Occupant traffic estimation through structural vibration sensing

Conference Paper · April 2016

DOI: 10.1117/12.2222024

Shijia Pan
Carnegie Mellon University

Mostafa Mirshekari
Carnegie Mellon University

Occupant Traffic Estimation through Structural Vibration Sensing

Shijia Pan[1*] Mostafa Mirshekari[2] Pei Zhang[1] Hae Young Noh[2]

[1]Dept. of Electrical and Computer Eng., Carnegie Mellon Univ., Pittsburgh, PA, USA, 15213;
[2]Dept. of Civil and Environmental Eng., Carnegie Mellon Univ., Pittsburgh, PA, USA, 15213

ABSTRACT

The number of people passing through different indoor areas is useful in various smart structure applications, including occupancy-based building energy/space management, marketing research, security, etc. Existing approaches to estimate occupant traffic include vision-, sound-, and radio-based (mobile) sensing methods, which have placement limitations (e.g., requirement of line-of-sight, quiet environment, carrying a device all the time). Such limitations make these direct sensing approaches difficult to deploy and maintain. An indirect approach using geophones to measure floor vibration induced by footsteps can be utilized. However, the main challenge lies in distinguishing multiple simultaneous walkers by developing features that can effectively represent the number of mixed signals and characterize the selected features under different traffic conditions.

This paper presents a method to monitor multiple persons. Once the vibration signals are obtained, features are extracted to describe the overlapping vibration signals induced by multiple footsteps, which are used for occupancy traffic estimation. In particular, we focus on analysis of the efficiency and limitations of the four selected key features when used for estimating various traffic conditions. We characterize these features with signals collected from controlled impulse load tests as well as from multiple people walking through a real-world sensing area. In our experiments, the system achieves the mean estimation error of ± 0.2 people for different occupant traffic conditions (from one to four) using k-nearest neighbor classifier.

Keywords: structural vibration, smart structure, indirect sensing, occupant traffic, signal mixture

1. INTRODUCTION

Occupant traffic information (i.e., the number of people passing by) for designated areas in buildings is useful for many smart building applications, such as occupancy estimation based energy management, indoor resource/space management, market research, etc. We gain this information through a variety of sensing methods, mainly device-based[5, 6, 11] or device-free.[13, 16, 23, 34, 38–40, 44, 45] Device-based methods utilize the devices that people carry with them (e.g., smart phone, GPS, etc.) to detect the presence of people, and thus they require subjects to carry the sensing devices with them all the time. However, in many scenarios, it is often impractical to rely on the subjects to bring their devices and provide access. Device-free techniques refer to the method that does not require the monitored subject to carry a sensing device. They are a better fit to our sensing goal, since we want to monitor occupants ubiquitously without any requirements on human interaction or carrying a device. However, these techniques, such as surveillance camera, motion sensors on the doorway, etc., often require intrusive installation of sensors at designated areas/positions, which make them costly to deploy on a large scale (e.g., installation at each doorway, line-of-sight, etc.).

More recent approaches utilize occupant footstep induced floor vibrations to infer pedestrians,[19, 21, 22, 27, 29] which can also be used as the indicator of the occupant traffic. Monitoring the occupant traffic indirectly through structural vibrations overcomes the installation barrier and sensing limitations faced by traditional device-free techniques. This approach has been shown feasible in other applications, including person identification,[29] person tracking/locating[19, 21, 22] and person-by-person occupancy estimation.[17, 27] However, prior work has focused on only single person walking scenarios, which may not be assumed in general.[17, 19, 21, 27, 29]

In this paper, we present a method to monitor multiple occupant traffic through sensing the ambient structural vibration. Our system achieves occupant traffic monitoring by acquiring signals from structural vibration sensors

* shijiapan@cmu.edu; phone: 650-335-2821

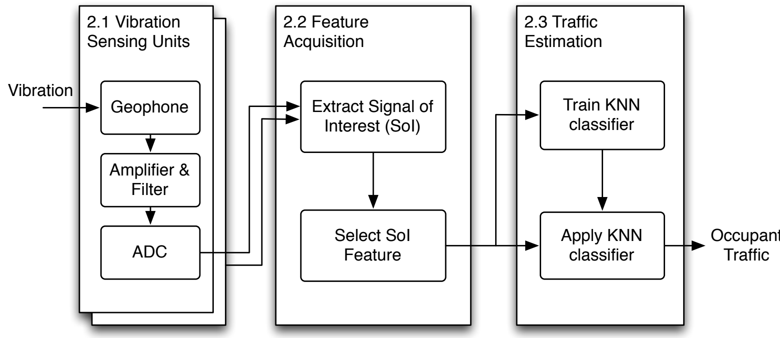

Figure 1. System overview.

and analyzing their features. Based on the study on interpersonal distance in group walking, when multiple people walk as a group, they tend to fragment in smaller units of one, two, or three members.[7] Therefore, fine-grained occupant traffic estimation for one to three people in a walking group is critical for a larger group of occupant traffic estimation. The major challenge faced by footstep-induced vibration sensing techniques is the complexity of signal mixing when multiple people walk through the sensing area in different ways. Different occupant traffic conditions cause different feature characteristics. We solve this problem by analyzing different signal mixing conditions and characterizing proper features for different traffic conditions. The contributions of this paper include:

- We analyze the characteristics of the four selected key features and present their effectiveness to infer the number of footsteps from induced floor vibration signals.
- We explored fine-grained occupant counting from one- to four-person groups with these selected features.
- We designed and conducted a series of field experiments to verify our signal modeling through both impact loading and human walking in a commercial building.

