WiHAR: From Wi-Fi Channel State Information to Unobtrusive Human Activity Recognition

Muhammad Muaaz, Ali Chelli, and Matthias Pätzold
Faculty of Engineering and Science, University of Agder, 4898 Grimstad, Norway.
Emails: {muhammad.muaaz, ali.chelli, matthias.paetzold}@uia.no

***Abstract*—A robust and unobtrusive human activity recognition system is essential to a multitude of applications, such as health care, active assisted living, robotics, sports, and tele-immersion. Existing well-performing activity recognition methods are either vision- or wearable sensor-based. However, they are not fully passive. In this paper, we develop WiHAR—an unobtrusive Wi-Fi-based activity recognition system. WiHAR uses the Wi-Fi network interface card to capture the channel state information (CSI) data. These CSI data are effectively processed, and then amplitude and phase information is used to obtain the spectrogram. In the subsequent step, the time-variant mean Doppler shift (MDS) caused by the human body movements in the radio signals before their arrival at the receiver is estimated. The MDS is used to extract time and frequency domain features that are needed to train the supervised learning algorithms (i.e., decision tree, linear discriminant analysis, and support vector machines (SVM)) to assess the performance of the WiHAR. Our results show that WiHAR combined with SVM achieves 96.2% recognition accuracy on the data set consisting of 9 participants where each participant performed four activities including: walking, falling, picking up an object from the ground, and sitting on a chair.**

***Index Terms*—wireless healthcare, machine learning, channel state information, spectrogram, mean Doppler shift**

I. Introduction

Human activity recognition (HAR) aims at classifying activities performed by an individual or a group of individuals based on sensor observations. In recent years, HAR has gained significant research interest due to the increasing number of applications that need to monitor human activities and behaviour in indoor environments. These applications include: fall detection in elderly or disabled people, health care, active assisted living (AAL) for physically or cognitively impaired individuals, rehabilitation, sports, and augmented reality. Existing HAR systems can generally be classified into three categories based on the type of sensor placed in the environment or used to capture the human activity data: (1) vision-based [1], [2], (2) wearable sensor-based [3], [4], and (3) radio frequency (RF)-based [5], [6]. Vision-based HAR systems exploit cameras to record videos or take images of users while performing activities. Later, image/video processing algorithms and learning techniques are applied to determine the activities. Although in recent years, vision-based systems have experienced great improvements (e.g., enhanced sensing capabilities of cameras and improvements in image and video processing algorithms), they still suffer from various challenges. For instance, they require the unobstructed view of the environment, their performance degrades significantly under low lighting conditions, and they must record video that may raise privacy concerns. Wearable sensor-based systems mostly use inertial measurement units (IMUs) to capture human activity data and later this data is processed to determine human activities. Unlike vision-based systems, the wearable sensor-based systems are cost-effective, preserve the user's privacy, and are not affected by the lighting conditions. However, they also suffer from some limitations. For instance, the wearable sensor-based systems only work if users wear the sensors, which might be uncomfortable for elderly and physically or cognitively impaired persons. Furthermore, wearable sensor-based systems require that sensors are placed on the human body according to the recommendations of the system provider, otherwise they might lead to inaccurate results.

RF-based HAR systems exploit the phenomenon that human bodies reflect RF signals, and human activities introduce variations in the RF signals present in the surrounding [6]. The RF-based HAR system requires that an RF transmitter and a receiver are deployed in the indoor environment and set to continuously transmit and receive RF signals. In this way, variations caused by human activities are recorded continuously. These received RF signals are used for activity recognition [6]. Unlike vision- and wearable sensor-based systems, RF-based approaches do not require users to wear sensors on their bodies or worry about the privacy concerns. Therefore, they are a suitable choice for activity recognition in health care or indoor environments. In recent years, researchers have actively explored and developed RF-based HAR techniques. These techniques either employ radar [7]–[9] or Wi-Fi-based [5], [6], [10], [11] techniques to capture the variations caused by human activities. Radar-based HAR systems on one side offer large bandwidth, which results in a higher spatial resolution, thus they can be used for fine grained human activity recognition. On the other side, radar-based systems require an expensive hardware setup, which limits their large-scale deployment. Unlike radar-based systems, Wi-Fi-based HAR systems do not require special hardware and can be integrated into readily available Wi-Fi infrastructures in homes, offices, or public places.

In Wi-Fi-based HAR systems, the attributes of the received RF signal (i.e., received signal strength indicator (RSSI) and the channel state information (CSI)) have been studied for

This work is carried out within the scope of WiCare project funded by the Research Council of Norway under the grant number 261895/F20.

the activity recognition task. Recent works reported that CSI-based systems yield significantly better results when compared to RSSI-based systems [6]. However, they suffer from a limitation, i.e., *they do not perform well due to changes in the environment* [12]. This is because the received RF signals contain information about the moving and non-moving objects present in the environment. Therefore, a CSI-based HAR system trained by using solely the amplitude variation of the CSI data collected in an environment may perform poorly when deployed in a different environment. One solution to this problem requires environment and subject-independent features to train HAR models [12]. However, this approach needs a lot of training data collected from a large number of subjects under different environmental settings. Another alternative is to use a semi-supervised learning approach, which requires manual labelling of new data upon detecting changes in the fingerprints of the activities due to chances in the environment [10]. However, this solution is not very practical because it requires user interaction which may not be feasible for elderly or cognitively impaired individuals.

