A Doppler-based Human Activity Recognition System using WiFi Signals

Yao Ge[1], Shibo Li[1,2], Minjian Shentu[2], Ahmad Taha[1], Shuyuan Zhu[2], Jonathan Cooper[1]
Muhammad Imran[1], Qammer Abbasi[1]
[1]University of Glasgow, Glasgow, UK
[2]University of Electronic Science and Technology of China, Chengdu, China

***Abstract*—WiFi-based human activity recognition has drawn a lot of attention in recent years due to the low cost and high popularity of WiFi devices. The wireless monitoring system is able to efficiently detect abnormal activities like falling and body shaking, without privacy invasion. In this paper, we propose a framework using Doppler Frequency Shift-based methodology to extract the features and classify different activities with channel state information collected from WiFi devices. The experimental results demonstrate the reliability of our method for the application of activity recognition.**

***Index Terms*—human activity recognition, channel state information, Doppler effect, WiFi sensing, deep learning**

I. Introduction

Human activity recognition (HAR) plays a significant role in healthcare monitoring. Most monitoring systems are developed based on video camera, which poses several privacy concerns. Wearable devices, on the other hand, do not pose significant privacy concerns, however they have to be mounted on the body, which is very challenged, especially for the elder and disabled persons. Moreover, the use of wearable devices means patients can simultaneously use the same device, which can induce the transmission of viruses and diseases like the current COVID-19 virus. Thereby, device-free wireless sensing is a viable alternative to be applied to the daily healthcare monitoring.

A. Related Work

For indoor monitoring task, frequency modulated continuous wave radar [1], [2], ultra wide band [3], and WiFi devices [4] have demonstrated their reliability and efficiency. WiFi-based sensing is becoming popular due to the wide applications of low-cost, commercial and off-the-shelf (COTS) WiFi devices. Several research studies have proposed the use of WiFi based wireless sensing for localization [5]–[8], tracking [9]–[13], activity detection [14]–[17], biometrics estimation [18], vital signs monitoring [19]–[21] and pose estimation [18], [22].

Doppler-based gesture recognition [15] was proposed to construct body velocity profile by using 3-6 WiFi devices. The recognition method correlates the Doppler frequency spectrum (DFS) collected from different devices to estimate body velocity along different directions and improves the robustness. Meanwhile, gait recognition [23], localizing and tracking [10], [12], and speed estimation [24], [25] all employed Doppler analysis. Besides, many systems have demonstrated that the deep neural network (DNN) is available in Doppler-based methodology to improve the performances. For instance, [15] also proposed a DNN learning model for gesture recognition. Moreover, the authors in [25], proposed a model based on convolutional neural network to obtain the mapping relationship between Doppler spectrum and velocity profile.

B. Technical Background

a) Channel State Information (CSI): In the wireless transmission path, the signal is interrupted by the physical environment, which results in reflection, scattering and multi-path fading, causing different signal strength and phase of the signal transmitted on each sub-carrier. The channel state information matrix of given subcarrier with frequency f and time t can be represented as [26]:

$$H(f,t) = e^{-j2\pi\Delta ft}(H_s(f) + \sum_{i=1}^{N_d} a_i(f,t)e^{-j2\pi d_i(t)\lambda}) \quad (1)$$

where $e^{-j2\pi\Delta ft}$ represents the phase shift of WiFi devices, H_s represents the CSI from static paths, containing the signals reflected from stationary objects and transmitted in the Line-of-Sight area. In the parenthesis, all the terms represent the summation of signals from all the signals influenced by dynamic objects. N_d is the index of dynamic path, $a_i(f,t)$ represents complex attenuation factor and initial phase of i^{th} path, $e^{-j2\pi d_i(t)\lambda}$ represents the phase change of i^{th} path, $d_i(t)$ is the length of i^{th} path and λ is wavelength of wireless signal. For each CSI packet, the structure is shown in the Fig. 1. , where M and N stand for the number of transmitter antennas and receiver antennas, respectively, and K represents the different subcarrier index.

