Channel Selective Activity Recognition with WiFi: A Deep Learning Approach Exploring Wideband Information

Fangxin Wang, *Student Member, IEEE*, Wei Gong, *Member, IEEE*, Jiangchuan Liu, *Fellow, IEEE*, and Kui Wu, *Senior Member, IEEE*

Abstract—WiFi-based human activity recognition explores the correlations between body movement and the reflected WiFi signals to classify different activities. State-of-the-art solutions mostly work on a single WiFi channel and hence are quite sensitive to the quality of a particular channel. Co-channel interference in an indoor environment can seriously undermine the recognition accuracy. In this paper, we for the first time explore wideband WiFi information with advanced deep learning toward more accurate and robust activity recognition. We present a practical Channel Selective Activity Recognition system (CSAR) with Commercial Off-The-Shelf (COTS) WiFi devices. The key innovation is to actively select available WiFi channels with good quality and seamlessly hop among adjacent channels to form an *extended channel*. The wider bandwidth with more subcarriers offers stable information with a higher resolution for feature extraction. Conventional classification tools, e.g., hidden Markov model and k-nearest neighbors, however, are not only sensitive to feature distortion but also not smart enough to explore the time-scale correlations from the extracted spectrogram. We accordingly explore advanced deep learning tools for this application context. We demonstrate an integration of channel selection and long short term memory network (LSTM), which seamlessly combine the richer time and frequency features for activity recognition. We have implemented a CSAR prototype using Intel 5300 WiFi cards. Our real-world experiments show that CSAR achieves a stable recognition accuracy around 95 percent even in crowded wireless environments (compared to 80 percent with state-of-the-art solutions that highly depend on the quality of the working channel). We have also examined the impact of environments and persons, and the results reaffirm its robustness.

Index Terms—Human activity recognition, deep learning, LSTM, channel hopping

1 Introduction

As an essential service of Internet of Things (IoT), human activity recognition has attracted significant attention in the past decade due to its great value in such applications as fall detection, health diagnosis and smart home. A rich set of sensing systems utilize cameras [1], wearable sensors [2], [3] and RFIDs [4], [5] for activity recognition. Recently, sensing using commodity WiFi devices has seen its success in this context as well [6], [7], [8], [9], [10], [11]; given the ubiquity, low cost and touchless operations of WiFi, it opens a promising venue toward pervasive and cost-effective activity recognition.

The basic idea for activity recognition using WiFi is that a human body movement will affect the surrounding WiFi signals, and the reflected WiFi signals by a particular activity exhibit distinct characteristics. Through analyzing the signal patterns, particularly the dynamics of the fine-grained channel state information (CSI), different activities can be distinguished. State-of-the-art works have demonstrated quite good recognition accuracy in experiments with clean WiFi channels [6], [7], [8], [9], [10], [11]. The real world WiFi channels however are far from being clean. Today, such indoor spaces as home, workplace and shopping mall are usually filled up with background wireless signals, including those from crowded public WiFi access points (APs), not to mention private APs. Nearby WiFi signals in the same channel can conflict with each other [12]. Existing systems mostly use a fixed WiFi channel for activity sensing and CSI collection, and as such their performance are sensitive to co-channel interference, which greatly affects the received signal quality and distorts the extracted features for recognition. Typical classification models used in the existing systems, such as hidden Markov model (HMM) [9] and k-nearest neighbors (kNN) [11], can be dramatically affected by such distortions when classifying the activities.

In this paper, we for the first time explore wideband WiFi information with advanced deep learning toward more accurate and robust activity recognition. We present a practical Channel Selective Activity Recognition system (CSAR)

- *F. Wang and J. Liu are with the School of Computing Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada. E-mail: fangxinw@sfu.ca, jcliu@cs.sfu.ca.*
- *W. Gong is with the School of Computer Science and Technology, University of Science and Technology of China, Hefei Shi, Anhui Sheng 230000, China and with the School of Computing Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada. E-mail: weigong@ustc.edu.cn.*
- *K. Wu is with the Department of Computer Science, University of Victoria, BC V8P 5C2, Canada. E-mail: wkui@uvic.ca.*

Manuscript received 15 Oct. 2017; revised 26 Feb. 2018; accepted 2 Apr. 2018. Date of publication 10 Apr. 2018; date of current version 5 Mar. 2020. (Corresponding author: Jiangchuan Liu.) Recommended for acceptance by W. Zhang. Digital Object Identifier no. 10.1109/TNSE.2018.2825144

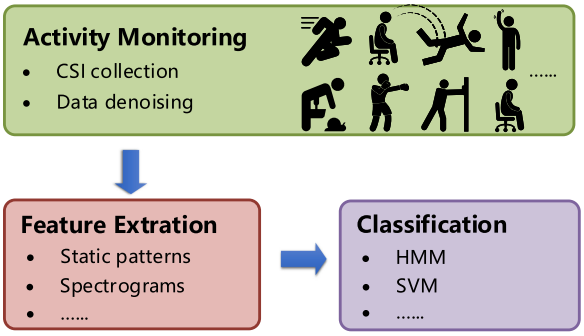


Fig. 1. The typical framework of WiFi-based activity recognition systems.

with Commercial Off-The-Shelf (COTS) WiFi devices. The key innovation is to constantly evaluate the quality of available WiFi channels and seamlessly hop among adjacent channels to form an *extended channel*. The extended channel, which has a wider bandwidth with more subcarriers, offers stable information with a higher resolution for feature extraction. Conventional classification tools, e.g., HMM and kNN, however, are not only sensitive to distortion but also not smart enough to explore the time-scale correlations from the extracted spectrogram. We explore advanced deep learning tools, in particular, long short term memory network (LSTM) [13], for this application context. We demonstrate an integration of channel selection and LSTM, which seamlessly combine the richer time and frequency features for activity recognition.

We have implemented CSAR using COTS WiFi devices (Intel 5300 WiFi cards, in particular). With the CSI tools developed by Halperin et al. [14], we have collected 3,350 real-world activity traces from volunteers in different environments, which, after feature extraction, have been used to train the learning engine. Our experiments have demonstrated that CSAR achieves a stable recognition accuracy around 95 percent even in crowded wireless environments. With the same setting, the state-of-the-art solution, which highly depends on the quality of the working channel, has an accuracy of 80 percent only. We have closely examined the performance of different system modules. The results suggest that both channel selection and deep learning improve the recognition accuracy, and their combination enables the largest performance gain. We have also examined the impact of environments and persons, and the results reaffirm the robustness of CSAR.

The rest of this paper is organized as follows. Section 2 introduces the background of activity recognition and the overview of our system. Section 3 describes the channel selection mechanism, including channel quality evaluation, channel hopping and channel combination. We introduce the data denoising and activity detection in Section 4, followed by the LSTM-based recognition method in Section 5. The implementation and evaluation are presented in Section 6. Finally, we list the related works in Section 7 and conclude the paper in Section 8.

2 Background and System Overview

As shown in Fig. 1, a typical WiFi-based activity recognition system consists of the following cascaded modules [9], [11]:

- *Channel status monitoring and filtering*. Given a pair of WiFi sender and receiver, a human body in between becomes as a reflector; his/her body movement will affect the WiFi signals, which can be observed by the pair. The first step of an activity recognition system is thus to monitor the raw signals and denoise them to reveal the changes caused by human activities. A low-pass filter and/or principal component analysis (PCA) is commonly used.
- *Feature Extraction*. Having the denoised signal, features related to human activities are to be extracted. This is typically done by such time-frequency transforms as short-time Fourier transform (STFT) or discrete wavelet transform (DWT), which convert the signal to a time-frequency dimension spectrogram. The wave shapes and patterns of different activities have also been used [8], [10].
- *Training and Recognition*. An initial training will be performed using the extracted features and the set of associated activities. Afterward, the newly extracted features will be used to identify new activities for recognition. Such machine learning tools for classification as k-nearest neighbors, support vector machine (SVM) and hidden Markov machine have been used in this context.

