An Improved CSI Based Device Free Indoor Localization Using Machine Learning Based Classification Approach

Tahsina Farah Sanam
Department of Electrical and Computer Engineering
Rutgers University
Email: tahsina.farah@rutgers.edu

Hana Godrich
Department of Electrical and Computer Engineering
Rutgers University
Email: godrich@rutgers.edu

***Abstract*—Indoor positioning system (IPS) has shown great potentials with the growth of context-aware computing. Typical IPS requires the tracked subject to carry a physical device. In this study, we present MaLDIP, a novel, machine learning based, device free technique for indoor positioning. To design the device free setting, we exploited the Channel State Information (CSI) obtained from Multiple Input Multiple Output Orthogonal Frequency-Division Multiplexing (MIMO-OFDM). The system works by utilizing frequency diversity and spatial diversity properties of CSI at target location by correlating the impact of human presence to certain changes on the received signal features. However, accurate modeling of the effect of a subject on fine grained CSI is challenging due to the presence of multipaths. We propose a novel subcarrier selection method to remove the multipath affected subcarriers to improve the performance of localization. We select the most location-dependent features from channel response based upon the wireless propagation model and propose to apply a machine learning based approach for location estimation, where the localization problem is shifted to a cell identification problem using the Support Vector Machine (SVM) based classifier. Experimental results show that MaLDIP can estimate location in a passive device free setting with a high accuracy using MIMO-OFDM system.**

I. Introduction

With the growth of pervasive, context aware and human-centric computing, development of an intelligent indoor environment is drawing increasing interest in literature. One of the central features of these systems is to be able to localize a person in indoors setting. This in return enables the development of various ubiquitous applications, such as rescue operations, building occupancy statistics, elder care, security enforcement, museum guidance for tourists, etc., to name a few. GPS systems are commonly used for positioning in outdoor environment. Designing feasible methods to carry out indoor localization is a challenge due to the poor performance of the GPS-based methods for indoors [1]. To this end, indoor localization based on wireless local area networks (WLAN) are getting more popular due to open access and lower cost of different wireless signals. For most of the existing works, the subject being tracked, may require to carry mobile devices [1]–[4]. However, accidental or intentional detachment of the device from the subject can terminate the tracking process. In this paper, a device-free localization technique is developed which can infer people's location (without wearing any device) by monitoring the changes in feature pattern of wireless signals available from the wireless links in the indoor environment.

Various WLAN-based indoor localization approaches have been studied in literature, estimating the location using "fingerprinting" approach, where a passive radio map is constructed utilizing the Received Signal Strength Indicator (RSSI) reported by the access points (AP). Later, in the presence of an entity in the area of interest, RSSIs from the Aps are measured and matched to the previously created radio map to infer the location [1], [2]. However, being the superposition of multipath components, RSSI not only varies over distance on the order of the signal wavelength but also fluctuates over time even at a static link, resulting in localization with lower accuracy [5].

Recently, fine grained PHY layer information of channel quality over all subcarriers for OFDM-based systems has gained interest for different applications due to its advantages over RSSI on several aspects. While RSSI provides MAC layer superposition of multipath signals, the PHY layer channel response is able to discriminate multipath characteristics [6], [7]. In [3], [4], authors perform CSI based location estimation using WiFi enabled device tagged with the subject. Device free indoor localization using CSI has been explored in [8], [9]. But these works considered CSIs from a Single Input Single Output (SISO) channel in OFDM system. In [8], authors perform CSI based device free localization through a probabilistic approach and achieve 85% accuracy where [9] adopts a power fading model based localization and achieve 90% accuracy with 1.5 meter localization error. Both of these works consider only SISO channel for CSI measurement.

