Wifi-Based Device-Free Gesture Recognition Through-the-Wall

Conference Paper · June 2021

DOI: 10.1109/ICASSP39728.2021.9414894

3 authors:

Sai Deepika Regani
University of Maryland, College Park
10 PUBLICATIONS **88** CITATIONS
SEE PROFILE

Beibei Wang
University of Maryland, College Park
108 PUBLICATIONS **5,492** CITATIONS
SEE PROFILE

K. J. Ray Liu
University of Maryland, College Park
858 PUBLICATIONS **29,405** CITATIONS
SEE PROFILE

ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-7281-7605-5/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICASSP39728.2021.9414894

WIFI-BASED DEVICE-FREE GESTURE RECOGNITION THROUGH-THE-WALL

Sai Deepika Regani, Beibei Wang, K. J. Ray Liu
University of Maryland, College Park, MD 20742, USA.
Origin Wireless Inc., 7500 Greenway Center Drive, Suite 1070, MD 20770, USA.
Email: rdeepika@terpmail.umd.edu, {bebewang,kjrliu}@umd.edu.

ABSTRACT

Device-free (passive) gesture recognition offers an enormous potential to simplify Human-Computer Interaction (HCI) in future smart environments. WIFI-based gesture recognition approaches have attained acclaim amongst others due to the omnipresence, privacy-preservation, and broad coverage of WIFI. However, there is no universal solution built on off-the-shelf devices that can accommodate an expandable set of gestures in a through-the-wall setting. In this work, we propose such a gesture recognition system that can recover information about the actual trajectory of the hand movement allowing an expandable set of gestures. Further, we leverage the rich multipath in a through-the-wall setting to develop a statistical model for the channel variations induced by a hand gesture. This model is used to derive a correspondence between the relative distance moved by the hand and the Time Reversal Resonating Strength (TRRS) decay. Based on this relation and the geometry of the gesture shape, we design feature extraction modules to enable gesture classification. We built a prototype of the proposed system on off-the-shelf WIFI devices and achieved a classification accuracy of 87% on a set of 6 uppercase English alphabets.

Index Terms— Gesture Model, Time Reversal Resonating Strength (TRRS), Channel State Information (CSI), WIFI sensing, HCI.

1. INTRODUCTION

Human-Computer Interaction (HCI) is becoming a notable portion of our routine as we are surrounded by an increasing number of smart devices and computers. Gesture recognition will enable diverse applications and enhance user convenience in the future of HCI and smart indoor environments. Many gesture recognition modalities are therefore being investigated, such as vision-based, device/sensor-based [1], acoustic-based [2], RF-based [3], WIFI-based [4], and so on. Among the many approaches, WIFI-based techniques are the most attractive due to its all-pervasiveness, ability to penetrate walls, accessible hardware, and privacy preservation. In this work, we built a device-free (passive) gesture recognition system using off-the-shelf WIFI devices that can work in through-the-wall scenarios and over an expandable set of gestures.

The earliest wireless channel descriptor used in WIFI-based approaches is the Received Signal Strength Indicator (RSSI). The patterns in the signal strength, such as the rising and falling edges, number of repetitions [5], and, more recently, angle-of-arrival based on the occluded signal strength [6], were used to achieve gesture recognition. RSSI-based approaches are highly sensitive to changes in the environment and location and do not provide fine-grained channel information. With the availability of the Channel State Information (CSI), a more detailed channel descriptor, new approaches were developed to achieve multiple wireless sensing applications [7, 8], including gesture recognition. However, due to the intense multipath in a typical indoor environment, it became difficult to isolate the signal of interest from the interleaved background/environment information. Most works used direct CSI matching by gathering prototypes in the training phase and associating them with those recorded in the testing phase. Similarity metrics such as Dynamic Time Warping (DTW) [9] or learning-based approaches such as machine learning [10] or deep learning [11] were used for comparison. Such techniques suffered from the "location-dependency" problem since the prototypes had a significant multipath contribution from the structure of the location itself and did not match those collected in a new environment. In this work, instead of using the actual value of CSI, we use the statistical relationships between CSIs, which are location independent.