State-of-the-art activity recognition systems using the flow above have achieved promising results; in the ideal experimental environment, more than 90 percent accuracy can be expected [11], [15]. In a real-world environment, however, the accuracy remains to be improved for practical use. We now try to identify the issues in each and every step, which also motivate our design of CSAR.

2.1 Signal Reflection Model for Activity Recognition

Typically, the *Channel State Information* is used to characterize the *Channel Frequency Response* (CFR) of a communication link [14], so for most WiFi-based activity recognition systems. CSI combines the information of the time of delay, amplitude attenuation, and phase shift for a signal propagated from a sender to a receiver [8]. Given the frequency domain representation $X(f,t)$ at the sender and $Y(f,t)$ at the receiver with frequency f and time t, we have

$$Y(f,t) = X(f,t) \times H(f,t), \tag{1}$$

where $H(f,t) = Y(f,t)/X(f,t)$ is the CFR of the corresponding wireless channel. Note that channel noise is implicitly included in it.

Modern WiFi devices with Multi-Input Multi-Output (MIMO) have multiple transmitting and receiving antennas. Each transmitting-receiving antenna pair (Tx-Rx) can transmit a signal that consists of multiple subcarriers based on the OFDM channelization. Most off-the-shelf WiFi cards, however, report only a subset of the subcarriers, e.g., 30 by Intel 5300 [14]. As in previous studies [9], we define the time-series CSI values from a given OFDM subcarrier of one antenna pair as a *CSI stream*. Let M_T and M_R be the number of transmitting and receiving antennas, respectively, and assume that each WiFi card reports N_{sc} subcarriers, there are a total of $N_{sc} * M_T * M_R$ CSI streams. By

grouping the CSI streams of all the reported subcarriers of a transceiver pair, we have a *CSI stream group*.

In an indoor environment, the paths that the received signal traverses include both *static paths* and *dynamic paths*. The static paths include the line of sight (LoS) path and the reflected paths by such static objects as walls, the ceiling and furnitures, whose lengths do not change in the presence of human activities; The dynamic paths include those reflected by the human action, whose lengths change with the human movement. $H(f,t)$ can be then represented as

$$H(f,t) = e^{-j2\pi\Delta ft}(H_s(f) + \sum_{k=1}^{P_d} a_k(f,t)e^{-j2\pi f\tau_k(t)}), \quad (2)$$

where $H_s(f)$ is the aggregate CFR of the static paths, P_d is the number of dynamic paths, and $a_k(f,t)$ and $\tau_k(t)$ are the complex channel attenuation and time of flight for path k, respectively.

Note that most COTS WiFi devices have carrier frequency offset (CFO) between the transmitting signal and the receiving signals [16], which is $e^{-j2\pi\Delta ft}$. For each path, the signal propagation distance is $d_k(t) = c\tau_k = f\lambda\tau_k(t)$, where λ is the wave length and c is the light speed. It follows $f\tau_k(t) = d_k(t)/\lambda = (\int_{-\infty}^{t} v_k(u)du)/\lambda$, where $v_k(u)$ is the rate of path length change. The CFR power can then be expressed as

$$|H(f,t)|^2 = |H_s(f) + \sum_{k=1}^{P_d} a_k(f,t)e^{\frac{-j2\pi}{\lambda}\int_{-\infty}^{t} v_k(u)du}|^2. \quad (3)$$

The CFR power can be used to mitigate the impact of CFO [9], [15]. The frequency characteristics of CFR power are a combination of sinusoids and depend on the signal path length change rate. The different path length changes are caused by different human activities, which have diverse speed. For example, the running activity has a high reflected path length change rate, showing wave features with high frequency, while the walking activity has a relatively low path length change rate, showing features with low frequency. As mentioned above, the feature extraction and activity recognition methods are then applied to distinguish different activities.

2.2 CSAR: Exploring Wider Band and Deep Learning

Most existing activity recognition systems employ a single WiFi channel for CSI data collection [8], [9], [10], [11], [15]. The recognition accuracy then highly depends on the quality of this particular channel. Even if a good quality channel is selected initially, it can get worse over time. Our experiments show that the state-of-the-art recognition system only achieves 80 percent accuracy using a channel with poor channel quality (more details are presented in Fig. 13 and Section 6.4). Hence, a channel-quality-aware collection in the early stage can greatly benefit the recognition in the later stage.

Besides the channel quality, the effectiveness of activity recognition is also affected by the WiFi channel bandwidth. For 802.11n, the typical bandwidth of each subcarrier is 312.5 KHz when the total channel bandwidth is 20 MHz [17]. Given a certain reported subcarrier number N_{sc}, the available CSI values are limited (e.g., even a channel can be divided into 64 spacing, only 30 available subcarriers are reported for Intel 5300 NICs). Even under the MIMO mode where $M_T * M_R$ antenna pairs can be constructed to obtain multiple CSI streams, these CSI streams are actually correlated: From the frequency aspect, different subcarriers have a very small difference in wavelength given the subcarrier frequency. For example, the biggest gap between two subcarriers in a 20 MHz channel is about 17 MHz, which is only a 0.3 percent difference in the 5 GHz band. Such a small difference actually provides limited information for activity recognition. From the spatial aspect, the antennas in one wireless card are closely placed, which often observe similar multipath reflections. Only if the transceivers communicate on multiple channels can we obtain truly richer CSI values for activity recognition.

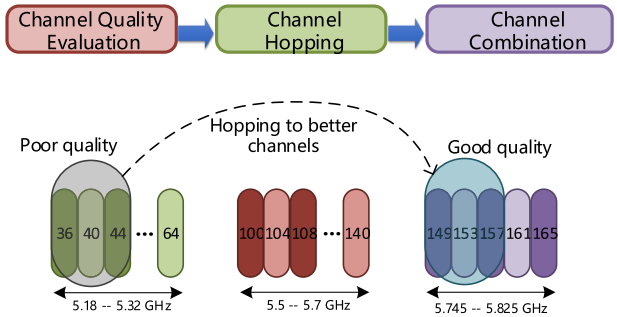


Fig. 2. Channel selection mechanism in CSAR system.

To overcome these limitations, we develop CSAR, a novel broadband Channel Selective Activity Recognition system. As illustrated in Fig. 2, CSAR extends the state-of-the-art activity recognition by incorporating channel quality evaluation, hopping and combination. It dynamically scans all the available channels and selects a set of adjacent channels with the best channel quality. It evaluates the qualities of selected channels from two aspects, the CSI amplitude and CSI stability, based on the collected CSI values. The channel hopping module then drives the transceivers to synchronously hop among the selected target channels, so as to obtain the CSI values from multiple channels. Their information is then combined for activity recognition. To our best knowledge, CSAR is the first to use information of multiple channels for activity monitoring with WiFi. It not only captures more clear and stable features with the best set of channels, but also constructs a wider *extended channel* with much richer information for further processing.

The richer information from the broad frequency and time domains however also calls for advanced machine learning beyond such conventional tools as kNN and HMM for classification. We demonstrate that a powerful deep learning tool, the long short term memory network [13], works best toward processing the extracted features from the time-frequency spectrogram of received signals. LSTM is effective with a massive scale of data input, i.e., the fine-grained spectrogram in CSAR. As a recurrent neural network, the current output in LSTM is decided by not only the current input but also the past states. As illustrated in Fig. 3, each network represents the processing of a time slot. The processed states in the previous stage can be stored and passed to the current stage. Connecting the past and the current events, it seamlessly integrates the features from the time dimension for more comprehensive classifications.

