Analyzing Shopper's Behavior through WiFi Signals

Yunze Zeng, Parth H. Pathak, Prasant Mohapatra
University of California, Davis
{zeng, phpathak, pmohapatra}@ucdavis.edu

ABSTRACT

Substantial progress in WiFi-based indoor localization has proven that pervasiveness of WiFi can be exploited beyond its traditional use of internet access to enable a variety of sensing applications. Understanding shopper's behavior through physical analytics can provide crucial insights to the business owner in terms of effectiveness of promotions, arrangement of products and efficiency of services. However, analyzing shopper's behavior and browsing patterns is challenging. Since video surveillance can not used due to high cost and privacy concerns, it is necessary to design novel techniques that can provide accurate and efficient view of shopper's behavior. In this work, we propose WiFi-based sensing of shopper's behavior in a retail store. Specifically, we show that various states of a shopper such as standing near the entrance to view a promotion or walking quickly to proceed towards the intended item can be accurately classified by profiling Channel State Information (CSI) of WiFi. We recognize a few representative states of shopper's behavior at the entrance and inside the store, and show how CSI-based profile can be used to detect that a shopper is in one of the states with very high accuracy ($\approx 90\%$). We discuss the potential and limitations of CSI-based sensing of shopper's behavior and physical analytics in general.

Categories and Subject Descriptors

C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design—*Wireless Communication*

Keywords

Physical Analytics; CSI; WiFi; Retail Store

1. INTRODUCTION

Analyzing shopper's behavior in retail stores and shopping malls can provide crucial insights in a variety of aspects such as browsing habits, shopping interests etc. These insights can be useful to the business owners in improving the recommendations, services and effectiveness of promotions. Such tracking and analysis of a shopper in a retail store is referred as physical analytics. Compared to online shopping where user's navigation patterns and other characteristics can be easily recorded, analytics in physical domain has received comparatively low attention in research. The complexity of physical analytics arise from many challenges. Simply relying on video surveillance to understand shopper's behavior is not scalable, given that deployment of video cameras and mining the video stream to extract information can have very high cost along with some serious privacy implications. Recent works [6, 7] have proposed to use a user-driven approach where inertial sensors and camera available on shopper's wearable devices are used for physical analytics. However, this approach requires the shoppers to carry such devices and also effectively communicate the acquired information to the business owner. Relying on store's infrastructure for sensing shopper's behavior is a preferred way to enable non-intrusive monitoring. Shopper localization using RF signals [1, 8, 11] or Visible Light Communication (VLC) [3] is an attractive alternative but it requires the shopper to actively engage in the process (e.g. connecting to store WiFi or exposing the smartphone camera to LED luminaires). Since many shoppers can choose not to participate in the process, it is difficult to enable accurate physical analytics using such methods. This motivates the need of a passive, non-intrusive, device-free, low-cost and privacy-preserving form of sensing shopper's behavior for accurate physical analytics.

In this paper, we propose to utilize WiFi signals for understanding shopper's behavior in retail stores. WiFi is an attractive choice due to its pervasiveness in shopping malls, super-markets etc. Although the primary purpose of the WiFi deployment is to provide low-cost internet access, we show in this work that it can provide an efficient and accurate way of sensing shopper's behavior. There has been a considerable amount of work in indoor localization using WiFi, and our system proposes to further extend the sensing through WiFi by identifying and classifying shopper's activities. The proposed system does not require the shopper to carry any device as the movements of the shopper is detected purely by observing the variations in CSI of WiFi. It is also low-cost as the retail store owner does not require any infrastructure other than WiFi for observing shopper's behavior. Depending on where a WiFi link is deployed, it can sense different shopper's activities. For example, when deployed close to the entrance, our system can monitor if a shopper is entering or leaving the shop. Additionally, our

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
WPA'15, May 22, 2015, Florence, Italy.
Copyright © 2015 ACM 978-1-4503-3498-3/15/05 ...$15.00.
http://dx.doi.org/10.1145/2753497.2753508.