Researchers later used the CSI multipath models to understand the correspondence between the movement of the hands and the observed CSI behavior. Significant works in this domain include the Fresnel zone model [12], phase-distance model [13], and the virtual antenna model [14]. However, these approaches had a strict line-of-sight (LOS) requirement that allowed focus on the signal reflected off the hand. Few other works used angle-of-arrival, time-of-flight, and Doppler shift information [15] to reconstruct the gesture shapes, which either tweaked the existing hardware or required strong LOS. Differently, we do not impose any such constraints and work directly with off-the-shelf hardware in a through-the-wall setup.

Few works exploited the secondary features of the gesture trajectory, such as speed and repetitions, to ease the de-

vice placement constraints. *WiGest* used the speed, relative distance, and repetitions to classify gestures [5]. Likewise, *WiGrep* used the number of repetitions in a gesture for classification in a through-the-wall setting [16]. Using secondary features seems to be a promising direction, yet the number of distinct gestures is limited. Instead, we aim to extract the primary features of the gesture trajectory, such as the number of constituted line segments, the angle between segments, and the relative location of intersection points. Such an approach allows an easy extension of the set of gestures.

In summary, there is no existing gesture recognition system based on commercial WIFI that can achieve all the four goals, namely, (i) Location/environment independence, (ii) Through-the-wall operation, (iii) Unconstrained device placement, and an (iv) Expandable broad set of gestures. This work aims to deliver all these goals, thereby enabling a practical and user-friendly gesture recognition system. The contributions of this work can be summarized as follows:

- We proposed a unique hand gesture model and derived a correspondence between the TRRS decay and the relative distance moved by the hand in the through-the-wall scenario.
- We have developed a complete pipeline for a gesture recognition system involving several feature extraction modules and the gesture classification methodology.
- We created a prototype of the proposed system on commercial WIFI chipsets and achieved a classification accuracy of 87% on a subset of the upper case English alphabets. The extension to a broader set of gestures is evident.

2. GESTURE MODEL

This section develops a statistical model to quantify the nature of channel variations induced by the hand gesture. For this, we adopt the Time Reversal Resonating Strength (TRRS) as the similarity measure between different channel instances. Let $\mathbf{h}_1$ and $\mathbf{h}_2$ be the Channel Impulse Responses (CIR) at times t_1 and t_2 respectively. The TRRS(η) between $\mathbf{h}_1$ and $\mathbf{h}_2$ is defined as [17]:

$$\eta(\mathbf{h}_1, \mathbf{h}_2) = \left| \frac{\sum_{l=0}^{L-1} h_1(l) h_2(l)^*}{\sqrt{\sum_{l=0}^{L-1} |h_1(l)|^2} \sqrt{\sum_{l=0}^{L-1} |h_2(l)|^2}} \right|, \quad (1)$$

where L is the number of CIR taps and $(.)^*$ is the complex conjugate operator. The range of η is from 0 to 1 with a value of 1 indicating the same channel.

Let $\mathbf{h}_0$ denote the CIR at $t = 0$ whose l^{th} tap can be explicitly written in terms of the multipath components (MPCs) as:

$$h_0(l) = \sum_{m \in M} \zeta_{l,m} e^{-j2\pi f_c \tau_{0,l}(m)}, \quad (2)$$

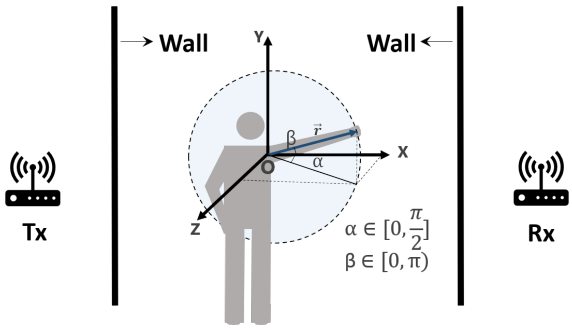


Fig. 1: Hand gesture with a stretched arm and pivoted about the shoulder.

