Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 198 (2022) 59–66

Procedia Computer Science

www.elsevier.com/locate/procedia

The 11th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2021)
November 1-4, 2021, Leuven, Belgium

Human Activity Recognition based on Wi-Fi CSI Data - A Deep Neural Network Approach

Andrii Zhuravchak[a,*], Oleg Kapshii[b], Evangelos Pournaras[c]

[a] *Ukrainian Catholic University, Lviv, Ukraine*
[b] *Infineon Technologies, Lviv, Ukraine*
[c] *University of Leeds, Leeds, UK*

Abstract

Using Wi-Fi Channel State Information (CSI) is a novel way of environmental sensing and human activity recognition (HAR). These methods can be used for several safety and security applications by (re)using Wi-Fi routers without the need for additional costly hardware required for vision-based approaches, known also to be particularly privacy-intrusive. This work introduces a full pipeline of a Wi-Fi CSI-based system for human activity recognition that assesses and compares two deep learning methods. We analyze how different hardware configurations affect WiFi CSI signals. We contribute a novel and more realistic data collection process, in which human activity recognition is seamlessly integrated in real-life, resulting in more reliable assessments of the model classification performance. We analyze how InceptionTime and LSTM-based classification models perform in human activity recognition. The source code and collected dataset are made publicly available for reproducibility and encouraging further research in the community.

Keywords: Wi-Fi; channel state information; human activity recognition; machine learning; digital signal processing;

1. Introduction

Nowadays, Wi-Fi routers are present nearly in every house and their capability to estimate a Wi-Fi channel state provides a complex, but powerful possibility to sense the surrounding environment. This work shows how a combination of Wi-Fi CSI and Machine Learning (ML) can be used for Human Activity Recognition. With such a system, we

* Corresponding author. Tel.: +3-809-897-94052
E-mail address: zhuravchak@ucu.edu.ua

10.1016/j.procs.2021.12.211

can determine if a person drops down with a heart attack, enhance security systems by detecting people present in the total darkness or even potentially check whether a person is breathing and decide whether to call for emergency.

Usually, home Wi-Fi routers are used exclusively for communication. But due to the environmental influence, connection quality can vary a lot. To analyze Wi-Fi channel states and tune router configuration if needed, RSSI (Received Signal Strength Indicator) and CSI (Channel State Information) mechanisms are used. Because a person in the area covered by a Wi-Fi router also distorts radio-waves during Wi-Fi communication between different devices, analysis of the distortion parameters gives a possibility to make some conclusion what activity is happening around.

The RSSI is a metric indicating the signal strength. It has been actively used for active localization based on the Wi-Fi fingerprinting technique [1] or as a metric for mobile devices passive tracking [2]–[4]. But, it is quite unstable, differs from vendor to vendor [5], and cannot accurately capture changes in signal due to human movements, especially if a person is not directly between an Access Point and a Wi-Fi router. The CSI tool gives more precise information about the channel state. For each antenna pair of transmitter and receiver at each sub-carrier frequency, it measures propagating wireless signals and provides amplitude and phase distortions for different sub-channels. In such a way, CSI variations in the time domain have different patterns for different humans, activities etc., and this can be used for human activity recognition [6].

There is a number of publications where Wi-Fi CSI-based systems for human activity recognition is discussed, showing that CSI data is sufficient to build such a system, unlike the RSSI. In particular, [7] and [8] show that the people detection problem using CSI could be done with a high accuracy. Systems reported in [9], [10] and [11] show an average accuracy of 20-40 cm for indoor localization based on WiFi CSI data. Systems such as [12] and [13] study a very similar problem to this paper with quite a success indicating the accuracy of 80-95%. However, human activity detection takes place in an over-controlled lab environment, where activities are studied in isolation and not as they emerge in everyday life.

In this work, we focus on developing the full CSI-based human activity recognition system taking into account hardware, data and model aspects. We conducted experiments to analyze the relationship between the Wi-Fi configuration and the CSI data. We show a data collection approach, where for a small amount of time, different activities can be performed in no strict order creating a more realistic "living lab" setup. We propose a data preprocessing workflow with which we build and compare InceptionTime and LSTM-based classification models. The whole codebase and dataset are publicly available for reproducibility and further research.