The rest of the paper is organized as follows: Section 2 introduces the method and apparatus, by which we acquire signals, extract features, and obtain the occupant traffic information. Next, in Section 3, we evaluate the system through experiments of multiple traffic conditions and analyze each feature's effectiveness. We then organize related work and analyze the pros and cons of our system in Section 4. Finally, we provide our conclusion in Section 5.

2. OCCUPANT TRAFFIC MONITORING SYSTEM

Our occupant traffic monitoring system consists of three components: the vibration sensing module (Section 2.1), the feature acquisition module (Section 2.2), and the traffic estimation module (Sectin 2.3). Figure 1 shows relations between different modules in the system: The vibration sensing units collect structural vibration signals at different locations and feed the signals to the feature acquisition module. Then the feature acquisition module pre-processes the collected vibration signals by marking the vibration signal induced by footsteps. The selected features are extracted from these footstep-induced signals and sent to the traffic estimation module. The traffic estimation module then processes the signal from each sensor, integrates the information from different sensors, and presents the number of occupants walked by.

2.1 Vibration Sensing

The vibration sensing unit consists of a geophone, amplifiers and filters, and analog-to-digital converters (ADCs). These units are placed on the floor to obtain structural vibrations through geophones. Geophones are used as vibration sensors in our system, which convert the velocity of the structure to voltage.[14] The geophones are selected instead of other vibration sensors because it is a low cost sensor and sensitive to the frequency range we focus on (around 0-100 Hz), which leads to high signal-to-noise ratio (SNR). The analog signal is amplified through the amplification module, and then digitized by ADC modules for further analysis. The effective sensing range (after amplification with a 10-bit ADC) in a commercial building setting is around 20 m diameter. We utilize a pair of sensing units with partially overlapping sensing areas to detect occupant traffic at each designated monitoring area. Each pair is then synchronized through wireless connection.

2.2 Feature Acquisition

In order to acquire signal features for occupant traffic estimation, the vibration signal induced by occupants' footsteps needs to be separated from the collected vibration signals. We refer to this as the signal of interest (SoI), which is explained in details in Section 2.2.1. The system then extracts the selected features from these SoIs for further occupant number estimation as discussed in Section 2.2.2. These features reflect the signal characteristics under different occupant traffic conditions.

2.2.1 Signal of Interest (SoI) Extraction

The SoI is defined as the ambient structural vibration signal induced by occupant footsteps. On the other hand, the vibration signal sensed by the system when there is no impulse applied on the structural within the system's sensing range is considered as noise.

Compared to one person walking scenario, when multiple people walk together, their footstep induced signal might overlap, which is referred to as signal mixture henceforth. We categorize the signal mixture conditions when multiple people walk together into three possible cases: 1) steps from different people are completely synchronized, in terms of the strike timing; 2) steps from different people are off-sync but induced vibration signals have temporal overlapping; and 3) steps are temporally staggered, therefore individual footstep signals are not overlapping. From our observation during the experiments, we notice that the case 1 and case 3 rarely happen due to the randomness in human motions. The degree of overlapping in case 2 varies and reflects the number of people walking together, which we will explain through the coming sections.

To extract the SoIs from a continuously detected ambient structural vibration signal, i.e., separate the SoIs from the noise, we utilize the anomaly detection method[17, 27] from our previous work. The anomaly detection method models the noise signal as Gaussian noise and detects signals that fall out of three standard deviations (3σ) from the mean (μ) as SoIs. A sliding window is used to segment the signal, and the signal energy of each window is calculated. The anomaly detection is conducted based on the signal energy of each sliding window.

However, the previous method[17, 27] extracts *fixed-length* step events, which is not suitable for signal mixture of multiple people. When their steps are not completely synchronized, the energy of the sliding window may stay high for a longer period due to the signal mixture and variation of signal offsets. In that case, the extracted signal using the fixed-length window may not contain the entire signal mixture. Therefore, in this work, the system keeps detecting the consecutive windows with high energy until the signal energy drops below the threshold level and combines all the detected windows together as a single SoI.

2.2.2 Feature Selection for Occupant Counting

Four features are selected for the system to estimate the occupant count, including cross-correlation between signals collected by sensors at different locations induced by the same footstep, cross-correlation between signals induced by consecutive footsteps collected by the same sensor, SoI duration, and SoI entropy. These features are selected based on 1) the understanding of the structural vibration signal attenuation,[27] 2) the assumption of linear addition mixture model for structural vibrations,[19] and 3) the understanding of the human gait consistency.[29] We focus on exploring the underlying principles for features and what would be the effective scenario to use such feature. In addition, the scenario where more than two people walking is explored, including the relation between the number of people and the features.