In contrast to prior work, in WiHAR, we apply the spectrogram approach to estimate the mean Doppler shift (MDS) from the CSI data associated with different human activities. The estimated MDS is used to extract time and frequency domain features for classification. Since, the MDS captures the impact of the amplitude and phase of the CSI, WiHAR is robust to changes in the environment. In other words, if the position of the fixed scatterers (e.g., furniture, walls) is modified, this has no impact on the recognition accuracy. Moreover, our investigation shows that WiHAR reports a high recognition accuracy, even at greater distances. For example, three out of the four activities tested are performed at a distance of 4 meters from the transmitter and the receiver. The main contributions of our paper are as follows:

- We collect CSI data from 9 participants using a network interface card (NIC) in an indoor environment in the presence of a single person performing four activities: walking, falling, sitting on a chair, and picking up an object from the ground.
- We apply the spectrogram approach to estimate the environment independent MDS associated with each collected activity sample.
- We extract time and frequency domain features from the MDS and evaluate the performance of the WiHAR by applying supervised machine learning classifiers namely: decision tree, linear discriminant analysis (LDA), and support vector machines (SVM).

The rest of the paper is organized as follows. Section II provides an overview of WiHAR. Section III provides a comprehensive view of channel modelling and demonstrates the relation of the CSI to the channel transfer function (CTF). Details about our experimental setup and data collection process are given in Section IV. In Section V, we explain steps to process the CSI data and estimating the MDS. This section also highlights the features extracted from the MDS. In Section VI, we apply the supervised machine learning technique to determine the performance of the WiHAR in terms of classification accuracy and discuss the results. Final conclusions are drawn in Section VII.

II. Overview of the WiHAR System

The WiHAR system consists of RF sensing and machine learning phases. In the RF sensing phase, a radio transmitter and a receiver are deployed in an indoor environment. The transmitter and receiver are two NICs that are configured to operate at 5.745 GHz Wi-Fi band. The transmitter continuously emits electromagnetic waves that propagate in the environment. When these waves hit static (e.g., furniture and walls) and moving objects (e.g., a moving person) present in the environment, they are reflected before their arrival at the receiver. The movements of human body parts cause Doppler shift in the received RF signal, and different types of movements produce different Doppler shift patterns. The received RF signal is regarded as a CTF, which is processed, and the spectrogram is computed. From the spectrogram, the MDS is determined which shows the variations in the Doppler frequency with respect to time as the person performs an activity. Thus, as the activity changes, the pattern of the MDS also changes. In the machine learning phase, the MDS can be regarded as a time-series from which various time and frequency domain features are computed. These features are fed to a machine learning algorithm to determine the type of activity performed by the person.

III. Channel Modelling Background

A. The Channel Transfer Function

As mentioned in the previous section, in the RF sensing phase, a radio transmitter and a receiver are deployed in an indoor environment. A single antenna is attached to the transmitter T_x and two antennas to the receiver R_x. The transmitter and receiver work in the Wi-Fi injector-monitor mode. Using the CSI tool [13], we inject random data packets in the RF channel, and the receiver receives these data packets and reports the estimated CSI along 30 (orthogonal frequency-division multiplexing) OFDM subcarriers. Complete details about the experimental setup are given in Section IV.

This CSI data is a matrix with a dimension of $N_{T_x} \times N_{R_x} \times K$, where N_{T_x}, N_{R_x}, and K stand for the number of transmitting antennas, the number of receiving antennas, and the number of OFDM subcarriers, respectively. Since we have used a single antenna on the transmitter side and two antennas on the receiver side, the dimension of our CSI data matrix is $1 \times 2 \times 30$. The measured CSI data corresponds to the CTF $H_{i,j}(f'_k, t)$ of the transmission link between the ith transmit antenna and the jth receive antenna, measured at the kth subcarrier f'_k. The expression of f'_k is given by $f'_k = f'_0 + k\Delta f'$, where f'_0, k, and $\Delta f'$ indicate the carrier frequency, the subcarrier index, and the subcarrier bandwidth, respectively. The time-series of the CTF values for a given antenna pair and a subcarrier is called a CSI stream. Thus, for each transmission link, (i.e., the transmit and receive antenna pair) there exist 30 CSI streams.

