September 11, 2017

FootprintID: Indoor Pedestrian Identification through Ambient Structural Vibration Sensing

Shijia Pan
Tong Yu
Mostafa Mirshekari
Jonathon Fagert
Amelie Bonde, et al.

FootprintID: Indoor Pedestrian Identification through Ambient Structural Vibration Sensing

SHIJIA PAN, TONG YU, MOSTAFA MIRSHEKARI, JONATHON FAGERT, AMELIE BONDE, OLE J. MENGSHOEL, HAE YOUNG NOH, and PEI ZHANG, Carnegie Mellon University

We present FootprintID, an indoor pedestrian identification system that utilizes footstep-induced structural vibration to infer pedestrian identities for enabling various smart building applications. Previous studies have explored other sensing methods, including vision-, RF-, mobile-, and acoustic-based methods. They often require specific sensing conditions, including line-of-sight, high sensor density, and carrying wearable devices. Vibration-based methods, on the other hand, provide easy-to-install sparse sensing and utilize gait to distinguish different individuals. However, the challenge for these methods is that the signals are sensitive to the gait variations caused by different walking speeds and the floor variations caused by structural heterogeneity.

We present FootprintID, a vibration-based approach that achieves robust pedestrian identification. The system uses vibration sensors to detect footstep-induced vibrations. It then selects vibration signals and classifiers to accommodate sensing variations, taking step location and frequency into account. We utilize the physical insight on how individual step signal changes with walking speeds and introduce an iterative transductive learning algorithm (`ITSVM`) to achieve robust classification with limited labeled training data. When trained only on the average walking speed and tested on different walking speeds, FootprintID achieves up to 96% accuracy and a 3X improvement in extreme speeds compared to the Support Vector Machine. Furthermore, it achieves up to 90% accuracy (1.5X improvement) in uncontrolled experiments.

CCS Concepts: •**Human-centered computing** →*Ubiquitous and mobile computing systems and tools;*

Additional Key Words and Phrases: ambient vibration sensing, structural vibration, pedestrian identification

ACM Reference format:
Shijia Pan, Tong Yu, Mostafa Mirshekari, Jonathon Fagert, Amelie Bonde, Ole J. Mengshoel, Hae Young Noh, and Pei Zhang. 2017. FootprintID: Indoor Pedestrian Identification t hrough A mbient S tructural V ibration S e nsing. *PACM I n teract. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 89 (September 2017), 31 pages.
DOI: http://doi.org/10.1145/3130954

1 INTRODUCTION

Indoor pedestrian identities are essential information for smart building applications such as patient/staff monitoring in hospitals, elder monitoring in retirement facilities, personalized services in smart homes, etc. For example, when the system detects a registered user walking in the kitchen, it locks the liquor cabinet if the user is underage. Various sensing methods have been explored, including vision [61, 68, 77], radio frequency

This work is partially supported by NSF (CNS-1149611, CMMI-1653550 and 1344768), Intel, Pennsylvania Infrastructure Technology Alliance (PITA), and Google.
Author's addresses: S. Pan, T. Yu, A. Bonde, O. J. Mengshoel and P. Zhang, Electrical and Computer Engineering, Carnegie Mellon University, NASA Ames Research Center, Moffett Field, CA 94035; M. Mirshekari, J. Fagert and H. Y. Noh, Civil and Environmental Department, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213.

DOI: http://doi.org/10.1145/3130954

[71, 76], mobile [17, 32, 38, 56, 64], acoustic, [2, 20] and structural vibration [16, 52] based methods. Each of these methods has limitations. Vision-based methods only work with a clear visual path, and this may limit sensor installation locations [6, 61, 77]. RF-based methods often need dense deployment to achieve high sensing accuracy [71, 76]. Mobile-based methods rely on the target carrying a specific device [17, 32, 38, 56, 64]. Acoustic-based methods are sensitive to ambient audible noise [2, 20]. Structural vibration-based methods, on the other hand, are easy-to-retrofit and provide sparse sensing ability [16, 52]. However, when applied to the pedestrian identification problem, they are often sensitive to many common condition variations, such as structural variation and individual walking speed changes. Hence, for scenarios where pedestrian walking speeds show high variation, such as staff in a nursing home or kids at home, a large amount of labeled training data that covers each condition is required for a pure data-driven approach. Unfortunately, this precondition is usually impractical in real world identification system.

In this paper, we present FootprintID, a sparse structural vibration sensing system that enables ubiquitous pedestrian identification. When a person walks, their footsteps induce the floor to vibrate and this vibration can be detected by vibration sensors (e.g., geophone, accelerometers). Since different people strike the floor differently [18], their unique walking patterns induce distinguishable vibration, which our system utilizes to identify the individual [9]. The challenge for vibration-based methods lies in their sensitivity to changing walking conditions, including walking speeds and stepping locations.

Signal changes due to walking speed and step location changes can be summarized as a few representative signal waveforms. These waveforms' distribution gradually changes with the individual walking speed and stepping location. With this physical insight, we present our algorithm that 1) selects step signals based on estimated step location, 2) selects classifiers (supervised learning v.s. transductive learning) based on detected walking speeds, and 3) applies our transductive learning algorithms Refined Transductive Support Vector Machine (`RTSVM`) and Iterative Transductive Support Vector Machine (`ITSVM`) when the tested walking speeds vary from those in the labeled training set.

Our `ITSVM` algorithm tailors the learning model with the unlabeled testing data (transductive learning) and updates the labeled data set model with unlabeled testing results based on physical insights and result confidence. Compared to supervised learning, `ITSVM` allows robust classification when the unlabeled testing data shows high variation from the labeled training data. `ITSVM` accomplishes this by utilizing physical insights to iteratively update the classification model with high certainty unlabeled data.

In this paper, we focus on scenarios in which one person walks through the sensing area at a time for two reasons: 1) in a home or a nursing home, the density of the population is relatively low, and one person walking through a sensing area is a typical case; and 2) identification of multiple pedestrians walking in an indoor environment can be built on the foundation of robust one-pedestrian identification. We evaluate the system through two sets of experiments with 10 pedestrians to explore the robustness of the system: 1) controlled experiments walking at several designated step frequencies, and 2) uncontrolled experiments walking at natural step frequencies in a more realistic setting.

In summary, the contributions of this paper are threefold:

- We introduce FootprintID, a system that identifies a pedestrian through their footstep-induced vibration.
- We characterize the walking speed and step location variation of footstep-induced structural vibration signals and utilize this information to achieve robust person identification.
- We present our transductive learning algorithm `RTSVM` and the improved `ITSVM` that dynamically updates the model of the labeled data based on the walking speed and step location of a person's footstep to extend the classifier and to handle extreme cases.

The rest of the paper is organized as follows: Section 2 introduces physiological information about human gaits. In Section 3, we describe the overall system functions, followed by Section 4, where we describe the sensing

unit used for signal acquisition. This signal is analyzed in Section 5 to guide the algorithm design, introduced in Section 6. Next, in Sections 7 and 8 we evaluate the accuracy of FootprintID under different scenarios. We discuss challenges that our system may encounter in Section 9 and then present past work in Section 10. Finally, we conclude our work in Section 11.

2 BACKGROUND

To provide a background understanding of our system, this section explores why people have different gaits and why the location and frequency of the steps affect the sensing signal.

2.1 Why Do People Have Different Gaits?

Gait describes a subject's walking pattern, i.e., the characteristics of the subject's limb locomotion [18]. From the moment the heel strikes the floor to when the toe leaves the floor, a series of muscle contractions cause the leg to move: the lengthening of the dorsiflexors of the ankle, the lengthening of the gastrocnemius-soleus muscles, and then the shortening of the gastrocnemius-soleus [18]. These muscles are attached to the bone structure, therefore the bone morphology also affects such motions [3, 10]. In different people, muscular ability, degrees of freedom in the joints, and bone morphology vary [10, 13, 45], all of which have an impact on gait. Another determining factor for human gait is neurological [39, 59, 67, 72]. An individual's actual movement is controlled by neurons connected to the spine and brain. People have varying neurological control over their limbs, which results in different gaits. Prior work demonstrated that gait variation among pedestrians can be detected in their footstep-induced vibration signals [52].

2.2 Structural Effects

Footstep-induced vibration waves travel through the structure before they are captured by the sensors and hence are affected by structural properties such as modulus of elasticity[1]. Therefore, spatially non-uniform structural properties can result in different vibration signal waveforms in different parts of a space (e.g., hallway). Specifically, two factors contribute to this non-uniform behavior: 1) material heterogeneity, and 2) structural layout.

In heterogeneous materials[2], as the wave propagates from one section of the material to the next one, it undergoes reflection and refraction[3] causing the signals in these two sections to show different waveforms [58, 73]. Furthermore, footstep location with respect to structural layout results in different wave reflection patterns and thus can change the signal characteristics[4] [1, 37]. For example, for the same footsteps, sensors at different distances from a wall receive signals of different shapes as the reflected part of the signals arrive at different times. In another example, the presence of structural joints[5] prevents the wave from traveling across adjacent regions.

2.3 Individual Variation: Step Frequency

Step frequency is one of the most critical parameters in the study of gait variation [15, 44, 55, 63]. The pedestrian's step frequency affects their step length and walking speed: the higher the step frequency, the longer the step length, and the faster the walking speed [44].

[1]Modulus of elasticity is a measure of the resistance of the material to elastic deformation. In general, a stiffer material has a higher modulus of elasticity [66].
[2]Materials with non-uniform distribution of properties.
[3]Change of wave propagation direction due to differences in structural properties
[4]In other words, different structural layout results in different boundary conditions.
[5]There are different types of joints in structures. For example, expansion joints are gaps between different parts of the structure to deal with concrete's volume change [14].

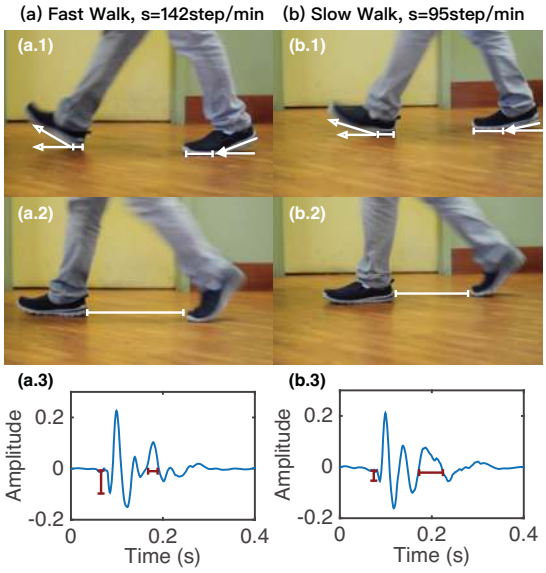

Fig. 1. Fast walk (a.1-3) v.s. slow walk (b.1-3). (a.1-2) show a person's gait phases when walking at the step frequency of 142 steps/min, and (b.1-2) show the same person's gait phases when walking at the step frequency of 95 steps/min. (a.3) and (b.3) show the Step Events extracted under these step frequencies.

In addition, the kinematic and kinetic parameters of a person's gait change with their step frequency [31, 46, 70]. As marked in Figure 1 (a.1) and (b.1), when a person walks fast, their heel strike movement shows a larger foot-ground angle. This is caused by the more powerful eccentric contraction of the dorsiflexors of the ankle when the subject tries to maintain a fast pace. This also leads to a longer step length as shown in Figure 1 (a.2) and (b.2). This means the footstep-induced vibration will also be affected by the person's step frequency.

Each individual may change their step frequency due to various reasons, including mood, health, urgency, personal preference, etc. Based on prior work on human gait parameters, the average step frequency is $\mu_M = 119$ steps/min with a standard deviation of $\sigma_M = 8$ steps/min for males between 20-29 [44]. For female participants between 20-29, the average step frequency is $\mu_F = 125$ steps/min with a standard deviation of $\sigma_F = 9$ steps/min [44]. Therefore, the system needs to take the variation of step frequency into account to conduct robust identification.

In summary, we discussed the reasoning behind the gait-based identification as well as the background of two major factors that affect vibration-based identification using gait information. We will further discuss variation in structural vibration characteristics caused by these factors in Section 5 after introducing our system.

3 FOOTPRINTID SYSTEM OVERVIEW

FootprintID is a vibration-based sensing system that identifies pedestrians from gait patterns. The system consists of four major modules as shown in Figure 2: the sensing module, the information extraction module, the structural variation handling module, and the step frequency variation handling module (our classifier that is robust to step frequency variations).

When a person steps on the floor, their step induces the floor to vibrate. The *sensing module* obtains the floor vibration signal (Section 4) and sends it to the information extraction module. The *information extraction module* first detects the footstep-induced signals, which we name as *Step Events* (SE), and then extracts the following information from these signals: 1) the location and frequency of the SEs (inter-step information), and 2) the signal

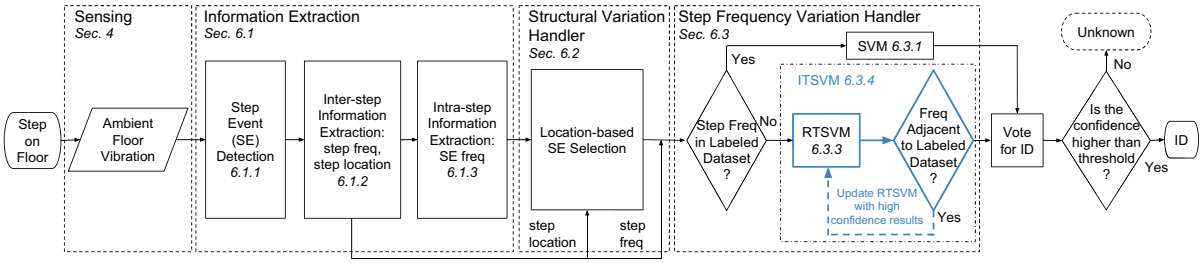

Fig. 2. System Overview.

feature of each SE (intra-step information). When a person passes through the sensing area, a sequence of SEs is detected, which we name as a trace. In this paper, we focus on one pedestrian walking through the sensing area scenarios, and we will discuss multiple pedestrians walking through the sensing area situation in Section 9.

The *structural variation handling module* estimates the relative location of each SE within a consecutive sequence, which we refer to as a *trace*, and selects the SEs that are closest to the sensor in each trace for further analysis. In this way, the system mitigates structural variation's effect on the SE signals. The *step frequency variation handling module* first trains both supervised classifiers and transductive classifiers in the bootstrap/learning phase (Section 8.3 and 9.1.1), and then conducts identification prediction in the classification phase. When predict the identity, FootprintID selects a classifier to obtain the identity of each SE based on the step frequency extracted. If the testing SE's step frequency is in the labeled training set, FootprintID selects the supervised learning method Support Vector Machine (SVM) to obtain the SE identity. On the other hand, if the testing SE's step frequency is not in the labeled training set, FootprintID selects our tailored transductive learning method `ITSVM` to estimate the SE identity.