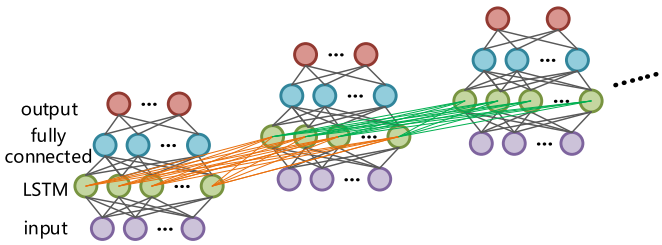


Fig. 3. The architecture of LSTM network.

3 Hopping-Based Channel Selection and Combination

In this section, we first describe how we evaluate the quality of a WiFi channel and select target channels for activity monitoring. Then we introduce the channel hopping and channel combination mechanisms.

3.1 Channel Quality Evaluation and Channel Selection

To select the best set of channels, we first need to evaluate the quality of channels. The first important consideration is CSI amplitude, which directly reflects the signal strength. Given N_T sample time points from T_s to T_e, we first collect all the CSI values across the selected channels within the time range. For a particular subcarrier s_j in channel h_i, we calculate the mean CSI amplitude value $\overline{Aij}$ as

$$\overline{A_{ij}} = \frac{\sum_t A_{ij}(t)}{N_T}, t \in samples\ in\ [T_s, T_e], \tag{4}$$

where $A_{ij}(t)$ is the CSI amplitude at time point t.

Since each channel reports N_{sc} subcarriers, we can calculate the average CSI amplitude $\overline{A_s}$ of a selected set of channels $\{h_1, \ldots, h_n\}$ as

$$\overline{A_s} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{N_{sc}} \overline{A_{ij}}}{N_{sc} * n}. \tag{5}$$

Fig. 4 plots the average CSI amplitude of subcarriers in two channel sets. We can see that the average amplitude of the good channel always keeps at 15. In contrast, the poor channel affected by the co-channel interference has a relatively low amplitude at about 12.3. With such a low signal strength, this channel is hardly useful for feature extraction and activity recognition.

Besides the CSI amplitude, the stability of CSI is also an important consideration. We observe that in some channels the received signals have explicit impulse noises or errors, which are hard to eliminate by conventional denoise

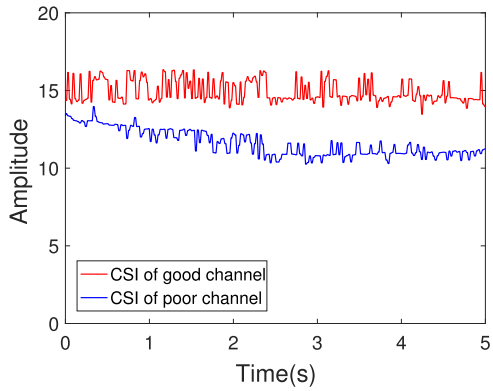


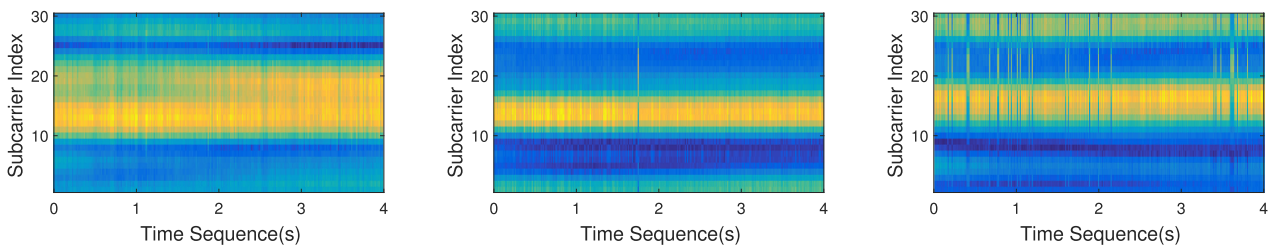
Fig. 4. The comparison of CSI values from two channels with different qualities.

mechanisms. Fig. 5 shows the CSI amplitude of 30 subcarriers of different channels in a time period in the same environment. We can observe that different subcarriers show different amplitude levels, e.g., the amplitude from 10th subcarrier to 16th subcarrier is high (indicated with brighter yellow color) while the amplitude from 1st subcarrier to 10th subcarrier is pretty low (indicated as darker blue). In Fig. 5a, the CSI values of every subcarrier are clean with good stability, offering reliable information for later recognition. On the contrary, the channel in Fig. 5c is very unstable, with impulse noises across all the 30 subcarriers, which can cause uncertain result in the recognition stage. Fig. 5b shows another channel with only one impulse noise. Such noise can be occasional and the corresponding channel is also acceptable.

CSAR considers both CSI amplitude and CSI stability as the quality indicators for a channel. Among these two metrics, CSI stability plays a more important role. The impulse noises in an unstable channel can generate abnormal high-energy frequency components in the subsequent spectrum analysis, which will shelter the normal features of activities and seriously undermine the recognition accuracy. For this reason, CSAR first selects those channel sets with the least impulse noises. When there are multiple channel sets that have the same CSI stability, CSAR then chooses the one with the highest average CSI amplitude as the final channel set for activity monitoring.

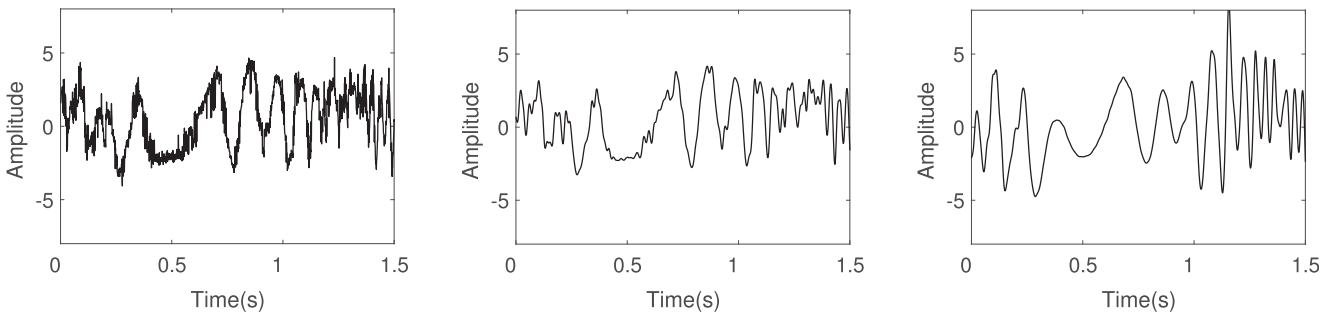
3.2 Channel Hopping and Combination

Once the channels are selected, CSAR circularly hops among these channels after sending one packet. Each packet contains the next hopping channel so that the transceivers



(a) Channel of good quality with no impulse noise. (b) Channel of medium quality with a few impulse noises. (c) Channel of poor quality with many impulse noises.

Fig. 5. CSI subcarriers of different channels with different quality.



(a) Original CSI values with much noise. (b) CSI through a low-pass filter. (c) CSI through PCA processing.

Fig. 6. Denoising results of the original CSI signals. We can observe that the original signal contains much high-frequency noise. The signal through low-pass filter is still not smooth. After PCA processing, the signal conserves most of common features and is smooth for processing.

can switch to the same destination channel. We have developed a complete channel hopping mechanism to guarantee the synchronous hopping between the transceivers. The hopping time from one channel to another channel is about 2 ms. With a sampling rate of 3,000 Hz using COTS WiFi devices, the system can at most use 3 channels for hopping, so that the sampling rate for each channel is enough for recognizing the common activities (e.g., running, falling down and waving hand).

