

TransitLabel: A Crowd-Sensing System for Automatic Labeling of Transit Stations Semantics

Moustafa Elhamshary^{†‡}, Moustafa Youssef[†], Akira Uchiyama[‡], Hirozumi Yamaguchi[‡], Teruo Higashino[‡]

[†] Wireless Research Center, Egypt-Japan University of Science and Technology (E-JUST), Alexandria, Egypt.

[‡] Graduate School of Information Science and Technology, Osaka University, Suita, Osaka, Japan.

{mostafaelhamshary, moustafa.youssef}@ejust.edu.eg, {uchiyama, h-yamagu, higashino}@ist.osaka-u.ac.jp

ABSTRACT

We present *TransitLabel*, a crowd-sensing system for automatic enrichment of transit stations indoor floorplans with different semantics like ticket vending machines, entrance gates, drink vending machines, platforms, cars' waiting lines, restrooms, lockers, waiting (sitting) areas, among others. Our key observations show that certain passengers' activities (e.g., purchasing tickets, crossing entrance gates, etc) present identifiable signatures on one or more cell-phone sensors. *TransitLabel* leverages this fact to automatically and unobtrusively recognize different passengers' activities, which in turn are mined to infer their uniquely associated stations semantics. Furthermore, the locations of the discovered semantics are automatically estimated from the inaccurate passengers' positions when these semantics are identified.

We evaluate *TransitLabel* through a field experiment in eight different train stations in Japan. Our results show that *TransitLabel* can detect the fine-grained stations semantics accurately with 7.7% false positive rate and 7.5% false negative rate on average. In addition, it can consistently detect the location of discovered semantics accurately, achieving an error within 2.5m on average for all semantics. Finally, we show that *TransitLabel* has a small energy footprint on cell-phones, could be generalized to other stations, and is robust to different phone placements; highlighting its promise as a ubiquitous indoor maps enriching service.

1 Introduction

With the fact that people spend most of their time at indoor spaces, indoor Location Based Services (LBSs) are being developed at a phenomenal rate with a variety of applications including mapping and navigation services, point-of-interest finders, geo-social networks, and advertisements. With the fact that people spend most of their time at indoor spaces, indoor LBSs are being developed at a phenomenal rate. A key requirement to indoor LBSs is the availability of indoor maps to display the user location on. These LBSs have still a huge potential for enhancement if rich semantic information is attached to indoor maps to support a wide class of indoor mapping applications (especially large public buildings that are visited daily by many people like railway stations, airports,

museums, etc). Realizing the economic value of this technology, a number of commercial navigation systems for indoor mapping have started to emerge. In late 2011, Google Maps started to expand its coverage by providing detailed floorplans for a few malls and airports in the U.S. and Japan as well as allowing buildings owners around the world to upload their indoor floorplans. Nevertheless, these maps are still limited in coverage to a small number of countries featuring only some major airports, shopping malls, etc. This limitation in coverage is due in part to the following reasons: (1) buildings owners may not allow sharing of their floorplans in public for privacy reason, (2) buildings internal structures often evolve over time, and/or (3) manual creation of these maps requires slow, labor-intensive tasks, and they are subject to intentional incorrect data entry by malicious users.

Railway stations, as an example of indoor places, are a key part of the day-to-day lives of people having millions of passengers every day (e.g., Shinjuku station in Japan has 3.64 million passengers/day on average¹). In highly populated countries, major stations have large indoor spaces (e.g., Shinjuku station in Japan has 36 platforms and over 200 exits¹). Therefore, a number of indoor navigation apps, e.g. the Tokyo station underground area navigation app², for stations have started to emerge. These applications, however, are built upon a manually created map of the building showing all important points of interest (e.g., fare collection gates, ticket vending machine, etc), which impedes their scalability to large scale deployments at different stations. For example, Google indoor maps covers less than 50 transit stations worldwide, which are a small fraction of the thousands of stations on Earth³. The lack of detailed digital floorplans for railway stations highlighting locations of various semantics limits passengers' experience, especially for foreigners or first-time visitors. Consequently, this sparks the need for the automatic construction of detailed indoor floorplans for transit stations.

To resolve this problem, the research community recently has embarked to address the problem of automatic construction of indoor floorplans by exploiting motion trajectories of mobile phone users [7, 21, 25]. These systems proved the feasibility of estimating the general layout of a building [7, 21, 25], identifying rooms shape and dimensions [7, 21], along with identifying other points of interest such as store entrances [7, 21]. Nevertheless, none of these approaches provide semantic-rich floorplans where various semantics are tagged on the floorplan that are necessary for many of today's map-based applications. For example, stations indoor navigation systems should rely on important semantics to better guide

¹https://en.wikipedia.org/wiki/Shinjuku_Station

²<http://en.rocketnews24.com/2016/02/17/tokyos-busiest-train-stations-have-a-new-free-english-compatible-navigation-app/>

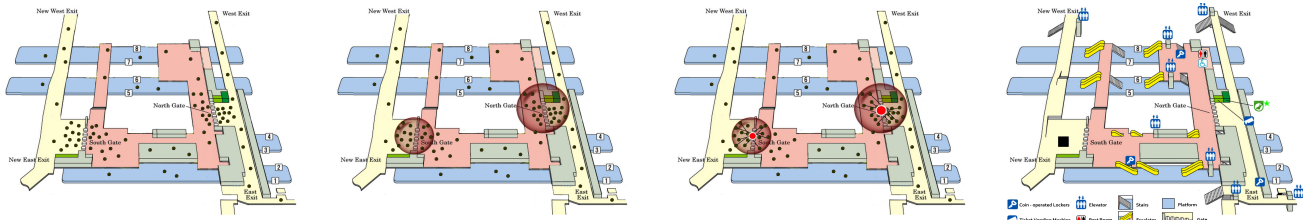
³<https://support.google.com/gmm/answer/1685827?hl=en>

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.

MobiSys'16, June 25-30, 2016, Singapore, Singapore

© 2016 ACM. ISBN 978-1-4503-4269-8/16/06...\$15.00

DOI: <http://dx.doi.org/10.1145/2906388.2906395>



(a) The locations of passing through fare collection gate activity as estimated by pedestrian dead-reckoning (PDR) are highlighted on the floorplan. (b) The output clusters from the DBSCAN based on crossing fare collection gates activity locations are highlighted on the floorplan. (c) The semantics locations (e.g., fare collection gates) are estimated as the center of mass of all samples within the output clusters. (d) The station indoor map with the discovered semantics locations, estimated by *TransitLabel*, are highlighted on the floorplan (*TransitLabel* output).

Figure 1: An example of *TransitLabel* in action to identify the location of ticket gates in a railway station.

passengers to their destinations; a station evacuation planning is ineffective if maps are not tagged with emergency exit stairs; a person with disability needs a map showing elevator-enabled routes; and an occasional passenger needs a map of important semantics that she must use to board the train (e.g., ticket gates, etc). Moreover, the discovered semantics can be leveraged to provide accurate calibration-free indoor localization, by providing opportunities for dead-reckoning error-resetting [3, 41]. Finally, fine-grained tracking of passengers’ activities and interaction with the different detected semantics opens the door for indoor analytics, which is of great business value.

In this paper, we present *TransitLabel* as a crowdsensing system that leverages the ubiquitous sensors available in commodity cell-phones to automatically enrich transit stations floorplans with different semantics. The core idea is that passengers perform many activities (e.g., crossing an entrance gate) that show identifiable signatures on the phone sensors. *TransitLabel* aims to recognize these high level activities and therefrom discover their uniquely associated semantics. Therefore, starting from an **unlabeled general** floorplan of a transit station, *TransitLabel* estimates the location of different semantics and tags their locations on the map accordingly to generate a detailed floorplan (Figure 1).

Translating this basic idea into a deployable system, however, involves addressing a number of challenges: First, identifying semantics signatures is based on the phone sensors during passengers’ activities; which are prone to human behavior artifacts. Second, current indoor localization technologies may require infrastructure support or prior calibration; and all have an average localization error in the range of few meters [42]. This can place the passenger in a location on the floorplan that is far from the actual one, affecting the accuracy of semantics’ location estimation. Finally, the system needs to be optimized for energy to avoid significant battery drainage.

To cope with these challenges, *TransitLabel* draws on a classifier-based approach based on the multi-modal sensors features to recognize passengers activities and thereby identify their associated semantics to address the first challenge. For the second challenge, *TransitLabel* relies on a DBSCAN clustering algorithm to cluster the correct crowdsensed samples of the same semantic to estimate its location and thus outlier locations are removed. To save energy, *TransitLabel* employs sensors with low energy footprint (i.e., inertial sensors) and the energy hungry sensor employed (i.e., sound) is turned on only when needed.

Implementation of *TransitLabel* over different Android phones shows that it can detect the fine-grained stations semantics accurately with 7.7% false positive rate and 7.5% false negative rate on average. In

addition, it can estimate locations of the detected semantics accurately, achieving an error of 2.5m on average using as few as 40 samples of each semantic. This comes with low energy consumption of 41 Joule on average for typical traces. We believe this could be a promising and a potential candidate for the real-world.

In summary, our contributions are three-fold:

- We present the *TransitLabel* system to automatically and unobtrusively crowdsense and identify transit stations semantics (e.g., ticket and drink vending machines, entrance gates, lockers, waiting (sitting) areas, restrooms, platforms and cars’ waiting lines, escalators, elevators and stairs) from the available sensors readings with minimal energy consumption.
- We provide a framework for extracting the features used to recognize high level user contexts (e.g., buying a ticket) from a sequence of temporal and spatial low level user states (e.g., walking, standing, etc) based on the phone sensors.
- We implement *TransitLabel* on Android phones, collect real data by 16 participants, and evaluate its accuracy, generalizability, robustness and energy-efficiency at eight different railway stations in Japan.