- **Cross-correlation between SoIs from Different Sensors for the Same Footsteps.**

  The normalized cross-correlation between SoIs from different sensors for the same footsteps, which later we refer as **spatio-different SoIs**, reflects the divergence of the signal attenuation observed by those sensors. When multiple footstep signals overlap (synchronized or have an offset less than the length of a footstep signal) within the sensing area, their distances to different sensors vary. Therefore, the attenuation of each footstep signal at different sensors are different. Due to the linear addition assumption, when these signals add up, the similarity between mixed signals observed by different sensors will be low. Therefore, the cross correlation between the spatio-different SoIs can be used for inferring the occupant count.

Assume there are n impulses, $S_1...S_n$ happening at approximately same time, and the observations at the two sensors are

$$\begin{cases} O_1 = a_{1,1}S_1 + a_{1,2}S_2 + ... + a_{1,n}S_n \\ O_2 = a_{2,1}S_1 + a_{2,2}S_2 + ... + a_{2,n}S_n \end{cases}$$

where $O_i, i = 1, 2$ are signals observed by sensors, $S_j, j = 1, 2, ..., n$ are signal impulse sources (i.e., footsteps), and $a_{i,j}, i = 1, 2; j = 1, 2, ..n$ is the signal decay coefficient. Ideally, when the **number of impulses** n increases, there is a higher chance that $a_{1,i} \neq a_{2,i}$, where $i = 1, 2, ...n$. This causes the signal observed by different sensors to exhibit less similarity, and therefore lower normalized cross correlation value. The relation between $a_{1,i}$ and $a_{2,i}$ is determined by the **impulse location** relative to different sensors.

- **Cross-correlation between SoIs for Consecutive Footsteps from the Same Sensor.**

  The normalized cross-correlation between different SoIs of the same sensor, which later we refer to as **temporal-different SoIs** reflects the consistency of waveforms, discarding the attenuation factor, induced by footsteps/impulses from a continuous footstep sequence (within a trace) and obtained by one sensing unit. When one person walks by a sensor, their footstep induced vibration signals are consistent (high cross correlation value).[29] However, when multiple people pass by, the inconsistent temporal offset between mixed signals and the inconsistent signal attenuation rate can break such consistence between the footstep signals. Therefore, cross correlation between the temporal-different SoIs can be used to infer the number of occupants. For each trace, the detected SoIs are compared pairwise (i.e., left foot signals are compared to left foot signals, same for the right foot signals). Since the SoIs can be of different length, for this feature, we find the highest peak within each SoI and then redefine the SoI as the signal with the specific length before and after the peak.

- **SoI Duration.**

  SoIs are extracted signal segments that contains multiple consecutive sliding windows of signal whose signal energy is anomaly compared to sliding window of noise signals. Different paces or number of people can cause overlapping signals, resulting in elongated event duration and increase in SoI duration variation, as indicated by the average and standard deviation. With the SoI detection algorithm introduced in Section 2.2.1 which extracts an SoI through multiple sliding windows, we define SoI duration as the length of these sliding windows added up, which means the sliding window size determines the SoI duration resolution.

- **SoI Signal Entropy.**

  Signal randomness is another feature we explore. The randomness can be quantified as the degree of order/disorder associated with a multi-frequency signal response,[32] which is carried by the wavelet entropy

  $$E(s) = -\sum_i s_i^2 \log(s_i^2)$$

  where s is the signal and s_i is the relative wavelet energy of the signal at resolution level i using orthogonal discrete wavelet transform of s.

  The noise signal is expected to be most random (i.e., high entropy). When there is only one impulse source, the randomness of detected signal will be low. However, when the number of mixing footstep induced signals is large enough, their distribution addition will converge to the Gaussian distribution by central limit theorem (CLT), which is used to model the noise. Therefore, we expect to see higher entropy when the number of mixing impulse increases.

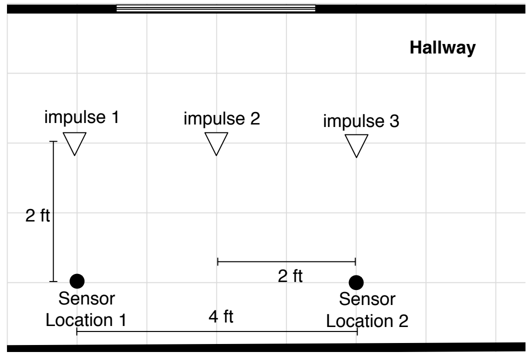

Figure 2. Ball drop experiment setting. The diagram indicates the relative locations of geophones (circles) and ball drops (triangles). The grid dimensions are $1' \times 1'$.

2.3 Traffic Estimation

Once the features are extracted, the system analyzes them to estimate the occupant traffic. i.e., the number of occupant passing by the sensing area. To estimate occupant number from the features, we conducted a prototyping classification using k-nearest neighbours.[8] The k-nearest neighbours is selected because it reflects the distribution of selected features directly. The classification procedure is equivalent to the following steps: 1) find the k neighbour points in the training set that are nearest to testing set; 2) find the labels of these neighbour points; and 3) assign the classification label based on the majority vote of its neighbors.[8]

The classification model for each case is learned in the training phase with the extracted feature set $F = [f_1, f_2, f_3, f_4]$, where f_i is the i^{th} feature value for each SoI detected from the vibration signals. Then we classify the testing cases by predicting the classification label using the k-nearest neighbor classifier model. The results and the conclusions we draw from the results are discussed in Section 3.