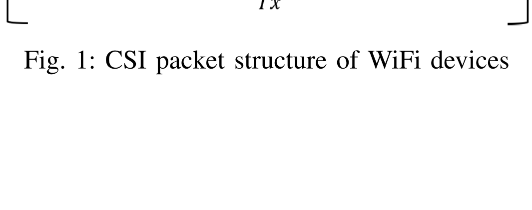


Fig. 1: CSI packet structure of WiFi devices

b) Transformer: Transformer is a classical natural language processing model proposed by the Google in 2017 [27]. It uses self-attention mechanism and does not use the sequential structure of recurrent neural networks. Alternatively, the model is trained in parallel and can acquire global information. The parallel processing capability of the Transformer model has demonstrated significant improvement in processing time as compared to mainstream sequential models. This was estimated to be 1.83 - 3.37 times faster in CSI-based activity classification, according to [28].

In this paper, we propose a DFS-based WiFi sensing framework using Transformer to correlate different DFS information with one transmitter and two receivers, to accurately classify different activities.

II. Methodology and System Design

We focus on the design of an indoor activity classification system using COTS WiFi devices. With multiple WiFi devices placed surrounding the monitoring area, wireless CSI data can be collected, but is distorted by human activities. In this section, we introduce the overall processing and classification procedures, including data collection, pre-processing, Doppler-based feature extraction and Transformer-based recognition. Accordingly, the proposed system can be divided into two modules, Doppler-based feature extraction module and the activity recognition module.

A. System Setup

a) System Implementation: The proposed system is implemented based on one WiFi transmitter and two WiFi receivers, equipped with Intel 5300 wireless network interface controller (NIC) and the relevant CSI driver [29]. Given few interfering radios, the driver is set with monitoring mode on the channel at 5 GHz with 40 MHz bandwidth. The transmitter is set to broadcast the rate of WiFi packets at $1,000$ packets per second with the receivers activated in a line, where there is a 5 cm (larger than the half-wave wavelength) distance to avoid the interference of closed antennas.

b) Evaluation Setup: Our system was tested in one indoor environment illustrated in Fig. 2a with a 2.4 m $\times$ 1.6 m rectangular monitoring area, as shown in Fig. 2b. Two volunteers participated in the designed experiments where each performed five different activities for data collection, to increase the inter-class variation and strengthen the AI model. In the experiment, all the volunteers were asked to finish one activity in three seconds and each data sample contains 3,000 packets CSI data. The overall dataset contains 500 samples of different activities including walking, jogging, leaning forward and back, putting both hands up and down, and waving left hand. The choice of activities was in relation to the envisioned applications and use cases of this system, which includes future healthcare sensing for in-home care (walking, jogging, leaning forward and back) and smart-home control (gesture recognition).

B. Pre-processing of WiFi CSI data

WiFi CSI data is not only influenced by high-frequency noise and phase offset but also disturbed by the static and dynamic motion from other sources, as indicated in Eq. 1. It is necessary to reduce the noise in the data before performing feature extraction. There are three noise components that require attention:

- Static components of CSI data from the environment.
- High-frequency noise from the communication hardware.
- Phase offset from the NIC.

Fortunately, a number of methods have been proposed to solve the above problem. For the phase offset, Widance system [30] provided a reliable phase offset removal methods. Given that the phase offset of all antennas should be the same which are produced from single NIC, the method uses the phase information from a selected antenna with lower sensitivity of dynamic information to unify the phase offset of the rest two antennas. IndoTrack [11] proposed an amplitude adjustment method to reduce the influence of static components in CSI data. Then, Butterworth band-pass filter and principle component analysis were applied to filter other noise [15]. Our proposed system applies all of the previously mentioned pre-processing techniques to extract the DFS.