The contributions of this paper are summarized as follows: (i) A more realistic data collection process for human activity recognition that is seamlessly integrated in the daily life of humans which should result in more reliable estimations of the classification models performance. (ii) A guide for WiFi network configuration for human activity recognition using WiFi CSI data. The guide was derived empirically from 5 experiments. (iii) A system workflow based on which a comparison between an InceptionTime and an LSTM-based classification model is made. (iv) An open dataset[1] and software code[2] for reproducability and encouraging further research in the community.

2. Background Information

Human movement within the range of a Wi-Fi network affect the multipath propagation [14]. The Wi-Fi channel also consists of a signal reflected by static objects, such as furniture. The CSI represents the signal propagation effect in the channel [15]. Additional reflections are caused by different activities as well.

CSI represents the change of the signal from a transmitter (denote it as x) to a receiver (denote it as y) and it is represented in the frequency domain as $y = Hx + n$, where H is a complex matrix consisting of CSI values and n is the channel noise. The CSI is estimated for each Orthogonal Frequency Division Multiplexing (OFDM) subcarrier links [16]. OFDM splits the total frequency spectrum into 56 or 114 frequency subcarriers for a channel bandwidth of 20 and 40 MHz respectively. The CSI for each subcarrier is: $h = |h|e^{j\theta}$, where $|h|$ represents the amplitude and θ the phase. To measure the CSI, the transmitter sends the Long Training Symbol (LTS), which contains pre-defined information

[1] Available at https://doi.org/10.6084/m9.figshare.14386892.v1 (last access: August 2021).
[2] Available at https://github.com/Retsediv/WIFI_CSI_based_HAR (last access: August 2021).

for each subcarrier. When the receiver receives the LTS, it estimates the CSI having the difference between the original and received LTS.

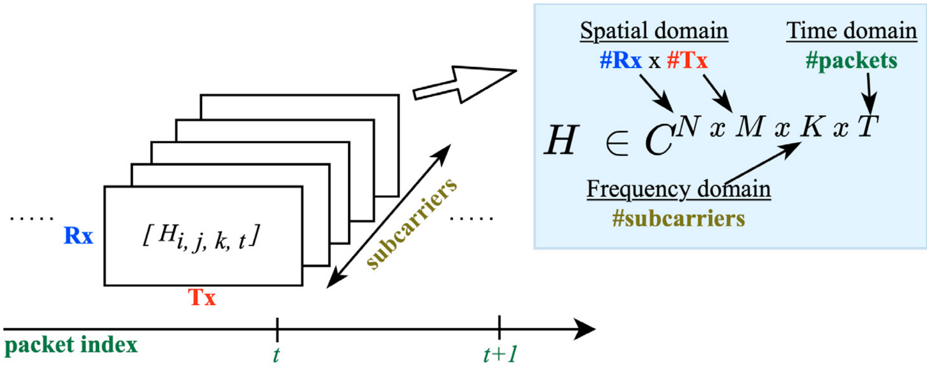

Fig. 1: 4D CSI tensor is a time series of CSI matrices of MIMO-OFDM channels (image inspired by [6])

In this work, the CSI matrix consists of 114 complex H matrices for 5 GHz frequency band or 56 for 2.4 GHz of dimension $N_{Tx}xN_{Rx}$. 56 and 114 are defined by the number of subcarriers the CSI extraction tool can handle (on 2.4 and 5 GHz). Its structure is shown in Fig. 1.

3. System setup

We have found three options to get CSI data from routers: Linux 802.11n CSI Tool [17], Atheros CSI Tool [18], and Nexmon Channel State Information Extractor [19]. All of them are custom firmware and Linux wireless drivers that should be installed on devices instead of vendor-provided.

Since Linux 802.11n CSI Tool requires obsolete Network Interface Controller and Nexmon CSI Extractor supports very limited number of devices, the Atheros CSI Tool is used. It is an open-source 802.11n measurement and experimentation tool. It enables the extraction of detailed wireless communication information from the Atheros Wi-Fi NICs, including the CSI, the received packet payload, the data rate, the timestamp, the RSSI of each antenna and others.

