b) Three Pairs

Fig. 13: **Accuracy under Different Number of Clusters:** We note that the overall accuracy for both of the one pair setting and three pair setting reaches the highest 73% and 90%, respectively, when the number of clusters is 17.

Recognized Activity

Ground Truth	WP1	WP2	RP1	RP2	SD	F	J	SU	NA
WP1	**0.819**	0.144	0.006	0.025	0	0.006	0	0	0
WP2	0.15	**0.825**	0.006	0.013	0	0	0.006	0	0
RP1	0.088	0.013	**0.675**	0.188	0.013	0.025	0	0	0
RP2	0.027	0.1	0.1	**0.763**	0	0	0	0	0
SD	0.013	0	0.038	0	**0.694**	0.069	0.125	0.063	0
F	0.098	0	0.132	0.064	0.149	**0.508**	0.05	0	0
J	0.037	0.038	0	0	0.113	0.044	**0.775**	0.013	0
SU	0	0	0	0	0.019	0.038	0	**0.673**	0.269
NA	0	0	0	0	0	0	0	0.1	**0.9**

Fig. 14: Confusion Matrix of One Pair

Recognized Activity

Ground Truth	WP1	WP2	RP1	RP2	SD	F	J	SU	NA
WP1	**0.838**	0.15	0.012	0	0	0	0	0	0
WP2	0.138	**0.856**	0	0.006	0	0	0	0	0
RP1	0.012	0.013	**0.85**	0.125	0	0	0	0	0
RP2	0.012	0.06	0.088	**0.838**	0	0	0	0	0
SD	0.006	0	0	0	**0.907**	0	0.081	0.006	0
F	0	0	0	0	0	**1**	0	0	0
J	0.025	0	0	0	0.062	0	**0.913**	0	0
SU	0	0	0	0	0	0	0	**1**	0
NA	0	0	0	0	0.05	0	0	0	**0.95**

Fig. 15: Confusion Matrix of Three Pairs

absence because our middleware can recognize human activity while keeping the sender and receiver LoS which means the people does not across the connection of the node pair, so the human movement does not hurt the PRR.

F. Comparison with State-of-the-art

As shown in Fig. 17, we compare our results with the fine-grained CSI approach from [9]. Even the RSS contains less information than the CSI, by using our middleware, we can achieve similar human activity recognition accuracy with the fine-grained CSI approach.

VII. Related Work

Human activity recognition and monitoring is an active research area, with many people investigating new methods for analyzing and identifying human activity through different

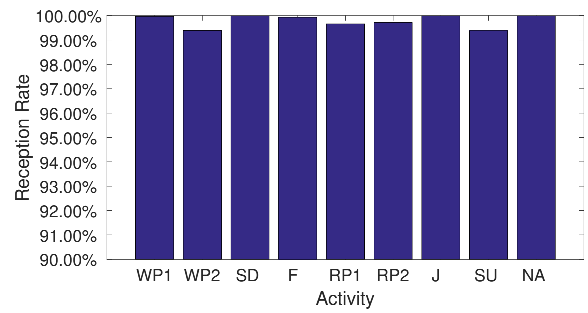


Fig. 16: **Packet Reception Ratio Under Different Activities:** The human activities do not hurt the PRR since all the PRRs under different human activities are higher than 99.38%.

Sampling Rate (S/s)	Harmony	CARM
800	90%	94.8%
400	88%	87%

Fig. 17: Compared with CARM [9], our *Harmony* can achieve similar performance with coarse-grained RSS

methods [13], [14], [15]. These approaches can be divided into two groups: using wearable devices and using passive devices. The first group requires the subjects carry a device on their body all the time, which typically allows for a higher accuracy in the tracking data, at the cost of user inconvenience and increased cost [16], [17]. Passive devices, however, allow for tracking without any user interaction, making them overall cheaper, but at the cost of the position or activity data being less accurate [18], [19]. Our work focuses on the passive approach, as the reduced interference with daily tasks and lower cost make it an appealing method to work with. Of the methods being currently explored, the most related work are those based on CSI, RSS, and Doppler Radar.

RSS-based Approaches. Although RSS has been well studied [22], [23], [24], [25] and is known to be extremely coarse-grained, little work has been conducted to exploit RSS for human activity recognition and monitoring. The most related work is [20], [21] that utilized the RSS for respiration rate monitoring. In the above work, the human object is stable, therefore, it is relatively easy to detect the human body movement. Different from this work, our work investigate how to use RSS to recognize human activities which is more challenging. Because the human movement will create multipath effects, which will affect the RSS values in different ways. Therefore, the human activities are more challenging to be recognized.

CSI-based Approaches. An alternative to RSS, CSI provides more detailed information. One of the benefits to using CSI is the ability to differentiate between sudden changes and normal link fluctuation, as the CSI model provides more information than just signal strength, however this also leads to an increased overhead requirement to filter and process the required information out [9], [2], [27], [28], [26]. Moreover, CSI is provided by very limited devices, making it extremely difficult to be widely adopted [26]. Instead of using CSI, our work leverage the exiting and widely available IoT devices' RSS information for human gesture recognition. Therefore, our work has the potential to be widely adopted due to the exponentially increasing number of IoT devices. Moreover, evaluation results show that our approach can achieve the similar accuracy as the latest CSI-based approach [9].