The `ITSVM` initially starts as a transductive learning model with the same labeled training data as the SVM. If the testing SE's step frequency is adjacent to that of the labeled training data, i.e., the testing SE's similarity to prior training SEs is high, FootprintID updates the refined transductive support vector machine (`RTSVM`) model's labeled data with the traces with high confidence results, otherwise, the model remains the same. Either way, `ITSVM` outputs the identities of the SEs. In the end, FootprintID calculates the identity of the trace based on the estimated identity of the SEs within the trace through a majority vote and takes the most frequent identity. The confidence of the decision is calculated and the system can reject low confident identities as 'unknown' by thresholding on this confidence value according to the application requirement.

4 SENSING

In this section, we introduce the sensing hardware that we use to measure footstep-induced structural vibrations [53]. To obtain the vibration signals, a sensing unit is placed on the floor. The sensing unit, as shown in Figure 3, consists of five main components: the geophone, the amplifier module, the processor board, the communication module (XBee radio), and the batteries. We fix the geophone to the floor with beeswax to help preserve high-frequency signals [7].

The geophone converts the velocity of the monitored surface to voltage for ambient structural vibration monitoring. We used model SM-24 for its low cost and sensitivity to the frequency range of interest in this work (0-200 Hz) [25]. The voltage change induced by vertical floor vibration when people walk is very small (approximately 10^{-6} to 10^{-5} m/s range). Therefore, the system needs to amplify the signal for identification purposes. FootprintID sensing nodes have three levels of amplification (approximately ×2000, ×4000, and ×6000) on the amplifier board [53]. The system selects the maximum amplification setting that does not cause clipping

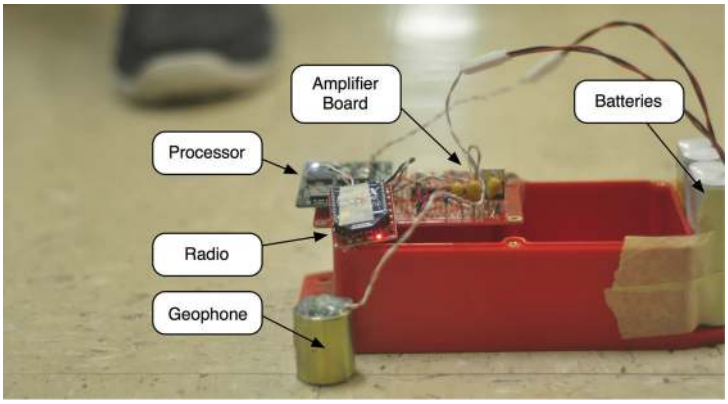

Fig. 3. A sensing unit consists of a geophone sensor, an amplifier board, a processor board with a communication module (Xbee radio), and batteries.

to obtain a high signal resolution for people with different stepping strengths. Note that in the deployment introduced in this paper, we use the amplification gain of ×2000 for the designated sensing area. This way, we achieve the lowest clipping rate while maintaining high signal resolution for the step signals selected for identification (we will further discuss the usability of the system in terms of sensing range in Section 9). Once the processor board obtains the amplified analog signal, it converts the signal into a digitized signal with a 10-bit ADC module sampled at 1000 Hz.

These amplified and digitized signals are then used for characterization and analysis. Figure 4 demonstrates two footstep-induced signals in both the time and frequency domain from two different pedestrians. We observe a clear difference in both the time and frequency domains between the two signals, which may be caused by their gait difference. For example, person #2 may have a longer heel to toe motion duration than Person # 1. In the time domain, the signal shows a larger lateral distance between the first and second peak, which we believe to be caused by the initial heel strike and the finishing toe pushing. We characterize the footstep-induced vibration signals in the next section.

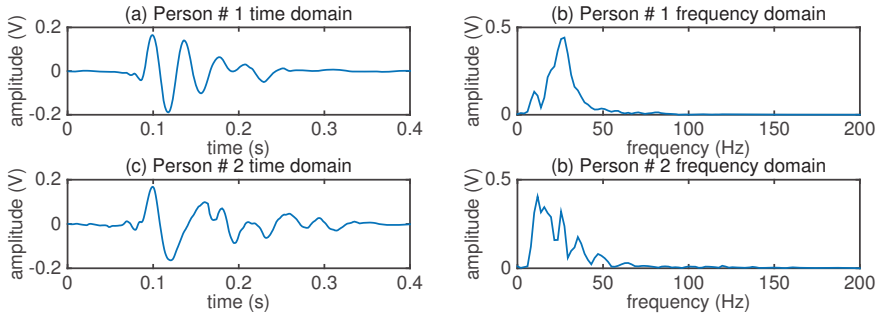

Fig. 4. Footstep-induced structural vibration signal examples. (a, b) show the time domain and frequency domain signal of a person # 1's footstep, and (c, d) demonstrate those of a person #2's footstep.

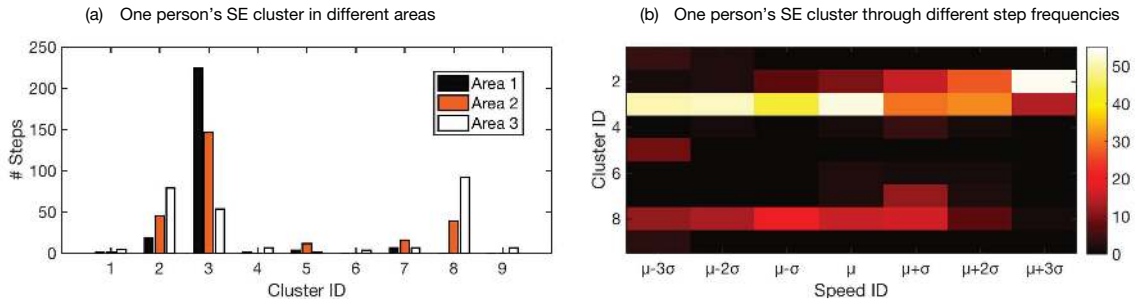

Fig. 5. Cluster example of Person # 1. (a) The x-axis is the cluster ID, and the y-axis is the number of steps in each cluster. The black, red, and white bars represent the SEs in different areas in a building. Different areas are dominated by different clusters. (b) The x-axis represents the relative step frequencies and the y-axis is the cluster ID. The grid values represent the number of SEs clustered to each cluster ID. Different step frequencies correspond to different clusters.

5 SIGNAL CHARACTERIZATION

To design a robust identification algorithm, we need to characterize the factors (discussed in Section 2) that cause variations in footstep-induced vibration signals. Prior work demonstrates the variation between Step Events from different people [52]. In this section, we explain how to quantify the variations within an individual's Step Events (Section 5.1) and what causes such variations (Section 5.2). We further discuss how to design the algorithm to account for these variations in Section 6.

5.1 Step Event Variation and Clustering

The Step Events (SEs) from an individual's dataset show variations that can be summarized by a few representative signal waveforms. In order to investigate the collected SEs in a systematic manner, we conduct hierarchical clustering [27] using pairwise cross-correlation values between time-domain signals of different SEs. We calculate the distance between each pair of SEs SE_i and SE_j as $distance = 1 - xcorr(SE_i, SE_j)$, where $xcorr$ calculates the peak cross correlation value of the signal pair. Then the pairs with the shortest distance are grouped into clusters made up of two objects. We repeat this process for the resulting clusters until all the objects in the original dataset are linked together in a hierarchical tree. To decide how many clusters to form for a specific tree generated from a person's data, we select a threshold value. The threshold value is empirically determined based on the experiments in the Load Test with Ball-drops (discussed in Section 7.1) to incorporate the structural variation within a designated area. Then we search the hierarchical tree and generate the clusters by the internal nodes of the tree. We use these clusters to explain the cause of the Step Event variation in the following section.

5.2 Causes of Step Event Variation

We have discussed the causes of Step Event (SE) variation in Section 2 from two angles: 1) a person may have multiple gait patterns that generate different waveforms when striking the floor at different step frequencies, and 2) when a person steps on different locations in the sensing area, the structural response of the same impulse changes due to the structural variation. These are the reasons that a pedestrian's SEs can be clustered. They also account for SEs with different distributions when location and walking speed vary. We present these clusters' distribution of Person #1's SEs as an example in Figure 5 and discuss these causes in detail in this section.

5.2.1 Structural Variation. A person's SEs in different locations within the sensing range may look different, which can be caused by the structural variation. Figure 5 (a) categorizes the clusters based on the estimated step

location. Area 2 is closest to the sensor location, while Area 1 and Area 3 are on either side of Area 2. We can observe that cluster 2 appears in Area 3 more often than in Area 2 and Area 1, and cluster 3 shows an opposite trend through the three areas. This indicates that the SE signals are also affected by step locations. This could be caused by the structural response variation or the slightly anisotropic characteristics of the floor [60]. We further show that such structural variation is gradual, which explains the repetitive appearance of the cluster at the same area in Section 7.1.

5.2.2 Step Frequency. A person's gait changes when they walk at different step frequencies as discussed in Section 2. When their gait changes, their SEs change correspondingly. Figures 1 (a.3) and (b.3) show the example step signal when a person passes by the sensor with a different step frequency. As we can see in the time domain signal of (b.3), the part of the signal around 0.2 seconds decays more smoothly than that of (a.3). This is caused by the slow stance phase (the foot remains in contact with the ground) in the gait cycle when people walk slowly. In addition, the first dip of the signal has a greater value in the fast walking SE than in the slow walking SE. This could be caused by the heavier heel strike when the walking speed is high. These SEs are repetitive in signals from the same person walking at the same designated step frequency. Therefore, we make the assumption that a person's footstep-induced vibration when walking through a designated area with different step frequency can be summarized into a few representative waveforms, i.e., different clusters.

Figure 5 (b) categorizes the clusters of Person # 1 by controlled step frequencies. A participant walking at a designated step frequency generates a footstep signal. We assign the designated step frequencies to 7 levels, including the average step frequency μ_{sf}, low frequencies from $\mu_{sf} - 3\sigma_{sf}$ to $\mu_{sf} - \sigma_{sf}$, and high frequencies from $\mu_{sf} + \sigma_{sf}$ to $\mu_{sf} + 3\sigma_{sf}$, where σ_{sf} is the step frequency standard deviation as discussed in Section 2 [6]. We introduce experiment details in Section 7.2.

In this figure, 9 clusters are generated from the SEs of 7 different step frequencies, and Clusters 2, 3, and 8 appear most frequently. Clusters 2, 3, and 8 appear more often in the high, low, and medium step frequency SEs respectively. From the low step frequency SEs to the high step frequency SEs, the portion of Cluster 3 decreased and the portion of Cluster 2 increased. This observation further verifies our conclusions on the relationship between step frequencies and gait changes.

6 ALGORITHM

The FootprintID algorithm is designed leveraging our study introduced in the last section and obtains a pedestrian's identity from their step-induced floor vibration signals. As discussed in Section 3, our algorithm consists of three modules: the information extraction (Section 6.1), the structural variation handler (Section 6.2), and the step frequency variation handler (Section 6.3).

6.1 Information Extraction

The information extracted from the raw ambient floor vibration includes two aspects: 1) the inter-footstep information (e.g., step frequency) and 2) the intra-footstep information (e.g., features of each SE). The system first needs to separate the vibration induced by the footsteps (SEs) from the ambient noise. Based on these SEs, the system further extracts the inter-footstep step frequency and relative location. Finally, for each SE, the system extracts its intra-footstep frequency components.

6.1.1 Step Events Detection. FootprintID detects the segments of vibration signals that are induced by footsteps and extracts them as Step Events [48]. When there is no foot striking the floor, the sensor detects the ambient structural vibrations, which we consider noises. Previous work characterizes such noise signal as Gaussian noise [48], therefore its signal energy follows a Chi-Square distribution. When a foot strikes the floor, the impact

[6]Note that in the rest of the paper, we refer to μ_{sf} and σ_{sf} as μ and σ for simplicity.

generates an impulse-like signal, which we call a Step Event. An SE contains higher signal energy than the ambient noise signals [48, 52]. Therefore, the system treats the footstep induced vibration signal as an anomaly from the noise signal and conducts anomaly detection based on the learned noise [29, 36, 48].

After it obtains the vibration signals, the system applies a sliding window to the signals. It calculates the mean μ_{wse} and standard deviation σ_{wse} of the windowed energy values of the noise signals. Then for each windowed signal, the system calculates the signal energy and compares this energy value with a threshold calculated from the noise model. If the energy value is over the threshold, the window is considered a candidate window. The consecutive candidate windows then form an SE. The threshold for anomaly detection is selected with statistical analysis to control the sensitivity of the detection algorithm as discussed in prior work. We choose the threshold of $\mu_{wse} + 3\sigma_{wse}$ to ensure a 99.7% step detection rate (because the threshold of $\mu_{wse} + 3\sigma_{wse}$ reduces the theoretical false positive rate to below 1%) [48].

6.1.2 Inter-Step Information Extraction. The system focuses on two types of inter-footstep information: the steps' relative locations to the sensor [48] and the step frequency of the walk. The step frequency is estimated based on the average time interval between consecutively detected SEs. To avoid noise affecting the step frequency estimation and causing the system to miss an SE, the system takes the mean excluding the highest and lowest K values of the time interval array within each trace ($K = 2$ in this work). The location of the step is estimated based on a heuristic tracking algorithm based on SE signal energy change trend [48]: the closer the step is to the sensor, the higher the signal energy of the SE. The details are discussed in Section 6.2 with location-based SE selection.

6.1.3 Intra-Step Information Extraction. In order to effectively represent a person's footstep, we conduct feature extraction on the selected SEs. First of all, for intra-footstep feature extraction, we normalize the signal energy to remove the footstep-sensor distance difference. From our preliminary experiments, the detected signal frequency is concentrated in the range between 0 to 200 Hz, which is sufficient to cover the characteristic frequency band of human footsteps. The frequency band is selected based on the sensor properties: the sensor's response frequency is $f_{response} \leq 240$ Hz, and the floor velocity response to footsteps has the characteristic frequency band between 20 and 90 Hz [16]. Therefore, we select the cut-off frequency at least twice as much as 90Hz and below 240Hz. In this work, for each SE, the 0 to 200 Hz power spectrum is selected as a feature to describe the characteristics of the signal.

