Fine-grained Adaptive Location-independent Activity Recognition using Commodity WiFi

Jianfei Yang[1], Han Zou[2], Hao Jiang[3], Lihua Xie[1]
[1]School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore
[2]Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, USA
[3]College of Electrical Engineering and Automation, Fuzhou University, Fuzhou, China
Email: {yang0478,elhxie}@ntu.edu.sg, hanzou@berkeley.edu, jiangh@fzu.edu.cn

***Abstract*—Device-free activity recognition is appealing in smart home applications. It not only is convenient, but also causes no privacy concern, as compared to other activity recognition techniques such as the vision based technique. Existing WiFi-based methods have achieved high accuracy in static circumstances but have limitations in adapting changes in environment and activities locations. In this paper, we propose a fine-grained adaptive location-independent activity recognition system (FALAR) which leverages WiFi signals to characterize and recognize common activities regardless of inconsistency of mutative surroundings. FALAR applies fine-grained channel state information (CSI) to achieve accurate recognitions. To address the issue of environmental changes, we present a Kernel Density Estimation (KDE) based motion extraction method and a coarse-to-fine search strategy for speedy processing. After a denoising scheme, we introduce Class Estimated Basis Space Singular Value Decomposition (CSVD) to efface the static path in the background, and use nonnegative matrix factorization to distinguish various activities by looking into the signal profiles. We evaluate FALAR using two commodity WiFi routers in a typical office environment. Our results show that it achieves remarkable performance.**

I. INTRODUCTION

Human activity recognition plays a crucial role in ubiquitous and mobile computing due to its wide applications such as smart home, public surveillance and health care. Pervasive methods apply radars, cameras [1] or wearable sensors [2] and achieve reasonable results. However, some limitations emerge with their practical applications. For instance, camera-based methods are highly influenced by illumination as well as occlusion, and often arouse privacy concerns. Wearable sensors [3], [4] solve these problems to some extent but human being do not necessarily carry them in daily life. Recently, WiFi-based approaches [5], [6] are proposed based on the analysis of multi-path distortions in WiFi signals caused by human activities. They preserve user privacy and perform better robustness against vision-based and sensor-based method.

Channel State Information (CSI), extracted from the physical layer of wireless communications, leverages high resolution and makes it possible for human activity recognition [7] as well as more challenging works such as crowd counting [8], [9] and person identification [10], [11]. Current CSI-based activity recognition systems have three key limitations. Firstly, during the propagation of WiFi signals, wireless signals are sensitive to environmental changes due to reflection, diffraction and scattering. The system needs to separate the environmental changes from human activities. Secondly, the same activity has different CSI fingerprints at different spots, which leads to low accuracy or large consumptions of collecting training datasets. Thirdly, in practice, more customized activities should be extended by users so it is important to design a semi-supervised approach to improve scalability. Generally, the calibrations of these new activities are not supposed to be complicated and time-consuming for user-friendly design. Although the current state-of-the-art approaches provide some solutions such as increasing training data, the challenges have not been addressed in a simple and practical way.

In this paper, we propose FALAR, a find-grained adaptive location-independent activity recognition system. It consists of two WiFi devices, used as transmitter and receiver. To detect the activities, we firstly develop a motion detection method based on Kernel Density Estimation (KDE) [12] to eliminate slight environmental changes. To speed up this accurate foreground learning approach, a coarse-to-fine search strategy is developed. After the motion is extracted as raw CSI data, we utilize a wavelet-based denoising scheme to eliminate the ambient noise of CSI. Then, to address the location-dependent issue, we reconstruct CSI data by Class Estimated Basis Space Singular Value Decomposition (CSVD). The reconstructed CSI waves preserve the characteristics of activity but discard most location information. Practically, the key issue is how to design a feature extraction and classification approach which reduces the calibration number of new activities. As such, we transform the extraction and classification process into matrix factorization and composition by means of the theory of Nonnegative Matrix Factorization (NMF) [13]. NMF decomposes the reconstructed CSI into a base matrix, which preserves discriminative features, and a coefficient matrix which is used as classifier in our method. This approach not only provides a robust and fast activity classifier, but also simplifies the calibration of new activities and thus makes our system user-friendly. The whole framework is shown in Figure 1. To implement FALAR on Commercial Off-The-Shelf (COTS) WiFi routers, we upgrade the Atheros CSI Tool [14] and develop a new OpenWrt firmware for the routers to directly get fine-grained CSI of all subcarriers. The experimental results show that FALAR is able to provide high activity recognition accuracy confronting environmental changes across distinct locations.