Doppler Radar-based Approaches. As dedicated purpose devices, radar based systems [29] provide much more precise measurements, a much higher sample rate than systems relying on existing hardware [2], [3]. The use of ultra-wideband signals offers other benefits as well, such as increased noise tolerance, and ability to work alongside of preexisting equipment without causing additional problems. Some systems utilizing this method are even able to track gestures and location from a NLOS (Non line of sight) location, allowing the device to be placed in a far larger variety of locations [4], [5], [6]. However, these approaches require specialized hardware. Therefore, these approaches are relatively hard to be widely adopted. Different from these approaches, our approach can use existing IoT devices' radios (e.g., WiFi, ZigBee, and Bluetooth), which are widely available.

VIII. CONCLUSION

With the exponentially increasing number of IoT devices, in this paper, we take the first attempt to investigate how to leverage the extremely coarse-grained RSS values from existing IoT devices for human activity recognition. Specifically, we addressed five technical challenges (detailed in Section III). We have conducted extensive evaluation. Results show that our middleware is generic, robust, and can achieve similar activity accuracy as the CSI-based approach (reported in [9]). *Harmony* can be deployed on top of existing wireless networks or wireless sensor networks to provide the human activity recognition services for different applications (e.g., fall detection and fitness).

ACKNOWLEDGMENT

This project is supported by NSF grants CNS-1503590 and CNS-1539047.

REFERENCES

[1] "http://www.cdc.gov/homeandrecreationalsafety/falls/adultfalls.html."

[2] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *MobiCom*, 2013.

[3] B. Lyonnet, C. Ioana, and M. G. Amin, "Human gait classification using microdoppler time-frequency signal representations," in *IEEE Radar Conference*, 2010.

[4] H.-l. Chang, J.-b. Tian, T.-T. Lai, H.-H. Chu, and P. Huang, "Spinning beacons for precise indoor localization," in *SenSys*, 2008.

[5] P. V. Dorp and F. C. A. Groen, "Feature-based human motion parameter estimation with radar," *IET Radar, Sonar Navigation*, 2008.

[6] C. Li, J. Ling, J. Li, and J. Lin, "Accurate doppler radar noncontact vital sign detection using the relax algorithm," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 3, pp. 687–695, 2010.

[7] A. T. Mariakakis, S. Sen, J. Lee, and K.-H. Kim, "Sail: Single access point-based indoor localization," in *Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '14. New York, NY, USA: ACM, 2014, pp. 315–328.

[8] L. Sun, S. Sen, D. Koutsonikolas, and K.-H. Kim, "Widraw: Enabling hands-free drawing in the air on commodity wifi devices," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '15. New York, NY, USA: ACM, 2015, pp. 77–89.

[9] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of wifi signal based human activity recognition," in *MobiCom '15*, 2015.

[10] "http://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/."

[11] 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.

[12] K. C. Pohlmann, *Principles of Digital Audio*, 4th ed. McGraw-Hill Professional, 2000.

[13] F. Adib, Z. Kabelac, D. Katabi, and R. C. Miller, "3d tracking via body radio reflections," in *NSDI*, 2014.

[14] Z. Zhou, Z. Yang, C. Wu, L. Shangguan, and Y. Liu, "Towards omnidirectional passive human detection," in *INFOCOM*, 2013.

[15] S. Sigg, S. Shi, and Y. Ji, "Rf-based device-free recognition of simultaneously conducted activities," in *UbiComp*, 2013.

[16] M. Azizyan, I. Constandache, and R. Roy Choudhury, "Surroundsense: Mobile phone localization via ambience fingerprinting," in *MobiCom*, 2009.

[17] H. Liu, Y. Gan, J. Yang, S. Sidhom, Y. Wang, Y. Chen, and F. Ye, "Push the limit of wifi based localization for smartphones," in *MobiCom*, 2012.

[18] N. Patwari, J. Wilson, S. Ananthanarayanan, S. K. Kasera, and D. R. Westenskow, "Monitoring breathing via signal strength in wireless networks," *IEEE Transactions on Mobile Computing*, vol. 13, no. 8, pp. 1774–1786, 2014.

[19] P. Bahl and V. N. Padmanabhan, "Radar: An in-building rf-based user location and tracking system," in *INFOCOM*, 2000.

[20] H. Abdelnasser, M. Youssef, and K. A. Harras, "Wigest: A ubiquitous wifi-based gesture recognition system," *CoRR*, vol. abs/1501.04301, 2015.

[21] J. Yang and Y. Chen, "Indoor localization using improved rss-based lateration methods," in *GLOBECOM*, 2009.

[22] O. J. Kaltiokallio, H. Yigitler, R. Jäntti, and N. Patwari, "Non-invasive respiration rate monitoring using a single cots tx-rx pair," in *IPSN*, 2014.

[23] 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*, vol. 13, no. 4, pp. 907–920, 2014.

[24] 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 *MoMM*, 2013.

[25] S. Sigg, U. Blanke, and G. Trster, "The telepathic phone: Frictionless activity recognition from wifi-rssi," in *PerCom*, 2014.

[26] C. Han, K. Wu, Y. Wang, and L. M. Ni, "Wifall: Device-free fall detection by wireless networks," in *INFOCOM*, 2014.

[27] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, and H. Liu, "E-eyes: Device-free location-oriented activity identification using fine-grained wifi signatures," in *MobiCom*, 2014.

[28] A. E. Kosba, A. Saeed, and M. Youssef, "Rasid: A robust wlan device-free passive motion detection system," in *PerCom*, 2012.

[29] N. Patwari, L. Brewer, Q. Tate, O. Kaltiokallio, and M. Bocca, "Breathfinding: A wireless network that monitors and locates breathing in a home," *J-STSP*, vol. 8, no. 1, pp. 30–42, 2014.