The system then combines the inter-step and intra-step information as features in the form of an array:

$$Feature(SE) = [f_1, f_2, ..., f_N] = [SE\_freq\_domain(1Hz), ..., SE\_freq\_domain(200Hz), step\_freq] \quad (1)$$

where N = 104. The first 103 features represent frequency domain signals between 0 to 200 Hz discretized evenly. The last feature is the step frequency of the walk discussed in Section 6.1.2. For each f_{num} ($num = 1...N$), the system conducts feature normalization using the corresponding feature values from labeled training data (f_{num_train}) to achieve uniform weight through all features, i.e.,

$$f_{num\_norm} = (f_{num} - \min(f_{num\_train}))/(\max(f_{num\_train}) - \min(f_{num\_train})) \quad (2)$$

Then the system generates the SE features $\boldsymbol{x}$ by normalizing each feature in the array and using it for SE classification in Section 6.3:

$$\boldsymbol{x} = [f_{1\_norm}, f_{2\_norm}, ..., f_{N\_norm}] \quad (3)$$

.

6.2 Structural Variation Handler: Location-based SE Selection

When a pedestrian walks through the sensing area, the structural variation through the sensing area causes the same pedestrian to have SEs with different waveforms despite a consistent gait. On the other hand, when the

consistent gait is applied at the same location, the SEs share a similar waveform (Section 5). This means that even similar foot strikes are only comparable when they are from the same area.

Therefore, to ensure the SEs are comparable, we select SEs that are from approximately the same area from each trace. The overall trend of the SE energy change can be used as an indicator of their relative locations to the sensor [48]. Therefore, we use a sliding window on the SE energy values to smooth the trend change in one individual's consecutive SE sequence, which we refer to as a trace. Then, for each window, we calculate the average value of the SE energy as the representative energy of the area covers these SEs. Finally, we select the peak of this sequence of calculated value. The selected SEs within that window are from the area that is closest to the sensor. We will further discuss the structural variation and effects of location-based SE selection in the load test Section 7.1, and step location evaluation Section 8.1.2.

6.3 Step Frequency Variation Handler

One of the key challenges for gait-based person identification through structural vibration sensing is that the footstep-induced structural vibration signal is sensitive to the gait changes induced by step frequency variations. Therefore, for a pure data-driven approach, a large amount of labeled training data collected from diverse walking speeds is required to achieve robust identification. If SEs of a particular walking speed range are missing from the labeled data, the identification accuracy for SEs in that range decreases. FootprintID handles this problem by choosing between supervised learning and transductive learning based on the detected footstep step frequency. As shown in Figure 2, for a dataset with step frequencies in the labeled training data, the system applies the supervised learning model directly. Otherwise, the system chooses the transductive learning model [19]. In this paper, we adopted a Support Vector Machine based transductive learning model for three reasons: 1) SVM-based methods [8] have proved an efficient classification method for decades, 2) compared to a neural network, SVM-based methods require less labeled data in general to achieve accurate modeling [22] and in our application the data labeling has high manual cost, and 3) our prior work on person identification [52] uses SVM and therefore we selected SVM-based learning method so that it is easier to compare the system performance with the prior work.

We present `ITSVM` (marked in blue in Figure 2), an iterative transductive support vector machine algorithm, to resolve the SE variance caused by different step frequencies. In this section, we will first introduce the traditional support vector machine (SVM, Section 6.3.1) and how it handles step frequency variation between training and testing datasets. Then we introduce the transductive SVM (`TSVM`, Section 6.3.2) and how we tailor `TSVM` (Section 6.3.3, 6.3.4) to incorporate the physical properties of our data.

6.3.1 Support Vector Machine (SVM). The Support Vector Machine (SVM) has been widely used to solve classification problems. Given two-class training data $(y_1, \boldsymbol{x}_1), \ldots (y_l, \boldsymbol{x}_l)$, it aims to find the maximum-margin hyperplane $\boldsymbol{w}$ by minimizing the following loss function:

$$\min_{\boldsymbol{w}, b} \quad \frac{1}{2}||\boldsymbol{w}||^2 + C \sum_{q=1}^{l} \max(1 - y_q(\boldsymbol{w}^T \phi(\boldsymbol{x}_q) + b), 0), \tag{4}$$

where $\boldsymbol{x}_q \in R^n, \forall q$ are training samples, $y_q = \pm 1, \forall q$ are corresponding labels, and C is a penalty parameter balancing regularization term ($\frac{1}{2}||\boldsymbol{w}||^2$) and training losses ($\sum_{q=1}^{l} \max(1 - y_q(\boldsymbol{w}^T \phi(\boldsymbol{x}_q) + b), 0)$) that control the generalization of the model. The kernel function $\phi(\cdot)$ in nonlinear SVM enables us to build models with high class-separability and generalization ability, even with a small amount of training data [11, 52]. After obtaining $\boldsymbol{w}$ with some optimization techniques, the sign of function value $f(\boldsymbol{x}_q) = \boldsymbol{w}^T \phi(\boldsymbol{x}_q) + b$ can help us decide the class test data $\boldsymbol{x}_q$ belongs to.

For multi-class problems, most existing multi-class SVM methods decompose the problem into several two-class classification problems [24]. The two commonly used methods are 1) the one-against-one strategy, which trains a

model for every two classes of training data, and 2) the one-against-rest strategy, which trains a model for every class against the rest of the classes.

The traditional nonlinear multi-class SVM achieved high accuracy in identifying the participants when they walk at a specific speed during a short amount of time [52]. However, when participants walk with different step frequency, we observe significant variations of Step Events across different step frequencies, as shown in Section 5. When this variation appears between training and testing dataset, the traditional supervised learning may fail. We further validate this conjecture with empirical results in the experiments in Section 7.2.

*6.3.2 Transductive SVM (*TSVM*).* Limited training data in some pedestrian step frequencies will cause variation between training and testing datasets. To avoid this problem faced by traditional SVM (Section 6.3.1), we investigate SVM in the transductive learning settings. That is, we leverage the structure of available unlabeled data (Section 5) to help improve prediction accuracy when the distribution of the available training data is different from the test data.

Instead of finding a decision boundary with maximal margins over labeled data by SVM, TSVM aims to train a boundary with maximal margins over both labeled data and unlabeled data. Therefore, compared to the loss function in Equation 4, TSVM has an additional loss function term representing the margin over all the unlabeled data. Given two-class training data $(y_1, \boldsymbol{x}_1), \ldots (y_l, \boldsymbol{x}_l)$ and unlabeled data $\boldsymbol{x}_{l+1}, \ldots \boldsymbol{x}_{l+u}$, TSVM aims to find the maximum-margin hyperplane $\boldsymbol{w}$ and bias term b by minimizing the following loss function:

$$\begin{aligned} \min_{\boldsymbol{w},b} \quad & \frac{1}{2}||\boldsymbol{w}||^2 + C_1 \sum_{q=1}^{l} \max(1 - y_q(\boldsymbol{w}^T \phi(\boldsymbol{x}_q) + b), 0) \\ & + C_2 \sum_{q=l+1}^{l+u} \max(1 - |\boldsymbol{w}^T \phi(\boldsymbol{x}_q) + b|, 0), \end{aligned} \tag{5}$$

where C_1 and C_2 are the penalty parameters balancing regularization term and training losses over labeled data and unlabeled data, respectively. Intuitively, TSVM tends to find boundaries in regions where there is less labeled and unlabeled data. Thus, it is one kind of low-density separation method studied in transductive learning problems [12].

By drawing from our knowledge of the multi-class SVM we discussed in Section 6.3.1, we extend TSVM to handle a multi-class problem using a one-against-one or a one-against-rest strategy. For a k-class problem, the one-against-rest strategy solves k binary SVM problems, each of which treats one class as positive and all the rest as negative, and its accuracy often yields to the one-against-one strategy [21, 24]. Hence, in the following discussion of TSVM on our dataset, we focus on the one-against-one TSVM to handle the multi-class problem.

6.3.3 Refined TSVM *(*RTSVM*).* One potential problem in the one-against-one setting for multi-class TSVM is irrelevant unlabeled data. For example, in a k-class problem, when we construct the binary model for class 1 and 2, the unlabeled data is from all k classes, while only the unlabeled data from 1 and 2 is relevant to the binary classification problem between class 1 and 2. The drawbacks are twofold. First, $(k-2)/k$ percentage of the unlabeled data is irrelevant and the binary TSVM can not effectively capture the information hidden behind the unlabeled data of class 1 and 2. Second, the training of TSVM can be slow because of a large amount of unlabeled data. Note that solving TSVM is a combinatorial optimization problem, where the size of search space grows exponentially when the number of unlabeled data increases. Therefore, the increasing amount of unlabeled data may significantly reduce the training speed. In this case, a selected unlabeled dataset, which contains less unlabeled test data for modeling, leads to a faster training speed.

In order to overcome the irrelevant unlabeled data problem, we design a mechanism to refine the relevant unlabeled data for the training of each binary TSVM in the one-against-one setting we call RTSVM. In RTSVM, every

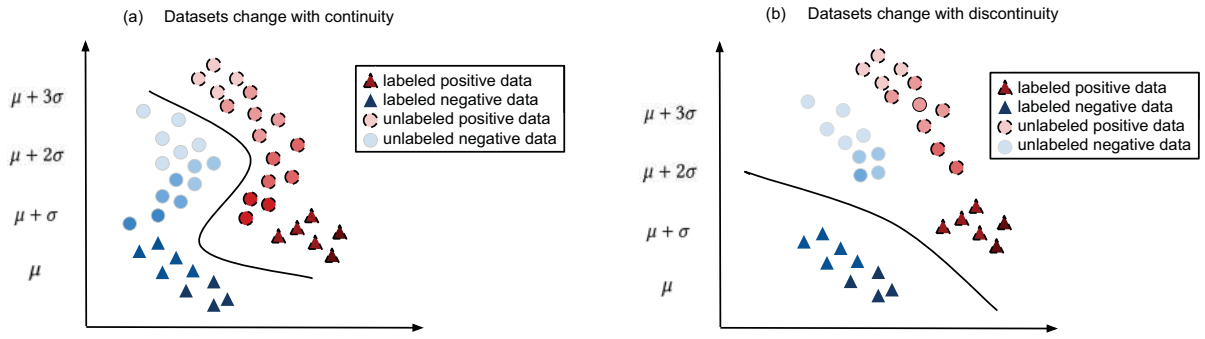

Fig. 6. The importance of continuously changing frequency in unlabeled datasets for TSVM. The triangles indicate the labeled data while the circles indicate the unlabeled data. The data points marked in blue are in the positive class while the points in red with dash outline are in the negative class. From bottom to top, the data frequency increases. (a) shows a dataset with a continuous increase and the TSVM hyperplane perfectly separating the two-class data. On the other hand, when there is one or more discontinuity in the same class data distribution, the low-density separation in TSVM may separate the data incorrectly in the regions where there are fewer points as shown in (b).

time the system trains a binary TSVM between class i and j, it utilizes supervised SVM to pre-select unlabeled SEs, which are 'most likely' to be class i or j. The multi-class SVM modeled from the labeled data is used to predict the identity of all unlabeled SEs. Then the RTSVM calculates the most frequently appearing class in each trace as the class of the trace and uses SEs from traces of class i or j for the binary TSVM modeling.

Based on the labeled SEs of class i and j and the selected unlabeled SEs, the system trains a TSVM for class i and j. After FootprintID obtains the binary TSVM for each class pair in the multi-class problem, it extends TSVM to solve the multi-class problem via one-against-one strategy. We use RTSVM as an intermediate step in our algorithm.

6.3.4 Iterative TSVM *(*ITSVM*).* One assumption in the RTSVM method is that for each one-against-one comparison, the unlabeled data and the labeled data are from the same two classes selected based on the SVM results. However, as discussed in Section 2.3, due to the step frequency variation, the individual gait varies. This causes supervised learning performance to reduce dramatically when detected step frequencies have a large gap in the labeled training data (e.g., step frequencies that are in the range of $\mu \pm 2\sigma$, $\mu \pm 3\sigma$). On the other hand, for data with step frequencies that have a relatively small gap in the labeled training data ($\mu \pm \sigma$), because the SVM classification results have higher accuracy, the RTSVM results built upon that have higher accuracy as well (we further analyze this in Section 8.1 and 8.2).

From the signal characterization in Section 5, we understand the dataset between rare and average step frequencies ($\mu \pm \sigma$) are a combination of Step Events with waveforms from those datasets (μ, $\mu \pm 2\sigma$, $\mu \pm 3\sigma$). Therefore, based on those observations, we can train our multi-class TSVM model in an iterative way, which we refer to as ITSVM:

(1) First, we construct multi-class RTSVM for the test data in the frequency of $\mu \pm \sigma$. This model can be used to label some unlabeled data in the frequency of $\mu \pm \sigma$. After this labeling procedure, the size of the 'labeled' dataset increases. Note that the 'labeled' dataset grows in a conservative way. When the RTSVM result confidence is higher than a threshold, the system labels the unlabeled data. The system calculates this confidence as the percentage of unlabeled SE's in a single trace predicted to be the same class by RTSVM. The updated labeled dataset is across the frequencies of μ and $\mu \pm \sigma$, which allows the SVM with rare step frequency data to achieve higher accuracy compared to the SVM with the original labeled dataset (μ).

(2) Second, based on the updated labeled dataset, we construct multi-class RTSVM with the test data with step frequencies of $\mu \pm 2\sigma$ and $\mu \pm 3\sigma$.

In the second step, we construct the transductive learning model with both the data of rare step frequencies ($\mu \pm 2\sigma$ and $\mu \pm 3\sigma$) and the data used in the first step (μ and $\mu \pm \sigma$). We use the data from the first step because of the low-density separation intuition behind TSVM [12], where a continuous changing of unlabeled data from different frequencies helps locate the true decision boundary. The low-density separation cases are illustrated in Figure 6. Figure 6 (a) shows the intuition data distribution of all unlabeled data (between μ and $\mu + 3\sigma$), and (b) demonstrates the intuition data distribution of rare step frequencies (between $\mu + 2\sigma$ and $\mu + 3\sigma$). If we exclude the data with frequency $\mu + \sigma$ (similarly, $\mu - \sigma$) in the unlabeled set when training TSVM, the sparsity of data in $\mu + \sigma$ region may place the TSVM's hyperplane in that region.

6.4 Trace Identity Calculation

Once FootprintID conducts identity estimation on each SE using the classification algorithm introduced above, it further combines the results of each SE in a trace to estimate the identity of each trace. Since each trace can be assigned to one person, utilizing multiple data points (SEs) would improve estimation accuracy. The FootprintID conducts a vote and selects the most frequent result as the representative of the trace ID. As we will discuss later in Section 8, the identification accuracy is based on the result of each trace instead of each SE.