The rest of the paper is organized as follows: Section 2 presents the system overview. We give the details of identifying station’s semantics from phone sensors in sections 3 and 4. Section 5 provides the evaluation of *TransitLabel*. Section 6 discusses the system limitations and possibilities for enhancement. Finally, sections 7 and 8 discuss related work and conclude the paper respectively.

2 The *TransitLabel* System

Figure 2 shows the *TransitLabel* system architecture. *TransitLabel* is based on a crowdsensing approach, where cell phones carried by users submit their data to the *TransitLabel* service running in the cloud. The data is first preprocessed to reduce the noise. Then, semantics are classified to separate the elevation change semantics (elevators, escalators, and stairs) from other railway stations exclusive semantics (ticket vending machines, entrance gates, etc). *TransitLabel* has two core components: one for extracting elevation change semantics and the other for extracting other stations exclusive semantics. *TransitLabel* takes a classifier approach to detect different semantics based on the extracted features from the collected sensor traces. In the rest of this section, we give an overview of the system architecture leaving the details for the semantics detection to sections 3 and 4.

2.1 Traces Collection

The system collects time-stamped sensor measurements including the available inertial sensors (accelerometer, gyroscope and mag-

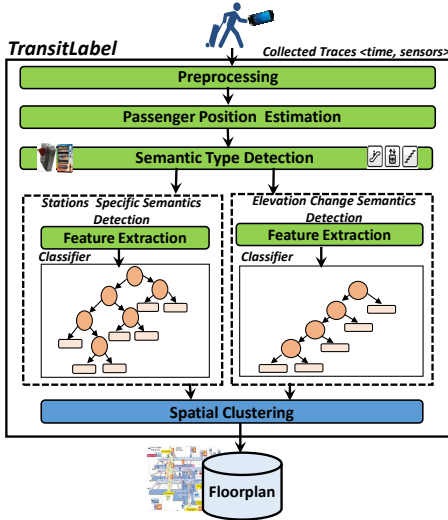


Figure 2: The *TransitLabel* system architecture.

netometer), barometer as well as the sound sensor (i.e., the microphone). Inertial sensors have a low-cost energy profile and they are already running all the time during the standard phone operation to detect phone orientation changes. Therefore, they consume zero extra energy. On the other hand, we use an adaptive sensor scheduling scheme called *triggered sensing* [31] to reduce the sound sensor energy consumption. The key idea is that sensors that are inexpensive in energy consumption (e.g., accelerometer) are used to trigger the operation of more expensive sensors (e.g., sound). Specifically, *TransitLabel* activates audio recordings only as soon as the passenger becomes stationary for a considerable time (4 seconds) and suspends it once she resumes walking. The intuition is that passengers traces inside railway stations are dominated by walking periods and they pause only to perform an activity (e.g., buying a ticket or a drink) which has a considerable stationarity time (more than 15 seconds). If the user does not resume walking after a certain time (60 seconds), the audio recordings will be suspended to save energy. The collected audio recordings during activities are used as a tie-breaker when other sensors (e.g., inertial sensor) fail to recognize certain activities.

Given the privacy implications of turning on the microphone, *TransitLabel* gives users full control over their own sensed data by means of a personalized privacy configuration. *TransitLabel* has different modes of operations (full sensor collection, privacy insensitive data only) that tailor the amount of data collected based on the user's preferences. In addition, according to a recent study [9], inertial sensors are enabled by most users and even the privacy-sensitive sensors (i.e., microphone) are enabled by about 78% of users. Finally, we **process audio data locally** on the user's device to further enhance the user privacy.

2.2 Preprocessing

This module is responsible for preprocessing the raw sensor measurements to reduce the effect of (a) phone orientation changes and (b) noise, e.g., small direction changes while moving. To handle the former, we transform the sensor readings from the mobile coordinate system to the world coordinate system leveraging the inertial sensors [32]. To address the latter, we apply a low-pass filter to raw sensors data using local weighted regression to smooth the data [10]. To filter out the noise in the employed frequency bands

(350Hz and 3kHz) in audio recordings, the standard sliding window averaging technique, with a window of 32 samples, is used.

2.3 Passengers' Position Estimation

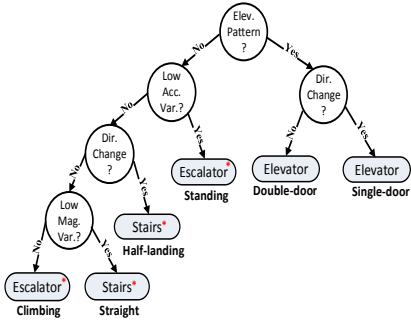
TransitLabel needs accurate passengers' locations during activities to estimate their uniquely associated semantics positions. To achieve this, *TransitLabel* employs the dead-reckoning technique to track the passenger's location starting from a reference point (e.g., the station entrance). We employ the step detection algorithm in [8] that takes into account the different users' profiles and gaits and apply them to the acceleration signal to detect the user steps. The user heading is estimated by incorporating the algorithm in [3,41] which leverages the correlation between compass and gyroscope to compensate gyroscope drift and compass interference errors to accurately estimate the user orientation. The estimated displacement and heading are fused to localize the user. However, the displacement error of dead reckoning is unbounded making it infeasible for indoor tracking. To alleviate this problem, *TransitLabel* incorporates the idea of SemanticSLAM [3, 41] by leveraging **amble and unique** physical points in the stations (i.e., semantics) to reset the accumulated error. Since dead-reckoning provides a rough location to the phone, it is also possible to roughly localize the semantics based on when the phone senses them. Now, since the floorplan is known, we can estimate the locations of all semantics in a crowd-sensing approach (as discussed later) by combining the rough estimates (i.e., the dead-reckoned positions) from multiple passengers' phones. These semantics, once detected based on their unique sensor signatures, can then be used to improve dead-reckoning of subsequent passengers, which in turn can refine the semantic locations. This recursive dependence between estimating the semantic location and the user location is similar to the Simultaneous Localization And Mapping (SLAM) framework.

2.4 Semantic Type Detection

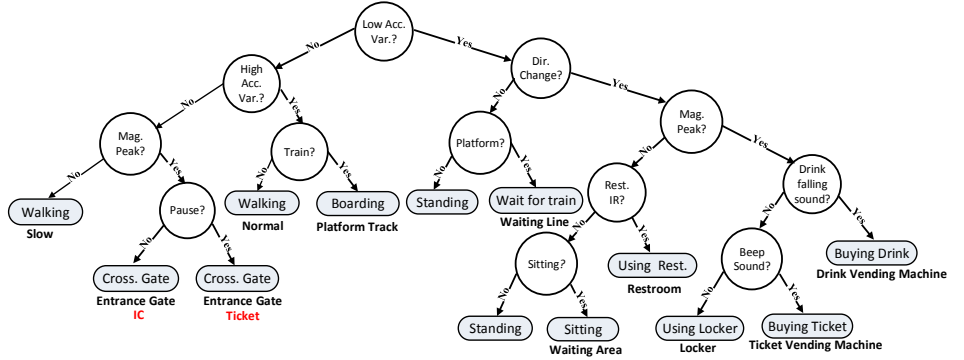
TransitLabel is designed to detect various stations semantics based on their unique usage patterns. This module separates between the two major types of semantics: elevation change semantics (elevators, stairs and escalators) and stations specific semantics (e.g., ticket vending machines, entrance gates, etc). The usage of elevation change semantics involves a noticeable change in the passenger's level (i.e., height) which is absent in other semantics (Figure 4). To separate them, we draw on the maximum difference among the relative barometer readings (i.e., pressure) in consecutive overlapping windows. The employed window size, 10 seconds, is small enough so that barometer readings are not affected by the environmental changes [33]. The intuition is that a change in pressure means a change in height which in turn means that the passenger is using one of the elevation change semantics. Moreover, the sign of the pressure difference indicates the direction of motion (up or down) which is useful for other purposes (e.g., detecting escalators direction). Evaluation of over 250 traces shows that the semantic type detection can achieve 0% false positive and negative rates. Later, the major two classes of semantics are further classified to their more fine-grained semantics.

2.5 Feature Extraction

In this section, we present the basic features used to identify the different semantics based on the data collected from the passengers' phone sensors. For instance, magnetic peak is a key feature to recognize many activities that involve direct interaction with electronic machines (e.g., vending machines). To extract it, we first apply a stream-based event detection algorithm to identify significant changes in the **magnitude** of the magnetic field. Once a sig-



(a) Elevation change semantics. Star marked semantics are further classified as floor or level change semantics leading to different floors or different levels within the same floor respectively.



(b) Stations specific semantics.

Figure 3: A decision tree classifier for detecting different types of semantics.

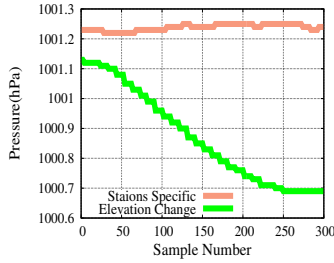


Figure 4: Comparing barometer readings of the usage of elevation change semantics (e.g., elevator) against station specific machines (e.g., ticket gate).

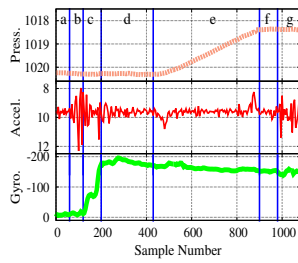


Figure 5: Elevator usage pattern: (a) waiting for it, (b) walking into it, (c) direction change, (d) stationary, (e) going up, (f) stopping, (g) walking out.

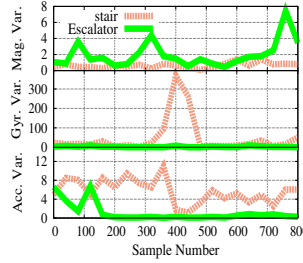


Figure 6: The sensor patterns that compare climbing up a half-landing stairs against standing on a moving up escalator.