According to the IEEE 802.11 standard, in a 20 MHz bandwidth channel, the occupied bandwidth is 16.6 MHz in total and the remaining 3.4 MHz is unused to avoid interference. Even there is a gap between subcarriers of two adjacent channel, we find that the bandwidth shift (3.4 MHz) is very small compared to the carrier frequency (in 5 GHz frequency band), which is only 0.07 percent and is even one-fifth of the gap between the first and last subcarrier in a channel. Such a small difference has a little impact on the channel combination. Thus, given the hopping channel list $\{h_1, \ldots, h_n\}$, for each hopping round, we stitch the corresponding CSI stream group collected from the same antenna pair together as $\tilde{h}$. In this way, we combine these adjacent channels together as an extended channel with higher bandwidth. Each CSI stream group of such channel has $n * N_{sc}$ subcarriers.

4 Data Denoising and Activity Detection

4.1 CSI Denoising

With the channel hopping and combination, CSAR can obtain the raw CSI values of the extended channel $\tilde{h}$. For this extended channel, we have $n * N_{sc} * M_T * M_R$ CSI streams, where n is the actual number of measured physical channels. The raw CSI streams collected from COTS devices, however, cannot be directly used for processing, for two reasons. First, each CSI stream contains too much noise, especially the high-frequency noise not related to human activities (see Fig. 6a for a sample segment of the raw CSI stream). The major noise of CSI streams comes from the internal state transitions of the transceivers, such as transmission power change and rate adaptation. Second, the raw collected data includes large scales of CSI values from many antenna pairs and different channels, which add a significant overhead to the later learning module. Even though the CSI streams in a CSI stream group have a slight difference from each other, they are actually correlated in spatial domain and frequency domain. Thus these CSI streams contain much similar information.

To address these issues, CSAR first utilizes a low-pass filter for CSI denoising. Based on the signal reflection model described in Section 2.1, the CSI amplitude changes caused by common human activities are mostly lower than 100 Hz. Hence, we only need to reserve those low-frequency signals. Before filtering, each CSI stream substracts the mean value to remove the direct-current component. Then we apply a low-pass filter (i.e., the Butterworth filter) to all CSI streams. Fig. 6b describes the same CSI stream in Fig. 6a after passing the low-pass filter with a cut-off frequency of 100 Hz. We can find that the filtered CSI stream contains much less noise but still not so smooth compared to the raw signal.

CSAR then uses PCA for data dimensionality reduction and extracts the common characteristics from multiple channels and subcarriers. In particular, we use 2 antennas for sending, 3 antennas for receiving, and set circular channel hopping number as 3. Given the reported subcarrier number as 30, we then have 540 dimensions for the signals. After PCA processing, we obtain a set of principal components. Given that the first principal component contains much of the noises [9], we calculate the average of the next three components as the representative component (denoted as p-stream) of the received signal. As illustrated in Fig. 6c, the processed p-stream contains little noise and extracts the wave features clearly.

4.2 Activity Detection

Before classifying activities from the denoised data, we first need to detect the start and the end of an activity. Given that the activity affects signal paths, the variance of the CSI amplitude during an activity is generally much higher than that with no activity. Such a high variance should last for a period, i.e., the time duration of the activity.

Based on such observations, we use the denoised p-stream of CSI for activity detection. CSAR obtains the denoised p-stream p_b and p_a when the target person is stationary and is active, respectively. Their average variances of a short time period are then calculated as δ_b and δ_a. We set an activity indication threshold as $\theta_H = (\delta_a + \delta_b)/2$. Besides, we also set a threshold θ_T, representing the minimum time duration that an activity should have. Thus we have the following activity detection strategy: 1) When the current state is *no activity*, if the p-stream of the signal starts having a variance over θ_H with a duration over θ_T, then we

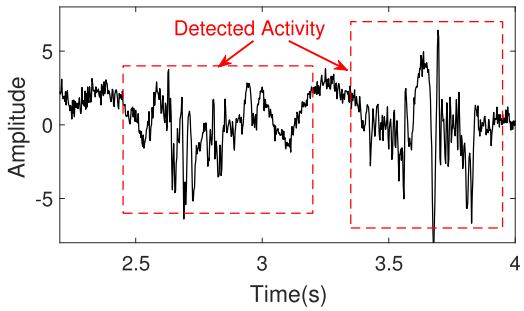


Fig. 7. An example to illustrate the detection process for the waving hand activity. The boxed part are detected as activity parts.

change the state of the following period as *in activity* and the corresponding time marks the start of the activity. 2) When the current state is *in activity*, if the p-stream of the signal starts having a variance below θ_H with a duration over θ_T, the state is then changed to *no activity* and the corresponding time is the end of the activity.

Fig. 7 shows the p-stream of the signals of a hand waving activity. We can observe that the signal has two obvious parts (as highlighted in the rectangles) with strong wave variance compared to other parts. In these two parts, the average variance is far beyond the threshold θ_H with time duration exceeding θ_T. From an empirical study on our dataset, we set the time threshold θ_T as 200 ms so that we can detect 98 percent of activities in our dataset. The detection mechanism can adapt to the environment changes and different activities by automatically adjusting the variance threshold δ_H. CSAR calculates the current variance of an activity and updates the threshold so that the threshold can fit different activities.

5 Deep-Learning-Based Activity Recognition

We now introduce the last piece of activity recognition, the learning model, starting from the feature extraction for the learning engine in CSAR.

5.1 Feature Extraction

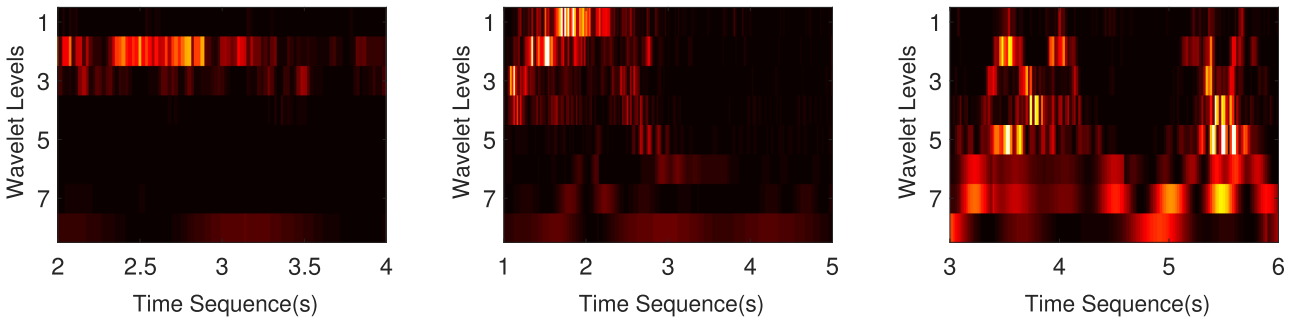
To feed a learning engine to classify different activities from the processed CSI values, we first need to extract the representative features. Although the p-stream compresses the data dimension and reveals the common wave patterns from the correlated CSI streams, it only reflects the time and amplitude of a waveform, whereas the frequency domain characteristics remain hidden. As such, it can hardly be directly used to classify activities. For example, running and walking can exhibit quite similar waveforms though running contains much higher frequencies.

CSAR chooses Discrete Wavelet Transform to extract the time-frequency features from the p-stream as in previous studies [7], [9], [18]. Extracting different frequency range at different time scale resolutions, DWT achieves a great trade-off between the frequency-scale and the time-scale resolution. Specifically, it groups the frequency component at different levels. In the high-frequency part, it provides high time resolution to distinguish subtle changes in a short time, and in the low-frequency part, it provides high frequency resolution to achieve a fine-grained classification for low-speed activities.