C. Doppler Frequency Shift Extraction

In this section, the Doppler method is used for feature extraction to get the profile of different persons' activities. The DFS generated by different limb movements can be used to identify the specific human activities. As the CSI data has been pre-processed, we applied the short-time Fourier transform on the dataset to get the Doppler spectrum, which represents the velocity variation of torso and limbs. There are two DFS representing two activities as shown in Fig. 3.

In Fig. 3a, the spectrum illustrates the activity of leaning forward and back, with a peak representing the fastest speed of the torso. In Fig. 3b, the abrupt change stands for the fast movement of the arms.

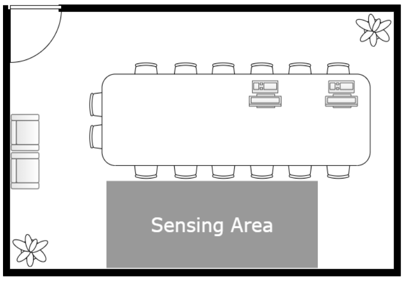
D. Activity Recognition

Our HAR monitoring system is achieved by a two-stream Transformer model, which performs human activity classification by learning the features of DFS. It involves extracting both the temporal and spatial features which contain the information within the entire sequence. The core of this model is the multi-scale convolution augmented transformer. The input DFS, with Gaussian encoding, is input into the two-stream self-attention layers to extract features, and finally pass through the feature fusion layer to obtain the prediction of human activity.

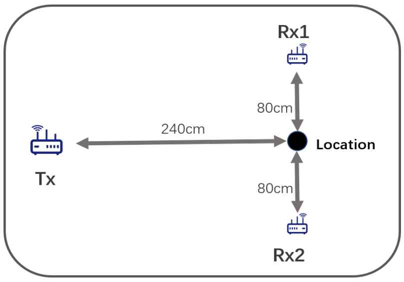
III. Evaluation

A. Network Training

The input of our network is the Doppler spectrum extracted from raw CSI data, as shown in Fig. 3. The input size of each sample is 121×3000, where the 121 represents the resolution of frequency from -60 Hz to 60 Hz, and 3000 represents the time series of milliseconds, where each sample lasts 3 seconds. To focus on the human activity monitoring, and improve the efficiency of the neural network, we adjusted the frequency

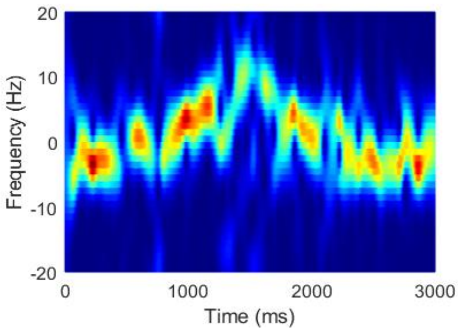


(a) Layouts of the evaluation environment

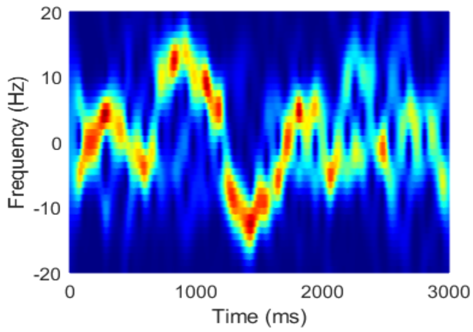


(b) Location of devices and the user

Fig. 2: Layout of experiment setup



(a) Leaning forward and back



(b) Put hands up and down

Fig. 3: Doppler spectrum of indoor activity profile

range to -20 Hz to 20 Hz, 3000 sets of CSI data was down-sampled to 500. Besides, due to the relatively small size of our dataset, we use a single temporal module to train Transformer. At first, the dataset of five activities from two people is training and then the third people is introduced. Finally an identity classification for the first two users is conducted. The whole dataset is divided into two sets, where 80% data are used as training set and 20% as test set.

B. Results

To quantitatively measure the performance of our system, we applied 5-fold cross validation, which divided sample data randomly into five parts, each time one is selected as the test set and the rest was used for training the system. The results of five activities of two volunteers achieve 87.6% accuracy (on average) in the 5-fold cross validation, with the best result of five achieves 92.7%. Table I shows the confusion matrix.