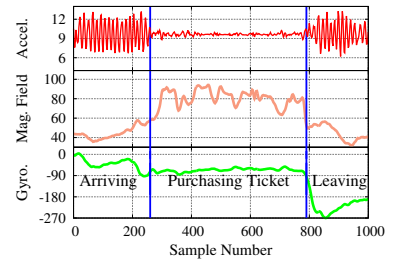


Figure 7: The acceleration, ambient magnetic field and gyroscope readings while using a coin operated machine.

nificant change (a $10\mu T$ increase in a window of 50 samples⁴) has been observed, we mark the corresponding time instant as the starting boundary of the peak area. We buffer subsequent measurements until a significant decrease in the magnitude of the magnetic field is observed. Once the starting and ending boundaries have been identified, we extract two features that characterize the peak area: the peak duration and strength. Moreover, some activities (e.g., buying tickets) are characterized by a sudden change in the user direction (i.e., surge in gyroscope readings) during or directly after the activity period. To detect this sudden change, we used the *approximate derivative* method: The derivative of sensor values within a time window are compared against a predetermined threshold (75° in a window of 60 samples) to detect the surge in sensor values. Finally, the variance of the acceleration is used to discriminate various passenger motion types (stationary, slow walking and normal walking) which contributes to the identification of many higher level passenger activities (e.g. buying a ticket). The acceleration variance values of 1.8 and 7 for a window of 200 samples are used as thresholds to separate stationary, slow walking and normal walking motion patterns respectively.

⁴We experimented with different values of thresholds and selected values that are robust to changes in the platforms/stations as confirmed by our experiments.

2.6 Station Semantics Extraction

As many semantics share some sensors patterns while having different patterns on other sensors, the hierarchical classification is an intuitive solution. Therefore, to identify semantics, *TransitLabel* relies on a tree-based classifier as it is easy to understand and to generate its rules. This classifier decomposes the task hierarchically into subtasks, proceeding from a coarse-grained classification (shared patterns) towards the distinction of fine-grained semantics (distinctive patterns) as detailed in sections 3 and 4.

2.7 Semantic Location Estimation

Whenever a semantic is detected by the semantic detection modules discussed later, *TransitLabel* needs to determine whether it is a new instance of a station semantic (i.e., not discovered before) or not as well as determine its location.

To do this, *TransitLabel* applies spatial clustering for each type of the extracted semantics. It uses the density-based clustering algorithm (DBSCAN [19]) which has a number of advantages as the number of clusters is not required before carrying out clustering; the detected clusters can be represented in an arbitrary shape; and outliers can be detected. The DBSCAN is applied to all samples of each discovered semantic to cluster all samples that are adjacent to each other in the spatial space. The parameter *Eps* specifies the radius of each cluster controlling the maximum distance among samples on the same cluster. After clusters are formed, the

locations of the newly discovered semantics are estimated as the weighted mean of the points inside their clusters. We weight the different locations based on their location accuracy reported by the localization approach. Specifically, in our position estimation approach, the longer the user trace from the last resetting point, the higher the error in the trace. Therefore, shorter traces have better accuracy. Based on the law of large numbers, the weighted average of independent noisy samples should converge to the actual location of the semantic. When a new semantic is identified, if there is an already discovered semantic within its neighborhood, we add it to the cluster and update its location. Otherwise, a new cluster is created to represent the new semantic. To reduce outliers, a semantic is not physically added to the floorplan until the cluster size reaches a certain threshold which is specified by the *Minpts* parameter (the minimum number of points that can form a cluster) of the DBSCAN algorithm. The DBSCAN parameters *Minpts* and *Eps* are selected empirically for each semantic type depending on its available number of samples, its physical dimension, and the average inter-distance among its physical instances in the real-world stations indoor maps. We do not state the DBSCAN parameters values for each semantic due to space constraints.

2.8 Practical Considerations

Sensor specifications are different from one phone manufacturer to another, which leads to different sensor readings for the same activity. To address this issue, *TransitLabel* applies a number of techniques including use of offset-independent features (e.g., variance), orientation independent features (e.g., magnitude of acceleration and magnetic field) and combining a number of features for detecting the same semantic. In addition, we experimented with various thresholds and select those leading to *high detection accuracy* with low false positive/negative rates while being robust to different users and stations. This is confirmed by experiment performed in the evaluation section.

TransitLabel does not also require real-time sensor data collection (i.e., it works offline); it can store the different sensor measurements and opportunistically upload them to the cloud for processing; allowing it to save both communication energy and cost. This is outside the scope of this paper.

3 Elevation Change Semantics Detection

To classify elevation change semantics⁵ into their fine-grained classes (elevators, escalators and staircases), we apply a decision tree classifier to the extracted features from passengers' phone sensors based on our observations of semantics usage scenarios and their physical structures (Figure 3a).

Elevators:

We begin by recognizing elevators as it is straightforward to distinguish their unique pattern. The typical usage scenario of an elevator consists of a normal working period, waiting for the elevator, walking into the elevator, changing direction to face the exit door, standing for a while, followed by a level change when it starts to move (Figure 5). This behavior is reflected to a unique pattern that consists of sequence of states: walking, stationary, stepping, direction change, level change, and Accelerate-Stationary-Decelerate emerging from the start and stop of the elevator. A set of features are extracted from accelerometer, gyroscope and barometer readings and fed to a Finite State Machine (FSM) to detect this multi-modal pattern. Starting from the initial state which represent the user waiting for an elevator (state (a) in Figure 5), the transitions

⁵We provide a thorough comparison with other related approaches in the related work section.

through all subsequent states (from (b) to (g)) have to occur to announce that a **single-door elevator** have been detected. However, for elevators having two doors, users do not need to turn around. Nevertheless, **two-doors elevators** can be recognized by the same FSM while skipping the direction change transition state.

Escalators (Standing):

The acceleration variance is used to decide whether a user is standing on an escalator or not. The intuition is that when a user keeps standing while carried by a moving staircase, the acceleration variance remains small compared to climbing stairs or escalators, which generates a high acceleration variance due to the vertical motion of the user. Conversely, if the acceleration variance is high, we cannot verify whether the user is climbing a stair or an escalator.

Half-landing Stairs (Climbing):

Half landing staircases have a turn in the middle forcing users to change their direction while straight stairs and escalators do not have any turns. Thus, if there is a surge in gyroscope readings (from the direction change) that took place in the middle of the elevation change period, it is affirmative that the user is climbing a half-landing stairs. Otherwise, if there is no direction change, we cannot verify whether a user is climbing a straight stair or climbing an escalator.

Escalators (Climbing):

The magnetic field variance, due to the escalator constant speed motors, can be used to reliably differentiate between climbing a straight stair and climbing an escalator (Figure 6). The value *100*, in a window of *200* samples, is used as the threshold of the variance of magnetic field.

Straight Stairs (Climbing):

After separating other elevation change semantics, the remaining samples are classified as straight stairs.

Level Change or Floor Change:

Many stations are multi-floor buildings with a typical floor height between 3.0 to 6.0 meters. The majority of elevation change semantics installed in railway stations move passengers from one floor to another (Floor change semantics). However, there exist some low height stairs and escalators which move passenger from level to another within the same floor (level change semantics). To classify the type of escalators and stairs (marked by red stars in Figure 3a), we rely on the magnitude of pressure difference during the elevation change period. Given that 1.0 meter height change corresponds to 0.12 hPa change in pressure, the pressure difference of *0.3* hPa is used as a threshold to separate level change escalators and stairs from floor change ones.

4 Station Specific Semantics Detection

Stations are rich with many exclusive semantics like ticket vending machines, entrance gates, drink vending machines, platforms, cars' waiting lines, lockers, restrooms, and waiting (sitting) areas. Based on our observations, these semantics force users to behave in predictable ways which are translated to unique sensor signatures that can be mined to identify them. For instance, a passenger crossing an entrance gate has to slow down her walking speed until she pauses to drop the ticket into the gate machine and then steps forward to grab it. Meanwhile her phone is experiencing a magnetic field distortion emanating from the gate machine electronics. *TransitLabel* draws on a decision tree classifier to recognize different passengers' activities (Figure 3b). The root of the decision tree separates activities into two main classes. The right branch of the tree comprises activities that require the passenger to be stationary during the service time (ticket and drink vending machines, lockers, restrooms, waiting (sitting) areas, etc). On contrast, the left branch comprises activities that do not force passengers to pause

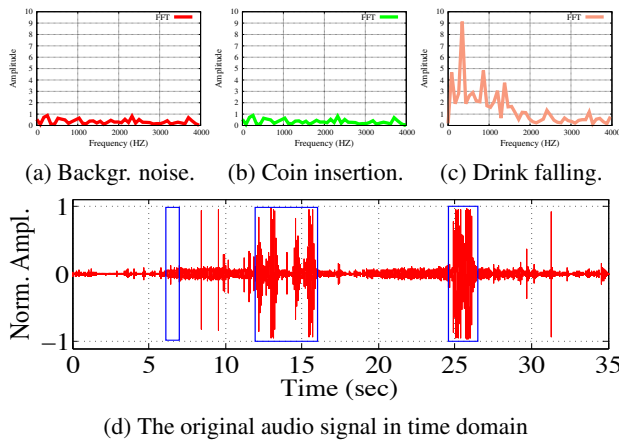


Figure 8: A time-domain sample of an audio signal depicting the usage of a drink vending machine is shown in (d) where three highlighted audio signal (bounded by blue boxes) corresponding to the background noise, the coin insertion sound, and the drink falling sound respectively. Figures (a), (b), (c) depict the frequency domain of these three audio signals respectively.