3. EVALUATION

To evaluate our system design, we conducted two types of experiments: 1) feature validation experiments implemented by the controlled impulse load test (Section 3.1), and 2) scenario analyses with footstep induced vibration signals collected when occupants were asked to walk through a hallway with designated number of occupant within the group (Section 3.2). Here we investigate the number of impulses/occupants from one to three, which is sufficient based on the research in group walking behavior that shows when multiple people walk together, they tend to fragment in smaller units of one, two, or three members.[7] The experiments are conducted based on the guideline approved by the CMU Institutional Review Board (IRB) review (Registration Number: IRB00000603).

3.1 Feature Validation through Load Test

Before conducting experiments on human, we conducted impulse load test with ball drops to understand the signal mixture conditions and their effects on the selected features. Impulse load tests with ball-drops often generate more stable vibration signals comparing to signals generated by human footsteps. In this section, we first introduce the load test experimental setup in Section 3.1.1, and then analyze the feature changes through different test conditions in Section 3.1.2.

3.1.1 Experimental Setup

Our experiments were conducted with a high resolution oscilloscope, which samples at 10000 Hz. The experiment setting is shown in Figure 2 with two sensors put four feet from each other allowing synchronized data collection by the oscilloscope. Here we concentrate on three different conditions where the number of simultaneous ball-drops are respectively one (impulse 1), two (impulse 1 & 3), and three (impulse 1 & 2 & 3) to demonstrate the feature changing trend when the number of mixed impulse signals increases. For each investigating case, we collected five ball-drops. The ball-drops are controlled by metronome at the speed of 45 impulses/min to achieve the impulses synchronization as much as possible. We investigate the selected features through these three conditions to understand the features' effectiveness under different signal mixture conditions.

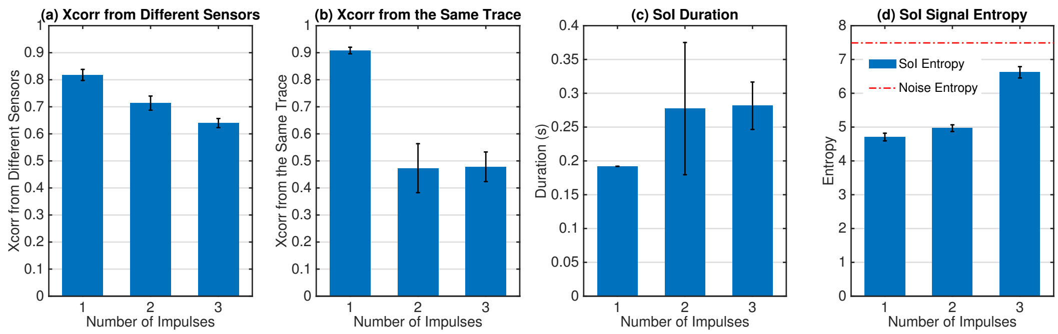

Figure 3. Impulse load test through three different traffic conditions. Error bars represents 1 standard deviation from the mean. Different number of impulses are applied to the sensing area with controlled rhythm by a metronome at 45/min. It can be observed that the normalized cross correlations (spatio and temporal) decreases while duration and entropy increase with the increasing number of impulses.

3.1.2 Feature Analysis

To understand each feature's effectiveness, we compare the features from three traffic conditions separated in Figure 3. The blue bars in Figure 3 display the average values of the features and the error bars indicate one standard deviations. The sub-figures (a) to (d) display results for cross correlation between spatio-difference SoIs, cross correlation between temporal-difference SoIs, SoI duration, and SoI signal entropy, respectively.

Normalized Cross Correlation between Spatio-different SoIs The impulses at different locations introduce variation in the signal mixture. This variation in signal mixture can be observed by different sensors. Figure 3 (a) shows the bar graph of the mean (respectively 0.82, 0.71, and 0.64) and the standard deviation (respectively 0.02, 0.03, and 0.02) of the results. The figure displays the decreasing trend of the normalized cross correlation value with increasing number of impulses. Cross correlation is therefore a feasible method for distinguishing between different numbers of impulses when their vibration signals mix. This suggests that the feature is useful when the observed footsteps have different locations relative to different sensors.

Normalized Cross Correlation between Temporal-different SoIs The smaller the occupant number is, the higher chance the footstep-induced signals have higher similarity. When multiple footsteps applied on the floor without perfect synchronization, their additivity at different phase will cause the temporal different SoI to decrease. Even when multiple impulses are applied to the floor at the same pace, it is difficult if not impossible to have the step perfectly synchronized (i.e., $10^{-3}s$ for 1000 Hz sampling rate). Figure 3 (b) demonstrated the comparison of the results from the impulse load test, with values 0.91, 0.47, and 0.48 respectively. The result shows a clear reduction when there is more than one impulse sources. This suggests that the feature is effective for distinguishing traffic conditions between one-impulse test v.s. multiple-impulse test.

SoI Duration The more impulses mix, the higher the maximum impact time offset two signals may have, leading to larger value of the SoI duration. Figure 3 (c) plots the average (0.19, 0.27 and 0.28 seconds) and standard deviation (0, 0.1, and 0.04 seconds) of SoI duration corresponding to different number of impulses generated following the metronome. The mean value of the SoI duration increases from one-impulse test to multiple-impulse test. This suggests that the SoI duration is useful for distinguishing one person walking v.s. multiple people walking. It can also be observed that the two-impulse test has the higher variation in duration than the three-impulse test. This could be because that the three-impulse test is performed after the two-impulse test. Therefore, people synchronize better after practicing from performing the two-impulse test. This suggests that SoI duration is sensitive to the synchronization of impulses. When the SoI duration is high, there is high probability that there are multiple people walking together, but when the SoI duration is low, we need to refer to other features to determine the number of people.

