Human Activity Recognition Using Wi-Fi and Machine Learning

Ali Chelli, Muhammad Muaaz, Ahmed Abdelgawwad and Matthias Pätzold

August 10, 2021

Human Activity Recognition Using Wi-Fi and Machine Learning

Ali Chelli[a], Muhammad Muaaz[a], Ahmed Abdelgawwad[a],and Matthias Pätzold[a]

[a]Faculty of Engineering and Science, University of Agder, P.O. Box 509, 4898 Grimstad, Norway.

ARTICLE HISTORY
Compiled November 5, 2020

ABSTRACT
In the era of Internet of things, access points will be deployed everywhere. The wireless signals offered by these access points can be used for more than just Internet connectivity. In fact, the human movement causes Doppler shifts in the received wireless signals. By combining signal processing techniques and machine learning, it is possible to recognize human activity from Wi-Fi signals. This paper builds on these ideas and develops a human activity recognition system that comprises two parts: radio-frequency sensing and machine learning. In the radio-frequency sensing part, we record the channel transfer function of an indoor environment in the presence of a participant performing three activities: walking, falling, and picking up an object. Using signal processing techniques, we estimate the mean Doppler shift of the channel, which contains the fingerprint of the user activity. The mean Doppler shift is used by a classifier to determine the type of performed activity. We assess the activity recognition performance of three classification algorithms: cubic support vector machine, K-nearest neighbor, and linear discriminant analysis. Our analysis shows that the cubic support vector machine, linear discriminant analysis, and K-nearest neighbor algorithms achieve an overall accuracy of 99.5%, 97.3%, and 95.1%, respectively.

KEYWORDS
Human activity recognition, machine learning, channel state information, channel transfer function, mean Doppler shift.

1. Introduction

With the rapid development of the Internet of things (IoT), various types of sensors have been embedded in indoor and outdoor environments. This offers the opportunity to collect useful data that can be utilized in environment monitoring, smart city, and human activity recognition (HAR). Traditional HAR techniques are mainly based on camera [1] or wearable devices [2]. However, with the widespread deployment of Wi-Fi access points, device-free HAR has attracted a lot of attention. As opposed to camera-based HAR, Wi-Fi-based HAR does not violate user's privacy. Besides, Wi-Fi-based HAR does not require the user to wear any sensing device, which allows avoiding user discomfort associated with wearable devices.

In Wi-Fi-based activity recognition systems, an electromagnetic wave emitted by the transmitter is reflected by the objects in the environment before reaching the receiver. If a person is moving in the vicinity of the transmitter and the receiver, this movement causes a Doppler shift in the received radio-frequency (RF) signal. The pattern of this

CONTACT Ali Chelli. Email: ali.chelli@uia.no

Doppler shift varies depending on the type of activity. Thus, it is possible to recognize the user activity based on the received RF signal. Wi-Fi-based HAR either uses the channel state information (CSI) [3–5] or the radio signal strength indicator (RSSI) [6] for activity recognition. CSI-based HAR systems have a better accuracy in recognizing human activity compared to RRSI-based HAR systems. The authors of [7] developed a software tool for gathering CSI from Wi-Fi data packets by means of a network interface card (NIC). The authors of [3], [4], and [5] utilize deep learning algorithms for activity recognition based on CSI data and achieve an overall recognition accuracy of 96%, 97.4%, and 97.6%, respectively.

In this paper, we use the software tool proposed in [7] to build a testbed for HAR using two laptops. One laptop acts as a transmitter, while the second laptop acts as a receiver. Seven participants are asked to perform three activities (walking, falling, and picking up an object) inside a 16 square meter room. The CSI data is collected while the participants carry out their activities. Using signal processing techniques, we estimate the time-variant mean Doppler shift (MDS) from the recorded CSI data. Subsequently, time and frequency domain features are extracted from the MDS and provided to a classification algorithm. This classifier must determine the type of performed activity. We test the performance of three classification algorithms: cubic support vector machine (CSVM), K-nearest neighbor (KNN), and linear discriminant analysis (LDA). Our results show that the CSVM, LDA, and KNN algorithms achieve an overall recognition accuracy of 99.5%, 97.3%, and 95.1%, respectively. With these results, we outperform most of the existing activity recognition systems in terms of overall accuracy.