Furthermore, for each trace, the system calculates the confidence of the identification decision. Then, based on the identification accuracy requirement for different applications, the system assigns the traces with confidence values lower than the application-based threshold as 'unknown'. To calculate the decision confidence, we first calculate the step-level prediction confidence, then the trace-level prediction confidence from the step-level confidence. We calculate the *step-level* prediction confidence for each SE from outputs of one-against-one SVM or ITSVM in FootprintID. For a k-class problem, a number of $k \times (k-1)$ pairwise decision confidence can be obtained. For an SE predicted as class i, there are k one-against-one (binary) SVM or ITSVM models. Each of these models predicts a class with the confidence c_j, where $j \in \{1, 2, \cdots k\}$. We calculate sample $\boldsymbol{x}$'s distance from the hyperplane $\boldsymbol{w}$ of j-th (binary) SVM or ITSVM model as $d_j = \frac{\boldsymbol{w}^T \phi(\boldsymbol{x}) + b}{|\boldsymbol{w}|}$, then we linearly normalize d_j as the confidence score c_j so that different pairwise classifiers can have confidence scores in the same range. The confidence of this data point being classified as class j is $C^{\text{step}} = \sum_{j=1}^{k} c_j$. Note that the scale range of different binary SVM or ITSVM's confidence c_j may be different for different $j \in \{1, 2, \cdots k\}$. Thus, in practice, for each binary SVM or ITSVM, we linearly normalize its prediction confidences into the same scale range $[0, 1]$. Then, we calculate the *trace-level* prediction confidence based on the *step-level* confidence of partial steps in this trace. Assume the system predicts a trace with m steps as class i after the vote, and there are n steps in this trace predicted as class i. For each step in these n steps, the corresponding confidence is C_p^{step}, where $p \in \{1, 2, \cdots, n\}$. Then the system calculates the *trace-level* decision confidence as $C^{\text{trace}} = \sum_{p=1}^{n} C_p^{\text{step}}$.

7 EXPERIMENTS

We conducted three experiments with varying parameter control levels to evaluate our FootprintID system: a load test with tennis ball-drops (Section 7.1), a human walking experiments with controlled step frequencies (Section 7.2), and an uncontrolled human walking experiments (Section 7.3). We designed the load test to investigate the sensing area's vibration response characteristics. The impulse source and location are controlled. We use the results to guide the selection of the clustering threshold introduced in Section 5. We designed the human walking experiments to investigate different system parameters and evaluate algorithm robustness. In the first walking experiment, the step frequency of each participant is controlled. The uncontrolled experiment is designed to evaluate the overall system robustness in realistic scenarios. The experiments are conducted based on the guideline approved by the CMU Institutional Review Board (IRB) review (Registration Number: IRB00000603).

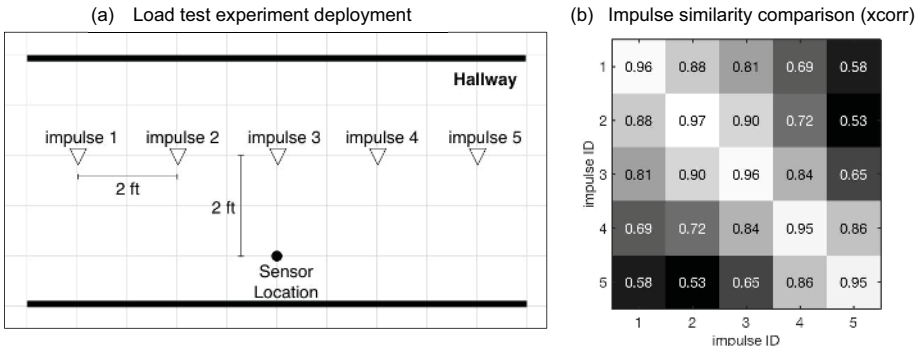

Fig. 7. Load test with ball drops. (a) Experimental setting: A sensing unit is placed on a side of the testing hallway and five impulse locations are selected as shown in the figure. Each neighbor pair has a two-foot sensor interval. (b) Similarity comparison results: The structural variation is gradual, and the nearby area (within 2 feet) has similar impulse responses.

7.1 Structural Variation: Load Test with Ball-drops

In the load test, we use ball drops (known impulse excitations) at different locations to understand location-based variation in the structural vibration responses. Human footsteps are subject to human behavior randomness. To understand the location-based variation in structural vibration responses, we employ controlled impulses generated by ball-drops from designated heights. At each location, we drop a ball 5 times from approximately 2ft above the floor to keep consistency. Therefore, when the impulse signals at different locations demonstrate variation, we can consider it the results of structural difference. Based on the structural difference-caused signal variation, we select the threshold to cluster the Step Events in the Signal Characterization Section indicating negligible differences caused by structural variation.

Figure 7 (a) shows the experimental setting for the load test, and the analysis results from the experiments are shown in Figure 7 (b). The variation between signals from the same location is relatively low (with an average cross-correlation value of 0.96) as compared to signals from different locations. In different locations, the signal change from one location to another can be large, e.g., location 1 and location 5 have an average cross-correlation value of 0.58. This means that the structural variation effects on footstep induced vibration data can be clearly observed in the area monitored by one sensor.

However, the average value of the cross-correlation between impulse signals two feet away from each other is 0.87. The signals that are within two feet of each other (the adjacent testing locations) always have over 0.84 cross correlation value, which we then use as the threshold for the Step Events clustering as we mentioned earlier. This means that the Step Events clustered together have less variation than the variation the same impulse applied a step-distance away might have from structural factors, i.e., the structural variation is negligible within a cluster.

These observations allow us to understand two situations we observed from the signal characterization in Section 5: 1) when a person walks by a sensor, their steps at different locations within the sensing range look different, even when their gait is similar and 2) if a person passes by a location multiple times, the footstep signal from different traces at locations in close proximity look similar. We will further discuss these two situations and their effects on the algorithm in Section 8.

7.2 Step Frequency: Human Test with Controlled Step Frequency

In order to test the system in a practical setting, we deployed a sensing unit in a hallway (shown in Figure 8 (a, b)) and asked the participants to walk through the hallway. 10 people (aged between 20 to 29 years) volunteered

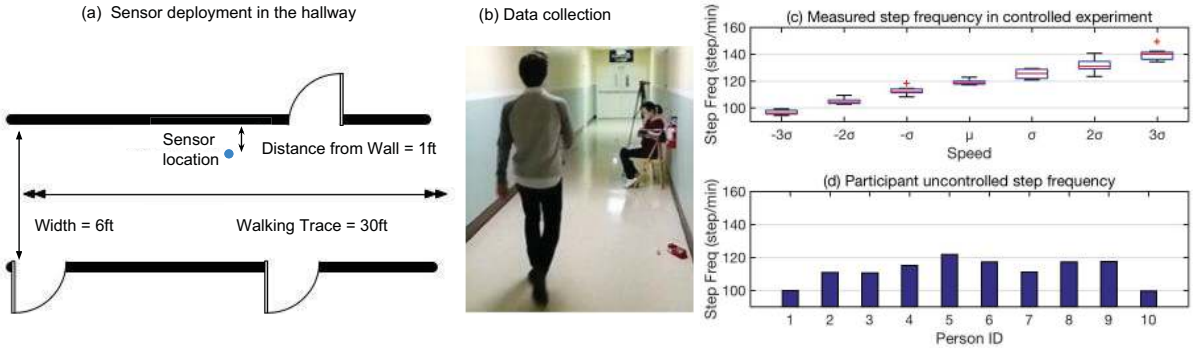

Fig. 8. The human walking experiment. (a) Experimental setting: a sensing unit is placed on the side of the hallway. (b) Data collection: human subjects are asked to walk along the hallway multiple times at the step frequency controlled by a metronome. (c) The measured step frequency in controlled experiments for the 10 participants. (d) Natural step frequency in uncontrolled experiments for the 10 participants.

to participate in the experiments, of which 8 are male and 2 are female. Their walking area is 30 ft × 6 ft. The deployed sensor can obtain between 10 to 15 steps' signals from each trace signal, and the system selects the 7 steps that are closest to the sensor from each trace for analysis.

We asked the participants to step with metronome beats to control their step frequency [44]. They were asked to wear flat bottom shoes that they are comfortable with when walking fast and for a long time. As introduced in the Signal Characteristics Section, we collected data for seven controlled step frequencies summarized in Table 1. Note that we require the participants to follow the metronome beats as consistently as possible. If the beats are too fast for them to follow while walking, they are encouraged to walk as fast as they can instead of jogging or running, because those are considered different gaits compared to walking and hence fall out of the scope of this paper.

Gender \ Step Freq	$\mu - 3\sigma$	$\mu - 2\sigma$	$\mu - \sigma$	μ	$\mu + \sigma$	$\mu + 2\sigma$	$\mu + 3\sigma$
Male	95	103	111	119	127	134	142
Female	98	107	116	125	134	143	152

Table 1. Metronome Frequencies (beats/min)

Furthermore, even when the participants are asked to follow the frequency of the metronome, their step frequency might still vary due to randomness in human behavior. We show the boxplot of detected step frequencies for SEs of each controlled step frequency in Figure 8 (c). The x-axis shows seven controlled step frequency levels corresponding to the first row in Table 1. The y-axis shows the detected step frequencies in each controlled step frequency experiment. We observe that although each step frequency level has variations, the variations are small enough that the 25th and 75th quartiles of each level do not overlap with those of others, i.e., each level is clearly distinguishable. This allows us to investigate the relation between the step frequency and the signal waveform discussed in Section 5.

7.3 Uncontrolled Experiments

We designed the uncontrolled experiments to evaluate system performance when the pedestrians walk at their natural speeds. The uncontrolled experiment was conducted the same day as the controlled experiments for

each participant. Therefore, the participant's physical condition, as well as their shoes, remain the same as the controlled experiments. They were asked to walk along the hallway in a way comfortable to them.

The experiment settings are the same as discussed in Section 7.2, except that the participants are asked to walk based on their natural step frequency. When they are asked to walk in their self-selected pace, their step frequency shows a different range as shown in Figure 8 (d). For example, Person 1's self-selected step frequency is around 100 step/min, which is a value between the controlled experiment frequency $\mu - 2\sigma$ and $\mu - 3\sigma$. As shown in Figure 5 (b), his Step Event cluster pattern for self-selected frequency is similar to that of $\mu - 2\sigma$ and $\mu - 3\sigma$, meaning at a specific step frequency, a person's gait is stable possibly due to the individual physical characteristics. The results are presented in Section 8.3.

8 RESULTS AND ANALYSIS

In this section, we analyze the results from three angles: the baseline experiments that evaluate three key factors of the system: the amount of labeled training data, the structural variation, and the step frequency variation (Section 8.1), the controlled experiments that mainly evaluate the robustness of the algorithm (Section 8.2), and the uncontrolled experiments that evaluate the overall performance of the system (Section 8.3).

8.1 Baseline: System Factors

The baseline experiments compare different training and testing data combinations with the SVM algorithm to understand the properties of the data distribution. All the baseline experiments produce the trace level identification accuracy, which is the voting result of 7 consecutive SEs from the area that is closest to the sensor, as the final results. There are three main aspects which affect the identification accuracy: the amount of the training data (Section 8.1.1), the location of the selected SEs (Section 8.1.2), and the step frequency of the training/testing datasets (Section 8.1.3).

8.1.1 Amount of the Training Data. The amount of labeled training data affects the SVM classification accuracy. The more labeled training data the system has access to, the higher accuracy the system achieves. Figure 9 (a) shows the identification accuracy of SVM when the number of labeled training data changes from 1 to 9. The system investigates three different types of accuracy: 1) step level accuracy, 2) selected step (closest to the sensor) accuracy, and 3) trace level accuracy. The results are for 10-fold cross-validation.

First, we compare the identification accuracy when the number of traces in the labeled training data increases. We observe an improving trend when the labeled training data size increases. When we only train on 1 trace, the system achieves 60% for step level accuracy, 62% for selected step accuracy, and 80% for trace level accuracy. When the training data increases to 6 traces, the step level accuracy rises to 83%, the selected step accuracy improved to 87%, and the trace level accuracy achieves up to 97%. These accuracy values show a converging trend when the amount of labeled training data is more than 6 traces per person.

The selected step shows a higher average accuracy than that of the overall steps. However, the trace level accuracy is significantly higher than both the selected step accuracy and the overall step level accuracy. The comparison between these three lines in Figure 9 (a) indicates that 1) within a trace, the SEs that are closest to the sensor achieve higher identification accuracy than the ones that are far away, and 2) by taking multiple SEs into consideration, the estimation accuracy has significant improvement. For a fixed number of traces, when the number of SEs used in a trace increases, we believe the accuracy increasing rate should gradually decrease and the accuracy value should converge to 100%, which is similar to the trend in Figure 9 (a) when the number of traces increases. We select 7 consecutive SEs in this evaluation for two reasons: 1) multiple decisions can help average the outlier decisions, compared to only select 1 specific SE to determine the identification; 2) in our experimental setup, 7 steps distance approximately covers a structural slab region, meaning the selected SEs will not show slab partition caused distortion.

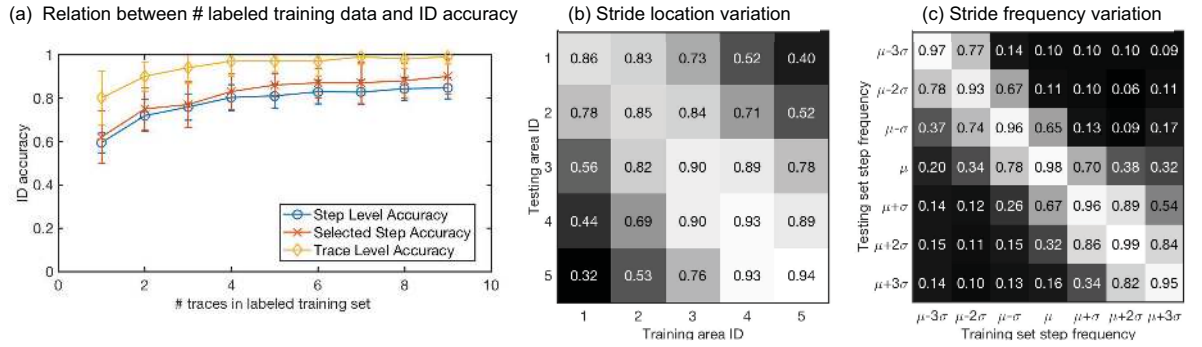

Fig. 9. Baseline experiments with SVM. (a) Identification accuracy comparison between the different amount of labeled training data in the unit of traces. The evaluated step frequency is the average step frequency μ. The blue line with circle markers indicates the step level accuracy. The red line with cross markers shows the selected (closest to the sensor) step accuracy. The yellow line with diamond markers shows the trace level accuracy. (b) Identification accuracy comparison between different sub-areas (3 SEs from each area). The diagonal accuracy values are higher than those on the side, which demonstrates the structural variation lower the identification accuracy. The diagonal values are lower than those in (a) and (c) due to the reduction in the number of SEs used in each trace. (c) Baseline step level results when training with one step frequency SEs. The x-axis is the training set step frequency, and the y-axis is the testing set step frequency. The values in the cube are the classification accuracy for 10 participants. The μ and σ are those discussed in Background Section, and the actual values are plotted in Figure 8 (c).