TABLE I: Confusion matrix for all five classes with Transformer-based activity classification

All Five Classes - Accuracy (87.6%)							
	Class		***Predict Class***				
			1	2	3	4	5
True Class	1	Walking	81	16	2	1	0
	2	Jogging	16	76	4	0	0
	3	Leaning forward and back	2	2	94	0	0
	4	Put the hands up and down	1	0	0	91	8
	5	Wave left hand	0	0	0	4	96

The table illustrates similarities between different actions. For example, jogging and walking classes are partially confused because the two activities have similar body movements. To explore the factors that influence activity recognition performance, the identity classification is conducted on a single walking behavior of user-1 and user-2. The dataset is randomly divided into 80% and 20% for training and testing, respectively. Our proposed system performs an overall accuracy of 77%, which shows that different identities influence the Doppler-based human activity classification result and it is possible to recognize the identity of different persons.

IV. Conclusion

In this paper, we develop an HAR system based on the Doppler feature extraction, exploiting low-cost WiFi devices operating at 5 GHz. The results demonstrate that the DFS information is highly related to each movement of individual. Using transformer neural network, the system is capable of classifying the activities as well as identifying the subjects' identity with an accuracy of 87.6% and 77%, respectively, in the test set. In our future work, the indoor gesture detection, the activity status detection such as walking and jogging, the abnormal activity detection such as leaning forward and back (regarded as falling) with larger dataset and more types of activities. This will satisfy the requirements of future smart-home design and healthcare monitoring.

References

[1] F. Adib, Z. Kabelac, and D. Katabi, "Multi-person localization via RF body reflections," in *12th USENIX Symposium on Networked Systems Design and Implementation*, Oakland, CA, May 2015, pp. 279–292.

[2] R. Zhang and S. Cao, "Real-Time Human Motion Behavior Detection via CNN Using mmWave Radar," *IEEE Sensors Letters*, vol. 3, no. 2, pp. 2019–2022, 2019.

[3] F. Khan, A. Ghaffar, N. Khan, and S. H. Cho, "An overview of signal processing techniques for remote health monitoring using impulse radio UWB transceiver," *Sensors (Switzerland)*, vol. 20, no. 9, 2020.

[4] Y. Ma, G. Zhou, and S. Wang, "WiFi sensing with channel state information: A survey," *ACM Computing Surveys*, vol. 52, no. 3, 2019.

[5] C. Chen, Y. Chen, H.-Q. Lai, Y. Han, and K. R. Liu, "High accuracy indoor localization: A wifi-based approach," in *2016 IEEE International Conference on Acoustics, Apeech and Signal Processing*. IEEE, 2016, pp. 6245–6249.

[6] C. Chen, Y. Chen, Y. Han, H. Q. Lai, F. Zhang, and K. J. Liu, "Achieving Centimeter-Accuracy Indoor Localization on WiFi Platforms: A Multi-Antenna Approach," *IEEE Internet of Things Journal*, vol. 4, no. 1, pp. 122–134, 2017.

[7] X. Guo, N. R. Elikplim, N. Ansari, L. Li, and L. Wang, "Robust WiFi Localization by Fusing Derivative Fingerprints of RSS and Multiple Classifiers," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 5, pp. 3177–3186, 2020.

[8] L. Li, X. Guo, and N. Ansari, "SmartLoc: Smart Wireless Indoor Localization Empowered by Machine Learning," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 8, pp. 6883–6893, 2020.

[9] J. Schmitz, F. Bartsch, M. Hernández, and R. Mathar, "Distributed software defined radio testbed for real-time emitter localization and tracking," in *2017 IEEE International Conference on Communications Workshops*. IEEE, 2017, pp. 1246–1252.