Note that this is the first work in the literature that uses the MDS for activity recognition. The MDS approach makes our system 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 reveals that our HAR system can maintain a high recognition accuracy even if the activity is carried out at a distance of 4 meters from the transmitter and receiver. Note that in most existing studies, the activity is performed very close (one meter) to the transmitter and receiver to achieve a recognition accuracy of over 90%. It is reported in [8] that the recognition accuracy drops below 85% if the participant is at a distance of 3 meters from the transmitter and receiver.

The rest of the paper is organized as follows. Section 2 provides an overview of the proposed activity recognition system and its building components. In Section 3, we describe the different data pre-processing steps that are applied to the collected CSI data to obtain the MDS. In Section 4, we assess the performance of the proposed HAR system and discuss the obtained results. Finally, Section 5 offers concluding remarks.

2. Overview of the CSI-Based HAR System

An overview of the CSI-based HAR system is provided in Fig. 1. The HAR system uses the CSI data of a Wi-Fi link to recognize human activities. The HAR system consists of two main stages: RF sensing and machine learning as shown in Fig. 1. In the RF sensing stage, a single person carries out an activity in an indoor environment. The RF data is collected by involving seven participants performing three different activities: walking, falling, and picking up an object. During the activity, the person's body parts move and cause a Doppler shift in the RF signal. The pattern of this Doppler shift varies from one activity to another and can thus be used for activity recognition. To capture this phenomenon, we use two laptops acting as a transmitter and receiver. Instead of the built-in RF antennas of the laptops, we use RF cables and connected external omni-directional antennas to the NICs of the laptops. A single transmit antenna A_T injects 1000 data packets per second in the wireless medium and two receive antennas A_{R1} and A_{R2} collect the injected packets. All the antennas are attached to the room ceiling. We use the software tool proposed in [7] to capture the Wi-Fi CSI from the received packets, while the users perform different activities.

The CSI data recorded during RF sensing is then transferred to the machine learning part. Using signal processing algorithms, the CSI data is calibrated, filtered, denoised, and the MDS associated with each CSI sample is computed. The MDS contains the fingerprints of the user activity and its pattern varies as the activity changes. The MDS is considered as a time-series from which we extract time and frequency domain features that form the feature vector. Based on this vector, the classification algorithm must determine the type of performed activity from three possible activities: walking, falling, and picking up an object (see Fig. 1).

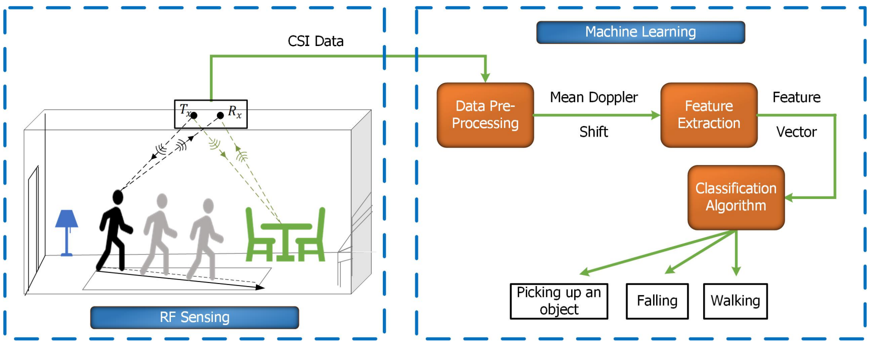

Figure 1. Architecture of the HAR system.

3. Data Pre-Processing: From CSI Data to MDS

The transmitter laptop T_x sends 1000 data packets per second in the wireless medium. These packets are transformed into an electromagnetic wave by the transmit antenna A_T. This wave travels in the indoor environment and is reflected by fixed (walls and furniture) and moving (body parts of the moving person) objects before arriving at the receive antennas A_{R1} and A_{R2} of the receiver laptop R_x. To capture the CSI of the received data packets, we use the software tool developed in [7]. The CSI data is an $N_{T_x} \times N_{R_x} \times K$ matrix, where N_{T_x} is the number of transmit antennas, N_{R_x} denotes the number of receive antennas, and K stands for the number of orthogonal frequency division multiplexing (OFDM) subcarriers.