(e.g., ticket gates, etc). In the balance of this section, we give the details of the classifier features that can differentiate the different station specific semantics (coin operated machines, entrance gates, platforms track, sitting areas, restrooms).

4.1 Coin Operated Machines

Nowadays, ticket vending machines are found in every station and drink vending machines exist in many transits to dispense items (e.g. beverages, etc) to customers automatically. In addition, coin operated lockers are widely installed in railway stations to allow passengers to leave their baggages for several hours respectively to visit the surrounding area freely (especially in major stations in the downtown areas).

To identify these machines, we observed that their typical usage traces consist of normal walking to the machine, followed by standing in front of it, inserting currency, beginning the service (choosing a drink or the ticket type in case of drink and ticket vending machines respectively or opening the locker door in case of lockers), finishing the service (grabbing the ticket or the drink in case of ticket and drink vending machines respectively or putting luggage into the drawer and locking it), and finally walking away (Figure 7). This usage scenario is translated to the following unique patterns on the sensors. First, the user is stationary during the machine usage. Second, there is a fluctuation in the magnetic field readings as soon as the user interacts with the machine. This fluctuation is due to the distortion from metals and electronic chips installed in these machines forming a peak in the magnetic field readings (detected by the peak detector). Finally, as these machines are usually mounted to walls, the passenger is forced to change her direction to walk away as soon as she finishes the service. This instantaneous change in the user direction is reflected to a surge in the gyroscope readings when the user starts to resume walking (detected by the surge detector). This unique patterns are leveraged by *TransitLabel* to separate this type of machines from other semantics (Figure 3b). Now, we will give the detail of how to discriminate the three classes of coin operated machines.

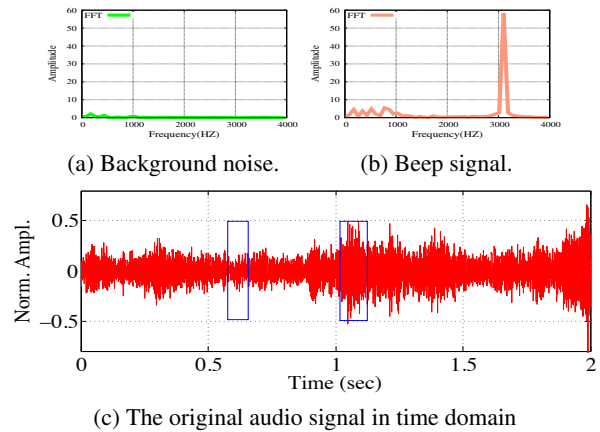


Figure 9: A sample of audio signal in the time domain depicting the usage of a ticket vending machine is shown in (c) where two different audio signal are highlighted (bounded by blue boxes) corresponding to the background noise and the beep audio signal respectively. Figures (a) and (b) depict the frequency domain of these two cropped signals respectively.

4.1.1 Drink Vending Machines

Based on our observation, they have a unique loud sound emitted when they are dispensing drinks to the customer. This sound is emanated when the drink is pulled down from the machine storage into its outlet.

To verify that we can recognize the unique drink falling sound in the ambience, in our preliminary experiment, we recorded an audio clip during the usage of a drink vending machine (Figure 8). The time-domain audio signal contains coin insertion sound existing in all coin operated machines, the drink falling sound, and the background noise respectively (highlighted by the three consecutive blue boxes in Figure 8d). The Fast Fourier Transform (FFT) of the audio signal shows a clear peak at the 350Hz frequency band in the drink falling audio clip (Figure 8c) while no peaks are evident at the 350Hz frequency band neither in the coin insertion nor the background noise clips due to the absence of this distinct acoustic signal (Figs. 8a, 8b). We observed that this frequency is consistent across all drink vending machines we experienced in the eight different stations in our dataset. *During the system operation*, we use an empirical threshold of three standard deviations (i.e., 99.7% confidence level of noise) to detect the drink falling acoustic signal in the ambient sound recorded during the usage of coin operated machines. If the received audio signal strengths in the 350Hz frequency band exceeds the threshold, indicating that the sound level is significant at this frequency band (as signal strength is jumped significantly at this frequency band), the system confirms the detection of a drink vending machine.

4.1.2 Ticket Vending Machines

Similarly, they emit a unique beep sound many times during the user interaction (e.g., pressing a button, indicating the end of transaction, etc). We envision that this beep signal can be leveraged as a reliable discriminator as it is absent in lockers where users insert coin, put luggage and lock the drawer without any distinctive sound. We incorporate the same acoustic detection algorithm used to identify drink vending machines to separate ticket vending machines from lockers. Figure 9c shows a raw audio recording collected during buying a ticket from a vending machine. We crop two

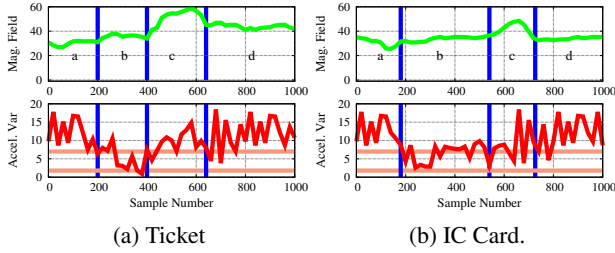


Figure 10: The sensor pattern of crossing a gate by a ticket and an IC card. Both consist of (a) normal walking, (b) deceleration near the gate, (c) acceleration accompanied by a peak on ambient magnetic field, and (d) normal walking.

sections from the original audio signal comprising the background noise and the beep audio signal respectively and convert these signals into the frequency domain by using FFT (Figs. 9a, 9b). We observed a clear peak around the frequency of 3kHz in the beep audio signal whereas no peaks are observed at the frequency of 3kHz in the background noise (Figs. 9a, 9b). When the ticket vending machine starts beeping, the signal strength in the 3kHz frequency band jumps significantly and therefore the ticket vending machine can be detected.

4.1.3 Lockers

To recognize lockers, we first attempted to identify the coin insertion sound in the ambience to avoid classifying all unrecognized activities as using lockers (i.e., catching all). However, we observed that while some coin sound signatures were visible, in many cases it was difficult to separate them from other frequency components (e.g., background noise). In addition, we find that the number of traces, other than coin operated machines ones, having a stationary period accompanied with a significant magnetic distortion on the user’s magnetometer followed by a direction change (i.e., surge in gyroscope readings) are small. So, once vending machines samples are separated, the remaining coin operated machines samples are classified as lockers.

4.2 Entrance Gates

Railway passengers have to pass through an automatic fare collection gate in their routes to the station’s platform. To cross a gate, there are two ways:

With a Ticket:

Passengers mostly pass by a ticket vending machine to get a ticket. Thereafter, as a passenger approaches the gate, a noticeable slows down in her walking speed is observed till she pauses in front of the gate to drop the ticket into the machine, then she steps forward to grab it from the machine, and finally she resumes normal walking (Figure 10a). This scenario translates to a unique motion pattern consisting of the following sequence: normal walking, deceleration, accelerating and normal walking. This unique motion pattern is detected by using the variance of acceleration (Figure 10a) where the two horizontal lines correspond to the thresholds used to separate different motion patterns. Simultaneously, crossing the gate heavily distorts the magnetic field by the gate ferromagnetic metals forming a distinct peak on the magnetometer reading (detected by a simple peak detector).

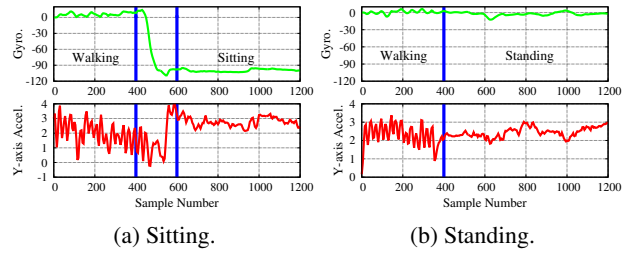


Figure 11: The effect of activity transition from walking to sitting against the effect of transition from walking to standing on the Y acceleration and gyroscope readings.

With an IC Card:

Nowadays, IC cards are commonly used for paying transit fees in many areas (e.g., Pasma⁶ in Japan, Ventra⁷ in Chicago). This gate entrance method has two differences from the ticket based one. First, passengers using IC cards do not have to pause as the card reader can recognize the card while it is in close proximity in users’ hand or wallet (acceleration variance still above the stationarity threshold (Figure 10b)). Second, mostly it is not preceded by the usage of a ticket vending machine activity (i.e., neither sequential nor dependent activities).

4.3 Waiting (Sitting) areas

Waiting areas are available in many stations platforms, especially those where trains inter-arrival time is long. Many people especially elderly people, pregnant women, people with disabilities and even normal passengers (e.g., in the winter) prefer to wait for trains in this area. We postulate that if there are many activities transitions from walking to sitting taking place within the premises of a certain area, then this area will be a waiting area with a high probability. To switch from walking to sitting, the passenger has to rotate first to be aligned with the seat and then sit down. The instantaneous surge in gyroscope reading (from the direction change) followed by a difference in relative magnitudes of Y acceleration values (from forward and backward motions while sitting) are used to characterize the transition from walking to sitting (Figure 11a). Conversely, the transition from walking to standing does not involve a noticeable difference on the readings of Y acceleration (Figure 11b). We leverage the change in direction as an affirmative feature to decrease the false positive rate in the sitting recognition given that the change in Y acceleration values may happen in other conditions (e.g., normal device bouncing while the user is walking).