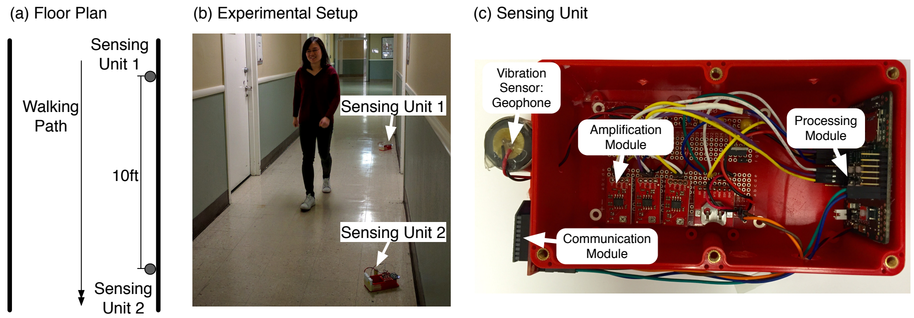

Figure 4. Sensing group deployment. Two sensing units are placed 10 ft apart from each other along the hallway. Participants are asked to walk along the hallway in a manner that is most comfortable for them.

SoI Entropy The more impulses mix, the more "complicated" the mixed signal will be. Figure 3 (d) shows that the entropy of a segment of noise signal is 7.49, higher than that of the SoIs. The average SoI entropy values for tests with respectively one, two, and three impulses at the same time are 4.7, 5.0, and 6.62, with the standard deviations of 0.11, 0.1, and 0.17. The results show a clear increasing trend in SoI entropy when the number of the impulses increase. This could be caused by the blending of multiple footstep impulse signals, which makes the mixture impulse more similar to the noise (the maximum randomness). The more footstep impulses mix together, the more random the signal will be. The results indicate that increasing number of impulses will yield higher randomness, meaning entropy can be used to infer the number of impulses in occupant traffic conditions. This feature performance may degrade for SoIs with low SNR because the noise in the signal may dominate the entropy value.

3.2 Scenario Analysis through Human Test

The human walking test aimed at testing the system in real occupant walking scenarios. In this section, we first of all introduce the experimental setting with human subjects walking through the sensing area in Section 3.2.1. Then we analyze and compare the characteristics of the features demonstrated in human walking tests in Section 3.2.2. At last, we display traffic estimation results with different traffic conditions in Section 3.2.3, and explain the observations with the impulse load tests results from Section 3.1. Considering the randomness in human walking test is more than that in impulse load test, we extended the number of occupant to four to verify the change trend of the features we observed in impulse load tests.

3.2.1 Experimental Setup

In our setup, we arranged sensing unit 1 and unit 2 (described in Figure 4 (c)) with a distance of 10 ft in a hallway as shown in Figure 4 (a) and (b). We choose such distance so that 1) in our testing location, the two sensing units will be close enough to be on the same floor plan, hence observing the same signal without other complicated structural effect such as beams; and 2) when an occupant walks by the sensors, the sensing units are far enough so that their footstep energy attenuation at each sensor varies. The amplification values are set empirically so that when multiple people pass by each sensor at a distance of two feet, their footstep induced structural vibrations do not clip. The sampling rate is set to 1000 Hz. When occupants walk by the sensor group, the footstep induced vibration signal containing sequences of SoIs is called a *trace*. The features are extracted for each SoI within a trace, and values from the same trace are averaged for classification. For each scenario, we ask people to walk pass through the sensing area eight times. When there are multiple people walking as a group, they are asked to walk in a manner that is most comfortable for them.

3.2.2 Feature Analysis

To evaluate the selected features in terms of occupant counting ability, we conducted experiments in the hallway with an assigned number of occupants walking in the same direction as a group with their natural walking patterns. When they walk as a group, the distance between each individual is less than three feet. When two people walk together, they are walking side-by-side with a similar pace. When three people walk together, one

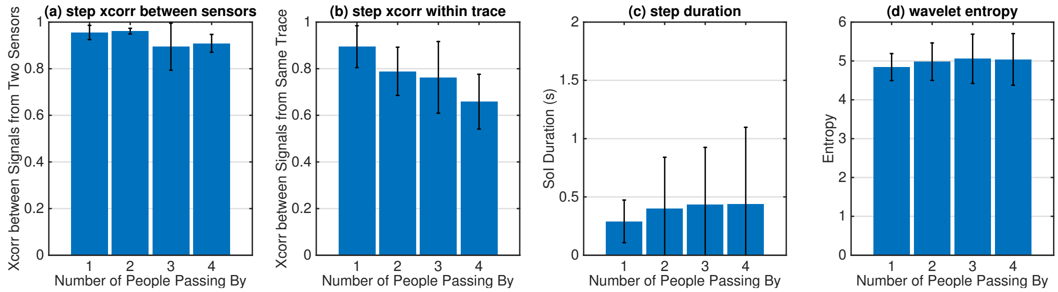

Figure 5. Human walking test through four different traffic conditions. Error bars represent 1 standard deviation from the mean. Different number of people were asked to walk with their natural speed as a group. Normalized cross correlations (spatio and temporal) decreases with increasing occupant number, while duration and entropy increase.

of the person walked slightly behind the other two due to the limited width of the hallway. When four people walk together, two of them walked behind the other two approximately two feet away.