Wireless technologies such as 802.11 a/g/n networks that use MIMO-OFDM, offer enhanced data throughput when compared with a SISO system in the presence of signal fading, interference, and multi-path. This advantage stems from exploiting the spatial diverse communication link [10]. The method proposed in this paper is different from existing CSI-based methods, in that it performs a device free localization by exploiting the spatial diversity of MIMO-OFDM system. It incorporates CSIs from multiple transmit and receive antenna

ISBN 978-90-827970-1-5 © EURASIP 2018

pairs to create a location-specific CSI fingerprint by leveraging wireless propagation model [11]. However, in the presence of rich multipath, it is difficult to model accurately the effect of a subject on the fine grained CSI. To overcome frequency selective fading, we propose a novel algorithm that adopts a threshold based subcarrier selection scheme from each MIMO link, choosing the ones which are less affected by multipath, resulting in higher accuracy in locating a subject. For extracting the most location-specific features from the pre-processed CSIs, we project them on a lower-dimensional linear subspace using Principle Component Analysis (PCA) based matrix factorization. Finally, we propose a machine learning based localization approach by shifting the localization problem to a cell classification problem using the SVM classifier. This facilitates the design of a novel device free indoor localization with superior performance over existing methods.

The rest of the paper is structured as follows. Section II presents the overall system design along with detailed methodology. Section III describes the experimental setup of the system and evaluates the performance of proposed method. Finally, the concluding remarks are discussed in Section IV.

II. System Model

A. Overview

MaLDIP consists of three hardware elements in a WLAN infrastructure: access points (AP), detecting points (DP) and a server. Each pair of AP and DP establishes a radio frequency (RF) link. Beacon messages are broadcast periodically by the AP. A WiFi compatible device is used as the DP that interacts with the AP and the server. Once the beacon message are received, the DP records the raw PHY layer CSIs across multiple subcarriers from the MIMO links and sends them to the server to store and process. Fig. 1 shows the system architecture of MaLDIP. Location estimation is performed in the centralized server through the following two phases:

B. Offline Phase

In the offline phase the CSI for each location are collected in a device free setting and a novel preprocessing is performed for location estimation.

1) CSI Fingerprint Generation: In this phase a location dependent CSI fingerprint is created with no active participation of device-based entity in the process. The area is considered as a grid of small square cells and there are C cells in that area of interest. Our goal is to use CSI fingerprint

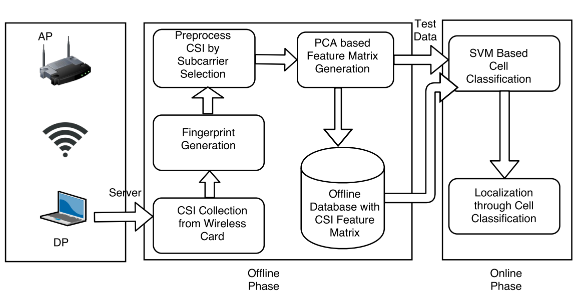

Fig. 1. System Architecture.

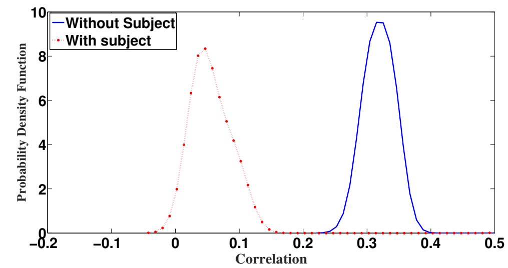

Fig. 2. Effect of Subject Appearance on CSI feature Shift.

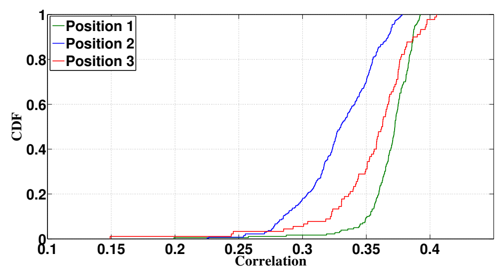

Fig. 3. Location-Specific CSI feature Variance.

to classify a testing entity with an unknown cell ID. From observation, it is seen that CSI over MIMO channel can reveal the change in the environment due to the appearance of an entity. Fig. 2 shows that there is a shift in the feature pattern of empirical probability distribution function of CSI correlation when a subject appears in an empty room. Moreover, CSI from MIMO channel can also identify entity at different location, as it reveals different feature pattern at different location as depicted in Fig. 3. This is because in MIMO-OFDM systems, for each transmit and receive antenna pair, CSIs over multiple subcarriers suffer from different scattering due to multi-path. This work exploits these two features of CSI in MIMO-OFDM system to generate unique fingerprint for each location.