where M is the set of multipath, ζ is the complex path gain, f_c is the carrier frequency and τ is the path delay. When the hand is moved by a short distance x, let the corresponding CIR be denoted by $\mathbf{h}_x$. Then,

$$h_0(l) h_x(l)^* = \Big(\sum_{m \in M} \zeta_{l,m} e^{-j2\pi f_c \tau_{0,l}(m)} \Big) \Big(\sum_{m \in M} \zeta^{\star}_{l,m} e^{j2\pi f_c \tau_{x,l}(m)} \Big). \quad (3)$$

Assuming a sufficiently large bandwidth such that the significant MPCs are captured on distinct CIR taps [17], we can approximate the numerator of $\eta(\mathbf{h}_0, \mathbf{h}_x)$ as:

$$\sum_{l=0}^{L-1} h_0(l) h_x(l)^* \approx \sum_{l=0}^{L-1} |\zeta_l|^2 e^{j2\pi f_c (\tau_{x,l} - \tau_{0,l})}. \quad (4)$$

In this work, we propose a unique way of performing a gesture with the arm stretched out, and the movement pivoted about the shoulder, as shown in Fig. 1. In such a motion, the hand can be modeled as composed of N_{D} dynamic scatterers whose displacement increases linearly from 0 to x with x being the displacement of the wrist. Further, in a through-the-wall setting, we can assume a uniform scattering environment [17] and that only a fraction of the total MPCs, a are affected by the movement of the hand. Under these assumptions, we can approximate Eq.(4) as:

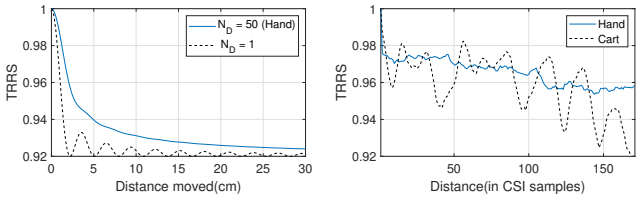
$$K \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \Big(\sum_{i=1}^{N_{\text{D}}} (1-a) + a e^{j2\pi f_c (\frac{2xi \text{Cos}\nu \text{Cos}\psi}{cN_{\text{D}}})} \Big) d\nu d\psi, \quad (5)$$

where K is a constant factor, ν is the angle between the direction of hand movement and normal to the reflecting surface, and ψ is the reflection angle of the moving scatterer (i.e., a part of the hand) [18]. Then, $\eta(\mathbf{h}_0, \mathbf{h}_x)$ can be simplified to:

$$\eta(\mathbf{h}_0, \mathbf{h}_x) = T_{\text{S}} + (1 - T_{\text{S}}) \frac{1}{N_{\text{D}}} \sum_{i=1}^{N_{\text{D}}} J_0^2 \Big(\frac{2\pi f_c x i}{c N_{\text{D}}} \Big), \quad (6)$$

where T_{S} indicates the contribution from the static MPCs and $J_0(.)$ is a zeroth-order Bessel function of the first kind [18]. For $N_{\text{D}} = 1$, η corresponds to the TRRS decay for a single

dynamic scatterer moving with one speed. Eq.(6) is numerically evaluated for $N_{\mathrm{D}} = 1$ and $N_{\mathrm{D}} = 50$ using $T_{\mathrm{S}} = 0.92$ for a distance of 30 cm. The theoretical TRRS decay curves are shown in Fig. 2a. In Fig. 2b, we show the experimental TRRS decay curves for the hand movement as proposed in this work vs. the motion of a cart over a distance of about 30 cm. Due to a lack of time-distance synchronization, we do not show the instantaneous value on the x-axis. However, it can be observed that the monotonous decay in the case of hand movement occurs over a much larger distance compared to the cart motion or when $N_{\mathrm{D}} = 1$. This monotonous decay helps us determine the relative distance moved by the hand, which, along with the gesture trajectory's geometric relationships, can be used to extract useful features for gesture classification.



(a) Theory (b) Experiment

Fig. 2: TRRS in case of hand movement (dynamic scatterers with increasing speed) and a cart movement (dynamic scatterers with one speed).