Shopper's state	Inferred activity	Useful in determining
Walking outside the entrance	Preparing to enter the store	Effectiveness of promotions outside the store
Walking at the entrance	Entering the store	Estimating store occupancy
Standing at the entrance	Observing close-to-entrance promotions	Effectiveness of in-store promotions like flyers etc.
Walking inside the store	Proceeding towards aisle of interest	Effectiveness of in-store arrangements

Table 1: Shopper's behavior around the entrance

Shopper's state	Inferred activity	Useful in determining
Walking fast	Walking towards pre-decided item	Opportunity for offering discount, additional items
Walking slowly	Browsing items of interest, exploring new items	Opportunity for advertisements, promotions
Standing	Closely observing item(s)	Factors affecting shopper's choice

Table 2: Shopper's behavior inside the store

presented way of using WiFi requires no modifications to off-the-shelf commercial WiFi hardware. We identify different states of a shopper in retail store and categorize them in two types of location in a retail store: (1) near the entrance (2) inside the store. It is then shown that each of the state exhibits a unique CSI signature which allows us to detect that a shopper is in a given state. Our experimental evaluation indicates that different states of a shopper near the door and inside the store can be classified with 90% and 96% accuracy respectively.

2. MOTIVATION AND BACKGROUND

2.1 Motivation

Accurate physical analytics entails analyzing shopper's behavior while meeting some important design goals. Improved understanding of the behavior will allow the brick and mortar stores to compete with the online stores by providing better service. For in-store analytics, it is necessary to answer various questions about shopper's behavior. The questions include how long a shopper remains in the store? How long the shopper stands at the entrance and looks at the advertisement board? What's the reaction of the shopper after looking at the advertisement board? Is the shopper entering or exiting the store? Is the shopper walking fast (with purpose, knows what she wants) or slowly (browsing)? The answers to these questions should be obtained while meeting the following requirements.

1. **Non-intrusive:** The shopping experience for the shopper should be as distraction-free as possible. This means that minimum input and interaction are expected from the shopper, and increasing reliance on passive monitoring is preferred. Such a requirement rules out techniques where user has to connect to a WiFi network or actively answer survey questions about her present state.

2. **Device-free:** An accurate estimation of shopper's behavior should not require her to wear/carry any devices such as smartphones or smartglasses. Relying on shopper's devices can introduce significant inaccuracies as shopper may not carry or use the devices as expected for physical analytics. Previous approaches like [6] assume that the shoppers wear such devices, however, in this work, our focus is to design a device-free technique for physical analytics.

3. **Low-cost:** It is always desirable that the physical analytics technique is low-cost and can reuse the existing infrastructure with minimum configurations and maintenance.

4. **Privacy-preserving:** The information made available through the use of analytics should not lead to privacy leakage for shoppers. For example, using video surveillance can reveal shopper's identity along with her behavior. Any such technique should be avoided to protect shopper's anonymity.

We show in this work that WiFi-based physical analytics can satisfy the aforementioned requirements. Given that most retail stores already have available WiFi infrastructure, our system does not incur any additional cost since it does not require any additional infrastructure. It purely relies on the changes in multipath observed through CSI variations to determine shopper's fine-grained behavior. Our system needs zero effort from the shopper and is implemented on the off-the-shelf commercial hardware. Since the WiFi signals can not be used to track the shopper's identity, it protects shopper's privacy without losing the functionality to capture shopper's behaviors in an anonymous way.

2.2 CSI Background

To measure wireless signal, Received Signal Strength Indicator (RSSI) is considered to be a coarse-grained way which characterizes the overall attenuation of radio signals from propagation. Using Orthogonal Frequency Division Multiplexing (OFDM) as the PHY layer, current 802.11a/g/n/ac standards can extract Channel Frequency Response (CFR) in the format of Channel State Information (CSI) from off-the-shelf commercial hardwares [2]. OFDM divides the channel into multiple subcarriers which suffer from independent flat fading. On one hand, this partitioning of the channel into subcarriers allow OFDM to combat the frequency selective fading caused by small-scale multipath. On the other hand, it provides us a way to get CFR in the format of CSI. CSI contains amplitude-frequency response and phase-frequency response in the granularity of each subcarrier. CSI for MIMO system is a $m \cdot n \cdot w$ complex matrix where m is the number of transmitter antennas, n is the number of receiver antennas and w is the number of subcarriers. For each spacial stream, at time t CSI can be accessed as a vector,

$$\mathbf{C} = [C_1, C_2, \ldots, C_s, \ldots] \quad (1)$$

where C_s is CSI value for subcarrier s. By further looking at each subcarrier s, C_s is a complex number which contains both amplitude and phase responses,

$$C_s = | c_s | e^{jp} \quad (2)$$

where c_s is the amplitude and p is the phase. For this work, we do not consider phase information and only use amplitude values for analysis.