In our system, the sampling rate for each physical channel is about 200 Hz given the channel hopping overhead. We choose the frequency scale level number to be 8, so that the extracted frequency ranges from 0.4 to 100 Hz (the first level is from 50 to 100 Hz, and the last level is from about 0.4 to 0.8 Hz). Such a frequency range is enough to cover common activities. Empirically, we group the value of every 20 ms of the DWT results to reduce the time-scale dimensions. For every 1 second CSI data, our extracted feature contains 8 frequency-scale dimensions and 50 time-scale dimensions. Fig. 8 shows the extracted DWT spectrograms of typical activities. We can observe that the spectrogram of walking in Fig. 8a has an obvious high energy part at level 2, corresponding to 25 to 50 Hz (speed from 0.75 to 1.5 m/s). On the other hand, in Fig. 8b, the spectrogram of falling presents a unique pattern where the frequency first increases to a high level and then suddenly drops to a low level. This feature matches well with the movement of falling activity, i.e., falling fast from standing posture to lying down posture and then keeping motionless for a while. Fig. 8c also shows the spectrogram of two consecutive pushing forward activities, where the two crests represent the two pushing motions.

5.2 Why LSTM (Long Short Term Memory)

Existing WiFi-based activity recognition systems mostly use conventional learning models such as hidden Markov model (HMM) [9] and k-nearest neighbours (kNN) [11] to process the extracted features. These models, however, face challenges when processing the spectrogram data from our channel hopping and combination.



(a) DWT spectrogram of walking activity. (b) DWT spectrogram of falling activity. (c) DWT spectrogram of pushing hand activity.

Fig. 8. Comparison of the DWT spectrograms of different activities.

First, as discussed, the DWT spectrograms of different activities present different features in both frequency dimensions and time dimensions. We need to comprehensively consider the spectrogram for a period of time rather than just a time slot. The kNN model [11] largely ignores the time dimension. The HMM model [9], though considers the time, is based on a strong assumption that the feature vectors of an activity are subject to a finite state machine, which changes state once a time.

Second, with channel hopping and combination, the data from the early stages of our CSAR contain much richer and fine-grained information. Existing works with the kNN model use only 10 features per virtual sample for classification [11]; those using the HMM model rely on coarse-grained data, e.g., 200 ms time interval [9]. They also require intensive parameter settings that are often experience-based, e.g, difference of each energy level in the DWT spectrogram and initial parameter estimation such as the state distribution and state transition matrix. These become quite difficult when the data size grows.

Third, different activities usually last various length of time duration, e.g., waving hand often lasts less than 1 s while walking usually takes more than 1 s. Even for the same activity, different people can also have different time durations. The kNN model needs to specify a fixed length of input, and HMM lacks flexibility in this aspect, too.

To address these challenges, CSAR uses a *Long Short Term Memory* network [13] to classify different activities from the extracted features. LSTM is an advanced deep artificial neural network, which has recently seen success in such real-world applications as hand-writing recognition, speech recognition and video recognition. Different from the *convolutional neural network* (CNN) that mainly exploits the spatial correlations of inputs, LSTM is designed to classify and predict the inherent relationships of time series. It combines the current inputs and the past states stored in the memory cell to exploit the time scale relationships and achieves a comprehensive classification. As an improved architecture of the *recurrent neural network* (RNN), it is able to learn long-term dependencies, aligning well in our activity recognition context.

Specifically, LSTM supports fine-grained data input and does not need initial state estimation. To build up our network model, we just need to construct the network architecture and specify how each layer is connected. Then we input the extracted features from DWT as described in Section 5.1. It also supports various length of input in time dimensions and can integrate such inputs together.

5.3 Building LSTM Model

In CSAR system, we build a typical four-layer LSTM network to classify different activities, as shown in Fig. 3. The first layer is the *input layer*, consisting of 8 input nodes, which corresponds to the number of frequency levels in the DWT processing. The next layer is the LSTM layer, consisting of 100 LSTM nodes. To fully explore the hidden time series features in the DWT spectrogram, here we use more than 10 times of LSTM nodes compared to the input layer. As is observed from the figure, the LSTM layer has a direct flow connecting to itself. Therefore, the decision of this network at time t is the integrated results of both the current input at time t and the previous temporal state at time $t-1$. The key of LSTM is the cell state with multiple control gates. These gates optionally let information pass through, deciding which information can be passed to the next time and which information should be blocked. With such control ability, LSTM can remember the past information even long-term events and combine them with the current input for a comprehensive classification.

The third layer is a fully connected layer with 200 nodes. Each node is fully connected to the previous LSTM layer. We use ReLU as the activation function instead of sigmoid to reduce the likelihood of vanishing gradient. And to prevent the overfitting in the training process, we randomly ignore nodes in the LSTM layer and the fully connected layer, and set the dropout rate as 0.5 in the experiment. The last layer is an output layer with 8 nodes, corresponding to the classification category. In the last layer, we calculate the cross-entropy of every activity and choose the one with the largest likelihood as the classification category. We set the maximum time length of the input as 2 s in the LSTM model based on the observation that all the activities in our dataset last less than 2 s.

6 Implementation and Evaluation

In this section, we introduce how we implement CSAR with COTS WiFi devices and evaluate its performance with diverse configurations.

6.1 Prototype Implementation

We implemented CSAR on COTS hardware, in particular, Intel 5300 Wireless Card with three antennas. Our prototype implementation consists of two DELL latitude D820 laptops installed Ubuntu system with Linux 4.2 kernel as the transceivers, each equipped with an Intel 5300 Wireless Card. For each laptop, we also install the CSI tools developed by Halperin et al. [14] to collect the CSI reported by the wireless cards.

We implemented the channel hopping mechanism based on the iwlwifi driver and Loss Of Radio CONnectivity (LORCON) project [19], which is an open source IEEE 802.11 packet injection library. We modify the library so that CSAR can achieve fine-grained channel hopping and packet handling in microsecond granularity. A key challenge for CSAR is to build a complete communication mechanism that enables the transceivers to always work on the same channel synchronously. Given the unreliable wireless environment, a transmitted packet in one particular channel can be lost. The transceivers will then lose connection if they don't know in which channel the other transceiver stays. To address this problem, we have designed an acknowledgement and retransmission mechanism to guarantee the synchronous transmission. For convenience, we call one transmitter as the client and the other as the server. The client first sends a packet indicating the next channel to hop to and meanwhile starts a timer. The server replies with a packet as acknowledgement and hops to the target channel also with a timer started. When the acknowledgement packet is received, the client then hops to the same channel as the server for next sending. Once any packet is lost, both the client and server turn to the default channel after a given time-out duration.

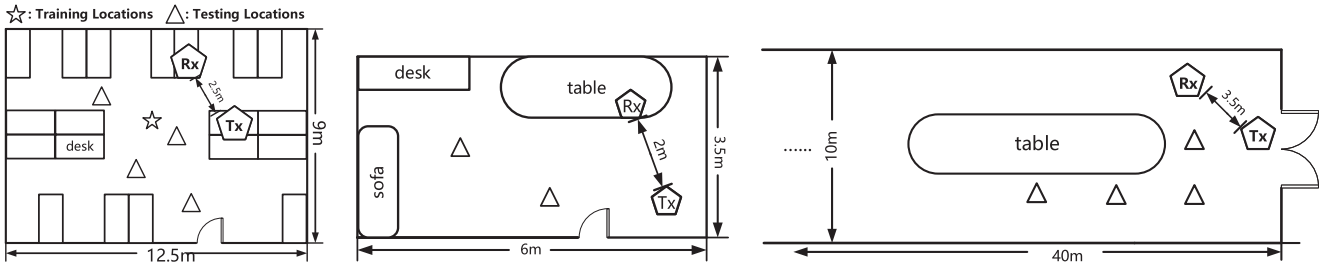


Fig. 9. The floor plans of different testing environments.

In our experiment, we use two antennas in the sender and three antennas in the receiver (i.e., $M_T = 2$ and $M_R = 3$). The transceivers in our system work and hop in the 5 GHz frequency band rather than 2.4 GHz frequency band, with the channel bandwidth of 20 MHz. This is because, in the 2.4 GHz frequency band, every channel has significant overlap with its neighboring channels, so that the system can only hop between some particular channels with no interference. Instead, the 5 GHz frequency band has much more effective channels for hopping. Considering the channel hopping time (about 2 ms in our measurement), we set the channel hopping number as 3 so that each channel has a sampling rate at 200 Hz, which is enough for recognizing common activities.