978-1-5386-1734-2/18/$31.00 ©2018 IEEE

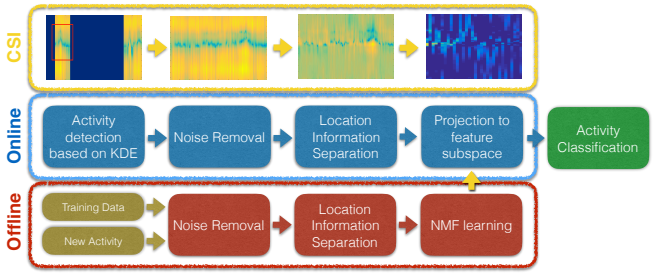

Fig. 1. The framework and its visual representation.

Contributions: We summarize the main contributions of FALAR as follows:

- To detect human motion accurately, we develop a KDE-based detection scheme. Using this novel detection algorithm, a coarse-to-fine search strategy is proposed, which is generally suitable for various activity recognition systems to use.
- By analyzing the characteristics of CSI, we reconstruct the CSI data by CSVD approach to remove location information associated with most static paths.
- We present an NMF-based method to extend the scalability of FALAR. To the best of our knowledge, this is the first effort using automatic way to accomplish feature extraction, classification and simple calibration of a class of activity for only several times. CSVD-NMF also accelerates the speed of offline training and online recognition, which makes FALAR user-friendly.
- We implement FALAR on COTS WiFi routers by mining fine-grained CSI on the innovative OpenWrt system. Our new firmware provides a new platform for CSI-based applications. The evaluation substantiates the robustness and accuracy of our system.

The rest of the paper is organized as follows. Section II analyzes the pros and cons of current methods. In Section III and IV, the detailed system design and experiment procedure are introduced respectively. We conclude our work and highlight some future trends in Section V.

II. RELATED WORK

Existing works on WiFi-based human activity recognition mainly employ two kinds of data: Received Signal Strength Indicator (RSSI) and CSI. The main difference between the two categories is the resolution of signal.

RSSI Based: Human activities bring about the signal strength variations, which is the motivation of RSSI-based methods. Such activity recognition systems [15], [16] cannot achieve high accuracy due to the low resolution of RSSI, but RSSI-based localization systems [17] are able to attain reasonable results.

CSI Based: CSI reflects multi-path effect by subcarrier information from the physical layer, and thus provides fine-grained signals. After CSI values were available in commercial Intel 5300 NICs [18], many researchers have developed many applications based on it. E-eyes [19] firstly examines channel features to identify in-place and moving activities in home environment. CARM [20] models the CSI and speed of person, which connects the movement speeds with activities. Comparatively speaking, our approach pays more attention on the detection under a complicated changing environment and recognition in different locations.

III. SYSTEM DESIGN

A. COTS WiFi Router Platform

Due to the environment and occupant activity, WiFi signal propagates through multiple paths, which leads to multi-path distortions mentioned before. Modern WiFi devices support IEEE 802.11n/ac standard that permits the use of multiple transmit and receive antennas for multiple input, multiple output (MIMO) communication. These WiFi devices perform maximal capacity of wireless channel by monitoring and changing transmit power allocations as well as rate adaptations for each individual MIMO stream. CSI represents the state of channel at the physical layer in terms of OFDM subcarriers, and thus has better resolution than RSSI. In OFDM, channel estimation consists of a vector of S complex elements representing channel in frequency domain with S-subcarriers. In this case, the WiFi signal can be modeled as Channel Impulse Response (CIR) $h(\tau)$ in frequency domain:

$$h(\tau) = \sum_{l=1}^{L} \alpha_l e^{j\phi_l} \delta(\tau - \tau_l) \tag{1}$$

where α_l and ϕ_l represent the amplitude and phase [21] of the l-th multi-path component respectively, τ_l is the time delay, L indicates the total number of multi-path and $\delta(\tau)$ denotes the Dirac delta function. The time domain channel response $h(t)$ can be derived by taking the Inverse Fourier Transform (IFFT) of the Channel Frequency Response (CFR). In reality, an $p \times q$ MIMO system in a narrow-band flat fading channel is given by the following equation:

$$Y_i = H_i X_i + N_i, \ \ i \in [1, S] \tag{2}$$

where p and q are the number of transmit and receive antennas respectively, N_i represents the noise and H_i represents the CSI for subcarrier i. Every transmitting of a preamble of OFDM symbols can generate a channel matrix estimation by WiFi devices. This channel matrix (CSI) is represented by:

$$H_i = \|H_i\| e^{j\angle H_i} \tag{3}$$

where both amplitude $\|H_i\|$ and phase $\angle H_i$ are included.

Conventional CSI-based systems employ Intel 5300 NIC tool [18] to extract CSI data from modified WiFi NIC card equipped with laptop or PC. These platform tools result in two limitations. Firstly, the system design is restricted because of the demand of laptops used as receivers. For example, the authors of [22] adopt more than 10 laptops as receivers in the indoor localization system. This leads to high cost, impeding the realization of large-scale systems. The second limitation is related to the number of subcarriers, which determines the resolution of CSI apart from antenna number. Intel 5300 NIC

tool only records CSI for 30 out of the total 56 subcarriers in 20Mhz bandwidth. To overcome these bottlenecks, we develop an innovative OpenWrt based firmware for kinds of commodity WiFi routers by means of upgrading the Atheros CSI Tool [14]. The new firmware allows the CSI measurements to be obtained in the routers rather than laptops, which can promote some applications to the next level in the future. Furthermore, our platform provides complete CSI data of 114 subcarriers for 40MHz bandwidth for each TX-RX pair when adopting 5GHz [14]. Fine-grained CSI data lays a foundation for the implementation of accurate activity recognition.

In FALAR, we only require two routers adopted as transmitter (TX) and receiver (RX). TX continues to send data packets to RX while RX monitors the channel and reports the CSI. Let N_T and N_R denote the number of transmit and receive antennas. At each time instant t, we can obtain $N_T \times N_R \times 114$ CSI streams, and for each TX-RX pair, the CSI data $\mathbf{h_t}$ is a vector in $\mathbb{R}^S$, which is used to extract the activity foreground in Section III-B.

B. Motion Detection based on Kernel Density Estimation

In activity recognition, the first step is to detect the happening activity with the interference of intrinsic noise in channel estimation and subtle environmental changes. To solve this issue, we investigate a similar topic, *foreground detection* in the computer vision field, where the problem is to subtract the background and identify the foreground object in a consecutive video frames [23]. In activity detection problem, the background is environmental noises and foreground is the fluctuation of CSI caused by human motions. The authors of Smokey [24] address the problem by Gaussian Mixture Model (GMM), but GMM requires too many empirical parameters, which reduces the availability in practice. Inspired by it, we propose a foreground detection method by means of kernel density estimation and coarse-to-fine search strategy.

CSI frame and pixel: At a time instant t, CSI frame is represented by a S dimension vector h_t mentioned in Section III-A. In our method, FALAR utilizes the amplitude of the i-th subcarrier CSI at time t, denoted by $x(t)_i, i \in [1, S]$ (namely a pixel). Our platform can provide $500Hz$ sampling rate so that we can obtain $x(t)_i$ every $T = 2ms$. For each subcarrier, we always have N_b background samples B in terms of the CSI frames from $t - N_b$ to $t - 1$.

KDE-based foreground detection. To separate the foreground and background, we model each pixel value by Gaussian distribution and update the background B in each frame. We introduce these essential observations as conditions: (1) The intrinsic background noise on a single subcarrier follows Gaussian distribution. (2) Human motions bring about variations of successive subcarriers of CSI. Here comes each step of KDE-based foreground detection.