8.1.2 Step Location of the Training Data. The SEs' locations in the training and testing data impact the evaluation accuracy. As discussed in Section 7.1, when impulses happen in the same area, their signals show consistency in waveforms, while in different areas, their signals show differences in waveforms. Similarly, when using SEs from the same area, the identification accuracy should be higher than those from different areas due to the consistency in waveforms.

For the SEs extracted from the same relative location in each trace signal, we further separate the investigating area into five sub-areas (with 3 consecutive SEs in each area) each approximately one step away from the next area. Figure 9 (b) shows the identification accuracy when the model is trained and tested on SEs from corresponding areas. When the training and testing datasets are from the same area based on the approximate localization (the average of diagonal values in the matrix in Figure 9 (b) is 90%), the identification accuracy is higher than when the areas are distant (32% and 40% respectively). Furthermore, we observe that areas closer to Area 5 have higher identification accuracy compared to those closer to Area 1. The average accuracy in Area 4 and Area 5 is 93.5% and the average accuracy in Area 1 and Area 2 is 85.5%, meaning some areas might be more homogeneous than others and allow higher classification accuracy. To summarize, selecting the SEs from the same area for training and testing allows higher identification accuracy.

8.1.3 Step Frequency of the Training Data. The SE's step frequencies in the training data determined the testing accuracy. Figure 9 (c) shows identification accuracy for SEs trained and tested at different step frequencies respectively. This result verifies our assumption that the closer the training and testing datasets' step frequencies are, the higher the step identification accuracy is. As we observe in Figure 9 (c), the diagonal values in the matrix are the highest through all the rows, reaching an average of 96% accuracy. For each evaluated step frequency, the one level lower/higher frequency shows a slight decrease in the identification accuracy, which has an average value of 76% but still has a significantly higher accuracy than those of the step frequencies that fall outside of

the $\mu \pm \sigma$ range. Those results indicate that the identification accuracy is higher when the step frequency gap between training and test data is smaller and further motivate us to build our model in an iterative way by ITSVM.

8.2 Controlled Experiments: Algorithm Analysis

The controlled experiments evaluate the algorithm robustness especially through testing data with different step frequencies discussed in Section 7.2. We mainly focus on two aspects of the evaluation: 1) how our method outperforms traditional SVM and how each component of the algorithm works (e.g., SVM v.s. TSVM v.s. RTSVM v.s. FootprintID), and 2) how biased datasets affect the algorithm performances. We use the software SVM^{light} [26] to run SVM and TSVM in our experiments. For SVM in (4), we use the RBF kernel, where $\phi(\boldsymbol{x}_q)\phi(\boldsymbol{x}_r) = exp(-\gamma|\boldsymbol{x}_q - \boldsymbol{x}_r|^2)$. The basic parameters from SVM with the RBF kernel are $\gamma = 1$ and $C = 16$. We select this parameter combination because it achieves the highest accuracy in the 5-fold cross-validation on our labeled dataset. For TSVM in (5), we also use the RBF kernel. Similar to SVM, the key parameters γ and C_1 are 1 and 16. The weight C_2 of the loss function introduced by the unlabeled data is 1, which follows the default setting in SVM^{light}. In this case, C_1 is larger than C_2, meaning that the model gives higher weight to the loss of the labeled data than that of the unlabeled data. For ITSVM, the threshold used to determine if unlabeled SEs in a trace can be labeled is 70%. It means that when over 70% of the RTSVM results predict the same class, the system labels them as this class. The threshold is determined empirically based on when the expansion of the supervised model has greater than 70% similarity.

8.2.1 Algorithm Components Comparison. To understand the contribution of each component in the algorithm, we evaluate four different algorithms mentioned in Section 6, including SVM, TSVM, RTSVM, and ITSVM, by comparing their runtime and identification accuracy. For each person at each step frequency, we collect 10 traces. For the traces of average step frequency μ, we select 6 traces as labeled training data and the remaining 4 as testing data. For the traces of other step frequencies, we select 4 out of 10 as testing data to match the dataset size. The cross validation is conducted 10 times for each test case.

Case I: Support Vector Machine (SVM). In this case, the system only applies SVM on the labeled training data, and the results are shown as the dark blue bars in Figure 11 (a). When the model is trained on the average step frequency μ, the identification accuracy on rare step frequencies ($\mu \pm 2\sigma$ and $\mu \pm 3\sigma$) data demonstrate tremendous decrease to the same level as a random guess. The test on average step frequency μ data shows as high as 98% accuracy, while the rare step frequencies like $\mu \pm 3\sigma$ drop down to 10% and 16% accuracy. The overall accuracy on datasets of all step frequencies shows an average of 43%. As discussed in Section 8.1.3, the more similar the training and testing step frequency, the higher the identification accuracy.

*Case II: Transductive SVM (*TSVM*).* In this case, the system applies the transductive SVM without selecting unlabeled data. The results are shown in Figure 11 with light blue bars. For datasets with step frequencies within the range of $\mu \pm \sigma$, the TSVM achieves 57% accuracy, which is 20% lower than that of the SVM. For datasets with step frequencies outside the range of $\mu \pm \sigma$, the TSVM achieves similar average accuracy compared to that of the SVM, which is 17%. As discussed in Section 6.3.2, the multi-class SVM is built upon binary-class SVM, and the irrelevant cases overfit the model. Therefore, this inaccurate model reduces the accuracy of the testing identification compared to use SVM.

*Case III: Refined TSVM (*RTSVM*).* In this case, the system applies the transductive SVM and refines the binary-class TSVM by only taking the selected subset of unlabeled data. This subset of unlabeled data is selected by applying SVM and choosing the data with corresponding classes. The results are shown as light green bars in Figure 11. For datasets with step frequencies within the range of $\mu \pm \sigma$, the TSVM achieves 77% accuracy, which is similar to that of the SVM. On the other hand, for datasets with step frequencies outside the range of $\mu \pm \sigma$, the TSVM outperforms the SVM by an average increment accuracy of 5%. Since the selective unlabeled data in

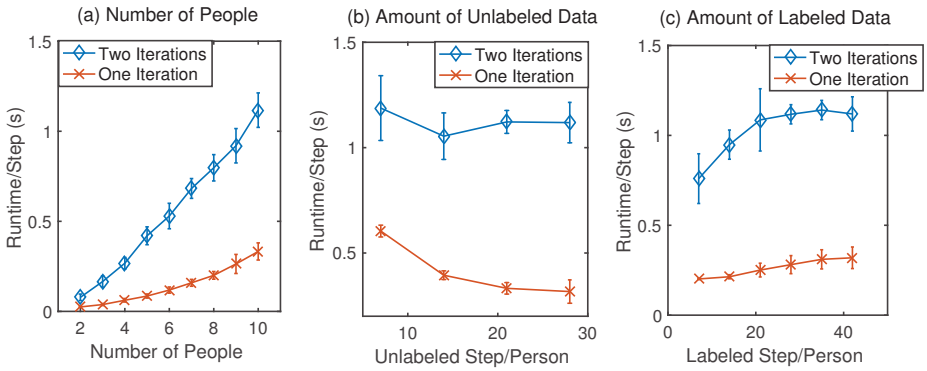

Fig. 10. Average runtime for each tested footstep. (a) shows the runtime analysis when the number of people (registered users) increases. Then for the case of 10 people, (b) shows the runtime difference when the amount of unlabeled data changes, and (c) shows the runtime when the amount of labeled data changes.

the binary-class transductive learning process has a higher precision than that in the TSVM, the final accuracy increases compared to the TSVM.

*Case IV: Iterative TSVM (*ITSVM*).* In this case, we evaluate the final settings of the presented algorithm (Figure 2). First of all, for traces with step frequencies in the range of the labeled training dataset, the system applies supervised learning (SVM) directly. Then, the system further selects the SEs with higher confidence in the first iteration of transductive learning to improve the accuracy of the additional labels. The accuracy of the rare step frequency ($\mu \pm 3\sigma$) datasets increases tremendously from 10% to 36% and 16% to 38%, approximately 2.85X average improvement. The overall accuracy on datasets of all step frequencies shows an average of 62%, 1.5X improvement compared to that of the SVM.

Algorithm Runtime Comparison. We compare the runtime of the algorithms listed above to understand their scalability. The results for 10 times cross-validation are shown in Table 2. The SVM algorithm is sensitive to the amount of labeled training data, as shown in the first two columns. When the labeled training data size increases $7X$, the runtime increases about $9.7X$. When we apply the TSVM, the labeled and unlabeled data, considered together, are of a similar size to the second case, in which the labeled training data includes all the step frequencies. When that happens, the runtime increases almost 40X compared to the second case. The RTSVM, on the other hand, decreases the runtime 5X compared to TSVM by refining the unlabeled data used. ITSVM conducts RTSVM on selected SEs and labels the SEs with high prediction confidence in the first iteration. It then conducts RTSVM on the rest of SEs in the second iteration. Therefore, ITSVM takes 2.8X the runtime compared to RTSVM.

Then for ITSVM, we further evaluate the average runtime to identify each SE, and study the effect data size has on the average runtime. To do this, we compare the average runtime per SE of the first iteration to that of the two iterations together. The runtime results are shown in Figure 10. We can see in Figure 10 (a) that

Algorithms	SVM (train on μ)	SVM (train on $\mu \pm 3\sigma$)	TSVM	RTSVM	ITSVM (FootprintID)
Runtime Avg. (s)	0.8724	9.7886	382.4303	74.7586	218.9920
Runtime Std. (s)	0.0496	0.9174	94.8886	8.3837	18.9385

Table 2. Runtime Comparison for 196 Testing Footsteps.

the average runtime increases almost linearly when the number of classes increases, and when there are two iterations, the increase ratio is almost 3 times that of the first iteration. In Figure 10 (b) the total runtime increases when the number of unlabeled data increases; however, the average runtime for each SE decreases for the first iteration and remains stable for two iterations. This is because that although the total runtime is a monotonic increasing function of the amount of unlabeled data, the average runtime for unlabeled SEs is not. The increasing or decreasing trend is determined by the ratio of labeled and unlabeled data. We will further analyze the time complexity later as a verification. On the other hand, in Figure 10 (c), for both the first iteration and both iterations together, the average runtime per SE increases with the increase of the labeled training data.

We also analyze the time complexity of different methods to validate the experimental observations. Note that the number of labeled data samples is l, and the number of unlabeled data samples is u. We further assume that the system classifies n users. Each time the system re-trains the model, the local search algorithm runs T repetitions in TSVM and runs k iterations in ITSVM. In SVM^{light}, optimizing the loss function of TSVM is a combinatorial optimization problem, which is solved by the local search algorithm. In each repetition of the local search, an SVM is trained by the quadratic programming. Therefore, the running time complexity of TSVM and RTSVM is typically $O(T(l+u)^2n^2)$. The time complexity of ITSVM is $O(kT(l+u)^2n^2)$. In contrast, the running time complexity of n-class SVM is $O(l^2n^2)$.

Based on the time complexity analysis, we further explain the runtime shown in Table 2. When compared to the n-class SVM, TSVM takes T repetitions as opposed to 1 for SVM, and the amount of unlabeled data u makes each repetition longer, therefore TSVM takes multiple times the runtime. RTSVM, on the other hand, has a smaller amount of unlabeled data compared to TSVM since its data is selected, resulting in a lower runtime. Compared to RTSVM, ITSVM has a higher runtime because ITSVM takes more iterations to train the final model. ITSVM needs less time than TSVM because ITSVM also only uses partial unlabeled data.

Although the calculation repetitions increase the transductive SVM's runtime, the algorithm saves more time in collecting the labeled data the supervised SVM needs, which can take the SVM an additional 1-2 hours per person. To achieve the identification accuracy shown in this paper, the system also needs a bootstrap phase to collect the unlabeled SEs with different step frequency, which we will introduce in the uncontrolled experiment as well as the discussion section.

Step Frequency Accuracy Comparison. We compare the identification accuracy with datasets of different step frequencies in Figure 11. Note that the step frequencies here are the 7 levels (from $\mu - 3\sigma$ to $\mu + 3\sigma$) assigned during the data collection process. We first compare the identification accuracy values of data in each step frequency. Then we further compare the identification accuracy of data in different ranges. In Figure 11 (a), we observe a tremendous drop in identification accuracy from $\mu \pm \sigma$ to $\mu \pm 2\sigma$ when supervised learning method (SVM) is used. The identification accuracy remains higher than 65% when the step frequency is between $\mu \pm \sigma$. The accuracy drops to 10% when the step frequency decreases to $\mu - 2\sigma$, and to 30% when the step frequency increases to $\mu + 2\sigma$. The TSVM and RTSVM have similar performances as the SVM at the extreme step frequencies ($\mu \pm 3\sigma$). The TSVM shows a lower accuracy compared to the SVM due to the confusion from multi-class unlabeled data, which we discussed in Section 6. The RTSVM shows the improvement compared to the SVM on the step frequencies that are not in the labeled training dataset. The ITSVM outperforms the rest of the algorithms in all evaluated step frequencies.