Two TP-link TL-WDR4300 routers were used as a receiver and a transmitter. The transmitter works in the Access Point (AP) mode, the receiver is configured to a Client mode. To create AP, we used OpenWRT (linux operating system targeting embedded devices) interface, which gives the ability to configure operating frequency, mode, channel bandwidth, and other parameters. After that, the client router has to associate with the created AP via an OpenWRT interface to stay in the same network. The custom Atheros CSI Tool firmware was built and installed there to enable PHY (Physical) layer information extraction.

Fig. 2: TP-Link TL-WDR4300 router

The system responsible for the Tx-Rx communication, CSI computing, and data capturing for further analysis was build, based on [20] and [21]. It consists of two programs, one responsible for sending a constant wi-fi packet and the other for receiving it and calculating the CSI. As a router does not have enough space to save a big amount of data, we upgraded the system as follows: (i) The transmitter sends data to the receiver. (ii) The receiver gets the data and computes CSI. (iii) The receiver sends raw CSI data to the user's laptop. (iv) The user's laptop handles data, stores it and visualizes on the fly. (v) Simultaneously, a camera makes a photo of a room for automatic scene parsing. These images are used for the validation of the human activity recognition.

4. WiFi Experimental Network Configuration

Based on the conducted experiments, we aim to recommend how to choose an effective Wi-Fi network configuration, meaning, the frequency, channel, bandwidth, number of antennas, which affect the Wi-Fi channel and its throughput, bandwidth, speed and coverage. The goal of the experiments is to test different system parameters and configurations to understand how they influence the human activity recognition system itself and its recorded data.

Our methodology[3] for the selection of an optimal configuration consist of comparing the amplitude and phase measured over different channels, bandwidths etc. for the same human motion pattern (walking). Then, plotting results on a graph and visually choosing the most distinguishable configuration so that the activity pattern can be easier observed.

Experiment 1: Wi-Fi channels. There is a number of frequency channels that can be used for the Wi-Fi communication. Usually it is recommended to use channels 1, 6, or 11, as they do not overlap with each other. Based on this fact, we assume, that the activity pattern can be better visible on one channel and worse on another. In this experiment we tested how the final signal looks like when the system works at different channels.

Experiment 2: Bandwidth 20 and 40 MHz. Working with IEEE 802.110n there is the possibility to use signal bandwidths of either 20 MHz, 40 or 80 MHz. A higher bandwidth corresponds to a higher data throughput. However, it reduces the number of channels that can be used. Activity recognition was tested on bandwidths 20 and 40 MHz (80 MHz bandwidth is not supported by our hardware)

Experiment 3: Frequency 2.4 and 5 GHz. Wi-Fi is able to work in two frequency ranges: 2.4 and 5GHz. In general, 2.4 GHz outshines 5 GHz as to the Wi-Fi coverage, but 5 GHz gives a much higher speed. Therefore it is important to check how the system behaves in each of the ranges.

Experiment 4: Different number of antennas - effect on human activity recognition not in the Line of Sight (LoS). The routers used in this research have three antennas (all three can work at the 5 GHz frequency, but only two at 2.4 GHz). Different antennas configurations (number of enabled antennas on the transmitter and receiver routers) were tested to understand how their number influences the ability to detect a walking activity. In this experiment, routers were located diagonally in the room. The person was walking next to one of the walls (not in the routers LoS).

Experiment 5: Detecting activity in another room with no Wi-Fi device. In this experiment, the transmitter and receiver, routers were located in different (adjoining) rooms. After that, “walking” activity was performed in both rooms. Then we performed the same experiments, but placed both routers in the same room.