3. GESTURE CLASSIFICATION

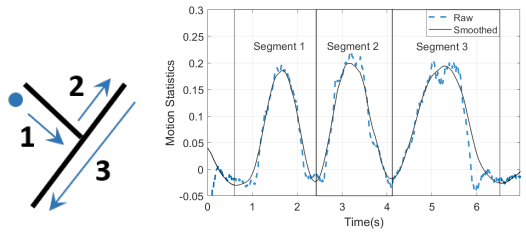
In this work, we consider gesture trajectories that are composed of straight-line segments only. Since we do not have information about the absolute direction of movement of the hand, we aim to recover the trajectory's relative shape irrespective of the direction/orientation. For example, the recovered gesture trajectories for the letters 'N' and 'Z' will be the same. Likewise, the letters 'W' and 'M' will appear the same without the absolute direction information.

Note that we obtain CSI from off-the-shelf WIFI chipsets instead of CIR. In the implementation, we exploit the frequency domain equivalent of TRRS calculation, as discussed in [19].

3.1. Feature Extraction

For describing the relative shape of a trajectory composed of straight-lines, it is sufficient to know the number of line segments present, the angle between two adjacent line segments, and the intersection points between line segments, if any. Let us briefly look at determining these features.

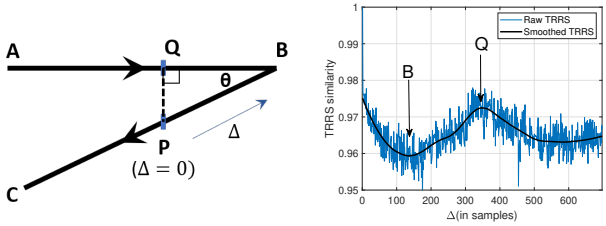
Gesture Segmentation: This module aims to identify the number of line segments in a gesture. When the hand changes direction from one segment to another, i.e., at the location of a turn, the speed of the hand is nearly zero. Such instances are detected using a motion detector [20]. We calculate motion statistics, which indicate the extent of movement in the environment and identify the turn locations from the locations of local minima/valleys. For instance, Fig. 3a shows the Y-shaped trajectory with the dot indicating the initial point, and Fig. 3b shows the corresponding motion statistics with the three identified partitions corresponding to the three segments.



(a) (b)

Fig. 3: (a) Y-shaped gesture with the dot indicating the initial location, and (b) Corresponding motion statistics and gesture segments.

Turn Angle Classification: This module classifies the angle between two segments in a gesture into $0°$, acute, and obtuse angles. For this, we leverage the TRRS monotonous decay that we derived in Section 2. Consider a turn in a gesture, as

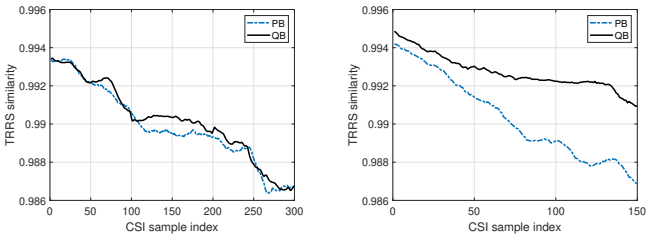


(a) (b)

Fig. 4: (a) Geometry of a turn in a gesture (b) TRRS decay feature.

shown in Fig. 4a. Let A be the initial point, B be the location of the turn, and C is the endpoint. Let P be any point on the second segment BC. Suppose we compute the TRRS between CSI corresponding to point P and the CSIs corresponding to points before P, along the gesture trajectory. In that case, we obtain a curve, as shown in Fig. 4b. The peak corresponds to point Q, and the valley corresponds to point B following the monotonous decay observation. B and Q instances are thereby derived from the curve, and the pairwise TRRS between the points P, B, and Q is computed. The angles are then classified according to the following rule: (i) If TRRS(PB)$\approx$ TRRS(QB) $\implies \theta = 0°$, (ii) If TRRS(PB)$<$TRRS(QB) $\implies 0° < \theta < 90°$(acute) and (iii) If Q does not exist, then

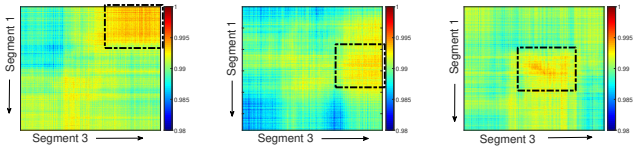
$\theta > 90°$(obtuse). We repeat the rule for different points in segment two and take a majority decision to classify the angle. Fig. 5 shows the TRRS of PB vs. QB for a $0°$ and an acute angle turn, as the point P moves along the second segment, i.e., from turn location, B to the endpoint, C.