Figure 11 (b) presents the average identification accuracy within the step frequency range in the x-axis to demonstrate the overall accuracy when unlabeled data with different step frequencies are involved. The supervised learning model (SVM) shows up to 98% identification accuracy when the test SEs' step frequencies are closer to the labeled training data, which is higher than that of the RTSVM (93%) in the same scenario. This is because the supervised model most accurately describes the SEs with a similar step frequency to the labeled training data. Therefore, for traces in which the step frequency falls into the step frequency range of the labeled data, the

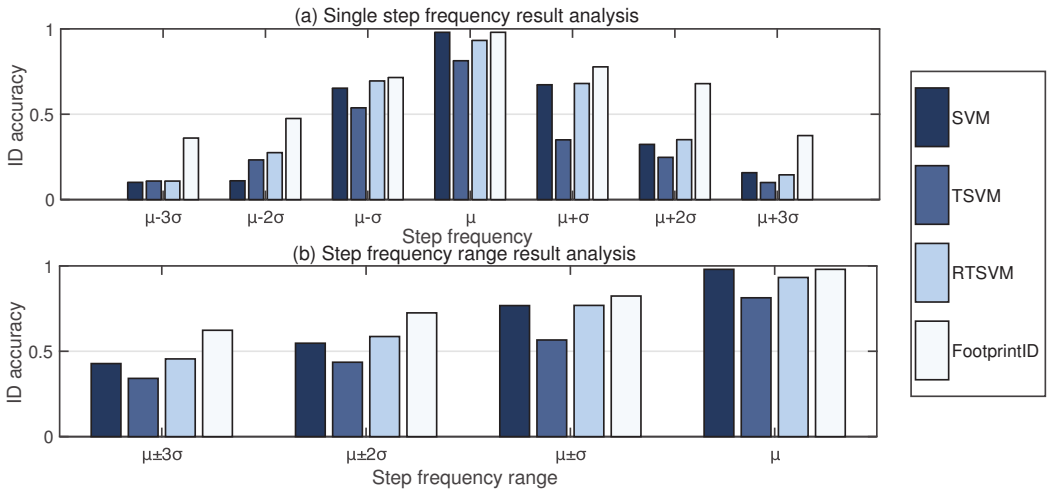

Fig. 11. Comparison between different algorithms. (a) identification accuracy when tested on signals of a specific step frequency; (b) identification average accuracy when tested on signals from a range of step frequencies. E.g., $\mu \pm 3\sigma$ in (b) indicates the range between $\mu - 3\sigma$ and $\mu + 3\sigma$, which is the average accuracy of all 7 levels of step frequency's results.

FootprintID (ITSVM) uses the supervised model and achieves the same high accuracy (98%). When the variety of the SE step frequencies increases, the advantage of our system is shown. When the test SE has a step frequency in the range between $\mu - 3\sigma$ and $\mu + 3\sigma$, the average identification accuracy for FootprintID is 62%, while that of the SVM is 43%. In general, ITSVM improves the most in comparison to SVM when the difference between testing and training data's step frequencies is larger (e.g., train on μ and test on $\mu \pm 3\sigma$).

Decision Confidence and Thresholding. In many applications, identifying users as unknown is more tolerable than making wrong predictions. Therefore, FootprintID utilizes the decision confidence calculated for each trace to determine whether to output an ID from the labeled dataset or output unknown. The *trace-level* decision confidence for a trace is calculated as discussed in Section 6.4.

Figure 12 demonstrates the *trace-level* decision confidence threshold's effect on the ITSVM results from Figure 11 (b). When the threshold is not applied, all trace decisions are considered confident and the identification accuracy is as discussed earlier in *Case IV: Iterative TSVM (*ITSVM*).* When the system increases the threshold value to 0.227, for the case of $\mu \pm 3\sigma$, the ratio of the traces labeled 'unknown' goes up to 50%, and the accuracy increases from the original 62% to 83%. When the unlabeled data range decreases, i.e., when not considering rare step frequency data, both the corresponding identification accuracy and confident ID ratio increase, as shown in Figure 12 with blue, green and purple lines. The error estimation in the rare step frequency data is of low confidence, and the thresholding effectively mitigates the error estimation.

8.2.2 Dataset with Biased Size. In this section, we discuss the system robustness when the unlabeled data used for updating the transductive learning model has biased size. This scenario could happen when the participants appear in the sensing area at different frequencies. The baseline for this experiment is SVM, and all the available traces are used. We evaluate the bias dataset effects using the accuracy reduction ratio compared to this baseline.

In general, the dataset of a specific step frequency of a participant can be insufficient or sufficient. The definition of insufficient and sufficient indicates the *degree of bias*. When the *number of people with sufficient dataset* is high,

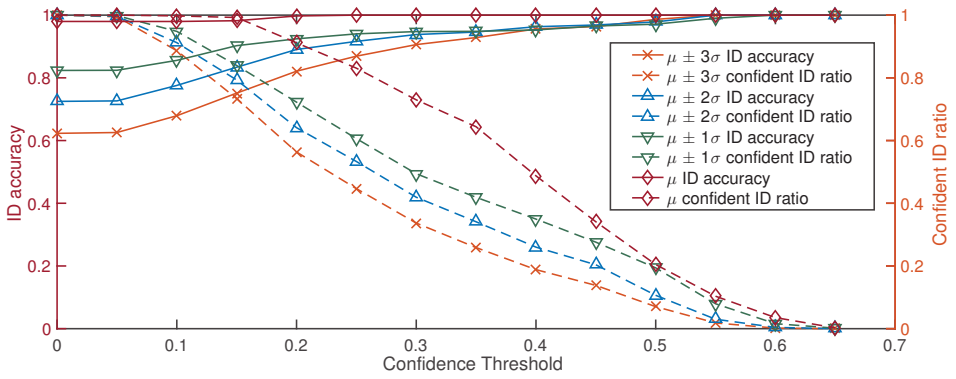

Fig. 12. Decision confidence thresholding on ITSVM results from Figure 11 (b). For example, the orange solid line with cross marker indicates the average identification accuracy of all controlled step frequencies when different threshold values are applied, and the corresponding dash line shows the confident ID ratio in that case. When the system classifies a trace with a confidence lower than the threshold, the system considers the pedestrian ID as 'unknown'. The higher the threshold, the higher the accuracy of the identified cases.

the identification accuracy should increase. When the dataset of different *step frequencies* has a different bias, e.g., only the rare step frequencies are biased, the identification accuracy varies too.

In the rest of this section, we, first of all, select the biased step frequencies based on different scenarios. Then we go through the different combination of degree of bias and the number of people with sufficient datasets.

Bias Variable I: Degree of Bias. We define the degree of bias as the definition of a sufficient and insufficient unlabeled dataset for each participant. That is, how much bias appears between the participants who show up in the sensing area most often? Here we investigate the cases when the unlabeled data ratios are 1:8, 2:7, 3:6, and 4:5 for people who show up less often (having insufficient dataset sizes) v.s. those who show up often (having sufficient dataset sizes).

Bias Variable II: Number of People with Sufficient Dataset. We consider the scenarios where some of the participants have a sufficient unlabeled dataset, while the rest have an insufficient unlabeled dataset. That is, how many people show up often (having sufficient dataset sizes)? For example, when only a small number of people have a sufficient unlabeled dataset, the system may have lower accuracy compared to when a large number of people have a sufficient unlabeled dataset.

Bias Variable III: Step Frequencies. When different step frequencies are biased differently, will the system performance be affected by the difference in the bias as well? That is, which iteration of biased datasets affects the system performance more?

Analysis. Figure 13 demonstrates the comparison of the identification accuracy regarding these three factors. As shown in both (a) and (b), when the gap between sufficient and insufficient are large (insufficient:sufficient dataset size is 1:8), the identification accuracy is lower for all cases. When even number people have insufficient and sufficient datasets, the accuracy actually drops when the degree of bias is large. This is because when the size of a dataset is too small, the estimation tends to have a higher error rate, which is expected from the analysis in Section 8.1.1.

We further analyzed the accuracy of insufficient and sufficient datasets respectively and made the following observations: 1) the sufficient dataset cases have similar accuracy; 2) for the insufficient dataset cases, the accuracy

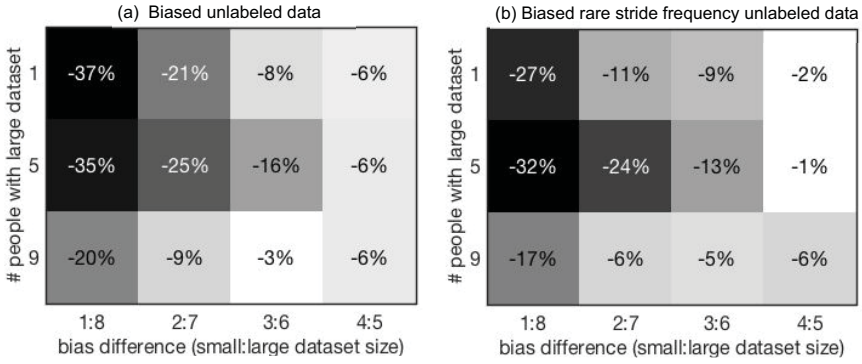

Fig. 13. Identification accuracy drop of biased datasets compared to fair datasets. (a) All step frequency unlabeled data is biased between different participants. (b) The rare step frequency unlabeled data is biased between different participants.

increases when the number of people with sufficient datasets increases. The latter happens because when the system identifies more people accurately, the error for the rest of the decisions decreases as well. Hence, due to the portion of insufficient and sufficient datasets, the second row (5 people with insufficient datasets and 5 people with sufficient datasets) has a lower error than the first row (1 person with a sufficient dataset and 9 people with insufficient datasets).

On the other hand, the third column (bias 3:6) in the third row (1 person with an insufficient dataset and 9 people with sufficient datasets) outperforms the second (bias 2:7) and the fourth column (bias 4:5) in the third row. This nonmonotonous trend could be caused by a combination effect of accuracy increase for insufficient datasets and accuracy decrease for sufficient datasets. As discussed in the baseline evaluation in Section 8.1.1, when the training data increases to 6 or more traces, the trace level accuracy remains high and does not change much. However, as the training data increases from 1 to 5, the trace level accuracy demonstrates a significant increase as shown in Figure 9 (a). In the situation we observe here, the accuracy from the population with a sufficient dataset may remain high till the dataset size drops to 5, while the accuracy from the population with an insufficient dataset keeps increasing. The ratio combination of the population and their accuracy decrease/increase makes the trend nonmonotonous.

When comparing the cases in which all unlabeled datasets are biased by a person (Figure 13 (a)) v.s. only the rare step frequencies dataset are biased by a person (Figure 13 (b)), the latter achieves higher accuracy especially when the degree of bias is high and the number of people with sufficient datasets is low. This is because when the first iteration of datasets is sufficient, the estimation has higher accuracy and therefore the new labeled data has a label with higher accuracy, hence the model describes the data more accurately.

When the dataset is biased, we can improve the identification accuracy by tuning the confidence threshold. As discussed in Figure 12, when the unlabeled dataset is large enough, the threshold used to select additional labels from the RTSVM results can be increased. The more accurate the first iteration model is, the more accurate the final results will be, as shown in the comparison between cases in Figure 13 (a) and (b).

8.3 Uncontrolled Experiments: Algorithm Robustness

We further evaluate the system with uncontrolled experiments where pedestrians walk through the sensing area one at a time with their natural walking form. As discussed in Section 7.2, when the participants are asked to walk with their natural form, their step frequencies vary as shown in Figure 8. When using ITSVM, the system determines the step frequency range that applies supervised SVM as between 116 and 122 steps/min, which is

Models	SVM	TSVM	RTSVM	ITSVM (FootprintID)
labeled: μ, unlabeled: $\mu \pm \sigma$, $\mu \pm 2\sigma$, $\mu \pm 3\sigma$, uncontrolled	56%	52%	52%	67%
labeled: μ, unlabeled: $\mu \pm \sigma$, $\mu \pm 2\sigma$, $\mu \pm 3\sigma$	56%	54%	51%	67%
labeled: μ, unlabeled: uncontrolled	50%	22%	22%	45%

Table 3. Runtime Comparison for 196 Testing Footsteps.

the step frequency range of the labeled data with walking speed μ shown in Figure 8 (c). We investigate the system practicality especially in terms of system bootstrapping from historical data with the combination of data from both controlled and uncontrolled experiments.

8.3.1 System Bootstrapping and Modeling. There are three modeling settings we investigated and listed in the table: 1) the system takes the labeled data with step frequency μ and unlabeled historical data with step frequency between $\mu \pm 3\sigma$ as well as the uncontrolled data to build the model, 2) the system takes the labeled data with step frequency μ and unlabeled historical data with step frequency between $\mu \pm 3\sigma$ to build the model, and 3) the system takes the labeled data with step frequency μ and the uncontrolled data to build the model. These three cases represent different scenarios of the system. The first modeling case represents the scenario when the system is in the bootstrap phase and there is enough historical data. In the bootstrap phase, the system rebuilds the model after collecting new unlabeled data. The second modeling case represents the scenario after the bootstrap phase. In the post-bootstrap phase, the system only applies the model; it doesn't rebuild it. The third modeling case represents the system in the bootstrap phase where there is not enough historical data yet.

8.3.2 Comparison and Analysis. We run the four algorithms compared in the controlled experiments, SVM, `TSVM`, `RTSVM` and `ITSVM`, on the uncontrolled experiment data and compare the average identification accuracy in Table 3. We can see that for the first two modeling cases, the `ITSVM` outperforms the other three algorithms. This

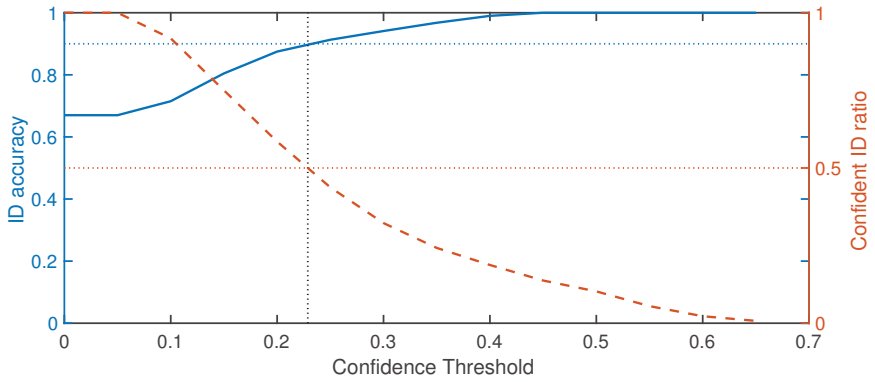

Fig. 14. The accuracy of the uncontrolled experiment in which people walk at their natural frequency. The x-axis is the decision confidence threshold value. The y-axis on the left (corresponding to the blue solid line) is the identification accuracy of the results after applying the threshold. The y-axis on the right (corresponding to the red dash line) is the ratio of the cases where the system successfully obtains an ID based on the decision confidence threshold. When the threshold discards half of the data (black dash line), the identification accuracy achieves 90%.

is because the detected step frequencies of the collected uncontrolled traces are spread out between $\mu - 3\sigma$ and $\mu + \sigma$. Therefore, the improvement in the ITSVM is mainly due to the higher identification accuracy on the low step frequency data. We can also see that when the model covers a large enough step frequency range, the extra unlabeled data does not necessarily increase the identification accuracy (67% for the first and the second model). In addition, when there is not enough gradually changing historical data like that shown in Figure 6, the system may end up with a lower accuracy when compared to SVM (45% for the third model).