[10] K. Qian, C. Wu, Z. Yang, Y. Liu, and K. Jamieson, "Widar: Decimeter-level passive tracking via velocity monitoring with commodity wi-fi," in *Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, 2017, pp. 1–10.

[11] X. Li, D. Zhang, Q. Lv, J. Xiong, S. Li, Y. Zhang, and H. Mei, "IndoTrack: Device-Free Indoor Human Tracking with Commodity Wi-Fi," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, pp. 1–22, 2017.

[12] K. Qian, C. Wu, Y. Zhang, G. Zhang, Z. Yang, and Y. Liu, "Widar2.0: Passive human tracking with a single Wi-Fi link," *MobiSys 2018 - Proceedings of the 16th ACM International Conference on Mobile Systems, Applications, and Services*, pp. 350–361, 2018.

[13] S. Shi, S. Sigg, L. Chen, and Y. Ji, "Accurate Location Tracking from CSI-Based Passive Device-Free Probabilistic Fingerprinting," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 6, pp. 5217–5230, 2018.

[14] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Gesture Recognition Using Wireless Signals," *GetMobile: Mobile Computing and Communications*, vol. 18, no. 4, pp. 15–18, 2015.

[15] Y. Zheng, Y. Zhang, K. Qian, G. Zhang, Y. Liu, C. Wu, and Z. Yang, "Zero-effort cross-domain gesture recognition with wi-fi," in *Proceedings of the 17th ACM Annual International Conference on Mobile Systems, Applications, and Services*, 2019, p. 313–325.

[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] Y. Wang, K. Wu, and L. M. Ni, "WiFall: Device-Free Fall Detection by Wireless Networks," *IEEE Transactions on Mobile Computing*, vol. 16, no. 2, pp. 581–594, 2017.

[18] F. Wang, J. Han, S. Zhang, X. He, and D. Huang, "Csi-net: Unified human body characterization and pose recognition," *arXiv e-prints*, pp. arXiv–1810, 2018.

[19] C. Uysal and T. Filik, "RF-Based noncontact respiratory rate monitoring with parametric spectral estimation," *IEEE Sensors Journal*, vol. 19, no. 21, pp. 9841–9849, 2019.

[20] D. Zhang, Y. Hu, Y. Chen, and B. Zeng, "Breathtrack: Tracking indoor human breath status via commodity wifi," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 3899–3911, 2019.

[21] X. Wang, C. Yang, and S. Mao, "On CSI-Based Vital Sign Monitoring Using Commodity WiFi," *ACM Transactions on Computing for Healthcare*, vol. 1, no. 3, pp. 1–27, 2020.

[22] F. Wang, S. Zhou, S. Panev, J. Han, and D. Huang, "Person-in-wifi: Fine-grained person perception using wifi," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 5452–5461.

[23] W. Wang, A. X. Liu, and M. Shahzad, "Gait recognition using WiFi signals," *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 363–373, 2016.

[24] H. Deng, D. Zhang, Y. Hu, and Y. Chen, "Estimating indoor human speed via radio signals," in *GLOBECOM 2020-2020 IEEE Global Communications Conference*. IEEE, 2020, pp. 1–6.

[25] Y. Chen, H. Deng, D. Zhang, and Y. Hu, "SpeedNet: Indoor Speed Estimation with Radio Signals," *IEEE Internet of Things Journal*, vol. 8, no. 4, pp. 2762–2774, 2021.

[26] 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*, 2015, p. 65–76.

[27] A. Waswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Neural Information Processing Systems*, 2017.

[28] B. Li, W. Cui, W. Wang, L. Zhang, Z. Chen, and M. Wu, "Two-stream convolution augmented transformer for human activity recognition," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 1, 2021, pp. 286–293.

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

[30] K. Qian, C. Wu, Z. Zhou, Y. Zheng, Z. Yang, and Y. Liu, "Inferring motion direction using commodity Wi-Fi for interactive exergames," *Conference on Human Factors in Computing Systems - Proceedings*, vol. 2017-May, pp. 1961–1972, 2017.