The individual path characteristics in a wireless propagation channel are modeled as a temporal linear filter, known as Channel Impulse Response (CIR). Assuming time invariant channel, CIR is defined as,

$$h(\tau) = \sum_{i=1}^{M} a_i e^{-j\theta_i} \delta(\tau - \tau_i), \tag{1}$$

where a_i, θ_i, and τ_i are the amplitude, phase, and time delay of the i-th path, respectively. M is the total number of multipath and $\delta(\tau)$ is the Dirac delta function. This CIR is characterized as Channel Frequency Response (CFR) in the frequency domain. Recently, in the commercial off-the-shelf WiFi devices, sampled versions of CFRs are revealed to upper layers in the format of Channel State Information (CSI) [7]:

$$H(f_k) = \sum_{i=1}^{M} a_i e^{-j2\pi f_k \tau_i}, \tag{2}$$

where f_k is the frequency of k-th subcarrier and $H(f_k)$ is CSI at k-th subcarrier. Each CSI depicts the amplitude and phase of a subcarrier as,

$$H(f_k) = |H(f_k)|e^{j\theta}, \tag{3}$$

where $|H(f_k)|$ is the amplitude and θ is the phase of each subcarrier. The received signal strength (in dB) at each subcarrier is proportional to the amplitude of CSI and can be expressed as,

$$\hat{H}(f_k) = 20log|H(f_k)|. \tag{4}$$

Hence CSI amplitude is a measure of the power of WiFi link between the transmitter and receiver. From (2), $H(f_k)$ can be written as,

$$H(f_k) = \sum_{i=1}^{M} a_i[\cos(2\pi f_k \tau_i) - j\sin(2\pi f_k \tau_i)]. \tag{5}$$

Hence CSI amplitude for each subcarrier is $|H(f_k)| = \sqrt{(\sum_{i=1}^{M} a_i \cos(2\pi f_k \tau_i))^2 + (\sum_{i=1}^{M} a_i[\sin(2\pi f_k \tau_i))^2}$. It is clear that CSI amplitude is dependent on the time delay τ_i and this time delay will be different when a human is present at different locations. Hence CSI amplitude as well as received signal strength at each subcarrier will be different at different locations due to the presence of human being. Based on wireless communication principles [12], the received power fades for attenuation due to propagation of a distance (Line of Sight path) between the transmitter and the receiver and for attenuation due to diffraction when a target is located in the area of interest. In static environment with no human present, the Line of Sight (LOS) path and the path reflected from ground and ceiling dominate all other multipath components. When a human steps into the area of interest, the received signals experience diffraction fading in addition to propagation fading. The affected signal power is calculated according to the radar equation [13]:

$$P_{sbj} = \frac{P_{Tx} G_{Tx} G_{Rx} \lambda^2 \sigma}{(4\pi)^3 r_1^2 r_2^2}, \tag{6}$$

where P_{Tx} and P_{sbj} are the transmitted and received signal power, respectively. The Antenna gains at the receiver and transmitter are denoted by G_{Rx} and G_{Tx}, respectively. λ is the wavelength of the transmitted signal. r_1, r_2 denote the Transmitter-human, Receiver-human distances, respectively. σ is the radar cross section of the person. Assuming common settings with all other factors unchanged, the signal strength difference between static environment and human presence is approximated as [13],

$$\Delta P \approx P_{sbj}. \tag{7}$$

Since the received signal strength is a function of CSI amplitude, to generate the location dependent fingerprint using CSI while the subject carries no detectable devices, we propose to exploit the signal strength difference of static and dynamic CSIs for each subcarrier. This enables us to capture the impact of human presence to certain changes of the received signal features at a particular location.