The activity movement speed perceived by different transceiver pairs can be slightly different. This is because the reflected signal is determined by both the locations and orientations of the transceiver pairs as relative to the target human [11]. Although the experiments involve only one transceiver pair, our CSAR can be easily scaled to multiple transceiver pairs for human activity recognition. With multiple transceivers, the activity features can be captured from different orientations for recognition.

It is also worth noting that CSAR requires channel hopping in the monitoring mode to select channels with good quality, and hence unable to reuse the data communication channels. However, since the monitoring process is independent, it will not affect the normal data communication and device (the laptop) usage. We just need to install another wireless card or use the wired cable for net service.

6.2 Evaluation Setup

In our experiment, we collect a total of 3,350 training samples by asking six volunteers to perform eight different activities in a pre-configured environment (a lab as illustrated in Fig. 9a). These volunteers are both undergraduate and graduate students varying in genders, heights and weights. The training samples are collected in the lab of size 12.5 $\mathrm{m} \times 9\ \mathrm{m}$, with the activity performing location being labeled as a star. The names and sample numbers for each activity is described in Table 1. When the volunteers begin to perform activities, the body movement will affect the reflected signals and we therefore can collect the corresponding sampling data for different activities. We then evaluate the recognition accuracy of CSAR in different environments, including the lab, a large lobby of size $25\ \mathrm{m} \times 8\ \mathrm{m}$ (as shown in Fig. 9c), and a small office of size $18\ \mathrm{m}^2$ (as shown in Fig. 9b). The testing locations are labeled in the corresponding figures as the triangle shapes.

TABLE 1
Activity Dataset and the Collected Samples

Activity	# Samples	Activity	# Samples
Walking (Wa)	470	Picking (Pk)	290
Falling (Fa)	220	Pushing (Ps)	540
Running (Rn)	440	Waving (Wv)	580
Sitting (St)	230	Boxing (Bx)	580

We conduct the deep learning process using a testbed equipped with a Nvidia Geforce GTX 1060 6GB Xtreme Gaming VR Ready Graphics Card. The total training time is about 12 minutes. Though the training time is slightly longer than such traditional methods as Hidden Markov Model [9], the training process is one-off and the time cost is negligible. Given the emerging online deep learning approaches [20], we also plan to implement the online deep-learning-based recognition in the future.

6.3 Sensitivity of Activity Detection

We first evaluate the activity detection in CSAR, in particular, the detection distance, which reflects the sensitivity of the detection mechanism. We treat a successful detection as correctly detecting both the start and the end of an activity. Fig. 10 plots the detection accuracy of two typical activities in the lobby where the configuration is illustrated in Fig. 9c. For each distance, we measured 20 samples and the test locations were randomly selected with the same distance to ensuring the generality. The general detection distance of walking is much farther than hand waving. Such a difference is mainly due to the different reflection areas: walking represents the torso-based activity that the reflected signal can be captured more clearly; hand waving represents the gesture-based activity generating

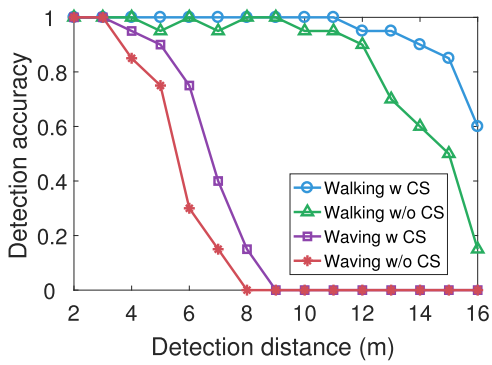


Fig. 10. The detection distance of the walking and the waving hand activities. Walking is a torso-based activity that has a large reflection area, while waving is a gesture-based activity that has a small reflection area.

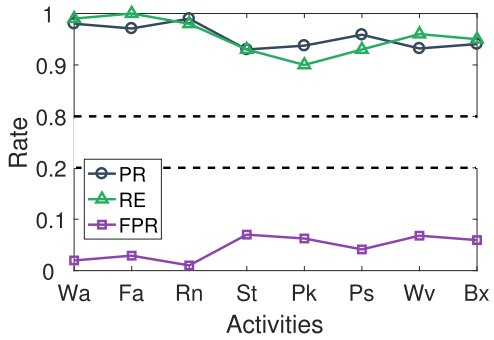


Fig. 11. Precision (PR), Recall (RE) and false positive rate (FPR) of all activities.

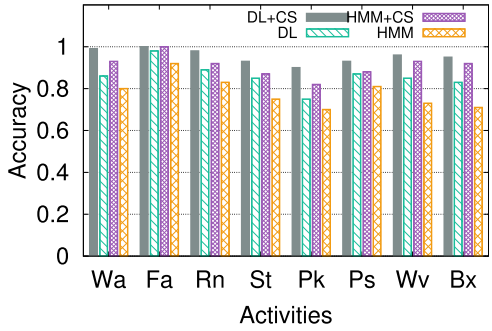


Fig. 13. Recognition accuracy using different models.

weaker reflected signals. On the other hand, our channel-selection-based detection method reveals a higher detection accuracy than that without the channel selection mechanism. For walking, our method can achieve an accuracy of 90 percent even at a distance of 14 m, while without channel selection, the accuracy falls quickly to only 60 percent at this distance. The result is same when detection the gesture-based activity with much smaller movement. This indicates that compared to the conventional activity approach, CSAR is able to utilize a better channel for activity detection, achieving more sensitive and reliable detection.

6.4 Accuracy of Activity Classification

We next evaluate the performance of CSAR for activity classification. To achieve a comprehensive analysis, we employ the following metrics widely used in statistics: 1) False positive rate (FPR), indicating the ratio of selecting negative samples as the positive samples, 2) Precision (PR), defined as $\frac{TP}{TP+FP}$, where TP (true positive) means the positive samples that are correctly predicted as the actual activities, and 3) Recall (RE), defined as $\frac{TP}{TP+FN}$ where FN is the positive samples that are falsely labelled as other activities. Fig. 11 shows the statistical results of the three metrics for the 8 activities using our deep learning and channel selection methods. We can observe that the false positive rates of all activities are below 10 percent. The walking and running activities have the lowest FPR, indicating that their features are very distinguishable. CSAR achieves an average precision of 94 percent with a standard deviation of 3 percent, which means that our system can correctly classify the majority of corresponding activities. The recall shows a large variance among different activities, e.g., the running activity achieves about 99 percent while the picking up activity is only 90 percent. Yet the average recall still keeps a high level at 95 percent. The statistical results indicate that CSAR has a complete and accurate recognition performance over all activities.

Actual Activity	Predicted Activity							
	Wa	Fa	Rn	St	Pk	Ps	Wv	Bx
Wa	0.99	0	0.01	0	0	0	0	0
Fa	0	1	0	0	0	0	0	0
Rn	0.02	0	0.98	0	0	0	0	0
St	0	0.01	0	0.93	0.06	0	0	0
Pk	0	0.01	0	0.07	0.9	0.02	0	0
Ps	0	0.01	0	0	0	0.93	0.03	0.03
Wv	0	0	0	0	0	0.01	0.96	0.03
Bx	0	0	0	0	0	0.01	0.04	0.95

Fig. 12. Confusion matrix of activity recognition.

To explore a fine-grained performance, we show the confusion matrix across the 8 activities in Fig. 12, where each element gives the ratio that we classify the actual activity (labeled by the corresponding row) to the predicted activity (labeled by the corresponding column). We can see that the correct classification ratio of each activity is more than 93 percent, indicating that CSAR achieves a high and stable classification accuracy over all activities. The falling activity achieves a 100 percent classification accuracy, as its features are quite distinguished from others. The sitting down activity and the picking up activity may be easily misclassified into each other as they are indeed similar in nature.