To understand the influence of the performed activities on the indoor RF channel, it is important to have a channel model that explains the contributions of fixed (i.e., walls, furniture, ceiling) and moving (i.e., moving person) scatterers to the CTF $H(f'_k, t)$[1]. During the measurement campaign, only a single person is moving inside the room and all other objects in the room are static. The static objects in the room are modelled by N fixed scatterers S_n^F $(n = 1, 2, \ldots, N)$. Whereas, moving objects, such as different body parts of the moving person (i.e., head, torso, arms, and legs) are modelled by M moving scatterers S_m^M $(m = 1, 2, \ldots, M)$. Therefore, the CTF of our indoor channel model can be expressed by summing the terms associated with fixed and moving scatterers as given below,

$$H(f'_k, t) = \sum_{m=1}^{M} c_m \exp\left(j\theta_{m,k}(t)\right) + \sum_{n=1}^{N} c_n \exp\left(j\theta_{n,k}\right) \tag{1}$$

where the path gains associated with the mth moving scatterer and nth fixed scatterer are represented by c_m and c_n, respectively. Here $\theta_{n,k}$ represents the phase associated with the nth fixed scatterer and the kth subcarrier. $\theta_{n,k}$ is an independent identically distributed (i.i.d.) random variable, which is uniformly distributed in the interval $[0, 2\pi)$. The phase $\theta_{m,k}(t)$ associated with the mth moving scatterer and the kth subcarrier is time-variant and can be expressed as [14]

$$\theta_{m,k}(t) \quad = \quad \theta_{m,k} - 2\pi f'_k \tau'_m(t) \tag{2}$$

where $\theta_{m,k}$ represents the initial phase at $t = 0$, which is uniformly distributed in the interval $(0, 2\pi]$. The symbol $\tau'_m(t)$ indicates the time-variant delay associated with the mth moving scatterer. The Doppler frequency due to the motion of the mth moving scatterer can be computed as [15]

$$\begin{aligned} f_m(t) = -f_{m,\max}(t) \Big\{ & \cos\left(\beta_m^v(t)\right) \Big[\cos\left(\beta_m^T(t)\right) \\ & \cos\left(\alpha_m^T(t) - \alpha_m^v(t)\right) + \cos\left(\beta_m^R(t)\right) \cos\left(\alpha_m^v(t) - \alpha_m^R(t)\right) \Big] \\ & + \sin\left(\beta_m^v(t)\right) \left[\sin\left(\beta_m^T(t)\right) + \sin\left(\beta_m^R(t)\right) \right] \Big\} \end{aligned} \tag{3}$$

where

$$f_{m,\max}(t) = \frac{v_m(t)}{c_0} f'_k. \tag{4}$$

The functions $\alpha_m^v(t)$ and $\beta_m^v(t)$ represent the azimuth and elevation angles of motion, respectively. The elevation angle of departure and elevation angle of arrival are represented by $\beta_m^T(t)$ and $\beta_m^R(t)$, respectively. The azimuth angle of departure, and the azimuth angle of arrival are represented by $\alpha_m^T(t)$, and $\alpha_m^R(t)$, respectively. The time-variant speed of movement of the mth moving scatterer is indicated by $v_m(t)$, and c_0 indicates the speed of light. The influence of the mth moving scatterer on the instantaneous channel phase can be expressed in terms of the Doppler frequency as given in (5) [16].

$$\theta_{m,k}(t) = 2\pi \int_{-\infty}^{t} f_m(u)du = \theta_{m,k} + 2\pi \int_{0}^{t} f_m(u)du. \tag{5}$$

The first term $\theta_{m,k}$ indicates the initial phase shift at time $t = 0$, which is modelled by a uniformly distributed random variable in the interval $[0, 2\pi)$ [16]. Once the instantaneous channel phases are obtained, the CTF comprising of fixed and moving scatterers can be modelled. As mentioned in the first paragraph of this section, every CSI stream $H_{i,j}(f'_k, t)$ corresponds to the CTF values associated with some antenna pair (i, j) and a subcarrier f'_k. Therefore, each CSI stream follows the mathematical model provided in (1).

B. Spectrogram and Mean Doppler Shift

To obtain the spectrogram of the CTF, we first compute the short-time Fourier transform (STFT) of the windowed CTF as given in (6) and subsequently the STFT is multiplied with its complex conjugate (see 7), which defines the spectrogram [17].

$$X_{H_k}(f, t) = \int_{-\infty}^{\infty} H_{i,j}(f'_k, t')g(t' - t)e^{-j2\pi f t'} dt' \tag{6}$$

$$S_{H_k}(f, t) = |X_{H_k}(f, t)|^2. \tag{7}$$

In (6), t', t, and $g(t)$ represent the running time, the local time, and the sliding window, respectively. Within the scope of this paper, we use a Gaussian window as expressed below

$$g(t) = \frac{1}{\sqrt{\sigma_w \sqrt{\pi}}} e^{-\frac{t^2}{2\sigma_w^2}}. \tag{8}$$