Based on the conducted experiments, we conclude that the proper hardware configuration is crucial for the human activity recognition system based on Wi-Fi CSI data. The recommendations based on each experiment are the following: *Experiment 1*: It is better to choose the least loaded channel from the recommended non-overlapping list. For example, for 2.4 GHz, it could be 1, 6, or 11. During our experiments, the pattern of the walking activity on channel 11 was more well distinguished comparing to other channels. *Experiment 2*: The experiment with the 40 MHz bandwidth shows better pattern visibility than for 20 MHz. However, the variance of amplitude and noise in 40 MHz is larger than in the 20 MHz. *Experiment 3*: We cannot conclude which of the 2.4 or 5 GHz Wi-Fi network performs better as the activity patterns are comparable. However, the amplitude variability at 5 GHz is smaller than at 2.4 GHz resulting in less data noise. *Experiment 4*: The higher the number of antennas, the better. While a person is not on the sight line, it is almost impossible to detect any activities and movements in the 1Tx 1Rx configuration. We can state, that at least two receiver and transmitter antennas are required to detect human activity when they are not on the sight line. *Experiment 5*: As well as we can access the Internet in different rooms, we can detect walking activity in another room if there is a receiver (client) there. However, we have not seen any clear signal changes caused by human activity in a room if the devices were located at another adjoining room.

[3] The corresponding code and visualizations can be found at the jupyter notebook in `https://github.com/Retsediv/WIFI_CSI_based_HAR/blob/master/model_for_har/notebooks/ExperimentsVisualization.ipynb`

5. Dataset Collection

To build and train a model for human activity recognition, a dataset is required. We found two publicly available datasets: [13] and [12]. Both were collected using Linux 802.11n CSI Tool with three antennas on the transmitter and receiver routers. However, we decided not to use them as the hardware is obsolete and cannot be bought. Moreover, they collected data by sequences, during which only one activity is performed. The sequences can be quite large in time, but they do not contain transitions between different actions, as well as quick and frequent change of actions during small periods of time. These limitations indicate a poor representation of real-world data and human behaviour. Last but not least, the data collection environment is a lab-like over-controlled environment.

Table 1: Collected dataset characteristics

Activities	walk, sit, stand, lying, get up/down, no activity
Size	1.2 Gb (no images), 9.1 Gb (with images)
Labels	activity and person bounding box
# of people involved	1
# of rooms used	3
WiFi router	TP-Link TL-WDR4300
Bandwidth	40 MHz
Channel	60
Frequency	5 GHz
Antennas	2Rx x 2Tx
№ of subcarriers	114

Therefore, we collected a new dataset. The workflow is described in Section 3. The activities performed are: walking, sitting, standing, lying, getting up, getting down, and absence of activity when there is no person in the room. Totally, 3 different rooms are used. Each CSI packet is labelled with an image, activity and bounding box of the person located on the image that may be applied for further work. The collection activity distribution is shown in Fig. 3. The description of the collected dataset is shown in Table 1.

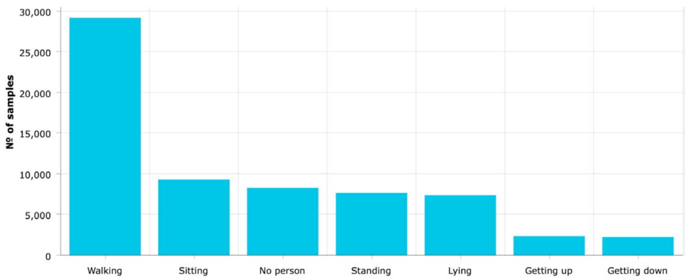

Fig. 3: Collected dataset: number of CSI packets for each activity

6. Human Activity Recognition Model

Even with the best possible hardware and configuration, CSI data contain noise and raw phase information cannot be directly used. This is caused firstly by the environment and reflected radio waves, secondly, by possible hardware instability. That is why the data preprocessing is an essential part of building a stable and accurate HAR system. The full workflow of the system, from the routers to model prediction and preprocessing is shown in Fig. 4.

6.1. Data preprocessing

Phase sanitization. In contrast to amplitude, raw phase information cannot be used. The problem is that it is affected by the carrier frequency offset (CFO) and sampling frequency offset (SFO). The CFO arises when the transmitter and receiver do not precisely synchronize their timing and phases before transmitting a packet. For example, for the 5 GHz band, the difference between the receiver and transmitter clocks can lead to a phase change in several π. And as the phase changes mainly less than 0.5π due to the human moving, the activity cannot be observed from phase change [6].