The cross-correlation between spatio-different SoIs reflects the signal mixture variation caused by different footstep locations. Figure 5 (a) displays the values of normalized cross correlation between spatio-different signals in different scenarios, whose averages over 8 traces are 0.96, 0.96, 0.89, and 0.91. The results show a general decrease trend when compare three or four occupant traffic condition to the one or two occupant traffic condition. This verifies the effectiveness of this feature when the footstep locations relative to different sensors vary. When two people walk together side-by-side (instead of one after another like the ball drop experiment setup in Section 3.1), their relative location to different sensors are similar, therefore the cross-correlation between spatio-different SoIs may not be able to distinguish the case clearly.

The cross correlation between temporal-different SoIs reflects the synchronization between different footsteps. The values of normalized cross correlation between temporal-different SoIs of different occupant traffic conditions are shown in Figure 5 (b) with average values of 0.89, 0.79, 0.76, and 0.66, demonstrating a decreasing trend. As discussed in Section 3.1, this feature is effective for distinguishing one-impulse v.s. multiple impulses, which is demonstrated here as a clear drop from one-footstep to multiple-footsteps conditions.

The SoI duration also reflects the synchronization between different footsteps. The SoI durations from human walking tests are shown in Figure 5 (c) with average values (over 8 traces for each traffic condition) of 0.29, 0.4, 0.43, and 0.44, and standard deviation values of 0.18, 0.44, 0.49, and 0.66. Note: the plotted bars in this sub-figure shows the average of the standard deviation of SoIs within a trace. When the footsteps are more synchronized, the standard deviation is low, as observed in Section 3.1 (i.e., the duration is consistent). Thus, the standard deviation of the SoI duration for each trace increases as the number of persons increase. This is due to less synchronous footsteps as the number of persons increase. Similarly, the mean value of the SoIs increases when the number of people increase. Therefore, when the steps are synchronized, this feature behaves similar to Section 3.1. However, because there are less synchronization in human behavior, we observe that standard deviation within a trace can be used for occupant traffic estimation as well.

The SoI signal entropy indicates the complexity of the signal. Figure 5 (d) shows the SoI entropy value of different occupant traffic conditions, where the average over 8 traces are 4.84, 4.98, 5, and 5.04. The increasing trend, as we modeled earlier in Section 3.1, is shown with corresponding increasing number of occupants in a walking group. The difference between these values are not as obvious as those in Section 3.1. This is possibly caused by the low signal-to-noise ratio of the footstep induced signal comparing to impulse load induced signal.

3.2.3 Classification Results

We tested the feature efficiency through the k-nearest neighbor classification described in Section 2.3 with five traces for training and three traces for testing through cross validation. This results in 12 data points per test. In order to understand the performance for each occupant traffic condition, we show the confusion matrix of classifying the number of occupant walking as a group from one to four in Table 1. The confusion matrix's (i, j) grid demonstrates the percentage of the case where number of i people walking case is classified as number of j people walking case.

Accuracy	1P	2P	3P	4P
1P	83.33%	16.67%	0%	0%
2P	25%	66.67%	8.33%	0%
3P	8.33%	25%	33.33%	33.33%
4P	0%	8.33%	0%	91.67%

Table 1. Confusion matrix for people counting (4 people walking).

Table 1 shows that 16.67% of the one occupant traffic condition is confused as two occupants traffic condition, but is not confused as three or four occupant traffic conditions. The two occupants traffic condition is mostly confused with one occupant traffic condition, which could be caused by the synchronized side-by-side walking that confuses feature cross correlation between spatio-different SoIs and the SoI duration. The three occupants traffic condition has the lowest accuracy and are confused with two and four occupants traffic conditions with up to 33% chance. As we discussed in Section 3.2.2, the cross correlation between spatio-different SoIs can confuse the three v.s. four occupant traffic condition. The cross correlation between temporal-different SoIs can confuse the two v.s. three occupant traffic condition. This results in the three occupant traffic condition being confused with two and four occupant traffic condition. Finally, the four occupants traffic condition estimation is fairly robust, only confused with the two occupants traffic condition in one instance.

We further calculated the mean estimation of the number of occupant passing by for the one to four occupants traffic conditions are 1.16, 1.83, 2.91, and 3.83 respectively. The result demonstrates less than 0.2 people mean estimation error for each condition.