In narrow band flat fading OFDM channel, the channel matrix, $\mathbf{H}$ estimated at the receiver, represents the PHY layer CSIs over multiple sub-carriers with the dimension $m \times n \times s$, where m and n are the number of transmit and receive antennas respectively and s denotes the number of subcarriers for each antenna pair. We group the subcarriers of CSI for each sample in their transmission/receiving antenna pairs as,

$$H = [H^1(f_1)\, H^1(f_2) \ldots H^1(f_s)\, H^2(f_1)\, H^2(f_2)\, \ldots H^j(f_s)], \tag{8}$$

Where $H^j(f_s)$ denotes the s-th subcarrier of j-th transmeter-receiver pair. Hence for each location, there are total L streams for each sample, where $L = m \times n \times s$. First, the system collects N samples of CSIs with no entity present in the area of interest, which is called static or ambient CSIs, $\mathbf{H_{amb}}$. Next, for each cell c, a set of CSI values, $\mathbf{H_{sbj}}$ are collected with the subject present in this cell, but carrying no device. By subtracting the signal strength of CSIs (using (4)) with a subject in a cell c from that of ambient CSIs, a CSI based fingerprint is generated for each cell as,

$$\hat{\mathbf{H}}_\mathbf{c} = \hat{\mathbf{H}}_\mathbf{amb} - \hat{\mathbf{H}}_{\mathbf{sbj},\mathbf{c}}. \tag{9}$$

$\hat{\mathbf{H}}_\mathbf{c}$ is a $N \times L$ matrix and quantifies how the presence of an entity affects the signal strength feature in a particular cell. Once CSI fingerprint for all the cells are generated, a preprocessing is performed to mitigate the multipath effect.

2) Preprocessing by Subcarrier Selection: Different subcarriers have different signal strength due to frequency-selective fading [7]. Thus, effect of multipath on different subcarriers is different. According to the diffraction theory [12], the amplitude of CSI measurement for all the subcarriers usually decrease when an entity appears in the RF link. But in indoor setting with rich multipath scenario, amplitude of CSIs of some of the subcarriers increase as depicted in Fig. 4 (subcarriers from 17 to 23), thus resulting in an inaccurate location estimation. In this work a subcarrier selection scheme is introduced to combat the frequency selective fading. The idea is to select the best subcarriers from all MIMO links which are less affected by multipath. To achieve this goal, the system propose to adopt a threshold to select the subcarriers for which the decrease in CSI amplitude is larger than the threshold. The idea is based on the fact that these selected subcarriers conform to the diffraction fading model [12]. The system defines the threshold δ as the averaged standard deviation of ambient CSIs, $\mathbf{H_{amb}}$ over all the s subcarriers:

$$\eta = \frac{1}{s}\sum_{k=1}^{s} \frac{f_k}{f_0} \times \eta_k, \tag{10}$$

where f_0 is the central frequency, and η_k is the standard deviation of the ambient CSI magnitude of the k-th subcarrier. Next, for each transmission/receiving antenna pair, only those subcarriers are kept for which the amplitude difference between $\mathbf{H_{amb}}$ and $\mathbf{H_{sbj}}$ for k-th subcarrier is greater than η. Therefore, for each cell, the system has a new set of subcarriers, S, such that,

$$S = \{k : (|\mathbf{H_{amb_k}}| - |\mathbf{H_{sbj_k}}|) > \eta; 1 < k < s\}. \tag{11}$$

$\hat{\mathbf{H}}_\mathbf{c}$ with filtered subcarriers are then used to generate location-specific feature matrix through Principal Component Analysis.