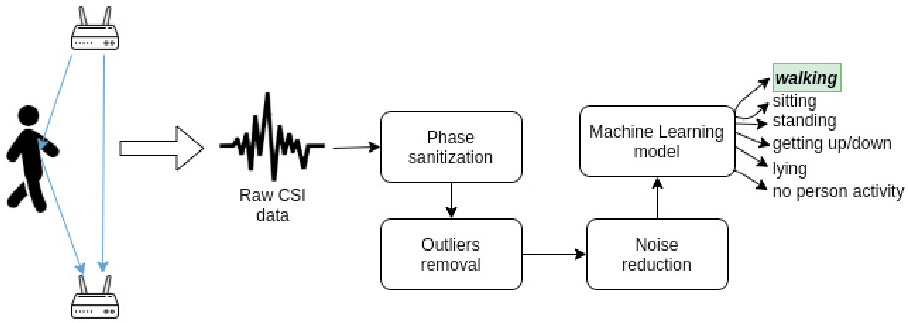

Fig. 4: System workflow from Wi-Fi router to the classification.

The SFO is caused by analogue to digital converter, and it also varies by subcarrier, so each of them has a different error in the end.

As we do not know both CFO and SFO, the raw phase information is not useful. But the linear transformation described in [22] aims to fix it. The results of the phase sanitization technique are shown in Fig. 5. As we can see, raw phase information in the left graphs does not give us any information about walking activity. However, after sanitizing, the phase data became less noisy, and we observe that during the walking activity, the phase changed.

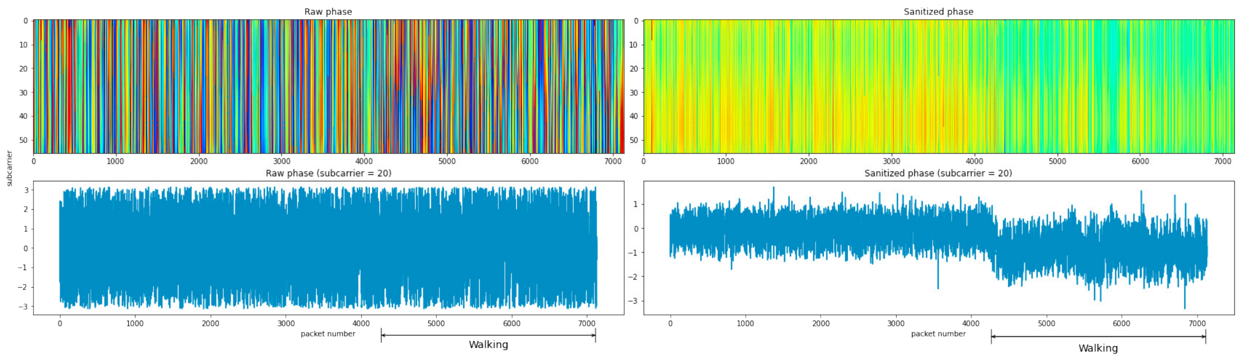

Fig. 5: Phase sanitization algorithm results: raw phase and sanitized.

Outliers removal technique. Amplitudes and phase data contain noise caused by the transition rate and power adaptations, thermal noise etc. As a result, outliers are introduced to the signal that are not caused by human actions. The Hampel Identifier algorithm [15] is used to eliminate this problem. It uses the median and median absolute deviation to detect the location and spread of outliers.

Noise reduction based on Discrete Wavelet Transform. Still, there is much noise in CSI data left and based on [15] the Discrete Wavelet Transform based noise reduction algorithm was implemented and applied. In a nutshell, the wavelet converts a signal into large-magnitude wavelet coefficients. Those coefficients, which are small are typically noise and we can get rid of them without losing the quality of the signal. After that, the inverse wavelet transform is used to reconstruct the data.

6.2. Machine Learning Model

It is hard to effectively analyze CSI data using conventional methods like SVM, decision tree, fuzzy rule-based classifiers, therefore we have chosen to use the neural network approach instead. The neural networks are the most effective tool for extracting hidden patterns and dependencies from complex data flows.

The dataset used for training and validation of the network is described in Chapter 5. As it is shown in Table 1, two antennas for the receiver and transmitter routers were used, so totally there are four antenna pairs. Each antenna pair has CSI data (phase and amplitude) for each of the 114 subcarriers. So, there are 4 * 114 * 2 = 912 features for each CSI packet we receive. Those features are further used for model training.