4. RELATED WORK

Indoor pedestrian sensing technologies enable various smart building applications.[20,35] The current sensing methods mainly fall into two categories in terms of mobility: the infrastructural sensing[31,46,47] and the mobile sensing.[28,37] The infrastructural sensing usually allows ubiquitous sensing without the requirement of human interaction, however the mobile sensing methods often can avoid the installation and maintenance of the deployed sensing nodes. To be more specific on the occupancy traffic information, it has been proven useful for various energy managements in smart building environments.[1,10] Many passive sensing methods and apparatuses have been proposed, including using cameras,[3,4,41,43] IR sensors,[15,18,34] RF sensors,[39,40] ultrasonic range finders,[13] etc. Optical based methods (camera and IR sensor), usually require installation in the designated area/position (e.g., on the ceiling pointing down, or ankle girth height in both side of doorway), which is costly in terms of installation and maintenance, esp. in large scale deployments. Similarly, ultrasonic range finder Doorjamb[13] based methods require designated sensor installation at each monitored doorway. RF based methods[39,40] also require extra radio transceivers in high density. Compared to these existing methods and apparatuses, our system utilizes easy-to-install structural vibration sensing units to obtain occupant traffic estimations, and faces less sensing constraints such as the requirement of line-of-sight.

Structural vibration were well known to be efficient for structural health/status monitoring.[24–26,42] In recent years, people start to explore the method to utilize structural vibration for human monitoring.[2,9,19,21,27,29,33] Structural vibration sensing has been used for various indoor occupant monitoring purposes,[9,19] including person localization,[2,21,33] occupancy estimation,[27] and occupant identification.[29] The prospect of research in this field inspired this work but most of the prior works focused on single occupant walking case. In addition, with more and more buildings equipped with structural vibrations system for either structural monitoring or human monitoring, our system can work without additional hardware installation.

Ambient structural vibrations contain useful information about excitations and the objects that generate them. Different sensing purposes involve different features,[12,36] and therefore feature selection and understanding is the key to utilizing signals in this field. Feature selection has been an important problem for all types of pattern recognition systems.[30] Existing approaches involve running as many features as possible[29] in trade of high computational power or selecting individual criteria to investigate on a deeper level.[30] Our work takes a different approach by conducting experiments to investigate the physical significance behind key feature characteristics. This understanding allows us achieve occupant traffic estimation and unlocks various opportunities.

5. CONCLUSION

In summary, we present a system that monitors the occupant traffic through ambient structural vibration sensing. The system achieves sparse sensing with two sensing units to cover each monitoring area, and utilizes four key features to conduct occupant traffic estimation. With load test experiments, we gained a further understanding of signal features when multiple impulse signals mixed, as well as limitations of utilizing those features in specific scenarios. Furthermore, with collected occupant footstep-induced structural vibrations, we evaluated selected features by comparing them through different categories. The occupant number detection for one to four people walking in the same direction achieves less than 0.2 people mean estimation error for each case. The ambiguity from different test cases could be further distinguished by learning historical data from different sensing groups and improved by a hierarchical classification method, which are the directions of our future research. This study of signal features is the key to enable further applications on step-based multiple occupants monitoring.

ACKNOWLEDGMENTS

This work is partially supported by National Science Foundation (NSF) under awards CNS-1149611, Department of Defense (DoD), Pennsylvania Infrastructure Technology Alliance (PITA), CMU-SYSU Collaborative Innovation Research Center (CIRC), Intel, Nokia, Google, and Renault.

REFERENCES

[1] Y. Agarwal, B. Balaji, R. Gupta, J. Lyles, M. Wei, and T. Weng. Occupancy-driven energy management for smart building automation. In *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, pages 1–6. ACM, 2010.

[2] R. Bahroun, O. Michel, F. Frassati, M. Carmona, and J. Lacoume. New algorithm for footstep localization using seismic sensors in an indoor environment. *Journal of Sound and Vibration*, 333(3):1046–1066, 2014.

[3] T.-Y. Chen, C.-H. Chen, D.-J. Wang, and Y.-L. Kuo. A people counting system based on face-detection. In *Genetic and Evolutionary Computing (ICGEC), 2010 Fourth International Conference on*, pages 699–702. IEEE, 2010.

[4] Y. Cong, H. Gong, S.-C. Zhu, and Y. Tang. Flow mosaicking: Real-time pedestrian counting without scene-specific learning. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1093–1100. IEEE, 2009.

[5] G. Conte, M. De Marchi, A. A. Nacci, V. Rana, and D. Sciuto. Bluesentinel: a first approach using ibeacon for an energy efficient occupancy detection system. In *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*, pages 11–19. ACM, 2014.

[6] A. Corna, L. Fontana, A. Nacci, and D. Sciuto. Occupancy detection via ibeacon on android devices for smart building management. In *Proceedings of the 2015 Design, Automation & Test in Europe Conference & Exhibition*, pages 629–632. EDA Consortium, 2015.

[7] M. Costa. Interpersonal distances in group walking. *Journal of Nonverbal Behavior*, 34(1):15–26, 2010.

[8] T. M. Cover and P. E. Hart. Nearest neighbor pattern classification. *Information Theory, IEEE Transactions on*, 13(1):21–27, 1967.

[9] K. Dobbler, M. Fišer, M. Fellner, and B. Rettenbacher. Vibroacoustic monitoring: Techniques for human gait analysis in smart homes. In *Ambient Assisted Living*, pages 47–58. Springer, 2014.

[10] V. L. Erickson, Y. Lin, A. Kamthe, R. Brahme, A. Surana, A. E. Cerpa, M. D. Sohn, and S. Narayanan. Energy efficient building environment control strategies using real-time occupancy measurements. In *Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, pages 19–24. ACM, 2009.