Here σ_w represents the spread of the window. The window function is positive, even, and has a unit energy. Finally, the MDS can be estimated from the spectrogram using the following expression

$$B_{H_k}^{(1)}(t) = \frac{\int_{-\infty}^{\infty} f S_{H_k}(f, t)df}{\int_{-\infty}^{\infty} S_{H_k}(f, t)df}. \tag{9}$$

IV. Data Collection and Experimental Setup

In this paper, we considered a scenario where four different human activities are monitored in an indoor environment as shown in Fig. 1. These activities include: walking, falling, sitting on a chair, and picking up a small object from the ground. We collected activity data from nine healthy participants. The demographics of the participants are given in Table I. We asked the participants to wait for a second before starting an activity and wait for another second after completing the activity. For the walking activity, the participants were asked to walk along a straight line from Point A to Point B and from Point B to Point A (see Fig. 1). Each participant repeated the walking activity five times in both directions. For the falling activity, a mattress was placed on the floor at Point B and

[1]For the ease of notation and simplicity, the subscripts indicating the antenna indices (i.e., i and j) are dropped in the rest of this subsection

participants were asked to stand still on the shorter side of the mattress and fall (towards and away from the antennas) on the mattress. Every participant repeated the falling activity five times in each direction. For the sitting activity, a chair without armrests was placed at Point B, and participants were asked to stand next to the chair and then sit on the chair as shown in Fig. 1. In the last activity, a small object was placed on the floor at Point B and the participants were asked to pick up this object. Every participant repeated the sitting and picking up the object activities five times facing towards the antennas.

We used the CSI Tool [13] to measure and parse the Wi-Fi CSI while users performed different activities. We installed the CSI Tool and necessary wireless drivers on two HP laptops (EliteBook 6930p, running Linux 14.04 LTS operating system) equipped with Intel 5300 wireless NICs. The NIC cards of the laptops were configured to run at a 5.745 GHz central frequency with a bandwidth of 20 MHz and in single input multiple outputs (SIMO) transmission mode. Instead of using the built-in RF antennas of the laptops, we used RF cables and connected external antennas to the NICs of the laptops. A single directional antenna is used on the transmitter (T_x) side to inject 1000 random data packets per second in the wireless medium. To collect the injected packets, two receive antennas (directional and omnidirectional) were used on the receiver (R_x) side. The antennas are attached to the table as shown in Fig. 1.

V. Data Processing

The collected CSI data consist of amplitude and phase information. Both, the amplitude and phase of the CSI data contain noise generated by the ambient environment, lack of synchronization between transmitter and receiver clocks, and automatic change in the transmission parameters (e.g., transmission power and transmission rate adaptation) [18]. The ambient noise and the automatic change in the transmission parameters affect both the amplitude and phase of the CSI data. Moreover, asynchronicity between transmitter and receiver clocks additionally introduces carrier frequency offset (CFO) and sampling frequency offset (SFO) to the phase.

Gender	Age				
-	21-30	31-40	41-50	51-60	61-65
Male	5	-	1	-	1
Female	-	2	-	-	-

TABLE I: The demographics of the participants of the study.

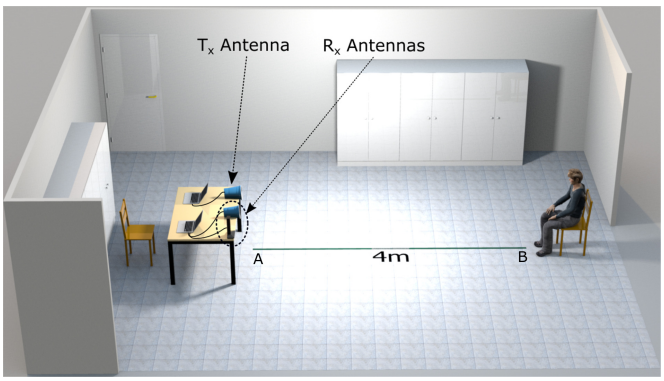
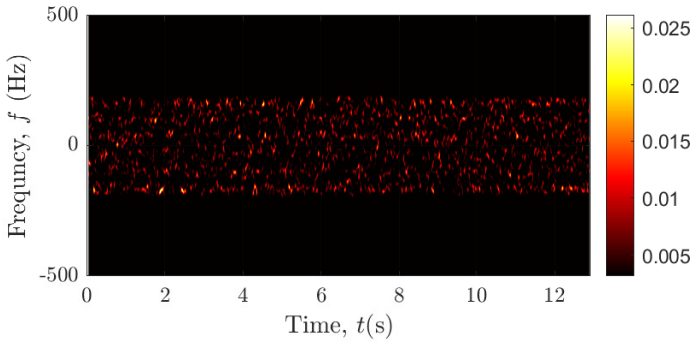


Fig. 1: Illustration of the experimental setup.