Along with that, we need to consider that people do not immediately change their activity, and this transition can take some time. Also, due to the environment characteristics and reflected waves, the model has to be able to learn patterns throughout time. The data is fed to the model as sequences with some step (window) between them. The target is to detect the activity performed at the end of that sequence.

Based on that, two methods were tested:

InceptionTime model. [23] proposes an architecture based on deep one-dimensional Convolutional Neural Network aimed at Time series classification problem. They show promising results in accuracy as well as in training time, which is smaller compared to similar architectures.

InceptionTime model is an ensemble of Inception networks followed by a Global Average Pooling layer and a Dense layer with a softmax activation function. Its input is the sequence of CSI data of length 1024. That sequence label is the activity class performed at the end of that sequence. The input data is divided by sequences with the step of 8 time points. Cross Entropy Loss was used as a loss function. The model was built based on the codebase that [23] provides.

After the training, the model showed accuracy on validation data 38.2% and 99.2% on the training. There is clearly an overfitting here, and the main reason is probably the insufficient amount of data for such an extensive and large architecture.

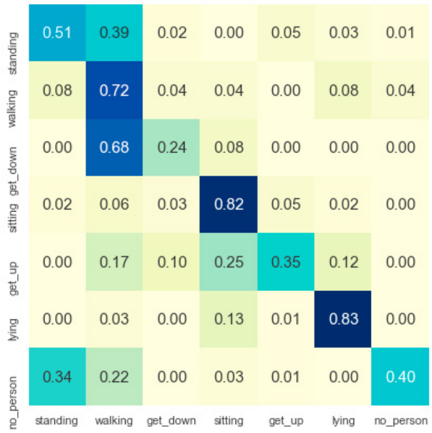

Fig. 6: Confusion matrix of LSTM-based model.

LSTM-based model. RNN works well with sequence data, while its LSTM improvement helps with long term dependencies. The model consists of a single bidirectional LSTM layer of the hidden dimension of size 256 and 4 dense layers (of sizes 512, 256, 128, 7) with ReLU as an activation function.

The input of the model is the sequence of CSI data of length 1024. That sequence label is the activity performed at the end of that sequence. Using that, we tried to simulate a real-time HAR system. The input data-array is processed in 16-time-points chunks with the 8-point overlapping between them. Cross Entropy Loss was used as a loss function. After the model training, its accuracy was 61.1%, precision 59% and recall 55% on validation data, 87.8% accuracy on testing data. The resulting confusion matrix is shown in Fig. 6. Here you can see a correlation between the amount of data collected for each activity and the accuracy we achieved. For example, activities "get up" and "get down" have so bad accuracy 35% and 24% respectively probably because there was a collected tiny dataset for them (accordingly to Fig. 3). For "standing" and "no person" activities it needs more analysis to understand why they are classifying so badly. This model also suffers from overfitting, but not as much as the previous one. Also, the accuracy is much higher overall. Based on that, we can state that LSTM-based approach works better in this specific case when there is not so much data available for training.

7. Discussion, Conclusion and Future Work

In this paper, we propose the full Wi-Fi CSI human activity recognition workflow from data collection to model development. We conducted experiments to understand how the Wi-Fi configuration, routers location and environment affect the CSI data. Based on the received results, we recommended a particular network configuration and provided a summary of our observations.

We proposed a new data collection approach integrated in human daily life, where during a small amount of time, different activities can be performed in any order. The dataset as well as data collection tools are publicly available and can be used for further research.

Our final model gives 61% classification accuracy for 7 activities (83% for lying recognition, 82% for sitting). Although the success is limited due to the limited volume of the dataset, these results are promising, and further research in this direction could be conducted with larger data sets.

We see possible areas for improvements in: (i) Extending the dataset with other locations, people and hardware variations. (ii) Enhancing features extractions to get the most from the available data: PCA, Power Spectral Density, signal skewness, Median Absolute Deviation, Autospectrum etc. (iii) Assessing data augmentation methods and use synthetic data to extend the current dataset.