ISBN 978-90-827970-1-5 © EURASIP 2018

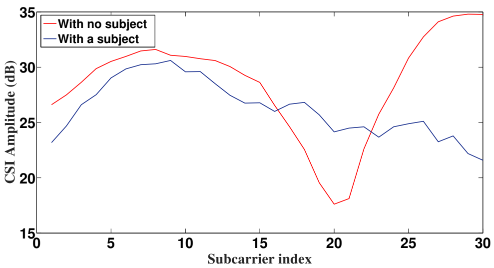

Fig. 4. Effect of Multipaths on Different Subcarriers.

3) Principal Component Analysis: From fingerprint generation phase, we get $\hat{\mathbf{H}}_\mathbf{c}$ for each cell, which is a $N \times d$ data matrix containing data samples as the rows and d is the number of effective CSI streams from all transmitter-receiver pairs after subcarrier selection. However, it is intuitive that the variation of the data may not be equally distributed over all dimensions. Therefore, it is sufficient to select a few key dimensions l with largest influence on data among the d dimensions to represent the data efficiently. We define the $d \times d$ positive semi-definite second-moment matrix $\mathbf{C}$ as,

$$\mathbf{C} = \hat{\mathbf{H}}_\mathbf{c}^T \hat{\mathbf{H}}_\mathbf{c}. \tag{12}$$

We perform Eigendecomposition on $\mathbf{C}$ to calculate the eigenvectors as $\mathbf{C}$=$\mathbf{V}\Lambda\mathbf{V}^\mathbf{T}$, where Λ is a diagonal matrix containing the eigenvalues of $\mathbf{C}$ and $\mathbf{V}$ is a matrix of eigenvectors corresponding to the eigenvalues. The top-l PCA subspace of $\mathbf{C}$ is the matrix $\mathbf{V_l}$= $[v_1 v_2 ... v_l]$, where v_i is the i-th column of $\mathbf{V}$. Given the top-l subspace, we construct the principal components using the equation $h_{c_i} = \hat{\mathbf{H}}_\mathbf{c} \times v_i$, where h_{c_i} is the i-th principal component. The time-varying correlations between CSI streams can be tracked by eigendecomposition and hence the extracted principal components of CSI streams are optimal to be used as features for cell classification.

C. Online Phase

In the online phase the CSI measurements for a test subject at a random cell are collected and pre-processed as described above. Next localization is performed through the machine learning based classification approach.

1) Location Estimation through Support Vector Machine: To identify the unknown label or cell ID of a subject, a machine learning based approach is implemented using SVM based classifier. A SVM is a supervised learning model commonly used for data analysis and pattern recognition. Multi-class SVM is an extended algorithm of SVM [14]. In this work a one-vs-all approach is adopted for solving multi-class learning problem where all the samples are divided into either the objective class or the non-objective class. In our experiments, CSIs of one cell belong to the objective class, and CSIs of all the other cells belong to the nonobjective class. In this way, the system builds the classification model for each cell. Then, SVM is applied to choose the class which classifies the test datum with greatest margin and thus identify the location/cell of the subject.

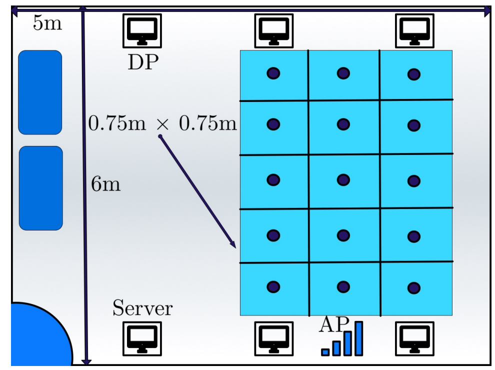

Fig. 5. The Layout of the Testbed in a Research Laboratory.