3. UNDERSTANDING SHOPPER'S BEHAVIOR

During a typical visit to a retail store, a shopper enters the store, purchases the intended products and leaves. This behavior can be further classified into fine-grained states as shown in Fig. 1. Each state of the shopper can be used to infer an activity related shopper's behavior. This mapping between the states and the inferred activities is shown in Tables 1 and 2. The tables also describe how the inferred activities are useful in analyzing various aspects of business strategies and feasibility of improvements. For example, for shopper's behavior near the entrance, determining the amount of time spent by a shopper walking outside the entrance can be an indication of the effectiveness of the promotions outside the store. Similarly, detecting that a shopper is walking fast inside the store means that she is walking towards known pre-decided items.

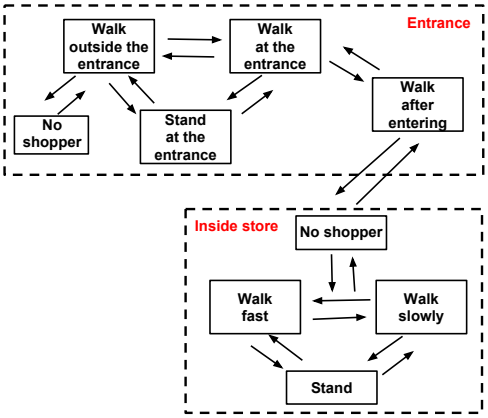


Figure 1: Shopper's states transition diagram

The characteristics of shopper's behavior includes not only the current state of the shopper but also how the shopper traverses through multiple states to accomplish what is intended in the visit to the retail store. This means that it will be useful to detect the transitions in shopper's state for a complete view of the behavior. In this work, we show how different states of a shopper (as shown in Fig. 1) can be determined using CSI. We will demonstrate how CSI variations can be used to fingerprint shopper's behavior with a very high accuracy. We note there can be many more fine-grained states of a shopper (such as fetching an item, checkout etc.) beyond the ones considered in this work. However, our objective in this work is to demonstrate the feasibility of WiFi based shopper behavior analysis for which we use fewer yet representative states.

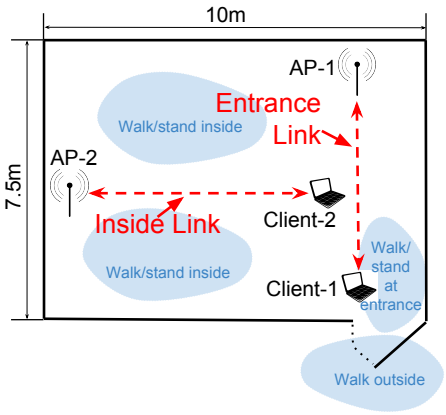


Figure 2: Experiment layout

4. ANALYZING SHOPPER'S BEHAVIOR USING CSI

4.1 Experiment Setup

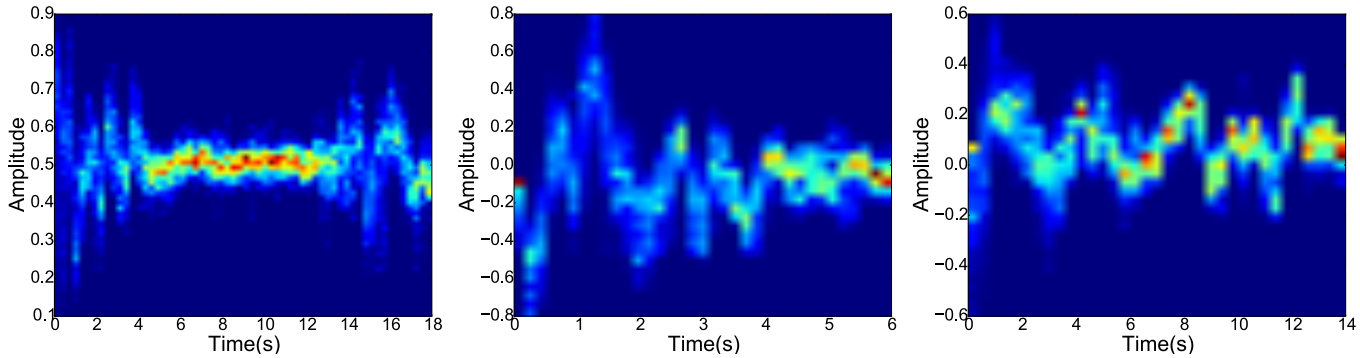
We implement our WiFi-based shopper behavior analyzing system using off-the-shelf commercial WiFi devices. Asus RT-AC66U WiFi router with 3 external omnidirectional antennas is used as the AP which operates in IEEE 802.11n mode in 2.4 GHz band. The WiFi client connecting to the AP is a Dell laptop with Intel 5300 802.11n WiFi NIC and 3 external antennas. We choose a large conference room in a university building to emulate a retail store scenario. Since in a typical large retail store, there can be a substantial distance between the entrance and the aisles, the "entrance" and the "inside store" scenarios use separate WiFi links as shown in Fig. 2. Note that in our setup, the experiments are done only on one link at a time to avoid interference.