4.4 Restrooms

Public restrooms are available in almost all stations. Normally, people must have stationary periods while they are at restrooms (e.g., hand washing). Moreover, due to the sensitive nature of public restrooms, their entry doors are faced by walls so users have to change their directions after crossing entry doors. However, the user stationarity and direction change patterns are not sufficient to efficiently separate the user being in a restroom from other contexts. To accurately identify restrooms, we incorporate the algorithm in [20] which detects restrooms by actively probing the acoustics of environment with the built-in speaker and microphone on the mobile phone. A probing sound is emitted by the phone and the impulse response (IR) is analyzed to detect the type of space

⁶<http://www.pasmo.co.jp/en/>

⁷<https://www.ventrachicago.com/>

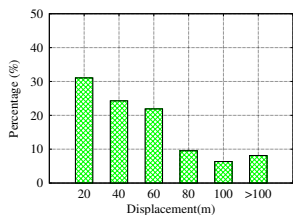


Figure 12: Passengers displacement distribution over the platform.

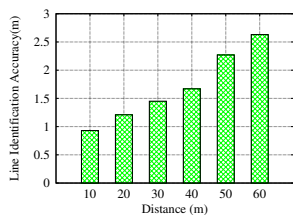


Figure 13: Lines location accuracy versus their distance from the platform access

(restroom or not). The acoustic characteristics of an environment depend on its dimension and its ability to absorb sound. Since, public restrooms have similar affordances (e.g., water resistance floors and wall, toilets and sinks), they have a unique absorption coefficient of sound and thus they can be detected easily. Since the sweep volume level does not affect the accuracy of the model significantly [20], *TransitLabel* leverages sweeps with lower volumes to avoid being invasive. Additionally, the model prediction performance is robust against the restroom occupancy and sounds generated by the occupants (e.g., flushing or hand-washing) [20].

4.5 Platform and Waiting Lines

To further enrich the semantics of *TransitLabel*, we also identify the platform area and the location of waiting lines for train cars. The platform is the area where passengers board the train. Therefore, their transportation mode changes from standing (waiting for train) to walking (into the train car) to be in a motorized transport (when the train moves) during a short time period. Transportation mode detection has been thoroughly studied in literature, e.g. [1, 23, 28, 36, 40]. We follow the approach proposed in [23] that provides high accuracy of transportation mode detection (walking, stationary or in a motorized transport) based on the energy-efficient accelerometer sensor. The short temporal consequence of standing, walking shortly and being in a train activities is leveraged to detect the platform track and its position is estimated from passengers' positions during train boarding.

Once the platform area is identified, the waiting line positions can be estimated from the passengers' positions reported while they are waiting for the trains (i.e., switching from walking to standing activity on the platform (Figure 11b)). We videotaped a sample of 3000 passengers' traces starting for their access to the platform (escalator, elevators and stairs) until they join a waiting line. Figure 12 shows that about 76% of passengers walk a short distance (less than 60m) on the platform where they tend to join the nearest uncrowded waiting line. This is partially due to the fact that platforms are typically designed to have multi-accesses to disperse the passengers load. In addition, as passengers use elevation change semantics to access the platform, their dead-reckoned derailed position will be curbed and calibrated to the access locations of these semantics on the platform when this semantic surface on the user's trace. This verifies that *TransitLabel* can localize the user accurately on the platform. Thus, waiting lines positions can be estimated by the clustering approach from passengers' locations when they are waiting in a line for the train. Once lines are detected, every neighbored collection of n lines (where n is the number of doors per car, which is a known constant) are representing a queuing area for a train car.

5 Evaluation

TransitLabel is evaluated through a deployment at eight different stations of different sizes in two different cities (Osaka and Kobe) in Japan. Table 1 shows the detailed description of the collected dataset. The stations are managed by different companies; having different buildings designs and sizes; and semantics placement. This emphasizes the scalable nature of *TransitLabel*. The average length and width of platforms are 160m and 12m on average respectively. Train cars are of 20m length with 3 doors and the average inter-door distance is 5.2m.

5.1 Data Collection Methodology

A group of 16 volunteers, of different ages (11 in their 20's and 5 in their 30's) and gender (12 males and 4 females), collected the necessary data for evaluation. The collected data consists of two datasets-*scenario based and free*- differing in how the data is collected. In the scenario-based dataset, 10 participants were assigned specific trajectories starting from the station entrance to different platforms. The trajectories were selected carefully to cover all possible routes that were exhibited by daily passengers while covering all available semantics at the same time. While heading to a specific platform, some participants purchased tickets to cross entrance gates; others crossed gates directly using different types of IC cards; some participants used drink vending machines while others used lockers; and all participants passed through elevation change semantics to access the platform. On the other hand, the free dataset is collected by six individuals from their everyday train commutation in different stations.

We have deployed two Android applications: The first application is a data collection tool that runs in the background to sample all inertial sensors, the barometer at 50Hz as well as recording audio. The second application is designed for ground truth collection and runs in the foreground to allow participants to manually tag their activities. The data collection was conducted in different times, different days and using different Android phones including Samsung Galaxy S5 and LG Nexus 5. Participants carry smartphones in different placements (in hand or in the trouser pocket). When the participants carry the logger phones in the trouser pocket, they carry another phone in their hands to annotate their activities. This captures the time-variant nature of semantics signatures and stations congestion level; generalization of the system over different stations; as well as the heterogeneity of users and devices.

5.2 Performance Results

In this section, we evaluate the semantics identification accuracy, semantics location estimation accuracy, power consumption, impact of different phone placement and finally quantify the generalization performance of *TransitLabel*.

5.2.1 Semantics Detection Accuracy

We evaluate the semantics detection accuracy based on the scenario-based datasets. The detection accuracy is measured by false positive⁸ and false negative⁹ rates.

Table 2 shows the confusion matrix for detecting various elevation change semantics. It shows that some elevation change semantics are easy to detect due to their unique patterns. This leads to zero false positive and false negative rates for the elevators (the coarse-grained category), half-landing stairs, and escalator (when users are standing) cases. However, climbing straight stairs sometimes are misclassified as climbing escalators when the stairs are

⁸Samples of other semantics classified as the current semantic.

⁹Samples of the current semantic that are not detected.

Table 1: The description of collected data.

Station	Osaka	Umeda	Senrihuo	Juso	Higashi-Umeda	Nishinomaya	Kobe-Sanyomia	Shin-Kobe
No. of Traces	108	78	71	69	61	59	53	48
No. of Users	10	9	7	6	6	6	5	4
Type and City	JR- Osaka	Subway- Osaka	Subway- Osaka	Hankyu- Osaka	Hankyu- Osaka	Hankyu- Osaka	Subway- Kobe	Subway- Kobe

very close to escalators so they have a similar magnetic distortion signature. Moreover, separating the two types of elevators generates some misclassifications which is due to the different elevator usage patterns (e.g., some users of single-door elevators do not turn around completely in one step). Nevertheless, as standing in or climbing up escalators activities are translated to the same semantic (escalator), *TransitLabel* can still achieve a high detection accuracy for elevation change semantics with 3.3% false positive and 4.1% false negative rates on average.

Table 3 shows the confusion matrix representing the detection accuracy of station specific semantics that force users to be stationary while using them (i.e., semantics in the right branch of the decision tree in Figure 3b). It shows that coin operated machines (the coarse-grained category) can be detected with near 100% accuracy using their unique inertial sensors pattern. To classify coin operated machines to their fine-grained categories (drink vending machines, ticket vending machines and lockers), the acoustic based detection scheme can achieve a good accuracy with 10.4% and 6.9% for false positive and false negative rates respectively on average. This is due to the responsive sound of vending machines that is universally used in many stations in Japan¹⁰, making it an identifiable signature for classification. On the other hand, standing and waiting for train activities are sometimes interchangeably misclassified. For example, when a user is standing on the platform (to answer a phone call), this may be interpreted as waiting for train and oppositely the user may be waiting for train but it is misclassified as standing when the system failed to recognize that the user is on the platform. *TransitLabel* lessens this effect by using the transportation mode detection algorithm to detect when the train has moved, which comes after the waiting for train inactivity, as opposed to any other type of inactivity.

The confusion matrix of classifying semantics from non stationary traces (i.e., semantics in the left branch of the decision tree in Figure 3b) is shown in Table 4. The table shows that *TransitLabel* can reliably detect crossing entrance gates using a ticket. The detection of crossing of entrance gates by IC cards is a bit challenging as some passengers either do not slow down their walking sufficiently or walk very slowly making their signature similar to ticket-based methods. Even worse, sometimes the card reader does not recognize the IC card and the passenger has to rollback and cross the gate again. However, since all gates have IC card readers and ticket slots integrated into the same machine, the two entrance methods (IC card and ticket) are aggregated into one semantic (entrance gate) that can be identified with 5.9% and 6% false positive and false negative rates, respectively. Moreover, the detection of train boarding is sometimes misclassified as crossing a gate by an IC card. The main reason is that train motors and electrical inverters emit large magnetic noise during the acceleration periods of the train which sometimes coincides with the pattern arising when the user crosses a gate using an IC card. To reduce this, *TransitLabel* leverages the sound emitted when the IC card touches the reader. Nevertheless, there is a trade-off between energy consumption and

¹⁰Vending machines around Japan are similar in their user interface and hardware to facilitate the Human-Machine interaction as well as to be able to recognize the same IC card types used to pay transit fees and drinks cost across the country.

the semantic detection accuracy in this case.

Figure 13 reports the accuracy of detecting the waiting line locations, as computed by *TransitLabel*. Aligned with our intuition, the accuracy of waiting line locations detection relies on their placement with respect to the platform accesses. Due to the limited area of platform (average dimension is 160m×12m), the short movement of a user on the platform (Figure 12), and the average inter-distance between waiting-line (5.2m); *TransitLabel* can estimate the line positions accurately, especially those near to the platform accesses where most user trails are short and thus the location accuracy is high [3,41].

Finally, *TransitLabel* can consistently detect the fine-grained classes of semantics accurately with 7.7% false positive rate and 7.5% false negative rates on average.