III. PERFORMANCE EVALUATION

A. Experimental Setup

The experiments are performed in a research laboratory with an area of $6m \times 5m$ during weekdays. The lab is equipped with typical office facilities like desks, shelfs, desktops, chairs etc. and hence is a subject of multipath. In our experiment, a TL-WR940N wireless router is mounted at a fixed location which works as the AP. The DP is a mobile device equipped with Intel 5300 Network Interface Card (NIC), which collects the CSI data using the Linux 802.11n tool [7]. We use 3×3 MIMO cahnnel with only one AP-DP link, where each transmission/receiving antenna pairs has 30 subcarriers. A host PC (Intel i7-4790CPU 3.60 GHz, 8GB RAM) serves as the centralized server for location estimation. We virtually partition the area into 15 uniform square grids/cells, each of which is $0.75m \times 0.75m$ in size, which is typical walking step size for adults. The layout of the testbed is shown in Fig. 5. In our experiments, we had 40 test measurements in each cell. We take measurements at the center of the cell and 0.10m, 0.20m, and 0.30m from the center of the cell, respectively. For each of these positions we take 10 measurements and calculated the mean value for performance evaluation.

B. Simulation Result

First we evaluate the accuracy of the proposed method and compare with Pilot, the CSI based device free localization method. We evaluate the performance in terms of mean cell estimation accuracy for two cases, with 3×3 MIMO channel and with SISO channel as shown in Fig. 6. Our evaluations show that proposed device free system with MIMO channel can localize a subject in a correct cell with 93.6% accuracy, which outperforms Pilot by 9%. With the SISO channel, accuracy of our system is 89.13%, but still achieves higher accuracy over Pilot.

In Table I, we investigated the performance of our system without the preprocessing technique and without the PCA features. Results show that localization using PCA features of the selected subcarrier improves the performance of localization.

in Fig. 7, we evaluated the performance for different number of samples. Results show that the localization accuracy of our system can be extended to 98.27% while we increase the

ISBN 978-90-827970-1-5 © EURASIP 2018

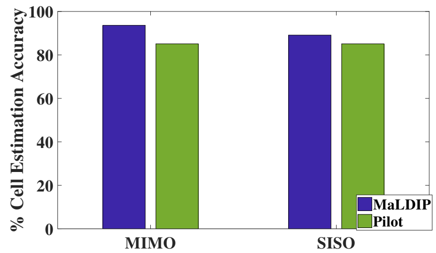

Fig. 6. Cell Estimation Accuracy for Different System

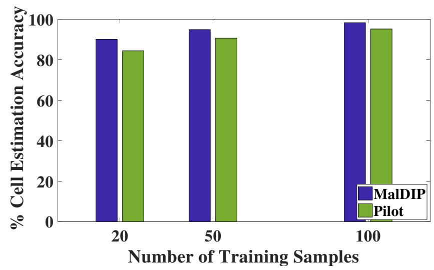

Fig. 7. Cell Estimation Accuracy for Different Sample Numbers.

sample size to 100. Moreover, Fig. 7 shows that proposed method with MIMO channel, always achieve higher accuracy than Pilot even with number of samples as low as 20. We attribute the better performance of proposed work due to the propagation model based CSI fingerprint selection using MIMO channel and PCA based feature matrix generation, which carry more implicit information compared to only correlation feature in Pilot. The novel subcarrier selection scheme in MaLDIP also contributes in increasing the accuracy.

Lastly, we investigate the impact of different number of principal components on mean cell estimation accuracy. With the reduction in the number of principal components the cell estimation accuracy gradually decreases and so as the processing time. Table II shows the average processing time and cell estimation accuracy with different degree of dimensionality reduction. The number of PCA components utilized in our system is empirically selected to achieve a good trade-off between computational complexity and classification performance.

TABLE I
COMPARISON OF CELL ESTIMATION ACCURACY IN DIFFERENT CASES

	w/o Preprocessing	w/o PCA	MaLDIP
MIMO	79.84 %	86.26 %	93.6 %
SISO	76.19 %	82.07 %	89.13 %

TABLE II
IMPACT OF DIFFERENT NUMBER OF PRINCIPAL COMPONENTS ON CELL ESTIMATION ACCURACY AND PROCESSING TIME

Ratio (l/d)	1/50	1/20	1/5	1/2	1
Cell Est. Accuracy	87.2%	93.6%	94.2%	95.1%	95.8%
Processing Time (ms)	59	106	298	492	601