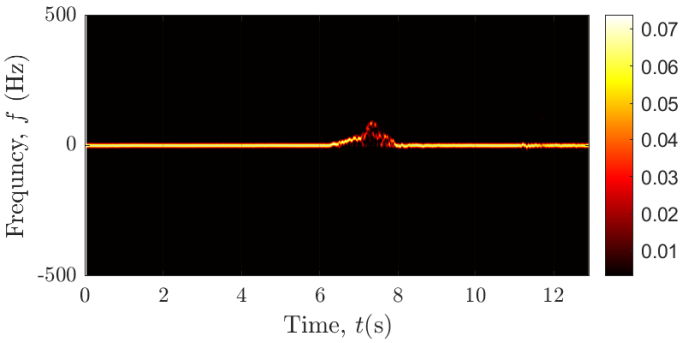
Traditional denoising methods could potentially reduce noise from the amplitude, but they do not effectively reduce the phase noise of the CSI data. Inspired by the previous work [19], we first apply the CSI ratio method that significantly reduces the influence of CFO and SFO on the phase information. The CSI ratio method requires that two antennas must be used on the receiver side to simultaneously collect CSI data $H_{1,1}(f'_k, t)$ and $H_{1,2}(f'_k, t)$ corresponding to each antenna. Thereafter, the ratio between the two CSI data is computed $H_{1,1}(f'_k, t)/H_{1,2}(f'_k, t)$. We have observed that, the spectrogram computed using the CSI ratio, shows clear fingerprints of any performed activity compared to the spectrogram computed by using the CSI data corresponding to any individual antenna, which shows nothing due to phase offset (see Fig. 2).

In addition to that, the CSI ratio technique does not require extra hardware like the back-to-back phase calibration technique [20]. For instance, the back-to-back phase calibration technique requires a splitter to establish a back-to-back channel. The splitter splits the transmitted signal in two identical signals and the signal received via the back-to-back channel is used to correct the phase of the signal received through the RF channel. Every single antenna on the receiver side reports the CSI measured along 30 subcarriers. This implies that CSI ratios along 30 subcarriers can be obtained. To reduce the number of correlated and redundant CSI ratios, we apply the principal component analysis (PCA) [21]. At this stage, we conducted several experiments to determine the influence of the number of PCA components on the spectrogram and the estimated MDS. Based on our experiments, we choose the first PCA component for the subsequent operations. A low pass filter is applied to the selected PCA component to further reduce the impact of out-of-band noise. The parameters of the low pass filter are chosen such that the Doppler shift caused by the person's activity is not distorted. After filtering, the spectrogram is computed as described in (6) and (8). Finally, we estimate the MDS from the spectrogram using (9).

We recall that during the data collection, the participant is inactive at the start and end of the activity. Therefore, we extract the segment from the MDS when the participant is active. To automatically detect the start and the end of the human activity, we monitor the variance of the MDS over a sliding time window of size 100 milliseconds. A variance equal to or higher than the threshold value marks the start of the active segment while a variance below the threshold value marks the end of the active segment. Once all segments are marked, we pick those segments which are equal to or longer than 1.5 seconds. This allows us to discard variations in the MDS that might have been introduced due to the small movements when the participant is inactive. Fig. 3 shows the segments that correspond to the MDS of the walking activity. Subsequently, Fig 4 shows the active MDS segments corresponding to four



(a) The spectrogram obtained with the CSI data of a single antenna



(b) The spectrogram obtained with the CSI ratio of two antennas

Fig. 2: The comparison of spectrograms obtained with and without applying the CSI ratio. It is important to note that both spectrograms were obtained, from the same trial, of the same activity, performed by the same participant. Furthermore, all other data processing steps described in Section V are kept the same.

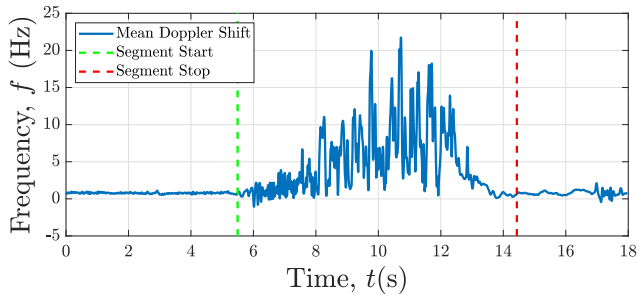
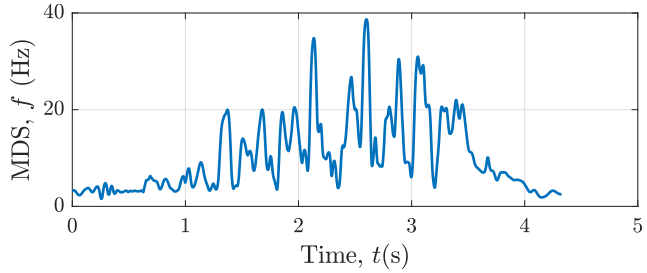
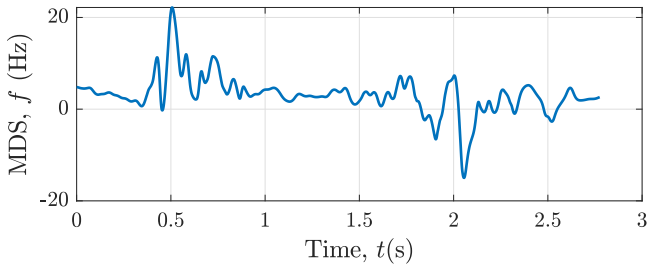