In addition, we took the results from the first modeling case and analyzed the 'unknown thresholding' changing and its effects on the identification accuracy using ITSVM. When the system considers the results only when the probability is beyond a pre-defined threshold, and otherwise the result of the identification is "unknown" [52], we have the corresponding identification accuracy demonstrated in Figure 14. As shown in the figure, when FootprintID thresholds on half of the high confidence trace, the identification accuracy rises from 67% to 90%.

9 DISCUSSION

We further discuss system practicality and potential extension here. For system practicality, we mainly discuss the runtime comparison and the system bootstrap phase, the sensing range, and the robust footstep detection against other types of excitation induced vibration. For system extension, in addition to the step frequency discussed in this paper, there are many other factors that await to be explored. Here we describe how sensor deployment location, different shoes people are wearing, and the number of people walking at the same time may impact our work.

9.1 System Practicality

We discuss the system practicality in three different ways, including the system runtime and bootstrap phase, the sensing range, and the robust footstep detection method.

9.1.1 System Runtime and Bootstrap Phase. In this section, we discuss our model training method's practicality based on the analysis on the system runtime in Section 8.2 and the system bootstrap in Section 8.3. As discussed in the runtime evaluation section, although the amount of unlabeled data and the calculation iteration increase the runtime, labeling data is the most time-consuming step in the system. It also shows that the runtime increases significantly when the unlabeled data increases because the size of search space grows exponentially when the number of unlabeled data increases. Furthermore, with the analysis in the uncontrolled experiments, when the size and distribution of the historical data grow to a certain degree, adding more data may not lead to a higher identification accuracy anymore. Therefore, the key to the system practicality lies in the control of the bootstrap phase regarding duration and unlabeled data size. For example, when the system is deployed, it will have a one-month bootstrap phase in addition to the labeled data collection. During the bootstrap phase, the system collects the unlabeled SEs and selects them based on their step frequencies to retrain the model iteratively. Then after the bootstrap phase, the system identifies the person based on the fixed model. Note that in this paper we systematically evaluate the controlled walking step frequency between $\mu \pm 3\sigma$ based on the survey paper [44]. However, when the pedestrian's walking step frequency exceeds the range of $\mu \pm 3\sigma$, the system will conduct more iterations so the model covering range can propagate until it includes those extreme step frequencies.

9.1.2 Sensing Range. The sensing range of each sensor is affected by multiple factors that mainly fall into two categories: structure and excitation signal. The structural factors include structural slab region sizes, material properties (stiffness, density, etc.), sensor locations on a slab region, etc. The excitation signal factors include the impact strength, which is affected by pedestrian physical characteristics, shoe material, analog amplifier gain, etc. We addressed the sensing range in our prior work [48, 52] with experiments done in hallways. The detection range can be as far as the 10m radius from the sensor [48, 52].

9.1.3 Robust Footstep Detection. In practical scenarios, there are many types of excitations that generate vibrations other than footsteps, such as doors opening and closing, objects falling, a chair dragging, etc. Our prior work has analyzed different types of excitation and the vibration signals they induce from both time and frequency domains [29, 36]. This work utilizes the one-class SVM to distinguish impulses that are induced by footsteps from other types of excitation when the signals are not overlapping. With this robust footstep detection, the system can apply the algorithm introduced in this paper on footstep induced vibration signals.

9.2 System Extension

We further discuss the variations that may affect the system in practical scenarios here, including step frequency change during a walk, sensor location variation, different shoe types, and scenarios with multiple pedestrians walking together.

9.2.1 Irregularity within a Walk. When a person walks in an indoor environment, their walking may differ from that discussed in this paper, e.g., their step frequency may change every few steps, they may walk away from the sensor instead of passing by it, etc. When these less predictable behaviors happen, our current system with only one sensor deployed may not be able to identify the person (e.g., not enough footstep detected). However, when the system adopts multiple sensors to achieve a larger coverage, the predictable behavior may be detected elsewhere. Hence the system can conduct the identification with the settings introduced in this paper.

For example, when a person's step frequency is continuously changing, the system can use the sensing area with most stable step frequency to identify the person. Similarly, if a person walks away from one sensor, they may walk pass by the other sensor in the next sensing area. However, when multiple sensors are involved, their location variation may become another challenge, which we will discuss in the following section.

9.2.2 Sensor Location Variation. When the sensing system is deployed at different locations, the vibration signal of the same person walking could look different due to structural variation. That is to ask if the system can identify an individual passing by one sensor with the training data collected from another sensor at a different location. In this paper, we assumed a small deployment area, such as a hallway, where each sensor can be trained individually. In scenarios with larger sensing areas and multiple sensors, we believe a transfer function that maximizes the similarity between Step Event features extracted by sensors at different locations could be extracted.

9.2.3 Shoe Types. In a practical scenario, each individual may have multiple shoes. When the same person wears different shoes, their footstep-induced vibration signal changes. There are two intuitive reasons behind this effect: 1) different shoes change a person's gait [75], and 2) different shoe soles are made of different materials with varying hardness, resulting in impulse responses of different magnitude, frequency, and duration [23].

We envision two major methods to handle the shoe type difference challenge: 1) modifying the model to achieve a shoe-invariant model, and 2) to extend the model for individuals that includes multiple shoe profiles. For the first method, we may revisit the feature selection to see if there are 1) shoe-invariant features from each Step Event, and 2) behavior level features from the obtained trace signal, e.g., left-right foot imbalance. For the second method, we may utilize the tracking information of the person to link the identity to the office or routine paths, which, in turn, can be used to label different shoe types and expand the existing models of individuals.

9.2.4 Multiple Pedestrian Scenario. In real-world situations, it is common for multiple people to walk within a sensing area at the same time. When this happens, footsteps of multiple people induce the floor to vibrate and the sensor will observe the mixture of these signals. Some works have started to explore the detection of multiple people walking together [51]. Based on the occupant traffic estimation, further signal separation can be explored. In addition, signal separation, localization, and identification are closely connected, and utilizing this information obtained concurrently may help the system achieve more accurate estimation in each task.

10 RELATED WORK

FootprintID is inspired by different research topics, including gait kinetic, person identification sensing methods, and structural vibration monitoring. This section discusses works that are closely related to this work.

10.1 Gait Kinetics

Kinetics of human gaits usually refers to the study on 1) forces passing through the joints, 2) force plate embedded in the floor records, and 3) ground reaction force vectors. Studies have shown that the force that a human subject applies to the ground can be used for identification of animals and human beings [62]. Further investigation on the ground's reaction to the force of indoor human footsteps also indicates the possibility of identifying pedestrians through the measurement of the force applied on the floor [33, 47]. The excitation force induces shape changing on the pressured part of the floor and vibrations of the structure [4]. The prior work focuses either on the gait's ground reaction force or the structural vibration induced by an excitation force, and the connection between these two are missing. We focus on the footstep-induced vibration sensing, which 1) allows sparse deployment due to the wave propagation properties in structure and 2) captures the gait kinetic information for identification purposes.

10.2 Sensing Methods Comparison

Various of sensing methods have been explored to obtain person identities in the indoor environment. They mainly fall into a few categories, including vision [49, 61, 65, 68], RF [69, 76], mobile [38, 64], inertial sensors [17, 32, 56], acoustic [2, 20], and vibration [16, 52] based methods.

Vision-based methods extract visual biometrics of an individual, including facial structure [5, 65], hand geometry [49, 57], and gait [61, 68]. Although the accuracy of the identification is high (over 95% accuracy for the mentioned biometrics) the systems often require a clear visual path. This makes it a challenge to utilize them in a surveillance scenario or ubiquitous sensing without requiring them to interact with a sensing system.

On the other hand, RF-based methods overcome such problems and have been explored to obtain a variety of information about people in an indoor environment [69, 76]. WiWho achieves an accuracy of 80% for identifying 6 people [76]. WiFiU achieves a false acceptance rate of 8.05% and a false rejection rate of 9.54% [69]. These methods often require the sensing target to be between a pair of transmitters, therefore they need dense deployment to achieve high accuracy information inquiry. Furthermore, such data-driven approaches usually require a large labeled training set, which makes the deployment of the system a challenge.

Mobile-based methods [38, 64] utilize the relation between the sensing target and the device with a unique ID and assume each target will carry the registered device. Another way to identify pedestrians with mobile devices uses inertial sensor based methods [17, 32, 56], which is through the gait information reflected on the motion. The recognition rates reported in these work are around 85% (Equal Error Rate between 6.7% and 7.3%). However, these mobile-based methods often assume that there are wearable devices on the pedestrian to obtain the gait information. Therefore, it might be difficult to use when deployed in applications such as elderly care or patient care, in which people may not carry their phones all the time.

Acoustic- and vibration-based methods have their similarities [16, 52]. Compared to acoustic sensing, vibration sensing is less sensitive to the audible noise in the environment. However, both are data-driven approaches and require a large amount of the labeled training data to achieve robust identification, especially when the walking speed varies [7]. In this paper, we presented our system FootprintID, a vibration-based method that takes physical insights into account and allows robust identification with a limited amount of labeled training data.

[7]We compare the accuracy of supervised SVM and FootprintID in Section 8.

10.3 Structural Vibration Sensing

Work has been done on structural vibration monitoring through sensor networks for various applications, including structural health monitoring [28, 40–43, 74]. The insight on these works provides a good understanding of the wave propagation and characteristics of the human-induced structural vibration. Recent research also starts to utilize such signals for indoor human monitoring [29, 30, 34, 35, 48, 54]. The sensing challenges of structural vibration monitoring for indirect human sensing include 1) noisy ambient noise [30], 2) multi-path of the vibration signal in an indoor environment [54], and 3) the distortion and low resolution of the sensing signal [50]. We adopted the solution proposed in our previous works to acquire high-resolution signals and avoid signal distortions. These prior works do not address the identification purpose utilizing structural vibration, whereas in this paper we present a robust identification method. A preliminary work explained the possibility of utilizing floor vibration for pedestrian identification purpose [52]. Compared to that, this work presents a low-cost sparse sensing system that can robustly identify pedestrians under different types of variations as discussed earlier in the Signal Characterization Section.

11 CONCLUSION

In this paper, we presented the FootprintID system, which identifies pedestrians passing through the sensing area through their footstep induced structural vibrations. Our algorithm overcomes the large amount of labeled training data needed for a data-driven approach to achieve pedestrian identification under various sensing conditions, especially variations in stepping frequency and structure. We characterized the variation of footstep-induced structural vibration signals and designed the `ITSVM` learning algorithm to achieve robust pedestrian identification for signals of rare step frequency. We conducted experiments with 10 people who each walked at seven different step frequencies. When the system is trained only on the average step frequency μ dataset, `ITSVM` achieves up to 96% accuracy, 1.5X improvement with test datasets of all step frequencies, and approximately 3X improvement on rare step frequency ($\mu \pm 3\sigma$) dataset when compared to traditional SVM. Furthermore, `ITSVM` demonstrates approximately 1.5X improvement on identification accuracy in uncontrolled experiments.

REFERENCES

[1] Jan Achenbach. 2012. *Wave propagation in elastic solids*. Vol. 16. Elsevier, Atlanta, GA, USA.

[2] David T Alpert and Martin Allen. 2010. Acoustic gait recognition on a staircase. In *World Automation Congress (WAC), 2010*. IEEE, New York, USA, 1–6.

[3] Thomas P Andriacchi, Seungbum Koo, and Sean F Scanlan. 2009. Gait mechanics influence healthy cartilage morphology and osteoarthritis of the knee. *The Journal of Bone & Joint Surgery* 91, Supplement 1 (2009), 95–101.

[4] Hugo Bachmann. 1992. Case studies of structures with man-induced vibrations. *Journal of Structural Engineering* 118, 3 (1992), 631–647.

[5] Jezekiel Ben-Arie and Dibyendu Nandy. 1998. A volumetric/iconic frequency domain representation for objects with application for pose invariant face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20, 5 (1998), 449–457.

[6] Chiraz BenAbdelkader, Ross Cutler, and Larry Davis. 2002. Person identification using automatic height and stride estimation. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, Vol. 4. IEEE, New York, USA, 377–380.

[7] Richard Bono. 2008. Transducer Mounting and Test Setup Configurations. In *IMAC XXVI TUTORIAL*. SEM, Orlando, Florida USA.

[8] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. 1992. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*. ACM, 144–152.

[9] Nikolaos V Boulgouris, Dimitrios Hatzinakos, and Konstantinos N Plataniotis. 2005. Gait recognition: a challenging signal processing technology for biometric identification. *signal processing magazine, IEEE* 22, 6 (2005), 78–90.

[10] Alessandra Carriero, Amy Zavatsky, Julie Stebbins, Tim Theologis, and Sandra J Shefelbine. 2009. Correlation between lower limb bone morphology and gait characteristics in children with spastic diplegic cerebral palsy. *Journal of Pediatric Orthopaedics* 29, 1 (2009), 73–79.

[11] Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)* 2, 3 (2011), 27.

[12] Olivier Chapelle, Bernhard Scholkopf, and Alexander Zien. 2009. Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]. *IEEE Transactions on Neural Networks* 20, 3 (2009), 542–542.