We also compare our method with the HMM model in CARM [9], which is served as the baseline. To understand the impact of different modules in CSAR, we consider three combinations, i.e., deep learning with channel selection (DL+CS), deep learning only (DL), and HMM with channel selection (HMM+CS). For DL only and the basic HMM, since they can not switch channels themselves, we randomly select five channels in the testing environment and use the average recognition accuracy of such channels as the result.

Fig. 13 shows the comparison of recognition accuracy between our CSAR (i.e., DL+CS) and others. DL+CS achieves an average accuracy of 95 percent, revealing an obvious advantage over the basic HMM, which has an 80 percent accuracy only. When HMM is enhanced with CS, the gap shrinks to 4 percent, showing the strength of channel selection. On the other hand, comparing DL with HMM (both with no channel selection), the difference in terms of recognition accuracy becomes 8 percent. This suggests that the advanced deep learning method better exploits the inherent relationship within the time series and frequency components of the extracted features. In short, both DL and CS play important roles, and their combination works the best for recognition than they are used individually.

6.5 Impact of Environment and People

To examine the impact of different environments, we evaluate the activity recognition accuracy of CSAR in the lab, lobby and office, respectively, as shown in Fig. 14. The lab is a trained environment while the lobby and the office are untrained environments. The average recognition accuracies in the lab, lobby and office are 95, 97 and 88 percent, respectively. The accuracies for both lab and lobby are quite

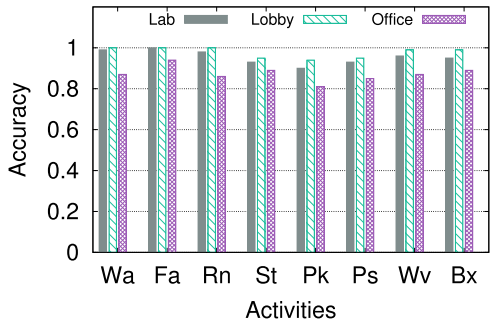


Fig. 14. Recognition accuracy in different environments.

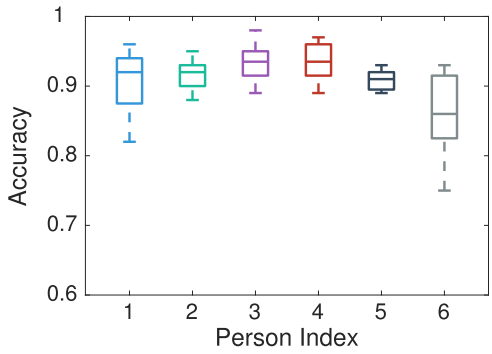


Fig. 15. Average recognition accuracy for different people.

high. This is because they are relatively open spaces with less multipath effect, and hence less impact on feature extraction. The compact office space incurs much stronger multipath; yet the accuracy is still acceptable, suggesting that CSAR is robust to environment change.

We further examine the impact of human diversity on activity recognition. We test the activity recognition accuracy for 6 volunteers at three random locations in the lab. Three of them (volunteer 1-3) participated in the training process and the other three (volunteer 4-6) are new for testing. Given the differences in speeds, ranges and styles, the average accuracy becomes diverse over different people. Fig. 15 shows the boxplot of the average activity recognition accuracy for these volunteers in the lab. We can observe that the overall accuracy for the 6 volunteers still stays in a high level with CSAR. The average accuracy of most volunteers are higher than 90 percent and only the 6th volunteer falls below 90 percent. The recognition accuracy does not show an obvious difference between trained volunteers and new volunteers. This result indicates that CSAR is robust to different people and can achieve a high and stable recognition accuracy.

7 Related Work

WiFi-based human activity recognition has attracted remarkable attention in recent years. Previous works examine the diverse impact of human activities on transmitted signals to distinguish different activities. Based on the usage of sensing devices, these activity recognition systems can be generally divided into two categories, i.e., specialized-hardware-based systems and COTS-device-based systems. The former utilizes such advanced devices as custom analog circuits and specialized antennas with software defined radios. In contrast, the COTS-device-based systems rely on commodity devices such as common wireless cards, laptops and normal antennas to capture the wireless signal changes. They are readily available but, compared to specialized hardware, face more challenges in accurately capturing signal changes during human activities.

Specialized-Hardware-Based System. Specialized hardware with software defined radios has been used to capture fine-grained signal metrics for activity recognition and other related applications. Pu et al. [6] developed a whole-home gesture recognition system that used USRP to capture the micro-level Doppler shifts in wireless signals. With two wireless sources in an indoor environment, it achieves an average of 94 percent recognition accuracy. Kellogg et al. [21] utilized a special low-power analog envelope-detection circuit to obtain the strength of received signals. By profiling the wave patterns caused by different gestures, AllSee achieves 97 percent accuracy over eight gestures within a very short range (less than 2.5 feet). Huang et al. [22] analyzed the reflected wireless signals to construct images of objects and humans. Adib et al. [23] utilized FMCW radios to detect the micro-movement of breathing and heart rate with a median accuracy of 99 percent. There have also been recent works on specialized hardware for human tracking [24], [25], [26].

COTS-Device-Based System. Solutions in this category have to accommodate the limitations imposed by COTS devices, including the limited and coarse-grained channel information that can be captured. In this context, CSI has been widely used. Wang et al. [8] explored the CSI wave pattern changes caused by human falling and analyzed such different patterns to detect the human falling. Wang et al. [10] proposed to recognize both in-home activities and walking movements by profiling the CSI changes across multiple subcarriers. Wang et al. [9] built up a novel model that correlates the human movement velocity and the CSI dynamics, and leverages the frequency level difference to distinguish different activities. Virmani and Shahzad [11] proposed a gesture recognition system with a translate function that can automatically estimate the target user's location and then classify the different gestures. Besides human activity recognition, COTS-WiFi-based sensing technology has also been widely used in many scenarios, such as recognizing the in-air drawing [27], hearing a predefined of spoken words [28], recognizing the keyboard typing [18], recognizing the dancing steps [29], passive tracking [30], [31] and multi-target localization [32].

Using COTS wireless devices for activity recognition has its inherent benefit in terms of cost-effectiveness and compatibility. Our interest in this paper also lies in this direction. We however reveal the challenges of state-of-the-art single-channel-based solutions in a dynamic real-world environment, and address them by a channel selective activity recognition system integrating the advanced LSTM network model to achieve accurate and reliable recognition.

8 Conclusion

In this paper, we proposed CSAR, a deep learning-based channel selective activity recognition system using COTS WiFi devices. The key novelty lies in its channel selection and combination mechanism, which automatically detects the quality of available channels and hops among a set of

adjacent channels. This effectively constructs an extended wideband channel, offering much richer and more stable channel status information for activity recognition. CSAR then employed an LSTM network for fine-grained activity recognition, which seamlessly integrates both the time dimension and the frequency dimension information. We have implemented CSAR in commodity WiFi devices and our extensive experiments showed that CSAR can achieve an average 95 percent recognition accuracy even in a crowded wireless environment.

Acknowledgments

This work was supported by a Canada Technology Demonstration Program (TDP) grant, a Canada NSERC Discovery Grant, and an NSERC E.W.R. Steacie Memorial Fellowship.