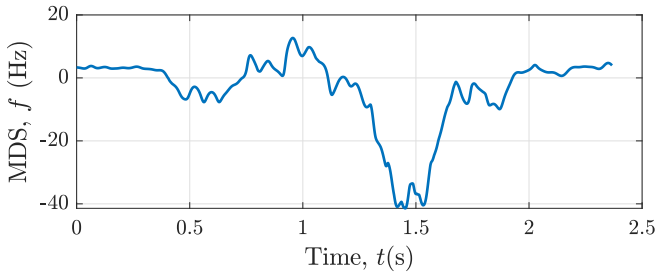
Fig. 3: Extraction of the active MDS segment from the MDS of a walking activity trial with inactive periods.



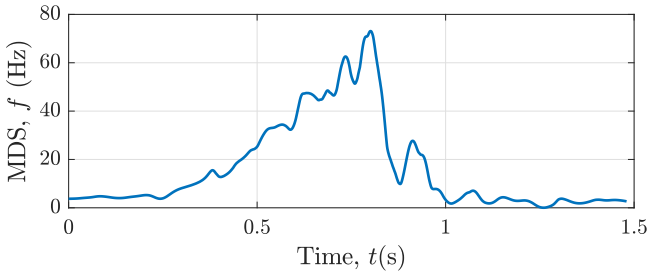
(a) Walking



(b) Picking up an object from the ground



(c) Sitting on a chair

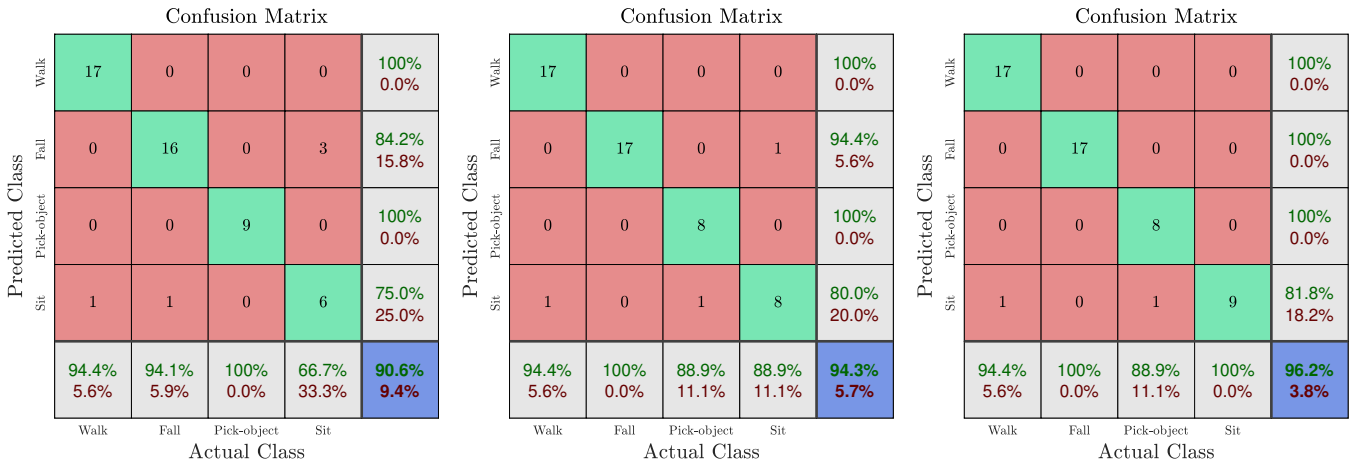


(d) Falling on a mattress

Fig. 4: The illustration of active MDS segments corresponding to four activities explored in this paper.

activities explored in this paper. Recall that each participant repeated the same activity several times. From every activity trial, we first estimate the MDS and then extract the active segment from the MDS and repeat this process for all collected data. Thereafter, we assign the labels (i.e., the type of activity) to each active MDS segment.

As we can see in Fig. 4, the length of active MDS segments varies from activity to activity. For instance, the falling activity lasts around 1.5 seconds and the walking activity is little longer than 4 seconds. Therefore, at first, we resize the segments to the same length by applying the cubic interpolation [22], and then extract time and frequency domain features. From each active MDS segment, we extract 63 features and concatenate them to form a feature vector. These features include simple statistical features (i.e., mean, variance, standard deviation, the difference of absolute standard deviation, root mean square (RMS), skewness, mean absolute value, zero-crossing, slop sign change, average amplitude change, simple square integral, maximum fractal length, the Willison amplitude, wavelength, log detector, slope, and autocorrelation of the MDS) as well



(a) Decision tree (b) LDA (c) SVM

Fig. 5: Confusion matrices of the classifiers used in this study.

as spectral features (i.e., spectral peak, energy in different frequency bands). Some of these features are single values and some consist of multiple values. The details of these features are not presented in this paper due to space limitations but a comprehensive discussion of all these features can be found collectively in [23] [24].