(a) (b)

Fig. 5: Turn angle classification feature in (a) $0°$, and (b) acute angle.

Matching Point Detection: This module aims to identify if two straight line segments in a gesture intersect. For this, we find the TRRS between every pair of points on the two segments to build a TRRS matrix and extract significant peaks from the matrix. The relative location of the peak can provide useful information for classifying gestures. For instance, Fig. 6a shows the TRRS matrix between the 1^{st} and 3^{rd} segments for a D-shaped gesture. In this, the matching point is located at the beginning of segment one and the end of segment three, as can be verified from the shapes in Fig. 7b. Likewise, P-shaped and X-shaped gestures have unique matching point locations that serve as features for classification.



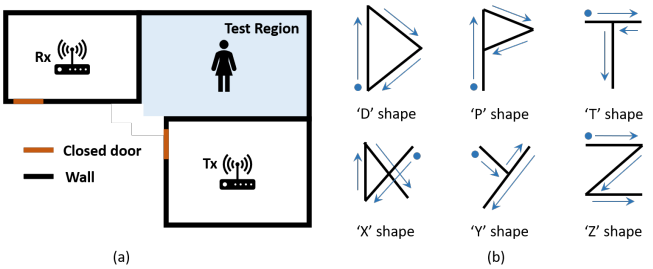
(a) (b) (c)

Fig. 6: Matching point detection for (a) 'D'-shaped, (b) 'P'-shaped, and (c) 'X'-shaped gestures.

The gesture classification is performed by gathering all the features derived in the previous modules and using simple conditional logic. For example, if a gesture consists of 3 segments, two acute angles, and the first and third segments intersect in their mid-regions, it is an X-shaped gesture. Such a simplistic approach for classification is the key to an expandable set of gestures.

4. PERFORMANCE EVALUATION

To evaluate the performance of the proposed system, we built a prototype on off-the-shelf commercial WIFI chipsets with a bandwidth of 80 MHz and 2×2 MIMO using a sampling rate of 300 Hz. The experimental setup is as shown in Fig. 7a., where two walls separate the transmitter (Tx), and the receiver (Rx), and one wall separates the test region and both the Tx and the Rx.



(a) (b)

Fig. 7: (a) A typical through-the-wall experimental setup, (b) Examples of gesture shapes with 3 segments.

Gesture Classification: On the dataset of D, P, T, X, Y, and Z shapes (Fig. 7b) with 50 samples each, the confusion matrix of the classification is as shown in Fig. 8, with an average accuracy of 87%. The higher accuracy for 'Y' and 'T' shapes can be attributed to the more unique features, i.e., both from the matching point detection and the $0°$ turn angle.

True label \ Predicted label	'D'	'P'	'T'	'X'	'Y'	'Z'
'D'	**0.78**	0.1	0	0	0	0.12
'P'	0.14	**0.76**	0	0	0	0.10
'T'	0.06	0	**0.88**	0	0	0.06
'X'	0.12	0	0	**0.80**	0.04	0.04
'Y'	0	0	0	0	**0.98**	0.02
'Z'	0	0	0	0	0	**1**

Fig. 8: Confusion matrix for gesture classification.

5. CONCLUSION

We have proposed and built a device-free gesture recognition system over a broad set of gestures using commercial WIFI chipsets in a through-the-wall setup. Using a gesture model, we have proved that the TRRS decay is monotonous with the distance moved by the hand. This observation, combined with the geometric relationships, allowed us to design algorithms that extract features for gesture classification. For a gesture set composed of 6 uppercase English alphabets, recognition accuracy of 87% is achieved. Moreover, the proposed system can be easily extended to a broader gesture set, enabling many future smart home environments and HCI applications.

6. REFERENCES

[1] S. Agrawal, I. Constandache, S. Gaonkar, R. Roy Choudhury, K. Caves, and F. DeRuyter, "Using mobile phones to write in air," in *Proceedings of the 9th international conference on Mobile systems, applications, and services*, 2011, pp. 15–28.