For the collection of CSI, the client pings the AP at 100 packets per second and the CSI is collected at the client side from the modified driver [2]. Since any WiFi link observes different static multipath depending on the location of endpoints, CSI profile for different shopper activities will be slightly different at different locations. This means that the training procedure has to be repeated for each new location of the WiFi endpoints.

4.2 CSI Processing

4.2.1 Understanding CSI Observation

In a stationary environment without any user movements, the CSI data captured between the two endpoints profiles the static multipath of the environment. When a shopper walks within the range of the link, the variations observed in the CSI can be used to profile how the shopper moves. However, it is challenging to create a unique signature of shopper's activity purely using the raw CSI data. This is because the raw CSI data contains a variety of noise introduced by the surrounding and high-frequency movements. In order to distill the underlying impact of the shopper's state, it is first necessary to remove the high-frequency noise from the raw CSI data. Since majority of the human activities exhibit lower frequencies [5], we use a band-pass filter with cut-off frequency of 2 Hz and 0.3 Hz to remove the high-frequency noise as well as the static component.



(a) CSI profile for continuous walking, standing and walking again

(b) CSI profile for walking fast

(c) CSI profile for walking slowly

Figure 3: CSI profiles when a shopper performs different activities

In order to understand how filtered CSI data can distinguish between different shopper's state, we perform two separate experiments. In the first experiment, a user is asked to walk, stand and walk again while the CSI is being captured. Fig. 3a shows the filtered CSI data for the experiment. Large CSI variations (with alternating peak-valley) are observed when the shopper is walking, while relatively smaller variations are observed during the time when the shopper is standing. In the second experiment, the shopper is asked to walk fast and slowly for a fixed distance on a given path. The filtered CSI profile is shown in Figs. 3b and 3c. Apart from the time taken to complete the walking, a simple visual inspection (amount of signal variation, number of variations cycles) can distinguish between the two types of walk. This shows us that CSI signal profile can be used to determine shopper's state and infer her activities for physical analytics.

4.2.2 Feature Calculation

To differentiate between different activities of the shopper using CSI, we calculate various statistical features from the CSI data and create CSI profiles of the activities. A sliding window based approach is used with 3 seconds window size, moving over the time-series CSI data at an interval of 1 second. The features are calculated for each of the 3 seconds windows. These features are adapted from the feature set used in activity recognition through motion sensors (e.g. accelerometer, gyroscope) [5]. We next describe some of the notable features we have used in our analysis.

- Mean/Absolute-mean/Max/Min/Median/Quartiles: We calculate these basic statistics after a bandpass filter with a cutoff frequency range with 0.3 Hz to 2 Hz. These features can describe the shape and the distribution of CSI amplitude in each time window.
- DC-Mean/DC-Area: After a lowpass filter with a cutoff frequency of 1 Hz, we calculate the mean and the sum of all amplitude. The DC band features can provide the static component in current environment which can be used to infer shopper's body posture.
- Variance/Range/Mean-crossing-rate: They can provide us the variation and fluctuation level of the CSI amplitude changing in the time domain.
- Skewness: Measuring the asymmetry of the CSI amplitude distribution.
- Kurtosis: Measuring the peakedness of the CSI amplitude distribution.
- Normalized-Entropy (H): It measures the disorder of the CSI amplitude samples in frequency domain. Let N be the window size and V_i be the normalized FFT coefficients, then H is calculated as