In our measurement campaign, we have one transmit antenna ($N_{T_x} = 1$), two receive antennas ($N_{R_x} = 2$), and 30 OFDM subcarriers ($K = 30$). The symbol f'_k refers to the carrier frequency of the kth OFDM subcarrier. We denote the elements of the CSI matrix by $H_{i,j}(f'_k, t)$, where the pair (i, j) indicates the indices of the transmit and receive antenna, respectively. Each element $H_{i,j}(f'_k, t)$ of the CSI matrix is known as a CSI stream. The link between the ith transmit antenna and the jth receive antenna is characterized by its channel transfer function (CTF) $H_{i,j}(f', t)$. Thus, the CSI stream $H_{i,j}(f'_k, t)$ is a discrete version of the CTF $H_{i,j}(f', t)$ sampled at frequency f'_k.

In our measurement campaign, the distance between the moving person and the antennas is short (less than 4 meters), which makes the angles of departure and arrival time variant. Besides, the speed of motion of different body parts is also time variant. This makes the underlying fading channel non-stationary, which implies that both the CTF and the CSI stream are non-stationary. Therefore, a time-frequency analysis tool such as the spectrogram is needed to analyse the behavior of the fading and characterize the fingerprint related to various user activities.

The spectrogram $S_{H_k}(f,t)$ of the CSI stream $H_{i,j}(f'_k,t)$ is obtained in two steps. First, we compute the short-time Fourier transform (STFT) $X_{H_k}(f,t)$ of the CSI stream $H_{i,j}(f'_k,t)$

$$X_{H_k}(f,t) = \int_{-\infty}^{\infty} H_{i,j}(f'_k,\tau)w(\tau - t)e^{-j2\pi f\tau}d\tau \tag{1}$$

where τ stands for the running time, t is the local time, and $w(t)$ is a sliding window. In our case, we use a Gaussian window function [9, Eq. (2.3.1)]. Second, the spectrogram $S_{H_k}(f,t)$ can be computed as $S_{H_k}(f,t) = |X_{H_k}(f,t)|^2$. The MDS $B^{(1)}_{H_k}(t)$ can be determined using the spectrogram as follows

$$B^{(1)}_{H_k}(t) = \frac{\int_{-\infty}^{\infty} fS_{H_k}(f,t)df}{\int_{-\infty}^{\infty} S_{H_k}(f,t)df}. \tag{2}$$

Since we use a commercial Wi-Fi NIC card for data acquisition, which measures the channel for each received data packet, the collected data suffers from several sources of error, such as noise, carrier frequency offset and sampling frequency offset [10]. To mitigate these errors, we use various signal processing algorithms for data calibration, filtering, and denoising. This improves the quality of the estimated MDS. Note that a poor estimation of the MDS leads to a low accuracy in activity recognition. For each MDS sample, we extract a feature vector that is provided to the classification algorithm. The latter must determine the type of activity based on the feature vector.

4. Experimental Results

In this section, we assess the performance of the proposed activity recognition system. First, the data is divided into training and test data sets representing 70% and 30% of the total data, respectively. A training example consists of the features extracted from a given MDS sample together with a label, which indicates the type of activity associated with the considered MDS sample. During the training phase, the classifier is exposed to labeled training data and learns the pattern associated with each activity. At the end of the training phase, the internal parameters of the classifier are tuned such that the trained algorithm can distinguish various activities with high accuracy. Subsequently, the performance of the trained classifier is evaluated using the test data set.

In our study, we compare the performance of three classification algorithms: CSVM, LDA, and KNN. The obtained results are illustrated in Table 1. This table shows that the CSVM algorithm achieves the best overall performance with an activity recognition accuracy of 99.5%, compared to 95.1% for KNN, and 97.3% for LDA. Table 1 provides the classification recall and precision of each algorithm for each activity. Note that the precision and the recall of the classification have two different meanings. The classification recall focuses on the actual activity and quantifies the percentage of successful classifications out of the *actual* samples belonging to a particular class. In contrast, the classification precision focuses on the predicted activity and indicates the percentage of correct classifications out of the samples *predicted* to belong to a certain activity. For instance, the classification recall for the activity falling for the CSVM algorithm is 100%. This implies that all the actual falls were correctly classified by CSVM. Whereas the classification precision of falling is 98.7% for CSVM, which means that 1.3% of the predicted falls are actually non-falling events.

Table 1. Performance of the classification algorithms KNN, LDA, and CSVM in terms of activity recognition.