(1) Background Subtraction: Considering that each subcarrier obeys a Gaussian distribution, we estimate each pixel individually. Let $x(t)_{i,j}$ be the j-th sample of i-th pixel in background B. Using these background samples, we can obtain an estimate probability density function of the pixel CSI value at any value using kernel density estimation. Given the observed CSI $x(t)_i$ at time t, we can estimate the probability of this observation as:

$$Pr(x(t)_i) = \frac{1}{N_b} \sum_{j=1}^{N_b} K_\sigma (x(t)_i - x(t)_{i,j}) \tag{4}$$

where K_σ is a kernel function with bandwidth σ. In FALAR, we choose Gaussian to be our kernel function due to the observation, and accordingly, the estimated probability can be derived by:

$$Pr(x(t)_i) = \frac{1}{N_b} \sum_{j=1}^{N_b} \frac{1}{\sqrt{2\pi}\sigma_i} \mathrm{e}^{\frac{(x(t)_i - x(t)_{i,j})^2}{2{\sigma_i}^2}} \tag{5}$$

Employing this probability, the pixel is classified as a foreground pixel if $Pr(x(t)_i) < th_i^1$, where th_i is a global threshold that controls the percentage of false detection. We can find that the major issue is the choice of suitable kernel bandwidth σ_i. A small bandwidth can result in a ragged estimate while too big will lead to a smoothed one.

To estimate the kernel bandwidth σ_i, we introduce the method developed in image processing work [12]. Calculate the median m_i of $\|x(t)_{i,j} - x(t)_{i,j+1}\|$ in the background samples. Assuming this distribution obeys Gaussian distribution $N(\mu, {\sigma_i}^2)$, then the distribution for the deviation $(x(t)_{i,j} - x(t)_{i,j+1})$ is deduced to be a normal distribution $N(0, 2{\sigma_i}^2)$. Then the median of the absolute deviation m_i is equivalent to the quarter percentile of the deviation distribution, which is

$$Pr(N(0, 2{\sigma_i}^2) > m_i) = 0.25 \tag{6}$$

The standard deviation of the distribution of the i-th pixel can be estimated as

$$\sigma_i = \frac{m_i}{0.68\sqrt{2}} \tag{7}$$

which is the kernel bandwidth.

(2) Probabilistic reduction of False Detection: The pure KDE-based approach can eliminate the intrinsic noise but what about some inconsequential fluctuations of CSI pixels caused by background changes or movements of people nearby? We can address this challenge by probabilistic suppression. If the pixel has been classified as a foreground pixel, we are supposed to check its neighborhoods, which represents the motion intensity and duration intuitively. The changes of only few consecutive pixels are mostly caused by inessential motion.

Let $x(t)$ be the observed value of a foreground pixel at time t. The pixel displacement probability $P_\mathcal{N}(x(t))$ can be defined to be the maximum probability in which $x(t)$ belongs to the background distribution of some point in the neighborhood $\mathcal{N}$ of $x(t)$

$$P_\mathcal{N}(x(t)) = \max_{y \in \mathcal{N}(x)} Pr(x(t) \| B_y) \tag{8}$$

where B_y is the background sample for pixel y and the calculation of probability estimation can be achieved by Equation (5). Then the upper bound restriction of $P_\mathcal{N}$ is able to eliminate some false detections by small motions.

On the whole, once a foreground pixel is detected at time t by $Pr(x(t)_i) < th_i^1 \wedge P_{\mathcal{N}}(x(t)_i) > th_i^2$, we regard this frame as a feasible frame. Once the number of consecutive feasible frames reaches L ($L = 1000$ in our method), we consider the $S \times L$ CSI matrix H as a feasible detection interval, which should be processed in the following subsections.

(3) Online Update and Coarse-to-fine Search Strategy: The main limitation is high calculation complexity when the background sample set B keeps updating. As a result, we develop an updating policy that the model is updated B every 500 samples ($1Hz$) until feasible frame comes up. The updating will be accelerated to every 50 samples ($10Hz$) during human activities until no feasible frames are obtained. In this way, the detection speed can reach 200fps in practice.

Our KDE-based detection model provides a coarse-to-fine search strategy for activity recognition. Conventional sliding window approach has fixed step size, which leads to some unnecessary calculation for activity recognition. In comparison, our method can detect an activity and then use advanced techniques to recognize it, which improves the efficiency and detection accuracy. This generic detection model can be utilized in other applications.

C. Noise Removal

Raw CSI data can be considered to be intrinsically noisy for recognition and thus needs preprocessing. Our denoise process is composed of interpolation and signal smoothing.