[13] Meng-Jung Chung and Mao-Jiun J Wang. 2010. The change of gait parameters during walking at different percentage of preferred walking speed for healthy adults aged 20–60 years. *Gait & posture* 31, 1 (2010), 131–135.
[14] Federal Facilities Council, National Research Council, and others. 1974. *Expansion Joints in Buildings: Technical Report*. National Academies Press, Washington, DC, USA.
[15] Frederic Danion, E Varraine, Mireille Bonnard, and Jean Pailhous. 2003. Stride variability in human gait: the effect of stride frequency and stride length. *Gait & Posture* 18, 1 (2003), 69–77.
[16] Alexander Ekimov and James M Sabatier. 2006. Vibration and sound signatures of human footsteps in buildingsa). *The Journal of the Acoustical Society of America* 120, 2 (2006), 762–768.
[17] Davrondzhon Gafurov, Einar Snekkenes, and Patrick Bours. 2007. Gait authentication and identification using wearable accelerometer sensor. In *Automatic Identification Advanced Technologies, 2007 IEEE Workshop on*. IEEE, Alghero, Italy, 220–225.
[18] James R Gage, Peter A Deluca, and Thomas S Renshaw. 1995. Gait analysis: principles and applications. *J Bone Joint Surg Am* 77, 10 (1995), 1607–1623.
[19] Alexander Gammerman, Volodya Vovk, and Vladimir Vapnik. 1998. Learning by transduction. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., Madison, Wisconsin, USA, 148–155.
[20] Jürgen T Geiger, Maximilian Kneißl, Björn W Schuller, and Gerhard Rigoll. 2014. Acoustic gait-based person identification using hidden Markov models. In *Proceedings of the 2014 Workshop on Mapping Personality Traits Challenge and Workshop*. ACM, Istanbul, Turkey, 25–30.
[21] King-Shy Goh, Edward Chang, and Kwang-Ting Cheng. 2001. SVM Binary Classifier Ensembles for Image Classification. In *Proceedings of the Tenth International Conference on Information and Knowledge Management (CIKM)*. ACM, Atlanta, GA, USA, 395–402.
[22] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. MIT Press. http://www.deeplearningbook.org.
[23] William G Halvorsen and David L Brown. 1977. Impulse technique for structural frequency response testing. *Sound and Vibration* 11, 11 (1977), 8–21.
[24] Chih-Wei Hsu and Chih-Jen Lin. 2002. A comparison of methods for multiclass support vector machines. *IEEE transactions on Neural Networks* 13, 2 (2002), 415–425.
[25] Input/Output, Inc. 2006. *SM-24 Geophone Element*. Input/Output, Inc. Rev. 3.
[26] T. Joachims. 1999. Making large-Scale SVM Learning Practical. In *Advances in Kernel Methods - Support Vector Learning*, B. Schölkopf, C. Burges, and A. Smola (Eds.). MIT Press, Cambridge, MA, Chapter 11, 169–184.
[27] Stephen C Johnson. 1967. Hierarchical clustering schemes. *Psychometrika* 32, 3 (1967), 241–254.
[28] Sukun Kim, Shamim Pakzad, David Culler, James Demmel, Gregory Fenves, Steven Glaser, and Martin Turon. 2007. Health monitoring of civil infrastructures using wireless sensor networks. In *Information Processing in Sensor Networks, 2007. IPSN 2007. 6th International Symposium on*. IEEE, Cambridge, Massachusetts, USA, 254–263.
[29] Mike Lam, Mostafa Mirshekari, Shijia Pan, Pei Zhang, and Hae Young Noh. 2016. Robust Occupant Detection Through Step-Induced Floor Vibration By Incorporating Structural Characteristics. In *IMAC XXXIV A Conference and Exposition on Structural Dynamics*. Vol. 54. SEM, Orlando, Florida USA, 1–13.
[30] Heyoung Lee, Jung Wook Park, and Abdelsalam Sumi Helal. 2009. Estimation of indoor physical activity level based on footstep vibration signal measured by MEMS accelerometer for personal health care under smart home environments. *Control Instrum* 5801, 5 (2009).
[31] Jennifer L Lelas, Gregory J Merriman, Patrick O Riley, and D Casey Kerrigan. 2003. Predicting peak kinematic and kinetic parameters from gait speed. *Gait & posture* 17, 2 (2003), 106–112.
[32] Jani Mantyjarvi, Mikko Lindholm, Elena Vildjiounaite, S-M Makela, and HA Ailisto. 2005. Identifying users of portable devices from gait pattern with accelerometers. In *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on*, Vol. 2. IEEE, Philadelphia, PA, USA, ii–973.
[33] Lee Middleton, Alex Buss, Alex Bazin, Mark S Nixon, and others. 2005. A floor sensor system for gait recognition. In *Automatic Identification Advanced Technologies, 2005. Fourth IEEE Workshop on*. IEEE, Buffalo, NY, USA, USA, 171–176.
[34] Mostafa Mirshekari, Shijia Pan, Adeola Bannis, Yan Pui Mike Lam, Pei Zhang, and Hae Young Noh. 2015. Step-level person localization through sparse sensing of structural vibration. In *Proceedings of the 14th International Conference on Information Processing in Sensor Networks*. ACM, Seattle, WA, USA, 376–377.
[35] Mostafa Mirshekari, Shijia Pan, Pei Zhang, and Hae Young Noh. 2016. Characterizing wave propagation to improve indoor step-level person localization using floor vibration. In *SPIE Smart Structures and Materials+ Nondestructive Evaluation and Health Monitoring*. International Society for Optics and Photonics, SPIE, Las Vegas, NV, USA, 98034–98034.
[36] Mostafa Mirshekari, Pei Zhang, and HaeYoung Noh. 2017. Calibration-Free Footstep Frequency Estimation using Structural Vibration. In *IMAC XXXV A Conference and Exposition on Structural Dynamics*. SEM.
[37] Friedrich Moser, Laurence J Jacobs, and Jianmin Qu. 1999. Modeling elastic wave propagation in waveguides with the finite element method. *Ndt & E International* 32, 4 (1999), 225–234.

[38] Le T Nguyen, Yu Seung Kim, Patrick Tague, and Joy Zhang. 2014. IdentityLink: user-device linking through visual and RF-signal cues. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, Seattle, WA, USA, 529–539.
[39] Jens Bo Nielsen. 2003. How we walk: central control of muscle activity during human walking. *The Neuroscientist* 9, 3 (2003), 195–204.
[40] H Noh, R Rajagopal, and AS Kiremidjian. 2013. Sequential structural damage diagnosis algorithm using a change point detection method. *Journal of Sound and Vibration* 332, 24 (2013), 6419–6433.
[41] Hae Young Noh, Dimitrios G Lignos, K Krishnan Nair, and Anne S Kiremidjian. 2012. Development of fragility functions as a damage classification/prediction method for steel moment-resisting frames using a wavelet-based damage sensitive feature. *Earthquake Engineering & Structural Dynamics* 41, 4 (2012), 681–696.
[42] Hae Young Noh, K Krishnan Nair, Anne S Kiremidjian, and CH Loh. 2009. Application of time series based damage detection algorithms to the benchmark experiment at the National Center for Research on Earthquake Engineering (NCREE) in Taipei, Taiwan. *Smart Structures and Systems* 5, 1 (2009), 95–117.
[43] Hae Young Noh, Ram Rajagopal, and Anne S Kiremidjian. 2012. Damage diagnosis algorithm using a sequential change point detection method with an unknown distribution for damage. In *SPIE Smart Structures and Materials+ Nondestructive Evaluation and Health Monitoring*. International Society for Optics and Photonics, SPIE, San Diego, CA, USA, 834507–834507.
[44] T Öberg, Alek Karsznia, and K Öberg. 1993. Basic gait parameters: reference data for normal subjects, 10-79 years of age. *Journal of rehabilitation research and development* 30 (1993), 210–210.
[45] Tommy Oberg, Alek Karsznia, and Kurt Oberg. 1994. Joint angle parameters in gait: reference data for normal subjects, 10-79 years of age. *Journal of rehabilitation Research and Development* 31, 3 (1994), 199–213.
[46] Sandra J Olney, Malcolm P Griffin, and Ian D McBride. 1994. Temporal, kinematic, and kinetic variables related to gait speed in subjects with hemiplegia: a regression approach. *Physical therapy* 74, 9 (1994), 872–885.
[47] Robert J Orr and Gregory D Abowd. 2000. The smart floor: a mechanism for natural user identification and tracking. In *CHI'00 extended abstracts on Human factors in computing systems*. ACM, The Hague, The Netherlands, 275–276.
[48] Shijia Pan, Amelie Bonde, Jie Jing, Lin Zhang, Pei Zhang, and HaeYoung Noh. 2014. BOES: building occupancy estimation system using sparse ambient vibration monitoring. In *SPIE Smart Structures and Materials+ Nondestructive Evaluation and Health Monitoring*. International Society for Optics and Photonics, SPIE, San Diego, CA, USA, 90611O–90611O.
[49] Shijia Pan, An Chen, and Pei Zhang. 2013. Securitas: user identification through rgb-nir camera pair on mobile devices. In *Proceedings of the Third ACM workshop on Security and privacy in smartphones & mobile devices*. ACM, ACM, Berlin, Germany, 99–104.
[50] Shijia Pan, Mostafa Mirshekari, Hae Young Noh, and Pei Zhang. 2015. Structural sensing system with networked dynamic sensing configuration. In *Proceedings of the 14th International Conference on Information Processing in Sensor Networks*. ACM, ACM, Seattle, WA, USA, 344–345.
[51] Shijia Pan, Mostafa Mirshekari, Pei Zhang, and Hae Young Noh. 2016. Occupant Traffic Estimation through Structural Vibration Sensing. In *SPIE Smart Structures and Materials+ Nondestructive Evaluation and Health Monitoring*. International Society for Optics and Photonics, SPIE, Las Vegas, NV, USA, 98035–98035.
[52] Shijia Pan, Ningning Wang, Yuqiu Qian, Irem Velibeyoglu, Hae Young Noh, and Pei Zhang. 2015. Indoor person identification through footstep induced structural vibration. In *Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications*. ACM, Santa Fe, NM, USA, 81–86.
[53] Shijia Pan, Susu Xu, Mostafa Mirshekari, Pei Zhang, and Hae Young Noh. 2017. Collaboratively Adaptive Vibration Sensing System for High Fidelity Monitoring of Structural Responses Induced by Pedestrians. *Frontiers in Built Environment* 3 (2017), 28.
[54] Jeffrey D Poston, Javier Schloemann, R Michael Buehrer, VVN Sriram Malladi, Americo G Woolard, and Pablo A Tarazaga. 2015. Towards indoor localization of pedestrians via smart building vibration sensing. In *Localization and GNSS (ICL-GNSS), 2015 International Conference on*. IEEE, Gothenburg, Sweden, 1–6.
[55] Boris I Prilutsky and Robert J Gregor. 2001. Swing-and support-related muscle actions differentially trigger human walk–run and run–walk transitions. *Journal of Experimental Biology* 204, 13 (2001), 2277–2287.
[56] Liu Rong, Duan Zhiguo, Zhou Jianzhong, and Liu Ming. 2007. Identification of individual walking patterns using gait acceleration. In *Bioinformatics and Biomedical Engineering, 2007. ICBBE 2007. The 1st International Conference on*. IEEE, Wuhan, China, 543–546.
[57] Raul Sanchez-Reillo. 2000. Hand geometry pattern recognition through gaussian mixture modelling. In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, Vol. 2. IEEE, Barcelona, Spain, Spain, 937–940.
[58] Mary Sansalone, Nicholas J Carino, and Nelson N Hsu. 1987. A finite element study of transient wave propagation in plates. *Journal of research of the National Bureau of Standards* 92, 4 (1987), 267–278.
[59] Pamela L Sheridan, Judi Solomont, Neil Kowall, and Jeffrey M Hausdorff. 2003. Influence of executive function on locomotor function: divided attention increases gait variability in Alzheimer's disease. *Journal of the American Geriatrics Society* 51, 11 (2003), 1633–1637.
[60] Martin L Smith and FA Dahlen. 1973. The azimuthal dependence of Love and Rayleigh wave propagation in a slightly anisotropic medium. *Journal of Geophysical Research* 78, 17 (1973), 3321–3333.
[61] Budi Sugandi, Hyoungseop Kim, Joo Kooi Tan, and Seiji Ishikawa. 2009. Real time tracking and identification of moving persons by using a camera in outdoor environment. *Int. J. Innov. Comput. Inf. Control* 5, 5 (2009), 1179–1188.

[62] David H Sutherland. 2005. The evolution of clinical gait analysis part III–kinetics and energy assessment. *Gait & Posture* 21, 4 (2005), 447–461.
[63] Rawesak Tanawongsuwan and Aaron Bobick. 2003. *A study of human gaits across different speeds*. Technical Report. Georgia Tech, Tech. Rep.
[64] Thiago Teixeira, Deokwoo Jung, and Andreas Savvides. 2010. Tasking networked cctv cameras and mobile phones to identify and localize multiple people. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*. ACM, Copenhagen, Denmark, 213–222.
[65] Y-I Tian, Takeo Kanade, and Jeffrey F Cohn. 2001. Recognizing action units for facial expression analysis. *IEEE Transactions on pattern analysis and machine intelligence* 23, 2 (2001), 97–115.
[66] Stephen Timoshenko, Stephen Prokofevich Timoshenko, Stephen Prokofevich Timoshenko, and Stephen Prokofevich Timoshenko. 1956. *Strength of materials*. Vol. 210. D. Van Nostrand Company, Inc., New York, USA.
[67] Joe Verghese, Richard B Lipton, Charles B Hall, Gail Kuslansky, Mindy J Katz, and Herman Buschke. 2002. Abnormality of gait as a predictor of non-Alzheimer's dementia. *New England Journal of Medicine* 347, 22 (2002), 1761–1768.
[68] Liang Wang, Tieniu Tan, Huazhong Ning, and Weiming Hu. 2003. Silhouette analysis-based gait recognition for human identification. *IEEE transactions on pattern analysis and machine intelligence* 25, 12 (2003), 1505–1518.
[69] Wei Wang, Alex X Liu, and Muhammad Shahzad. 2016. Gait recognition using wifi signals. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 363–373.
[70] David A Winter. 1984. Kinematic and kinetic patterns in human gait: variability and compensating effects. *Human Movement Science* 3, 1 (1984), 51–76.
[71] Chenren Xu, Bernhard Firner, Robert S Moore, Yanyong Zhang, Wade Trappe, Richard Howard, Feixiong Zhang, and Ning An. 2013. Scpl: Indoor device-free multi-subject counting and localization using radio signal strength. In *Information Processing in Sensor Networks (IPSN), 2013 ACM/IEEE International Conference on*. IEEE, Philadelphia, USA, 79–90.
[72] Jaynie F Yang and Monica Gorassini. 2006. Spinal and brain control of human walking: implications for retraining of walking. *The Neuroscientist* 12, 5 (2006), 379–389.
[73] P Constance Yang, Charles H Norris, and Yehuda Stavsky. 1966. Elastic wave propagation in heterogeneous plates. *International Journal of solids and structures* 2, 4 (1966), 665–684.
[74] Hae Young Noh, K Krishnan Nair, Dimitrios G Lignos, and Anne S Kiremidjian. 2011. Use of wavelet-based damage-sensitive features for structural damage diagnosis using strong motion data. *Journal of Structural Engineering* 137, 10 (2011), 1215–1228.
[75] Lee Yung-Hui and Hong Wei-Hsien. 2005. Effects of shoe inserts and heel height on foot pressure, impact force, and perceived comfort during walking. *Applied ergonomics* 36, 3 (2005), 355–362.
[76] Yunze Zeng, Parth H Pathak, and Prasant Mohapatra. 2016. WiWho: wifi-based person identification in smart spaces. In *Proceedings of the 15th International Conference on Information Processing in Sensor Networks*. IEEE Press, Vienna, Austria, 4.
[77] Wei-Shi Zheng, Shaogang Gong, and Tao Xiang. 2011. Person re-identification by probabilistic relative distance comparison. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, Colorado Springs, CO, USA, USA, 649–656.

Received February 2017; revised May 2017; accepted July 2017