$$H = -\sum_{i=1}^{N/2} V_i \cdot \log_2 (V_i) \tag{3}$$

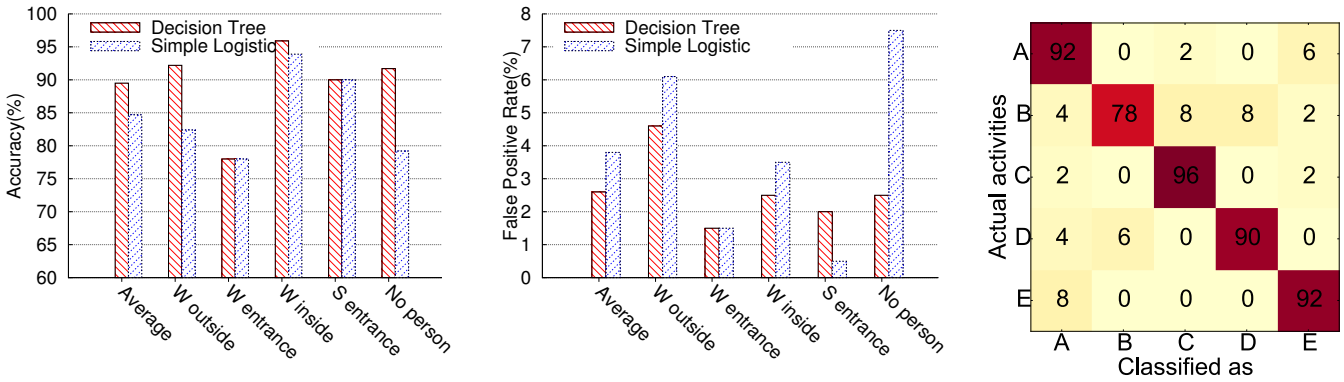
- Normalized-Energy (E): It measures the sum of energy without the DC component in the frequency domain. Let N be the window size and V_i be the normalized FFT coefficients, then E can calculated as

$$E = \sum_{i=1}^{N/2} V_i^2 \tag{4}$$

- FFT-Peak: Selecting the largest FFT coefficient without the DC component. It can reflect the dominant activity frequency in the current window.
- Dominant-Frequency-Ratio: We divide the largest FFT coefficient by the sum of all FFT coefficients. It can reflect the ratio of the dominant frequency to the sum of all frequencies.

4.2.3 Classifier Introduction

As mentioned before, our CSI-based sensing of shopper's behavior is highly dependent on the static multipath profile. This means that it is necessary to retrain the machine learning classifier if the location of the WiFi endpoints between which the CSI is measured changes. For a given location of the WiFi link, the classifier can be trained by a user with moving to different states as shown in Fig. 1. Note that we perform our training separately for "Entrance" and "Inside store" in the layout shown in Fig. 2. It is not required to retrain the classifier for different shoppers. Two machine learning algorithms - decision tree and simple logistic regression - are used to train and test the classifier.



(a) Accuracy of shopper's behavior classification(W is "walk" and S is "stand")

(b) False positive rate of shopper's behavior classification(W is "walk" and S is "stand")

(c) Confusion matrix of the behavior classification

Figure 4: Performance around the entrance (In the confusion matrix: A is "walk outside", B is "walk at entrance", C is "walk inside", D is "stand at entrance" and E is "no person")

4.3 Performance Evaluation

4.3.1 Activities Around the Entrance

For evaluating the performance of detecting shopper's states at the entrance, we first ask 3 users to be in the following states (refer to Figs. 1 and 2) - walk outside the entrance, walk at the entrance, walk after entering the store, stand near the entrance. Another state "no person" is also considered to identify absence of any shopper. Each shopper repeats these activities 20 times where each round of the activity lasts around 10 seconds.

The classification accuracy results for identifying shopper's state at the entrance are shown in Fig. 4. Average accuracy of nearly 90% and 85% can be achieved using the decision tree and simple logistic regression-based classifier respectively. We observe that decision tree-based classifier always performs better or equal compared to the simple logistic regression-based classifier. Fig. 4b shows that average false positive rate for classification is approximately 2.5% and 3.8% for decision tree and simple logistic classifiers respectively. This shows that shopper's state can be accurately identified using the CSI with a low false positive rate. It is observed that walking outside the entrance and walking after entering are typically better classified compared to walking near the entrance. Fig. 4c shows the confusion matrix for each state at the entrance. We can see that walking near the entrance is often mis-classified walking inside and standing at the entrance, due to their similarity in location and activity. Similarly, walking outside the entrance is often misclassified as no shopper being present. This is due to the fact that while walking outside, a shopper may walk far away from the WiFi link, which leads to the misclassification as no user present.