References

[1] A. Tahat, G. Kaddoum, S. Yousefi, *et al.*, "A Look at the Recent Wireless Positioning Techniques With a Focus on Algorithms for Moving Receivers," *IEEE Access*, 2016.

[2] J. Wilson and N. Patwari, "Radio Tomographic Imaging with Wireless Networks," *IEEE Transactions on Mobile Computing*, journal 9, number 5, pages 621–632, 2010.

[3] E. Pournaras, "Proof of witness presence: Blockchain consensus for augmented democracy in smart cities," *Journal of Parallel and Distributed Computing*, journal 145, pages 160–175, 2020.

[4] F. Musciotto, S. Delpriori, P. Castagno, *et al.*, "Mining social interactions in privacy-preserving temporal networks," IEEE, 2016.

[5] F. Adib, Z. Kabelac, D. Katabi, *et al.*, "3D Tracking via Body Radio Reflections," *Usenix NSDI*, december 2013.

[6] Y. Ma, G. Zhou, and S. Wang, "Wifi Sensing with Channel State Information: A Survey," *ACM Comput. Surv.*, journal 52, number 3, Jun. 2019.

[7] K. Qian, C. Wu, Z. Yang, *et al.*, *ACM Transactions on Embedded Computing Systems (TECS)*, 2018.

[8] J. Lv, D. Man, W. Yang, *et al.*, "Robust WLAN-Based Indoor Intrusion Detection Using PHY Layer Information," *IEEE Access*, journal 6, 2018, ISSN: 2169-3536.

[9] K. Qian, C. Wu, Z. Yang, *et al.*, "Widar: Decimeter-Level Passive Tracking via Velocity Monitoring with Commodity Wi-Fi," New York, NY, USA: Association for Computing Machinery, 2017.

[10] X. Li, D. Zhang, Q. Lv, *et al.*, "Indotrack: Device-free indoor human tracking with commodity wi-fi," journal 1, number 3, 2017.

[11] D. Vasisht, S. Kumar, and D. Katabi, "Decimeter-level localization with a single wifi access point," march 2016.

[12] S. Yousefi, H. Narui, S. Dayal, *et al.*, "A survey on behavior recognition using wifi channel state information," *IEEE Communications Magazine*, journal 55, number 10, pages 98–104, 2017.

[13] L. Guo, L. Wang, J. Liu, *et al.*,

[14] H. Wang, D. Zhang, Y. Wang, *et al.*, "Rt-fall: A real-time and contactless fall detection system with commodity wifi devices," journal 16, number 2, pages 511–526, 2017.

[15] T. Chowdhury, "Using Wi-Fi Channel State Information (CSI) for Human Activity Recognition and Fall Detection," Ph.D. dissertation, april 2018.

[16] "IEEE Standard for Information technology — Telecommunications and information exchange between systems Local and metropolitan area networks — Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," 2016.

[17] D. Halperin, W. Hu, A. Sheth, *et al.*, "Tool release: Gathering 802.11n traces with channel state information," *SIGCOMM Comput. Commun. Rev.*, page 53, january 2011.

[18] 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*, Paris, France, 2015, 53–64.

[19] M. Schulz, D. Wegemer, and M. Hollick. (2017). "Nexmon: The C-based Firmware Patching Framework," [Online]. Available: https://nexmon.org.

[20] NovelSense, *Atheros-csi-tool-userspace-app*. [Online]. Available: https://github.com/NovelSense/Atheros-CSI-Tool-UserSpace-APP.

[21] Y. Xie, *The user-space applications for atheros-csi-tool*. [Online]. Available: https://github.com/xieyaxiongfly/Atheros-CSI-Tool-UserSpace-APP.

[22] S. Sen, B. Radunovic, R. R. Choudhury, *et al.*, "You Are Facing the Mona Lisa: Spot Localization Using PHY Layer Information," New York, NY, USA: Association for Computing Machinery, 2012.

[23] H. Ismail Fawaz, B. Lucas, G. Forestier, *et al.*, "InceptionTime: Finding AlexNet for Time Series Classification," *Data Mining and Knowledge Discovery*, journal 34, Sep. 2019.