[2] M. Chen, P. Yang, J. Xiong, M. Zhang, Y. Lee, C. Xiang, and C. Tian, "Your table can be an input panel: Acoustic-based device-free interaction recognition," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 1, pp. 1–21, 2019.

[3] J. Wang, D. Vasisht, and D. Katabi, "RF-IDraw: Virtual touch screen in the air using RF signals," *ACM SIGCOMM Computer Communication Review*, vol. 44, no. 4, pp. 235–246, 2014.

[4] H. F. T. Ahmed, H. Ahmad, and C. Aravind, "Device free human gesture recognition using Wi-Fi CSI: A survey," *Engineering Applications of Artificial Intelligence*, vol. 87, p. 103281, 2020.

[5] H. Abdelnasser, K. Harras, and M. Youssef, "A ubiquitous WiFi-based fine-grained gesture recognition system," *IEEE Transactions on Mobile Computing*, vol. 18, no. 11, pp. 2474–2487, 2018.

[6] L. Sun, S. Sen, D. Koutsonikolas, and K.-H. Kim, "Widraw: Enabling hands-free drawing in the air on commodity wifi devices," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, 2015, pp. 77–89.

[7] K. J. R. Liu and B. Wang, *Wireless AI: Wireless Sensing, Positioning, IoT, and Communications*. Cambridge University Press, 2019.

[8] B. Wang, Q. Xu, C. Chen, F. Zhang, and K. J. R. Liu, "The promise of radio analytics: A future paradigm of wireless positioning, tracking, and sensing," *IEEE Signal Processing Magazine*, vol. 35, no. 3, pp. 59–80, 2018.

[9] M. A. A. Al-qaness and F. Li, "WiGeR: Wifi-based gesture recognition system," *ISPRS International Journal of Geo-Information*, vol. 5, no. 6, p. 92, 2016.

[10] W. He, K. Wu, Y. Zou, and Z. Ming, "WiG: WiFi-based gesture recognition system," in *2015 24th International Conference on Computer Communication and Networks (ICCCN)*. IEEE, 2015, pp. 1–7.

[11] Q. Bu, G. Yang, X. Ming, T. Zhang, J. Feng, and J. Zhang, "Deep transfer learning for gesture recognition with WiFi signals," *Personal and Ubiquitous Computing*, pp. 1–12, 2020.

[12] S. Ren, H. Wang, L. Gong, C. Xiang, X. Wu, and Y. Du, "Intelligent contactless gesture recognition using WLAN physical layer information," *IEEE Access*, vol. 7, pp. 92 758–92 767, 2019.

[13] D. Wu, R. Gao, Y. Zeng, J. Liu, L. Wang, T. Gu, and D. Zhang, "FingerDraw: Sub-wavelength Level Finger Motion Tracking with WiFi Signals," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 1, pp. 1–27, 2020.

[14] Z. Tian, J. Wang, X. Yang, and M. Zhou, "WiCatch: A Wi-Fi based hand gesture recognition system," *IEEE Access*, vol. 6, pp. 16 911–16 923, 2018.

[15] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *Proceedings of the 19th annual international conference on Mobile computing & networking*, 2013, pp. 27–38.

[16] S. D. Regani, B. Wang, M. Wu, and K. J. R. Liu, "Time Reversal Based Robust Gesture Recognition Using WiFi," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 8309–8313.

[17] F. Zhang, C. Chen, B. Wang, H.-Q. Lai, Y. Han, and K. J. R. Liu, "WiBall: A time-reversal focusing ball method for decimeter-accuracy indoor tracking," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 4031–4041, 2018.

[18] S. Thoen, L. Van der Perre, and M. Engels, "Modeling the channel time-variance for fixed wireless communications," *IEEE Communications letters*, vol. 6, no. 8, pp. 331–333, 2002.

[19] C. Chen, Y. Chen, Y. Han, H.-Q. Lai, F. Zhang, and K. J. R. 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, 2016.

[20] F. Zhang, C. Wu, B. Wang, H.-Q. Lai, Y. Han, and K. J. R. Liu, "WiDetect: Robust Motion Detection with a Statistical Electromagnetic Model," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 3, pp. 1–24, 2019.