Interpolation: Although the transmitter sends the package at a very high frequency, the receiver cannot report CSI with the same frequency due to packet loss and lack of support for high-level task in OpenWrt. To obtain uniform time sequence of CSI, we employ linear interpolation to ensure the stationary interval of consecutive frames.

Signal smoothing: Due to variation of transmission power, some high frequency noise exists in the time domain of all subcarriers. To address this problem, we utilize a Discrete Wavelet Transform (DWT) based denoising approach to remove it. Wavelet transform uses various transforms $\psi_{a,b}(t) = a\psi(t-b)$ obtained from a basic function through scaling and translation in terms of parameter a and b respectively. The wavelets coefficients of our CSI time sequence $r(t)$ can be calculated by:

$$W(a,b) = \int_{\infty}^{-\infty} \psi_{a,b}^{*}(t) r(t) dt \tag{9}$$

where $*$ is the complex conjugate and *Symlet* is chosen to be the basic function ψ. After DWT, we filter the wavelet coefficients and get the sanitized CSI data $\hat{H}$ by the inverse transformation. By virtue of denoising scheme, fine-grained CSI data is ready for activity recognition.

D. CSVD-NMF Scheme for Activity Recognition

(1) Activity Extraction via Class Estimated Basis Space Singular Value Decomposition: We aim to extract activity from CSI data wherever it happens. The information of human activity mainly refers to CSI caused by multi-path effect. In previous analysis of CSI data [25], it is proved that CSI dominant component of CSI lies in CSI line-of-sight path, which does not help. In this case, we develop a CSVD-based reconstruction method to eliminate main components of static paths in CSI data.

For each detection CSI matrix H in $\mathbb{R}^{S \times L}$ space, we can compute the singular value decomposition of it, $U\Sigma V^T = H$. Note that $U^T U$ and $V^T V$ both yield the identity matrix so they both have orthonormal columns. The matrix Σ is a diagonal matrix whose diagonal elements $\{\sigma_1, \sigma_2, ..., \sigma_\gamma\}$ are singular values in decreasing order, $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_\gamma \geq 0$. γ is the number of singular values. The prominent property of the SVD is that when we reconstruct an approximation to a matrix M, the singular values perform decreasing importance. Explicitly, we define rank-K approximation to have first K singular values in Σ but others are zero. It can be proven that rank-K approximation is always the closest rank-K matrix to M. That is $M_K = argmin_A \|M - A\|_F$, where the operation $\|\cdot\|_F$ is the Frobenius norm.

In our application, to eliminate the main component, we specify the first singular value to be zero and reconstruct a location-independent approximation of CSI as follows:

$$H_{\sigma_1=0} = \sum_{i=2}^{\gamma} \sigma_i u_i {v_i}^T \tag{10}$$

where u_i and v_i are the singular vectors corresponding to the singular value σ_i. Intuitively, we can find the similarity of $H_{\sigma_1=0}$ of the same activity in different locations. In the training phase, we calculate the average CSI of each class and reconstruct the CSI for NMF classification.

(2) Automatic Feature Extraction and Classification using Nonnegative Matrix Factorization: We adopt nonnegative matrix factorization to conduct automatic feature extraction and classification. We discover that most obvious features of the CSI are determined by part of pixels in the whole frame. Although hand-crafted features can be mined in time and frequency domain, different activities may have discriminative feature. Therefore, we develop this method to learn a robust representation for classification automatically. Surprisingly, it also permits fast extension of new types of activities with only several times of calibration.

Training phase: NMF employs non negative constraint to learn part-based representation. After normalization to guarantee non negative property, we can decompose each class of training data obtained by CSVD, $V_{S \times L} = [v_1, v_2, ..., v_L]$, to its subspaces, i.e. $V \approx WH_c$, where $\sum_{i=1}^{n} v_{ij} = 1, j = 1, 2, ..., L;\ v_{i,j} \geq 0$, $W \in \mathbb{R}^{S \times K}$ is the base representation of CSI and $H_c \in \mathbb{R}^{K \times L}$ denotes the coefficient. K is the number of basis of feature subspace W. Each column $v_i \in \mathbb{R}^{n=S \times L}$ represents a training sample for one activity. This problem can be solved using divergence approximation $Y = WH_c$. Then it is equivalent to the optimization problem:

$$\begin{aligned} \min_{W,H_c} D(V\|WH_c) &= \sum_{i,j} (v_{ij} log_n \frac{v_{ij}}{y_{ij}} - v_{ij} + y_{ij}) \\ s.t.\ \sum_{i=1}^{n} w_{ij} &= 1, \forall j \ and \ W, H_c \geq 0 \end{aligned} \tag{11}$$