IV. CONCLUSION

In this paper, a novel device free indoor localization is proposed for MIMO-OFDM system using machine learning based classification approach. To design a device free system, a novel location-specific fingerprint is generated based on wireless propagation model. In addition, the fingerprints are generated by exploiting the frequency diversity of CSI along with spatial diversity of MIMO-OFDM system. The proposed system is implemented with commercial low cost NICs to obtain CSIs from 802.11n transmission. A subcarrier selection scheme is proposed to combat frequency selective fading. Finally, a cell classification is performed for estimating the unknown location of a random subject using SVM. Experimental results demonstrate that our system can successfully localize a subject with cell estimation accuracy as high as 98.27% with only one AP-DP link. In addition, it can maintain an accuracy over 90% with a substantial reduction in number of training samples by exploiting the MIMO-OFDM technology. No active participation of devices tagged with the subject is necessary through out the process, which enables to develop a device free indoor localization system.

REFERENCES

[1] M. Seifeldin, A. Saeed, A. E. Kosba, A. El-Keyi, and M. Youssef, “Nuzzer: A large-scale device-free passive localization system for wireless environments,” *IEEE Transactions on Mobile Computing*, vol. 12, no. 7, pp. 1321–1334, July 2013.

[2] P. Bahl and V. N. Padmanabhan, “Radar: an in-building rf-based user location and tracking system,” in *Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Cat. No.00CH37064)*, 2000, vol. 2, pp. 775–784 vol.2.

[3] X. Wang, L. Gao, S. Mao, and S. Pandey, “Deepfi: Deep learning for indoor fingerprinting using channel state information,” in *2015 IEEE Wireless Communications and Networking Conference (WCNC)*, March 2015, pp. 1666–1671.

[4] J. Xiao, K. Wu, Y. Yi, and L. M. Ni, “Fifs: Fine-grained indoor fingerprinting system,” in *2012 21st International Conference on Computer Communications and Networks (ICCCN)*, July 2012, pp. 1–7.

[5] N. Patwari and J. Wilson, “Spatial models for human motion-induced signal strength variance on static links,” *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 791–802, Sept 2011.

[6] K. Kleisouris, Y. Chen, J. Yang, and R. P. Martin, “The impact of using multiple antennas on wireless localization,” in *2008 5th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*, June 2008, pp. 55–63.

[7] Daniel Halperin, Wenjun Hu, Anmol Sheth, and David Wetherall, “Predictable 802.11 packet delivery from wireless channel measurements,” in *Proceedings of the ACM SIGCOMM 2010 Conference*, New York, NY, USA, 2010, SIGCOMM ’10, pp. 159–170, ACM.

[8] J. Xiao, K. Wu, Y. Yi, L. Wang, and L. M. Ni, “Pilot: Passive device-free indoor localization using channel state information,” in *2013 IEEE 33rd International Conference on Distributed Computing Systems*, July 2013, pp. 236–245.

[9] J. Wang, H. Jiang, J. Xiong, K. Jamieson, X. Chen, D. Fang, and B. Xie, “Lifs: Low human effort, device-free localization with fine-grained subcarrier information,” in *22nd Annual International Conference on Mobile Computing and Networking*, Oct 2016, pp. 243–256.

[10] D. A. Hall, “Understanding benefits of mimo technology,” in *Microwaves and RF*, 2013.

[11] Zheng Yang, Zimu Zhou, and Yunhao Liu, “From rssi to csi: Indoor localization via channel response,” *ACM Comput. Surv.*, vol. 46, no. 2, pp. 25:1–25:32, Dec. 2013.

[12] A.F. Molisch, *Wireless Communications*, Wiley - IEEE Series. Wiley, 2011.

[13] D.M. Pozar, *Microwave Engineering*, Wiley, 2004.

[14] Chih-Chung Chang and Chih-Jen Lin, “Libsvm: A library for support vector machines,” *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, May 2011.