5.2.2 Discovered Semantics Location Accuracy

In this subsection, we study how much data is enough for *TransitLabel* to estimate semantics locations accurately as in crowdsensing-based systems the accumulation of more samples will enhance the system performance. Figure 14 quantifies the effect of the number of crowd-sensed samples on the accuracy of semantics (apart from waiting lines). The figure shows that even if some semantics have some outliers, the system can achieve a good accuracy in estimating their locations. This stems from the fact that independent correct samples of the same semantic are in adjacent locations and tend to cluster while erroneous samples are widely scattered in the spatial space and do not form a cluster. In addition, *TransitLabel* works offline so as a user encounters a semantic, *TransitLabel* learns her errors, and therefore can track back and partly correct her past trail thus the effect of the cold start problem is mitigated. Finally, as stations are rich with semantics, the localization error grows and sharply drops at semantics curbing the localization error and in its turn enhances the semantic location accuracy. Even though the instantaneous PDR error still has an effect on the semantics locations estimation, especially when the number of samples of semantics are small, it is evident from the figure that this error will drop quickly as the number of crowd-sensed samples increases. *TransitLabel* can consistently achieve the accuracy of 2.5m using as few as 40 samples for each discovered semantic type. Thus, *TransitLabel* converges reasonably quickly. However, we note that it needs to be periodically run to handle dynamic environment changes.

5.2.3 Energy Consumption

Figure 15 shows the energy consumption of *TransitLabel* averaged over typical traces from entering the station to boarding trains. For this, we run an application that samples the GPS every second to show the contrast in power consumption (*GPS is neither available in all locations in stations nor able to detect semantics but it is used as a baseline system*). The energy is calculated using the PowerTutor profiler [22] and the Android APIs using the HTC Nexus One cell phone. *TransitLabel* leverages the inertial sensors for passengers' activity recognition and position estimation. Since inertial sensors are indeed used during the normal phone operation, to detect the phone orientation change or estimate the user location for any indoor LBS, *TransitLabel* practically consumes little extra sensing power in addition to the standard phone operation. In

Table 2: Confusion Matrix for classifying different **elevation change** semantics.

	Elevator (Single)	Elevator (Double)	Stairs (Straight)	Stairs (Half-land.)	Escalator (Stand.)	Escalator (Climb.)	Escalator (Over.)	FP	FN	Σ
Elevator (Single)	100	5	0	0	0	0	0	6.7%	4.8%	105
Elevator (Double)	7	79	0	0	0	0	0	5.8%	8.1%	86
Stairs (Straight)	0	0	111	0	0	9	9	0%	7.5%	120
Stairs (Half-land.)	0	0	0	51	0	0	0	0%	0%	51
Escalator (Stand.)	0	0	0	0	114	0	-	-	-	114
Escalator (Climb.)	0	0	0	0	0	107	-	-	-	107
Escalator (Over.)	0	0	0	0	-	-	221	4.1%	0%	221
Total								3.3%	4.1%	583

Table 3: Confusion Matrix for classifying different stations specific semantics discovered from **stationary** traces.

	Drink Vending Machine	Ticket Vending Machine	Locker	Restroom	Waiting (Sitting) Area	Standing	Cars' Waiting Lines	FP	FN	Σ
Drink Vending Machine	123	1	6	0	0	0	1	7.6%	6.1%	131
Ticket Vending Machine	9	260	12	0	0	0	0	3.2%	7.5%	281
Locker	1	8	119	0	0	0	0	20.3%	7%	128
Restroom	0	0	4	87	2	4	0	7.2%	10.3%	97
Waiting (Sitting) Area	0	0	0	4	88	6	3	11.9%	12.9%	101
Standing	0	0	3	3	7	140	5	11.4%	11.4%	158
Cars' Waiting lines (waiting for a train)	0	0	1	0	3	8	125	6.6%	8.8%	137
Total								9.7%	9.1%	1033

Table 4: Confusion Matrix for classifying different stations specific semantics discovered from **non stationary** traces.

	Walking	Platform Track (Boarding)	Entrance Gate (Ticket)	Entrance gate (IC Card)	Entrance gate (Overall)	FP	FN	Σ
Walking	170	7	5	9	14	10.5%	11%	191
Platform Track (Boarding)	3	135	2	11	13	15.2%	10.6%	151
Entrance Gate (Ticket)	9	2	264	6	-	-	-	281
Entrance gate (IC Card)	8	14	9	234	-	-	-	265
Entrance gate (Overall)	17	16	-	-	513	4.9%	6.0%	546
Total						10.2%	9.2%	888

Table 5: The semantics classification accuracy in a **one-station-out** cross validation.

	Elevator(S.)	Elevator (D.)	Stairs (Str.)	Stairs (Half.)	Escalator	Drink Vd. Mch.	Ticket Vd. Mch.	Lock.	Restr.	Sit. Area	Wait. Lines	Plat. Tr.	Entr. gate	Σ
FP	8.6%	4.6%	0%	0%	5.4%	9.1%	3.9%	21%	8.2%	10.8%	8%	9.2%	5.3%	7.2%
FN	3.8%	10.4%	10%	0%	0%	7.6%	8.2%	8.6%	9.3%	15.8%	10.2%	11.9%	8.6%	8.0%

addition, the sound sensor; which has a higher energy footprint; is activated only shortly during the time of activities (average activity duration is short- 31seconds- excluding the restroom). This also avoids the impact of false positive triggers of the sound sensor (e.g., sensor being activated while the user is standing doing a phone call). Finally, as soon as the user waits for the train on the platform, we suspend the activation of the sound sensor. All lead to a low energy consumption of *TransitLabel* that is significantly (44%) less than the GPS consumption.

5.2.4 Impact of the Phone Placement

To demonstrate that our approach is robust to different device placements, we carried out experiments on different phone placements (in a user hand or in a trouser pocket). Figure 16 shows the semantics detection accuracy as measured by the F-measure¹¹ in both phone placements. The results suggest that the semantic accuracy is not significantly dependent on device poses, with *TransitLabel* achieving a high accuracy of about 80% at least (in the case of non-stationary traces which is the most tricky case due to the effect of the legs movement on the phone in the pocket reading). *This robustness mainly comes from the transformation of sensor readings to the real world coordinates and the extraction of placement-independent features from phone sensors.* For example, the magnetometer readings features are not largely attenuated by human bodies (i.e., in pocket). While the acceleration can be affected by the motion noise to some extent, its impact is effectively mitigated

¹¹F-measure is the harmonic mean of precision and recall and represents a single number for comparison.

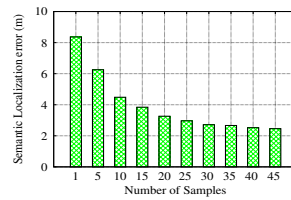


Figure 14: Effect of the number of samples on the accuracy of semantics location estimation.

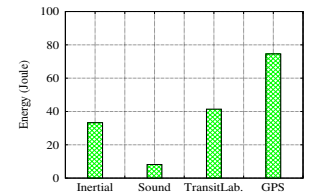


Figure 15: Energy footprint of *TransitLabel*.

by smoothing the raw acceleration and averaging the acceleration variance over a long time window. Finally, the collected audio data is not much derailed by the phone placement in pocket as passengers are forced to be close to semantics at the usage time. Note that *TransitLabel* uses different thresholds values to identify passenger's activities depending on the phone placement (pocket or hand). Nonetheless, the light proximity sensor can be used to detect the phone position and, accordingly, decide which thresholds value to be used. Other phone positions can be handled in a similar manner, which is a subject of future work for space constraints.

5.2.5 Generalization of *TransitLabel*

We based our semantics identification on their typical usage pattern/signature by the majority of train commuters. However, some users may have different usage patterns (regular versus occasional travelers) and some stations have different building structures and

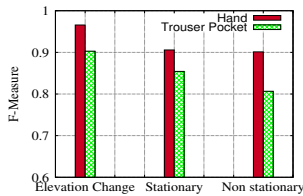


Figure 16: The Impact of phone placement on the semantics detection accuracy.

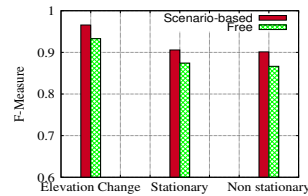


Figure 17: The generalization experiment of *TransitLabel* semantic detection.

different machines hardware (e.g., vending machines) that may lead to some semantics misclassifications. Thus, to demonstrate that *TransitLabel* could generalize over various users and stations, we consider an experiment where a group of daily train commuters are asked to collect data *freely (without prescribed scenarios)* in different stations in two different cities managed by different train companies in Japan. Figure 17 shows the accuracy of the semantic detection by *TransitLabel*. The results demonstrate that *TransitLabel* can still achieve a comparable semantic detection accuracy to the scenario-based experiments. This robustness is mainly due to the uniform nature of passengers’ activities at railway station. Specifically, most passengers follow similar routes, exhibit uniform behavior as they have a common target (boarding the train) starting from getting a ticket to boarding the train. In addition, *TransitLabel* fuses multiple features (e.g., accelerometer, gyroscope and magnetometer based features are fused to identify the coin operated machines as shown in the classification tree in Figure 3b) to identify the same semantic, reducing the sensitivity to specific scenarios or machines.

Per-Station Accuracy: To understand the classification accuracy on a per-station basis, we show a one-station-out cross validation (one station data is selected as the validation data while all data collected at other stations is used as the database) results in Table 5. Evident from the table, the classification accuracy does not deviate much between different stations due to fusing many features and the uniform railway passengers’ behaviors. This further emphasizes the generalization ability of *TransitLabel*.

6 Discussion and Limitations

We discuss some crowd-sensing challenges addressed by the current version of *TransitLabel* along with our ongoing work.