TABLE I
COMPARISON OF TRUE POSITIVE RATE (TPR)

Methods	Sitting (%)	Walking (%)	Running (%)	Overall
E-eyes	94.13	88.91	85.06	89.37
CARM	97.58	92.82	91.74	94.05
FALAR	96.21	99.52	96.33	97.35

The optimization result W, H_c are regarded as feature subspaces matrices for one activity and we train every activity sample in the same manner. Note that training samples number can be very small and the extension of new class does not require extra training of classifier for the whole dataset, which provides convenience for the training of new activity.

Classification phase: For a new test matrix X, we project it to the feature subspace of k-th activity feature subspace by $H_x = W_k^{-1}X$ and H_x represents the reconstruction of X using the basis of the new subspace W_k. Then we compare the coefficients H_x with original one H_k and use the normalized error as the classification criteria:

$$\varepsilon_k = \frac{\|H_x - H_k\|}{\|H_x\|}, \ k = 1, 2, ..., c \tag{12}$$

where c is the total number of training samples. This measures the similarity between detected CSI and each activity sample and then we regard it as the distance for K Nearest Neighbors (KNN) algorithm. X belongs to the most probable class in terms of its nearest neighbors. Thus, the online classification only needs matrix multiplication rather than costly hand-crafted feature extraction.

IV. EVALUATION

We implemented a prototype of FALAR based on two TP-LINK N750 wireless dual band routers, of which one with single antenna acts as TX while the other one with triple antennas is RX. Having upgraded their firmware to our OpenWrt system, we can directly obtain the fine-grained CSI measurement from routers. TX was set to 802.11n AP mode at 5GHz with a 40MHz bandwidth and the client, RX, was connected to this WLAN network. In a frame, the size of CSI streams is 3×114. The experiments were conducted in a normal office in Figure 2, where staffs were working in their positions. This situation leads to environmental interference for our system. Our evaluation consists of three parts in order to validate the recognition accuracy in different locations, extension of new activity and robustness. The empirical parameters are set as: $S = 3 \times 114$, $L = 1000$, $th^1 = 0.1, th^2 = 0.4$ in KDE, $K = 5$ in NMF and four nearest neighbors in KNN.

We adopt two metrics to evaluate the recognition accuracy. **True Positive Rate (TPR):** the ratio of the number of times for correctly recognizing an activity to the total number of activity. **False Positive Rate (FPR):** the fraction of cases that are incorrectly recognized as an activity among all testing samples other than activity. **Confusion Matrix:** each cell represents the recognition accuracy in terms of one activity.

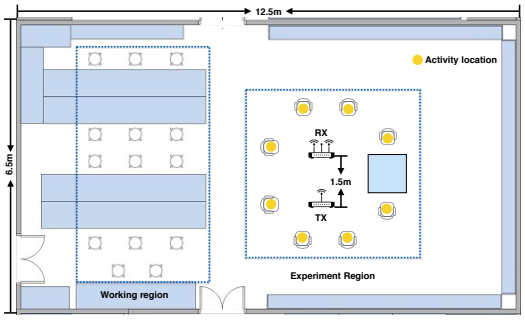

Fig. 2. The layout and system setup of experiment.

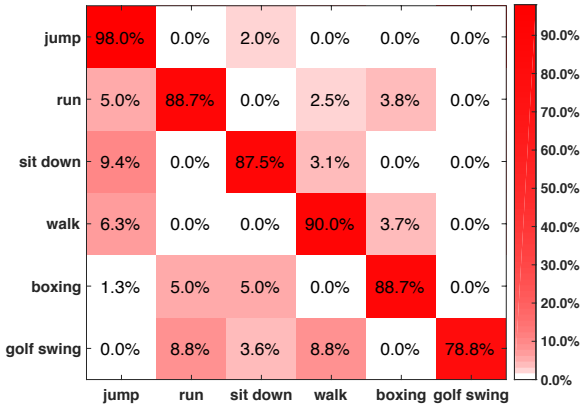

Fig. 3. Confusion matrix of recognition accuracy.