References

[1] J. K. Aggarwal and M. S. Ryoo, "Human activity analysis: A review," *ACM Comput. Surveys*, vol. 43, no. 3, 2011, Art. no. 16.
[2] 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 *Proc. 9th ACM Conf. Embedded Netw. Sensor Syst.*, 2011, pp. 274–287.
[3] K. Yatani and K. N. Truong, "BodyScope: A wearable acoustic sensor for activity recognition," in *Proc. ACM Conf. Ubiquitous Comput.*, 2012, pp. 341–350.
[4] X. Fan, W. Gong, and J. Liu, "TagFree activity identification with RFIDs," in *Proc. ACM Interactive Mobile Wearable Ubiquitous Technol.*, 2018, Art. no. 7.
[5] X. Fan, W. Gong, and J. Liu, "i2tag: RFID mobility and activity identification through intelligent profiling," *ACM Trans. Intell. Syst. Technol.*, vol. 9, no. 1, 2017, Art. no. 5.
[6] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *Proc. 19th Annu. Int. Conf. Mobile Comput. Netw.*, 2013, pp. 27–38.
[7] H. Abdelnasser, M. Youssef, and K. A. Harras, "WiGest: A ubiquitous WiFi-based gesture recognition system," in *Proc. IEEE Conf. Comput. Commun.*, 2015, pp. 1472–1480.
[8] Y. Wang, K. Wu, and L. M. Ni, "WiFall: Device-free fall detection by wireless networks," *IEEE Trans. Mobile Comput.*, vol. 16, no. 2, pp. 581–594, Feb. 2017.
[9] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of WiFi signal based human activity recognition," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, 2015, pp. 65–76.
[10] 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 *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw.*, 2014, pp. 617–628.
[11] A. Virmani and M. Shahzad, "Position and orientation agnostic gesture recognition using WiFi," in *Proc. 15th ACM Int. Conf. Mobile Syst.*, 2017, pp. 252–264.
[12] G. L. Stüber, *Principles of Mobile Communication*, vol. 2. Berlin, Germany: Springer, 2001.
[13] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[14] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Predictable 802.11 packet delivery from wireless channel measurements," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 40, pp. 159–170, 2010.
[15] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Device-free human activity recognition using commercial WiFi devices," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 5, pp. 1118–1131, May 2017.
[16] D. Vasisht, S. Kumar, and D. Katabi, "Decimeter-level localization with a single WiFi access point," in *Proc. 13th USENIX Symp. Netw. Syst. Des. Implementation*, 2016, pp. 165–178.
[17] IEEE 802.11n Working Group, et al., "Draft amendment to standard for information technology-telecommunications and information exchange between systems-local and metropolitan networks-specific requirements-part 11: Wireless LAN medium access control and physical layer specifications: Enhancements for higher throughput," IEEE Standard P802. 11n/D1. 0, 2006.
[18] K. Ali, A. X. Liu, W. Wang, and M. Shahzad, "Keystroke recognition using WiFi signals," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, 2015, pp. 90–102.
[19] Lorcon project. (2009). [Online]. Available: https://github.com/kismetwireless/lorcon
[20] D. Sahoo, Q. Pham, J. Lu, and S. C. Hoi, "Online deep learning: Learning deep neural networks on the fly," *arXiv preprint arXiv:1711.03705, CoRR*, vol. abs/1711.03705, 2017.
[21] B. Kellogg, V. Talla, and S. Gollakota, "Bringing gesture recognition to all devices," in *Proc. 11th USENIX Symp. Netw. Syst. Des. Implementation*, 2014, pp. 303–316.
[22] D. Huang, R. Nandakumar, and S. Gollakota, "Feasibility and limits of Wi-Fi imaging," in *Proc. 12th ACM Conf. Embedded Netw. Sensor Syst.*, 2014, pp. 266–279.
[23] F. Adib, H. Mao, Z. Kabelac, D. Katabi, and R. C. Miller, "Smart homes that monitor breathing and heart rate," in *Proc. 33rd Annu. ACM Conf. Human Factors Comput. Syst.*, 2015, pp. 837–846.
[24] F. Adib, C.-Y. Hsu, H. Mao, D. Katabi, and F. Durand, "Capturing the human figure through a wall," *ACM Trans. Graph.*, vol. 34, no. 6, 2015, Art. no. 219.
[25] F. Adib, Z. Kabelac, D. Katabi, and R. C. Miller, "3D tracking via body radio reflections," in *Proc. 11th USENIX Symp. Netw. Syst. Des. Implementation*, 2014, pp. 317–329.
[26] F. Adib and D. Katabi, "See through walls with WiFi!," in *Proc. ACM SIGCOMM Conf.*, 2013, pp. 75–86.
[27] L. Sun, S. Sen, D. Koutsonikolas, and K.-H. Kim, "WiDraw: Enabling hands-free drawing in the air on commodity WiFi devices," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, 2015, pp. 77–89.
[28] G. Wang, Y. Zou, Z. Zhou, K. Wu, and L. M. Ni, "We can hear you with Wi-Fi!," *IEEE Trans. Mobile Comput.*, vol. 15, no. 11, pp. 2907–2920, Nov. 2016.
[29] K. Qian, C. Wu, Z. Zhou, Y. Zheng, Z. Yang, and Y. Liu, "Inferring motion direction using commodity Wi-Fi for interactive exergames," in *Proc. CHI Conf. Human Factors Comput. Syst.*, 2017, pp. 1961–1972.
[30] D. Wu, D. Zhang, C. Xu, Y. Wang, and H. Wang, "WiDir: Walking direction estimation using wireless signals," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2016, pp. 351–362.
[31] K. Qian, C. Wu, Z. Yang, Y. Liu, and K. Jamieson, "WiDar: Decimeter-level passive tracking via velocity monitoring with commodity Wi-Fi," in *Proc. 18th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2017, Art. no. 6.
[32] J. Wang, D. Fang, Z. Yang, H. Jiang, X. Chen, T. Xing, and L. Cai, "E-HIPA: An energy-efficient framework for high-precision multi-target-adaptive device-free localization," *IEEE Trans. Mobile Comput.*, vol. 16, no. 3, pp. 716–729, Mar. 2017.

Fangxin Wang (S'15) received the BS degree from the Department of Computer Science of Technology, Beijing University of Post and Telecommunication, Beijing, China, in 2013 and the MS degree from the Department of Computer Science and Technology, Beijing, China, in 2016. He is currently working toward the PhD degree in the School of Computing Science, Simon Fraser University, Burnaby, British Columbia, Canada. His research interests include wireless networks, big data analysis, and machine learning. He is a student member of the IEEE.

Wei Gong (M'14) received the BE degree in computer science from the Huazhong University of Science and Technology, the ME degree in software engineering and the PhD degree in computer science both from Tsinghua University. He is a professor of the School of Computer Science and Technology, University of Science and Technology of China. His research focuses on wireless networks, Internet-of-Things, and distributed computing. He had also conducted research with Simon Fraser University and the University of Ottawa. He is a member of the IEEE.

Jiangchuan Liu (S'01-M'03-SM'08-F'17) received the BEng (Cum Laude) degree from Tsinghua University, Beijing, China, in 1999, and the PhD degree from the Hong Kong University of Science and Technology, in 2003, both in computer science. He is currently a full professor (with the University Professorship) with the School of Computing Science, Simon Fraser University, British Columbia, Canada. He is a Steering Committee member of the *IEEE Transactions on Mobile Computing*, and associate editor of the *IEEE/ACM Transactions on Networking*, the *IEEE Transactions on Big Data*, and the *IEEE Transactions on Multimedia*. He is a co-recipient of the Test of Time Paper Award of IEEE INFOCOM (2015), ACM TOMCCAP Nicolas D. Georganas Best Paper Award (2013), and ACM Multimedia Best Paper Award (2012). He is a fellow of the IEEE and an NSERC E.W.R. Steacie Memorial fellow.

Kui Wu (S'98-M'02-SM'07) received the BSc and MSc degrees in computer science from Wuhan University, China, in 1990 and 1993, respectively, and the PhD degree in computing science from the University of Alberta, Canada, in 2002. He joined the Department of Computer Science, University of Victoria, Canada, in 2002, where he is currently a full professor. His research interests include smart grid, mobile and wireless networks, and network performance evaluation. He is a senior member of the IEEE.