System Robustness

Our evaluation, quantified in Section 5, is carried out by different users using different phones in different placement through different data collection methodologies and spanned different stations and cities in Japan. These extensive experiments verified *TransitLabel* robustness under a wide range of scenarios. We believe that this robustness is based on a number of factors/design decisions including reorienting the phone sensors data, harnessing offset-invariant and orientation-invariant features, multi-modal sensor fusion, and combining a number of low level activities to identify the higher level passengers’ activity. All allow *TransitLabel* to generalize to different stations run by different operators, users, devices, and operation scenarios.

Scalability

We have tested *TransitLabel* extensively in eight railway stations by 16 users under different scenarios. Scaling *TransitLabel* to a worldwide scale is directly based on location clustering and lever-

aging the cloud. In particular, the semantics of each station can be processed independently of other stations based on the location of the collected traces. This spatial clustering lends itself nicely to the processing in the cloud, further enhancing the system scalability.

Preserving User’s Privacy

TransitLabel gives users full control over their sensed data and processes the audio data locally on the user’s phone. Nonetheless, some users may opt to turn off the sound or other sensors for privacy concerns. However, since *TransitLabel* is a crowd-based system, it will still be able to identify these semantics from the samples submitted by other users.

Handling Dynamic Changes

The station internal structures may evolve over time and, accordingly, the state of the different semantics (location changed or removed). To address this, *TransitLabel* periodically rerun its clustering algorithm across time windows of different granularity to detect the removal of specific semantics. Specifically, the lack of a cluster with a specific size at the location of a previously detected semantic indicates the removal of this semantic. If clusters of correct samples of a semantic type formed in consecutive windows are being mapped to a location that vary substantially from old instances locations, this is indicative of a change in the environment. To classify the change type, we monitor newly uploaded samples of each instance of a semantic type. The change of location of a specific semantic is treated as a simple removal of this landmark and detecting it at a new location. We leave the evaluation of this aspect to future work due to space constraints.

Other Indoor Environments

Although *TransitLabel* is designed for railway stations, it can be customized to other indoor environments, e.g., airports, that have similar semantics (e.g., elevators, stairs, escalators, vending machines, lockers, security gates, and automated boarding-pass printing machine). Moreover, *TransitLabel* can be extended to use semi-supervised learning techniques [3, 41] to extend to new environments without any prior knowledge about what activities are expected. Specifically, new automatically learned semantic classes can be presented to a human user to provide the label to them, significantly reducing the overhead of extending *TransitLabel* to new environments.

7 Related Work

The ideas in *TransitLabel* are built on three threads of research: mobile phone localization, human activity recognition and floor-plan construction. We survey the most relevant work in each thread in the interest of space.

Mobile Phone Localization

Mobile phone localization has been well-studied with a variety of approaches so far [42]. As GPS signal is not available in many stations (e.g., subway stations), an indoor localization technique is needed to estimate the passenger’s location. The most ubiquitous indoor localization techniques are either WiFi-based or dead-reckoning based. WiFi-based techniques, e.g., [12, 14, 24, 44–47], require calibration to create a prior wireless map for the building. However, the calibration process is time consuming, tedious, and requires periodic updates; leading to the emergence of new calibration-free techniques [13]. Dead-reckoning based localization techniques, e.g., [26, 41], leverage the inertial sensors on mobile phones to dead-reckon the user starting from a reference point [26]. However, dead-reckoning error quickly accumulates leading

to complete deviation from the actual path. Therefore, many techniques have been proposed to reset the dead-reckoning error including snapping to environment anchor points, such as elevators and stairs [3, 41] and matching with the map information either indoor [34] or outdoor [4, 5].

TransitLabel employs the basic concept in [3, 41] as it provides accurate, energy-efficient localization, and does not require an infrastructure support. However, *TransitLabel* discovers novel, activity-based, fine-grained and richer set of semantics targeting railway stations (e.g., vending machines, lockers, entrance gates, platforms, waiting (sitting) areas, among others) to reset the accumulated localization error frequently.

Human Activity Recognition

Activity recognition literature has demonstrated the ability to recognize user behavior using phone equipped sensors. Accelerometer data was used to detect the user transportation mode (walking, stationary, being in motorized transport) [1, 23, 28, 36, 40] and it can classify more fine-grained activities and attributes like running, breathing rate, climbing up the stairs, biking, cleaning kitchen, vacuuming, and brushing teeth [2, 6, 18, 29, 35]. Moreover, accelerometer data is used to detect more complex human activities like biking, lying, cleaning kitchen, cooking, sweeping, washing hands, and medication [11]. Ambient sensors like temperature, humidity, pressure, and light have been used to label user's location directly as being in kitchen, bedroom, bathroom and living room [30]. Moreover, the AmbientSense system [37] can recognize 23 different contexts (e.g., coffee machine, raining, restaurant, dishwasher, toilet flush, etc) by analyzing ambient sounds sampled from phone. In addition, the RoomSense system in [38] uses active sound probing to classify the type of room (e.g., corridor, kitchen, lecture room, etc) where the user is located. Ref. [20] actively probes the environment and then analyzes the impulse response on the phone to separate restroom from other rooms. Recently, RF-based device-free activity tracking and recognition has been used to detect different activities and the location of the person using standard RF networks [27, 39, 43].

TransitLabel recognizes a higher level and novel set of passenger activities at railway stations. The crowd-sensed locations of these activities are then mined to discover their uniquely associated stations semantics.

Automatic Floorplan Construction

Recently, a number of systems have been proposed that employ pedestrian motion traces to automatically construct indoor floorplans [7, 15–17, 21, 25]. For instance, CrowdInside [7] processes inertial motion traces using computational geometry techniques to extract the overall floorplan shape as well as corridors and room boundaries. It also identifies a variety of points of interest in the environment such as elevators and stairs. However, their semantic detection method neither targets stations specific semantics (e.g., entrance gate, etc) nor it provides fine-grained classes of elevation change semantics (e.g., stair types and elevator types). In addition, their elevator detection algorithm leverages only the motion pattern (Accelerate-Constant-Decelerate) which may coincide with normal human walking patterns. This cannot happen in our method as normal walking traces are separated beforehand using the semantic type detection module. Finally, different passengers' behaviors (e.g., climbing up or standing in escalators) makes the low acceleration variance, used in their method, is not a reliable discriminator between climbing stairs and escalators. Jigsaw [21] uses a computer vision approach to extract the position, size, and orientation of landmark objects from images taken by users. It then com-

bines user mobility traces and locations where images are taken to produce the hallway connectivity and the room size. The system proposed in [25] leverages Wi-Fi fingerprints and user motion information to determine which rooms are adjacent in the building and estimating their sizes. It then orders them along each hallway and adjusts the room sizes to optimize the overall floorplan layout. *Nevertheless, all previous systems did not attach any semantic information to the floorplan layout.* Finally, Ref. [33] assumes that there is a difference in time to change floors using elevators, escalators, and stairs and thus relies on the rate of height change (i.e., pressure) to separate them. However, this hypothesis is neither robust to different users walking speeds nor to different elevators/escalators motion speeds.

The closest work to ours is the automatic enrichment of indoor floorplans with semantic names technique in [17]. It exploits phone sensors data collected from users during their normal check-ins to location-based social networks (LBSNs) (e.g., Foursquare) and combines them with data extracted from the LBSNs databases to associate a place name with its location on an unlabeled floorplan. This system, however, can be applied only to indoor environments where the check-ins granularity level matches the semantic names needed to enrich the map. For example, in shopping malls, user check-ins with venues names which can be leveraged by Sensesense [17] to attach a venue name to each room in the floorplan. However, users at railway stations usually do check-ins with stations names (they don't check-in using fine-grained station indoor semantics names) which is not sufficient to infer and locate the in-station semantics.

TransitLabel assumes in its operation the availability of an unlabeled station floorplan by using one of these approaches. It then enriches the input floorplan with different semantics based on data collected from users' phones.

8 Conclusion

We presented the *TransitLabel* system for automatically enriching transit indoor maps via a crowdsensing approach based on standard cell phones. For energy efficiency, *TransitLabel* leverages low-energy phone sensors and sensors that are already running for other purposes (e.g., inertial sensors). We presented the *TransitLabel* architecture as well as the features and classifiers that can accurately recognize different passenger's activities in railway stations which are mined to detect their uniquely associated semantics.

We implemented *TransitLabel* using commodity mobile phones running the Android operating system and evaluated it at different railway stations in Japan. Our results show that *TransitLabel* can detect stations fine-grained semantics accurately with 7.7% false positive and 7.5% false negative rates on average leading to high accuracy in semantics location estimation. Finally, *TransitLabel* can be generalized over various stations running by different operators and user groups; and is robust to different phone placements while having a significantly small energy profile.

Currently we are expanding *TransitLabel* in multiple directions including inferring more station semantics, handling dynamic changes in the environment, deployment of *TransitLabel* in other indoor environments, among others.

9 Acknowledgement

We sincerely thank the anonymous reviewers for their invaluable feedback which helped improve the paper.

This work is supported in part by JSPS KAKENHI grant numbers 26220001, 15H02690, and 26700006 to Osaka University and in part by a Google Research Award to E-JUST.