A. Recognition accuracy:

The data of activities, including *jump, run, sit down, walk, box*, was collected in eight fixed locations marked in Figure 2 and each class has 20 samples in each location (totally 160 samples). We randomly use 20 samples in five positions for training while other 140 samples in eight positions for testing. The result is performed by confusion matrix. To compare its performance with two state-of-the-art CSI based activity recognition methods, E-eyes [19] and CARM [20] in terms of *sitting, walking, running*, we have to keep the same setting as these methods. We use 5 samples in all positions for training and others as testing. As shown in Figure 3, the accuracy of *jump* reaches 98% and the other four activities achieve about 90% accuracy. The total average accuracy is calculated as 90.6% with only 20 training samples. In table I, we can see that our method improves the overall TPR by 3.3% and 7.98% over E-eyes and CARM respectively. The reason is that they adopt limited hand-crafted features of coarse CSI data with 30 subcarriers information.

B. Extension evaluation of new activities

To demonstrate the performance of extension ability, we add a new challenging activity *Golf swing*. The average accuracy still reaches more than 75% in terms of only 5 training samples shown in the last row in Figure 3, while the advanced methods such as [26] utilize 150 samples of each activity to train the model. We also investigate the relationship between number of

TABLE II
COMPARISON OF FALSE POSITIVE RATE (FPR)

Methods	Sitting (%)	Walking (%)	Running (%)	Overall
E-eyes	5.53	10.93	12.33	9.60
CARM	2.24	6.48	8.22	5.65
FALAR	2.35	0.84	3.63	2.73

samples and accuracy. The accuracy increases as the number of training samples are enlarged, but over 30 samples, the enhancement is not obvious and may bring about overfitting.

C. Robustness:

Table II demonstrates the comparison of FPT between three activities. Our method achieves best FPR of 2.73% over E-eyes of 9.60% and CARM of 5.65%. The robustness of our system is because of effective features and KDE-based detection that distinguishes activity from the background. Without it, the system might wrongly detect motions caused by environmental changes rather than human beings.

V. CONCLUSION

In this paper, we investigate the properties of CSI data to propose an indoor sensing framework, FALAR. To get fine-grained CSI, we develop a new firmware for OpenWrt. To overcome inconsequential changes, we firstly utilize KDE to accurately extract the motions. To address the problem of activity locations, we use CSVD to eliminate most static paths. To improve the scalability of FALAR, we design an automatic feature extraction phase based on nonnegative matrix factorization, which allows the training of new activity is absolutely independent. The evaluation of FALAR demonstrates the average accuracy of more than 90% in challenging environment. The limitation of FALAR is that the activity in the line-of-sight has not been included and investigated. In the future, WiFi-based, vision-based and wearable device-based method can be complementary to achieve seamless human activity recognition architecture.

REFERENCES

[1] R. Bodor, B. Jackson, and N. Papanikolopoulos, "Vision-based human tracking and activity recognition," in *Proc. of the 11th Mediterranean Conf. on Control and Automation*, vol. 1, 2003.

[2] Z. Chen, Q. Zhu, C. S. Yeng, and L. Zhang, "Robust human activity recognition using smartphone sensors via ct-pca and online svm," *IEEE Transactions on Industrial Informatics*, 2017.

[3] B. Huang, G. Qi, X. Yang, L. Zhao, and H. Zou, "Exploiting cyclic features of walking for pedestrian dead reckoning with unconstrained smartphones," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*. ACM, 2016, pp. 374–385.

[4] H. Zou, Z. Chen, H. Jiang, L. Xie, and C. Spanos, "Accurate indoor localization and tracking using mobile phone inertial sensors, wifi and ibeacon," in *Inertial Sensors and Systems (INERTIAL), 2017 IEEE International Symposium on*. IEEE, 2017, pp. 1–4.