4.3.2 Activities Inside the Store

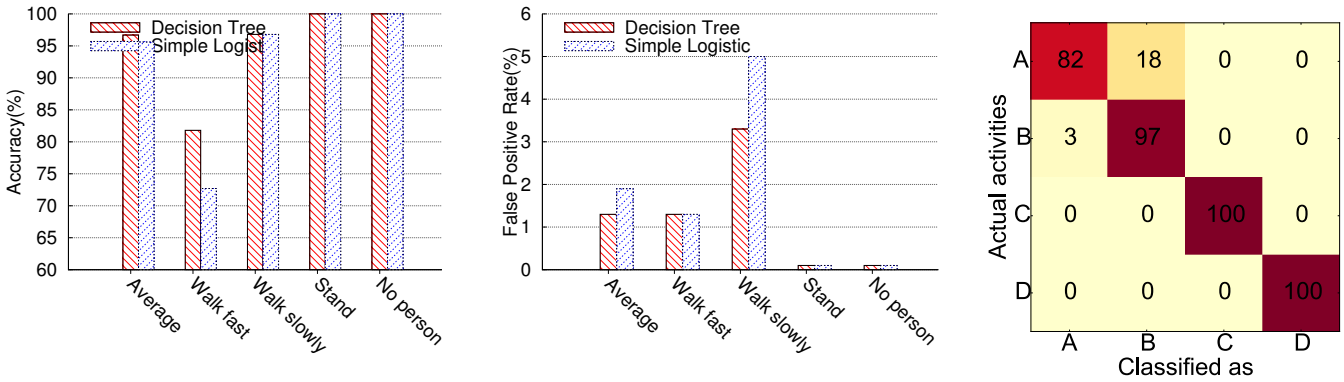
After entering the store, the shopper will have different activities which can be used to infer different shopping behavior. Here we consider different states during a typical shopping (shown in Fig. 1) - shopper is walking fast, walking slowly or standing. To conduct the experiments and collect CSI profile, we ask three persons to walk (fast and slowly) and stand inside the store. The walking/standing locations are shown in Fig. 2. Same as the entrance case, each person repeats the activities for 20 rounds and each round is recorded for around 10 seconds.

Fig. 5 shows the results of classification accuracy for shopper's states inside the store. A higher average accuracy of classification (decision tree - 96%, simple logistic - 95%) is observed compared to the entrance scenario. The false positive rate is also observed to be as low as 2% in Fig. 5b. The lowest classification accuracy is observed for the walking fast state. Based on the confusion matrix shown in Fig. 5c, the walking fast is often misclassified as walking slowly. Depending on different users involved in experiments, walking speeds for fast and slow walking can vary substantially, leading to higher mis-classification rate.

5. RELATED WORK

Physical Analytics: The growing interest in tracking user's movements, locations and activities has attracted both industry and research community towards physical analytics. WiFi-based indoor localization has been studied extensively [1, 8, 11] in recent years. Other ways of localization such as using visible light and LEDs [3] are proposed to overcome the accuracy limitations of RF-based localization methods. The localization techniques only provide shopper's location but no other information such as shopper's behavior, browsing pattern etc. are available. In a recent work, Rallapalli et al. [6] proposed the use of smart-glasses for analyzing shopper's fine-grained behavior such as gaze, fetch etc. The limitation of such approach is that the information is available to the shopper and not the business owner, requiring active engagement from the shopper to share the information. Instead, in this work, we have focused on detecting shopper's behavior only with passive monitoring using WiFi.

WiFi-based Sensing: With the availability of CSI in commercial off-the-shelf hardware using tools such as [2], WiFi-based sensing has attracted considerable attention in recent years. The CSI information has been used for gesture recognition [4,12] and in-home activity recognition [10]. Our work builds on this research to further improve the accuracy of fine-grained activity recognition (such as fast walking vs.



(a) Accuracy of shopper's behavior classification

(b) False positive rate of shopper's behavior classification

(c) Confusion matrix of the behavior classification

Figure 5: Performance inside the store (in the confusion matrix: A is "walk fast", B is "walk slowly", C is "stand" and D is "no person")

slow walking), and identifies important challenges of WiFi-based sensing in physical analytics applications.