10 References

- [1] A. M. AbdelAziz and M. Youssef. The diversity and scale matter: Ubiquitous transportation mode detection using single cell tower information. In *Proceedings of the 81st Vehicular Technology Conference (VTC Spring)*, pages 1–5. IEEE, 2015.
- [2] H. Abdelnasser, K. A. Harras, and M. Youssef. UbiBreathe: A Ubiquitous non-invasive WiFi-based Breathing Estimator. In *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*, pages 277–286. ACM, 2015.
- [3] H. Abdelnasser, R. Mohamed, A. Elgohary, M. Farid, H. Wang, S. Sen, R. Choudhury, and M. Youssef. SemanticSLAM: Using environment landmarks for unsupervised indoor localization. *IEEE Transactions on Mobile Computing*, (1):1–1.
- [4] H. Aly, A. Basalamah, and M. Youssef. Map++: A crowd-sensing system for automatic map semantics identification. In *Proceedings of the Eleventh Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, pages 546–554. IEEE, 2014.
- [5] H. Aly and M. Youssef. Dejavu: An accurate energy-efficient outdoor localization system. In *Proceedings of the 21st ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL)*, pages 154–163. ACM, 2013.
- [6] H. Aly and M. Youssef. Zephyr: Ubiquitous accurate multi-sensor fusion-based respiratory rate estimation using smartphones. In *Proceedings of IEEE Conference on Computer Communications (INFOCOM)*. IEEE, 2016.
- [7] M. Alzantot and M. Youssef. CrowdInside: Automatic construction of indoor floorplans. In *Proceedings of the 20th International Conference on Advances in Geographic Information Systems (SIGSPATIAL)*, pages 99–108. ACM, 2012.
- [8] M. Alzantot and M. Youssef. UPTIME: Ubiquitous pedestrian tracking using mobile phones. In *Proceedings of Wireless Communications and Networking Conference (WCNC)*, pages 3204–3209. IEEE, 2012.
- [9] Y. Chon, N. D. Lane, Y. Kim, F. Zhao, and H. Cha. Understanding the coverage and scalability of place-centric crowdsensing. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing (UbiComp)*, pages 3–12. ACM, 2013.
- [10] W. S. Cleveland and S. J. Devlin. Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association*, 83(403), 1988.
- [11] S. Dernbach, B. Das, N. C. Krishnan, B. L. Thomas, and D. J. Cook. Simple and complex activity recognition through smart phones. In *Proceedings of 8th International Conference on Intelligent Environments (IE)*, pages 214–221. IEEE, 2012.
- [12] K. El-Kafrawy, M. Youssef, A. El-Keyi, and A. Naguib. Propagation modeling for accurate indoor WLAN RSS-based localization. In *Proceedings of the 72nd Vehicular Technology Conference Fall (VTC -Fall)*, pages 1–5. IEEE, 2010.
- [13] R. Elbakly and M. Youssef. A robust zero-calibration RF-based localization system for realistic environments. In *Proceedings of the Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. IEEE, 2016.
- [14] A. Eleryan, M. Elsabagh, and M. Youssef. Synthetic generation of radio maps for device-free passive localization. In *Proceedings of the Global Telecommunications Conference (GLOBECOM)*, pages 1–5. IEEE, 2011.
- [15] M. Elhamshary, A. Uchiyama, H. Yamaguchi, and T. Higashino. LandmarkSense: A Mobile Sensing System for Automatic Detection of Railway Stations Landmarks. Technical report, IPSJ SIG, 12 2015.
- [16] M. Elhamshary and M. Youssef. CheckInside: A fine-grained indoor location-based social network. In *Proceedings of the International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*, pages 607–618. ACM, 2014.
- [17] M. Elhamshary and M. Youssef. SemSense: Automatic construction of semantic indoor floorplans. In *Proceedings of the 6th International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pages 1–11. IEEE, 2015.
- [18] M. Elhamshary, M. Youssef, A. Uchiyama, H. Yamaguchi, and T. Higashino. Activity recognition of railway passengers by fusion of low-power sensors in mobile phones. In *Proceedings of the 23rd International Conference on Advances in Geographic Information Systems (SIGSPATIAL)*, page 57. ACM, 2015.
- [19] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, volume 96, pages 226–231, 1996.
- [20] M. Fan, A. T. Adams, and K. N. Truong. Public restroom detection on mobile phone via active probing. In *Proceedings of the 2014 ACM International Symposium on Wearable Computers (ISWC)*, pages 27–34. ACM, 2014.
- [21] R. Gao, M. Zhao, T. Ye, F. Ye, Y. Wang, K. Bian, T. Wang, and X. Li. Jigsaw: Indoor floor plan reconstruction via mobile crowdsensing. In *Proceedings of the 20th annual international conference on Mobile computing and networking (MobiCom)*, pages 249–260. ACM, 2014.
- [22] M. Gordon et al. PowerTutor: a power monitor for android-based mobile platforms, 2013.
- [23] S. Hemminki, P. Nurmi, and S. Tarkoma. Accelerometer-based transportation mode detection on smartphones. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems (SenSys)*, page 13. ACM, 2013.
- [24] M. Ibrahim and M. Youssef. A hidden markov model for localization using low-end GSM cell phones. In *Proceedings of IEEE International Conference on Communications (ICC)*, pages 1–5. IEEE, 2011.
- [25] Y. Jiang, Y. Xiang, X. Pan, K. Li, Q. Lv, R. P. Dick, L. Shang, and M. Hannigan. Hallway based automatic indoor floorplan construction using room fingerprints. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing (UbiComp)*, pages 315–324. ACM, 2013.
- [26] Y. Jin, H.-S. Toh, W.-S. Soh, and W.-C. Wong. A robust dead-reckoning pedestrian tracking system with low cost sensors. In *Proceedings of IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 222–230. IEEE, 2011.
- [27] A. E. Kosba, A. Abdelkader, and M. Youssef. Analysis of a device-free passive tracking system in typical wireless environments. In *Proceedings of the 3rd International*

- Conference on New Technologies, Mobility and Security (NTMS)*, pages 1–5. IEEE, 2009.
- [28] J. Krumm and E. Horvitz. LOCADIO: Inferring motion and location from Wi-Fi signal strengths. In *Proceedings of the Second Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services (MobiQuitous)*, pages 4–13, 2004.
- [29] J. R. Kwapisz, G. M. Weiss, and S. A. Moore. Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter*, 12(2):74–82, 2011.
- [30] S. Mazilu and G. Troster. A study on using ambient sensors from smartphones for indoor location detection. In *Proceedings of 12th Workshop On positioning, navigation and communication (WPNC)*. IEEE, 2015.
- [31] P. Mohan, V. N. Padmanabhan, and R. Ramjee. Nericell: Rich monitoring of road and traffic conditions using mobile smartphones. In *Proceedings of the 6th ACM conference on Embedded network sensor systems (SenSys)*, pages 323–336. ACM, 2008.
- [32] N. Mohssen, R. Momtaz, H. Aly, and M. Youssef. It’s the human that matters: Accurate user orientation estimation for mobile computing applications. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitous)*, pages 70–79. ICST, 2014.
- [33] K. Muralidharan, A. J. Khan, A. Misra, R. K. Balan, and S. Agarwal. Barometric phone sensors: More hype than hope!. In *Proceedings of the 15th Workshop on Mobile Computing Systems and Applications (HotMobile)*, pages 12–18. ACM, 2014.
- [34] A. Rai, K. K. Chintalapudi, V. N. Padmanabhan, and R. Sen. Zee: Zero-effort crowdsourcing for indoor localization. In *Proceedings of the 18th annual international conference on Mobile computing and networking (MobiCom)*, pages 293–304. ACM, 2012.
- [35] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman. Activity recognition from accelerometer data. In *AAAI*, volume 5, pages 1541–1546, 2005.
- [36] S. Reddy, M. Mun, J. Burke, D. Estrin, M. Hansen, and M. Srivastava. Using mobile phones to determine transportation modes. *ACM Transactions on Sensor Networks (TOSN)*, 6(2):13, 2010.
- [37] M. Rossi, S. Feese, O. Amft, N. Braune, S. Martis, and G. Troster. AmbientSense: A real-time ambient sound recognition system for smartphones. In *Proceedings of IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, pages 230–235. IEEE, 2013.
- [38] M. Rossi, J. Seiter, O. Amft, S. Buchmeier, and G. Tröster. RoomSense: An indoor positioning system for smartphones using active sound probing. In *Proceedings of the 4th Augmented Human International Conference (AH)*, pages 89–95. ACM, 2013.
- [39] M. Seifeldin and M. Youssef. A deterministic large-scale device-free passive localization system for wireless environments. In *Proceedings of the 3rd International Conference on Pervasive Technologies Related to Assistive Environments (PETRA)*, page 51. ACM, 2010.
- [40] T. Sohn, A. Varshavsky, A. LaMarca, M. Y. Chen, T. Choudhury, I. Smith, S. Consolvo, J. Hightower, W. G. Griswold, and E. De Lara. Mobility detection using everyday GSM traces. In *Proceedings of international conference on Ubiquitous Computing (UbiComp)*, pages 212–224. Springer, 2006.
- [41] H. Wang, S. Sen, A. Elgohary, M. Farid, M. Youssef, and R. R. Choudhury. No need to war-drive: unsupervised indoor localization. In *Proceedings of the 10th international conference on Mobile systems, applications, and services (MobiSys)*, pages 197–210. ACM, 2012.
- [42] M. Youssef. Towards truly ubiquitous indoor localization on a worldwide scale. In *Proceedings of the 23rd International Conference on Advances in Geographic Information Systems (SIGSPATIAL)*, page 12. ACM, 2015.
- [43] M. Youssef. A Decade Later - Challenges: Device-free passive localization for wireless environments. In *Proceedings of the Fifth IEEE COSDEO Workshop, IEEE PerCom*, 2016.
- [44] M. Youssef, M. Abdallah, and A. Agrawala. Multivariate analysis for probabilistic WLAN location determination systems. In *Proceedings of the Second Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services (MobiQuitous)*, pages 353–362. IEEE, 2005.
- [45] M. Youssef and A. Agrawala. The Horus WLAN location determination system. In *Proceedings of the 3rd international conference on Mobile systems, applications, and services (MobiSys)*, pages 205–218. ACM, 2005.
- [46] M. Youssef and A. Agrawala. Location-clustering techniques for WLAN location determination systems. *International Journal of Computers and Applications*, 28(3):278–284, 2006.
- [47] M. Youssef and A. Agrawala. The Horus location determination system. *Wireless Networks*, 2008.