VI. Results and Discussion

We apply the supervised machine learning technique to classify the features vectors extracted from the active MDS segments of different activities. In supervised learning, the labelled training data are used to train a model or a classifier. The labelled training data consist of training examples. Each training example contains features which are extracted from the MDS and an output (i.e., a label corresponding to the type of activity). In the training phase, the classifier uses the labelled training data to build a model that can learn to map features to the labels. Here, the goal is to approximate the model so well that it predicts the labels of future examples. In our experiments, we used the holdout validation technique to train and test the models. Therefore, we divided the data into training and test data sets representing 80% and 20% of the total data, respectively. The training data are used to train three supervised learning algorithms: the decision tree, the LDA, and the SVM. After training each classifier, we used the test data to estimate the performance of each classifier. The confusion matrices of the three considered algorithms are illustrated in Fig. 5.

The rows of every confusion matrix of Fig. 5 indicate the predicted class and the columns correspond to the actual class. In every confusion matrix of Fig. 5, the green cells indicate the number of correctly classified instances of each activity. The off-diagonal red cells indicate the number of wrongly classified instances. The true positive rate of each class is given in the row at the bottom of each confusion matrix. The positive predictive value of each class is given in the far-right column of every confusion matrix. The blue diagonal cell of every confusion matrix indicates the overall accuracy of the algorithm.

The confusion matrix of the decision tree algorithm is provided in Fig. 5a. We observe that the decision tree classifier achieves an overall accuracy of 90.6% as shown in the blue diagonal cell of the confusion matrix in Fig. 5a. The green cells of the confusion matrix indicate the number of correctly classified activities. For instance, the top diagonal cell shows that 17 walking scenarios are correctly classified, whereas falling, picking up an object, and sitting are successfully recognized in 16 and 9, 6 cases, respectively. The first column indicates that there are 18 walking cases, of which 1 case, which in reality belongs to the walking activity, is misclassified as the sitting activity resulting in a recall of 94.4%. Similarly, the recall of the falling, picking up an object, and sitting activity is 94.1%, 100%, and 66.7%, respectively. The first row of Fig. 5a tells us that the decision tree has predicted the walking activity in 17 cases, and none of the examples from the other activities is falsely accepted as the walking activity. Thus, the walking activity has 100% precision.

The confusion matrices for the LDA and SVM are provided in Figs. 5b and 5c, respectively. From these figures, we see that the overall accuracy for the LDA classifier is 94.3% and for the SVM is 96.2%. The recall of the LDA classifier for the walking, falling, picking up an object, and sitting activity is 94.4% 100%, 88.9%, and 88.9%, respectively. The precision of the LDA classifier for the activities walking, falling, picking up an object, and sitting is 100%, 94.4%, 100%, and 81.8%, respectively.

The SVM algorithm has the best performance in terms of overall accuracy. the SVM was trained using a third-degree polynomial kernel and the cost parameter was optimized using the grid search algorithm. The SVM algorithm outperforms decision tree and LDA algorithms by 5.8% and 1.9%, respec-

tively. The precision of the SVM classifier is 100%, 100%, 100% and 81.8% for the activities walking, falling, picking up an object, and sitting, respectively. The recall of the SVM classifier is 94.4%, 100%, 88.9%, and 100% for the walking, falling, picking up an object, and sitting activities, respectively.

VII. Conclusion

In this paper, we have developed a HAR system that use standard Wi-Fi devices. It comprised of two phases: RF sensing and machine learning. In the RF sensing part, we utilized two laptops acting as a transmitter and receiver equipped with one transmit and two receive antennas, respectively. The CSI data were collected by involving nine participants who performed four activities i.e., walking, falling, picking up an object from the ground, and sitting on a chair. The captured CSI data were processed to remove the impact of noise and other sources of errors (e.g., CFO and SFO). Thereafter, the processed data is used to compute the spectrogram which was used to estimate the MDS for each data sample. Afterwards, we extracted a feature vector of length 63 from each MDS sample. We used the supervised learning approach with the hold out validation technique, where we trained three classification algorithms (decision tree, LDA, and SVM) using 80% of total data. After the training phase, the performance of these algorithms was evaluated by using 20% hold out data. Our results showed that the SVM algorithm performs better than decision tree and LDA in terms of overall accuracy, precision, and recall. The decision tree, LDA, and SVM algorithms achieved an overall accuracy of 90.6%, 94.3%, and 96.2%, respectively. Although the results in this work are quite good, we believe that there is still much room for improvement. In our future work, we will explore further features and include a feature selection algorithm to improve the accuracy of WiHAR without increasing the computational load. Moreover, we will also increase the size of the data set by increasing the number of activities and including more trials per activity. Finally, we are also aiming to conduct more tests to determine the RF sensing range of the WiHAR and quantitatively evaluate its performance in different environments.