6. POTENTIAL AND LIMITATIONS

Cost, Scalability and Privacy: Since WiFi-based behavior analysis can reuse the existing WiFi infrastructure, it is more cost-effective and scalable compared to video-based analysis which incurs substantial deployment and processing cost for a large retail store. WiFi-based sensing also preserves shopper's anonymity as shopper's identity can not be revealed through sensing.

Number of shoppers: One major limitation of our proposed approach is that it assumes there is only one shopper in the range of any WiFi link. Since, in practice, there can be many shoppers in a retail store, this assumption would require many WiFi links to be deployed in the retail store. To sense the activities of multiple shoppers within one WiFi link, sectors can be generated using beamforming and shoppers in different sectors can be monitored in parallel.

Fine-grained shopper activities: Another challenge in WiFi-based sensing is that it becomes increasingly difficult to detect fine-grained activities of the users. For example, detecting if a shopper is reading the item label or puts the item in a cart requires fine-grained CSI fingerprinting. In recent work such as [9, 10], such fine-grained activity recognition is shown to be feasible with CSI, however, further work is required in the context of physical analytics.

Hybrid Sensing: Although the CSI-based sensing of shopper's behavior protects her identity, in some cases, it can be beneficial to identify the shopper uniquely for profiling purposes. To address this, CSI-based sensing can be combined with activity recognition through shopper's wearable/mobile devices [6, 7]. Such hybrid approach can overcome some of the limitations of CSI-based analytics by improving classification with multiple shoppers and fine-grained activity recognition.

Type of retail store: The shopper's behavior is likely to different depending on the type of the retail store. For example, the same shopper may walk differently in a furniture store and a grocery store. Hence, it is necessary to train the WiFi-based sensing of user's behavior for a given retail store to account for business specific characteristics.

7. CONCLUSIONS

In this work, we present a novel physical analytics approch which leverages CSI from WiFi network to infer shopper's behavior. Our proposed system is a non-intrusive, device-free, low-cost and privacy-preserving way to perform physical analytics. It can achieve around 90% accuracy to classify different states of the shopper during a typical in-store visit. We also discuss various limitation and potential of our system.

8. REFERENCES

[1] K. Chintalapudi, A. Padmanabha Iyer, and V. N. Padmanabhan. Indoor localization without the pain. In *ACM MobiCom*, 2010.

[2] D. Halperin, W. Hu, A. Sheth, and D. Wetherall. Tool release: Gathering 802.11n traces with channel state information. *ACM SIGCOMM CCR*, 41(1):53, Jan. 2011.

[3] P. Hu, L. Li, C. Peng, G. Shen, and F. Zhao. Pharos: Enable physical analytics through visible light based indoor localization. In *ACM Hotnets*, 2013.

[4] P. Melgarejo, X. Zhang, P. Ramanathan, and D. Chu. Leveraging directional antenna capabilities for fine-grained gesture recognition. In *ACM Ubicomp*, 2014.

[5] E. Munguia Tapia. *Using machine learning for real-time activity recognition and estimation of energy expenditure*. PhD thesis, Massachusetts Institute of Technology, 2008.

[6] S. Rallapalli, A. Ganesan, K. Chintalapudi, V. N. Padmanabhan, and L. Qiu. Enabling physical analytics in retail stores using smart glasses. In *ACM MobiCom*, 2014.

[7] S. Sen, D. Chakraborty, V. Subbaraju, D. Banerjee, A. Misra, N. Banerjee, and S. Mittal. Accommodating user diversity for in-store shopping behavior recognition. In *ACM ISWC*, 2014.

[8] S. Sen, J. Lee, K.-H. Kim, and P. Congdon. Avoiding multipath to revive inbuilding wifi localization. In *ACM MobiSys*, 2013.

[9] G. Wang, Y. Zou, Z. Zhou, K. Wu, and L. M. Ni. We can hear you with wi-fi! In *ACM MobiCom*, 2014.

[10] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, and H. Liu. E-eyes: device-free location-oriented activity identification using fine-grained wifi signatures. In *ACM MobiCom*, 2014.

[11] Z. Yang, C. Wu, and Y. Liu. Locating in fingerprint space: wireless indoor localization with little human intervention. In *ACM MobiCom*, 2012.

[12] Y. Zeng, P. H. Pathak, C. Xu, and P. Mohapatra. Your ap knows how you move: fine-grained device motion recognition through wifi. In *ACM HotWireless*, 2014.