[5] H. Zou, Y. Zhou, J. Yang, W. Gu, L. Xie, and C. Spanos, "Freedetector: device-free occupancy detection with commodity wifi," in *Sensing, Communication and Networking (SECON Workshops), 2017 IEEE International Conference on*. IEEE, 2017, pp. 1–5.

[6] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, "A survey on behavior recognition using wifi channel state information," *IEEE Communications Magazine*, vol. 55, no. 10, pp. 98–104, 2017.

[7] H. Zou, Y. Zhou, J. Yang, W. Gu, L. Xie, and C. Spanos, "Multiple kernel representation learning for wifi-based human activity recognition," in *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Dec 2017, pp. 268–274.

[8] W. Xi, J. Zhao, X.-Y. Li, K. Zhao, S. Tang, X. Liu, and Z. Jiang, "Electronic frog eye: Counting crowd using wifi," in *Infocom, 2014 proceedings ieee*. IEEE, 2014, pp. 361–369.

[9] H. Zou, Y. Zhou, J. Yang, W. Gu, L. Xie, and C. Spanos, "Freecount: Device-free crowd counting with commodity wifi," in *Global Communications Conference (GLOBECOM), 2017 IEEE*. IEEE, 2017.

[10] Y. Zeng, P. H. Pathak, and P. Mohapatra, "Wiwho: wifi-based person identification in smart spaces," in *Proceedings of the 15th International Conference on Information Processing in Sensor Networks*. IEEE Press, 2016, p. 4.

[11] H. Zou, Y. Zhou, J. Yang, W. Gu, L. Xie, and C. Spanos, "Wifi-based human identification via convex tensor shapelet learning," in *AAAI*, 2018.

[12] A. Elgammal, R. Duraiswami, D. Harwood, and L. S. Davis, "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance," *Proceedings of the IEEE*, vol. 90, no. 7, pp. 1151–1163, 2002.

[13] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[14] Y. Xie, Z. Li, and M. Li, "Precise power delay profiling with commodity wifi," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*. ACM, 2015, pp. 53–64.

[15] Y. Zhao and N. Patwari, "Noise reduction for variance-based device-free localization and tracking," in *Sensor, Mesh and Ad Hoc Communications and Networks (SECON), 2011 8th Annual IEEE Communications Society Conference on*. IEEE, 2011, pp. 179–187.

[16] G. Wang, Y. Zou, Z. Zhou, K. Wu, and L. M. Ni, "We can hear you with wi-fi!" *IEEE Transactions on Mobile Computing*, vol. 15, no. 11, pp. 2907–2920, 2016.

[17] H. Zou, H. Jiang, J. Yang, L. Xie, and C. Spanos, "Non-intrusive occupancy sensing in commercial buildings," *Energy and Buildings*, vol. 154, pp. 633–643, 2017.

[18] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11 n traces with channel state information," *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 1, pp. 53–53, 2011.

[19] 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 *Proceedings of the 20th annual international conference on Mobile computing and networking*. ACM, 2014, pp. 617–628.

[20] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of wifi signal based human activity recognition," in *Proceedings of the 21st annual international conference on mobile computing and networking*. ACM, 2015, pp. 65–76.

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

[22] J. Wang, H. Jiang, J. Xiong, K. Jamieson, X. Chen, D. Fang, and B. Xie, "Lifs: Low human effort, device-free localization with fine-grained subcarrier information," in *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*. ACM, 2016, pp. 243–256.

[23] K. Patwardhan, G. Sapiro, and V. Morellas, "Robust foreground detection in video using pixel layers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 4, pp. 746–751, 2008.

[24] X. Zheng, J. Wang, L. Shangguan, Z. Zhou, and Y. Liu, "Smokey: Ubiquitous smoking detection with commercial wifi infrastructures," in *Computer Communications, IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on*. IEEE, 2016, pp. 1–9.

[25] J.-Y. Chang, K.-Y. Lee, Y.-L. Wei, K. C.-J. Lin, and W. Hsu, "Location-independent wifi action recognition via vision-based methods," in *Proceedings of the 2016 ACM on Multimedia Conference*. ACM, 2016, pp. 162–166.

[26] M. A. A. Al-Qaness, F. Li, X. Ma, Y. Zhang, and G. Liu, "Device-free indoor activity recognition system," *Applied Sciences*, vol. 6, no. 11, p. 329, 2016.