References

[1] H. Li and I. Cohen, "Inference of human postures by classification of 3D human body shape," in *2003 IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG)*, vol. 00, 10 2003, p. 74.

[2] M. Harville and D. Li, "Fast, integrated person tracking and activity recognition with plan-view templates from a single stereo camera," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, ser. CVPR'04. Washington, DC, USA: IEEE Computer Society, 2004, pp. 398–405.

[3] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," *SIGKDD Explor. Newsl.*, vol. 12, no. 2, pp. 74–82, Mar. 2011.

[4] K. Yatani and K. N. Truong, "Bodyscope: A wearable acoustic sensor for activity recognition," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, ser. UbiComp '12. New York, NY, USA: ACM, 2012, pp. 341–350.

[5] S. Sigg, M. Scholz, S. Shi, Y. Ji, and M. Beigl, "RF-sensing of activities from non-cooperative subjects in device-free recognition systems using ambient and local signals," *IEEE Transactions on Mobile Computing*, vol. 13, no. 4, pp. 907–920, April 2014.

[6] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Device-free human activity recognition using commercial WiFi devices," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 5, pp. 1118–1131, May 2017.

[7] F. Adib, Z. Kabelac, and D. Katabi, "Multi-person motion tracking via RF body reflections," in *NSDI'15 Proceedings of the 12th USENIX Conference on Networked Systems Design and Implementation*, Oakland, USA, May 2015, pp. 279–292.

[8] F. Adib, Z. Kabelac, D. Katabi, and R. C. Miller, "3D tracking via body radio reflections," in *NSDI'14 Proceedings of the 11th USENIX Conference on Networked Systems Design and Implementation*, Seattle, USA, Apr. 2014, pp. 317–329.

[9] P. V. Dorp and F. C. A. Groen, "Feature-based human motion parameter estimation with radar," *IET Radar, Sonar Navigation*, vol. 2, no. 2, pp. 135–145, Apr. 2008.

[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 *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '14. New York, NY, USA: ACM, 2014, pp. 617–628.

[11] B. Wei, W. Hu, M. Yang, and C. T. Chou, "From real to complex: Enhancing radio-based activity recognition using complex-valued csi," *ACM Trans. Sen. Netw.*, vol. 15, no. 3, Aug. 2019. [Online]. Available: https://doi.org/10.1145/3338026

[12] W. Jiang, C. Miao, F. Ma, S. Yao, Y. Wang, Y. Yuan, H. Xue, C. Song, X. Ma, D. Koutsonikolas, W. Xu, and L. Su, "Towards environment independent device free human activity recognition," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*. ACM, 2018, pp. 289–304.

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

[14] M. Pätzold and C. A. Gutierrez, "Modelling of non-WSSUS channels with time-variant Doppler and delay characteristics," in *7th IEEE International Conference on Communications and Electronics (ICCE 2018)*, Hue, Vietnam, Jul. 2018, pp. 1–6.

[15] A. Abdelgawwad and M. Pätzold, "A framework for activity monitoring and fall detection based on the characteristics of indoor channels," in *87th IEEE Vehicular Technology Conference (VTC Spring 2018)*, Porto, Portugal, Jun. 2018, pp. 1–7.

[16] ——, "On the influence of walking people on the Doppler spectral characteristics of indoor channels," in *28th IEEE Int. Symp. on Pers., Indoor, and Mobile Radio Commun. workshops (PIMRC'17)*, Montreal, QC, Canada, Oct. 2017.

[17] B. Boashash, *Time-Frequency Signal Analysis and Processing – A Comprehensive Reference*, 2nd ed. Elsevier, Academic Press, 2015.

[18] 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, Oct. 2017.

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

[20] N. Keerativoranan, A. Haniz, K. Saito, and J.-i. Takada, "Mitigation of CSI temporal phase rotation with B2B calibration method for fine-grained motion detection analysis on commodity Wi-Fi devices," *Sensors*, vol. 18, no. 11, 2018. [Online]. Available: https://www.mdpi.com/1424-8220/18/11/3795

[21] J. Jackson, *A User's Guide to Principal Components*. New York, USA: John Wiley & Sons, 2003.

[22] E. Maeland, "On the comparison of interpolation methods," *IEEE Transactions on Medical Imaging*, vol. 7, no. 3, pp. 213–217, Sep. 1988.

[23] A. Chelli and M. Pätzold, "A machine learning approach for fall detection and daily living activity recognition," *IEEE Access*, vol. 7, pp. 38 670–38 687, 2019.

[24] E. J. Rechy-Ramirez and H. Hu, "CES-513 stages for developing control systems using EMG and EEG signals: A survey," University of Essex, Colchester, Tech. Rep., 2011. [Online]. Available: http://repository.essex.ac.uk/id/eprint/14259