		Algorithms		
		KNN	LDA	CSVM
Overall Accuracy %		95.1	97.3	99.5
Recall %	Walking	98.4	93.7	98.4
	Falling	89.7	98.7	100
	Picking up an object	100	100	100
Precision %	Walking	93.9	98.3	100
	Falling	98.6	95.1	98.7
	Picking up an object	91.3	100	100

The classification recall of the KNN classifier for the activities walking, falling, and picking up an object is equal to 98.4%, 89.7%, and 100%, respectively. Whereas for the LDA algorithm, the classification recall is 93.7%, 98.7%, and 100% for the activities walking, falling, and picking up an object, respectively. The CSVM algorithm reaches 100% recall for the activities falling and picking up objects and 98.4% recall for walking. The classification precision of the KNN classifier for the activities walking, falling, and picking up an object is equal to 93.9%, 98.6%, and 91.3%, respectively. While for the LDA algorithm, the classification precision is 98.3%, 95.1%, and 100% for the activities walking, falling, and picking up an object, respectively. The CSVM algorithm reaches 100% precision for the activities walking and picking up objects and 98.7% precision for falling.

5. Conclusion

In this paper, we have developed a HAR system, which consists of two building components: RF sensing and machine learning. The RF sensing component is a testbed used to collect RF data of indoor channels, while a human participant carries out three activities (walking, falling, and picking up an object). This RF testbed comprises two laptops acting as a transmitter and receiver. The transmitter laptop sends 1000 packets per second. At the receiver laptop, a software tool captures the CSI data in each received packet. This CSI data is a discrete version of the CTF sampled at 30 frequencies. The CSI data is collected from seven participants. Using signal processing algorithms, we reduce the noise in the CSI data and estimate the MDS associated with each CSI sample. A feature extraction algorithm is applied to each MDS sample to extract a feature vector. The data is then split into training and test data. The training data contains both the feature vector and a label indicating the type of the performed activity. We adopt a supervised learning approach, where the classifier is trained using the training data and then its performance is assessed using the test data. We evaluate the performance of three classification algorithms: CSVM, LDA, and KNN. Our experimental results show that the CSVM algorithm achieves the best performance with an overall accuracy of 99.5%, while LDA and KNN have an overall accuracy of 97.3% and 95.1%, respectively.

Acknowledgement

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

References

[1] Sehairi K, Chouireb F, Meunier J. Comparative study of motion detection methods for video surveillance systems. Journal of Electronic Imaging. 2017 Apr;26(2):26–29.

[2] Chelli A, Pätzold M. A machine learning approach for fall detection and daily living activity recognition. IEEE Access. 2019 Mar;7:38670–38687.

[3] Wang F, Gong W, Liu J. On spatial diversity in WiFi-based human activity recognition: A deep learning-based approach. IEEE Internet of Things Journal. 2019 Apr;6(2):2035–2047.

[4] Zou H, Zhou Y, Yang J, et al. DeepSense: Device-free human activity recognition via autoencoder long-term recurrent convolutional network. In: 2018 IEEE International Conference on Communications (ICC); May; Kansas City, MO, USA; 2018. p. 1–6.

[5] Zou H, Zhou Y, Yang J, et al. Towards occupant activity driven smart buildings via WiFi-enabled IoT devices and deep learning. Energy and Buildings. 2018 Oct;177:12 – 22.

[6] Abdelnasser H, Youssef M, Harras K. WiGest: A ubiquitous WiFi-based gesture recognition system. In: IEEE Conference on Computer Communications (INFOCOM 2015); Apr.; Hong Kong, China; 2015. p. 1472–1480.

[7] Halperin D, Hu W, Sheth A, et al. Tool release: Gathering 802.11N traces with channel state information. ACM SIGCOMM Comput Commun Rev. 2011 Jan;41(1):53–53.

[8] Wang Y, Wu K, Ni LM. WiFall: Device-free fall detection by wireless networks. IEEE Transactions on Mobile Computing. 2017 Feb;16(2):581–594.

[9] Boashash B. Time-frequency signal analysis and processing a comprehensive reference. 2nd ed. Elsevier, Academic Press; 2015.

[10] Yousefi S, Narui H, Dayal S, et al. A survey on behavior recognition using WiFi channel state information. IEEE Communications Magazine. 2017 Oct;55(10):98–104.