[11] G. Fierro, O. Rehmane, A. Krioukov, and D. Culler. Zone-level occupancy counting with existing infrastructure. In *Proceedings of the Fourth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, pages 205–206. ACM, 2012.

[12] C. Geiß, P. A. Pelizari, M. Marconcini, W. Sengara, M. Edwards, T. Lakes, and H. Taubenböck. Estimation of seismic building structural types using multi-sensor remote sensing and machine learning techniques. *ISPRS Journal of Photogrammetry and Remote Sensing*, 104:175–188, 2015.

[13] T. W. Hnat, E. Griffiths, R. Dawson, and K. Whitehouse. Doorjamb: unobtrusive room-level tracking of people in homes using doorway sensors. In *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*, pages 309–322. ACM, 2012.

[14] Input/Output, Inc. *SM-24 Geophone Element*, 2006.

[15] J. Kerridge, A. Armitage, D. Binnie, L. Lei, and N. Sumpter. Monitoring the movement of pedestrians using low-cost infrared detectors: initial findings. In *Proceedings US Transport Research Board Annual Meeting*, 2004.

[16] J. Krumm, S. Harris, B. Meyers, B. Brumitt, M. Hale, and S. Shafer. Multi-camera multi-person tracking for easyliving. In *Visual Surveillance, 2000. Proceedings. Third IEEE International Workshop on*, pages 3–10. IEEE, 2000.

[17] M. Lam, M. Mirshekari, S. Pan, P. Zhang, and H. Y. Noh. Robust occupant detection through step-induced floor vibration by incorporating structural characteristics. In *IMAC XXXIV A Conference and Exposition on Structural Dynamics*.

[18] G.-G. Lee, H.-k. Kim, J.-Y. Yoon, and J.-J. Kim. Pedestrian counting using an ir line laser. In *Convergence and Hybrid Information Technology, 2008. ICHIT'08. International Conference on*, pages 482–485. IEEE, 2008.

[19] H. Lee, J. W. Park, and A. S. Helal. Estimation of indoor physical activity level based on footstep vibration signal measured by mems accelerometer in smart home environments. In *Mobile Entity Localization and Tracking in GPS-less Environnments*, pages 148–162. Springer, 2009.

[20] V. Lenders, E. Koukoumidis, P. Zhang, and M. Martonosi. Location-based trust for mobile user-generated content: applications, challenges and implementations. In *Proceedings of the 9th workshop on Mobile computing systems and applications*, pages 60–64. ACM, 2008.

[21] M. Mirshekari, S. Pan, A. Bannis, Y. P. M. Lam, P. Zhang, and H. Y. Noh. Step-level person localization through sparse sensing of structural vibration. In *Proceedings of the 14th International Conference on Information Processing in Sensor Networks*, pages 376–377. ACM, 2015.

[22] M. Mirshekari, S. Pan, P. Zhang, and H. Y. Noh. Characterizing wave propagation to improve indoor step-level person localization using floor vibration. In *SPIE Smart Structures and Materials+ Nondestructive Evaluation and Health Monitoring*.

[23] S. Narayana, R. V. Prasad, V. S. Rao, T. Prabhakar, S. S. Kowshik, and M. S. Iyer. Pir sensors: characterization and novel localization technique. In *Proceedings of the 14th International Conference on Information Processing in Sensor Networks*, pages 142–153. ACM, 2015.

[24] H. Y. Noh, D. G. Lignos, K. K. Nair, and A. S. Kiremidjian. 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):681–696, 2012.

[25] H. Y. Noh, K. K. Nair, A. S. Kiremidjian, and C. Loh. 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):95–117, 2009.

[26] H. Y. Noh, R. Rajagopal, and A. S. Kiremidjian. 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*, pages 834507–834507. International Society for Optics and Photonics, 2012.

[27] S. Pan, A. Bonde, J. Jing, L. Zhang, P. Zhang, and H. Y. Noh. Boes: building occupancy estimation system using sparse ambient vibration monitoring. In *SPIE Smart Structures and Materials+ Nondestructive Evaluation and Health Monitoring*, pages 90611O–90611O. International Society for Optics and Photonics, 2014.

[28] S. Pan, A. Chen, and P. Zhang. 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*, pages 99–104. ACM, 2013.

[29] S. Pan, N. Wang, Y. Qian, I. Velibeyoglu, H. Y. Noh, and P. Zhang. Indoor person identification through footstep induced structural vibration. In *Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications*, pages 81–86. ACM, 2015.

[30] H. Peng, F. Long, and C. Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(8):1226–1238, 2005.

[31] A. Purohit, Z. Sun, S. Pan, and P. Zhang. Sugartrail: Indoor navigation in retail environments without surveys and maps. In *Sensor, Mesh and Ad Hoc Communications and Networks (SECON), 2013 10th Annual IEEE Communications Society Conference on*, pages 300–308. IEEE, 2013.

[32] O. A. Rosso, S. Blanco, J. Yordanova, V. Kolev, A. Figliola, M. Schürmann, and E. Başar. Wavelet entropy: a new tool for analysis of short duration brain electrical signals. *Journal of neuroscience methods*, 105(1):65–75, 2001.

[33] J. Schloemann, V. S. Malladi, A. G. Woolard, J. M. Hamilton, R. M. Buehrer, and P. A. Tarazaga. Vibration event localization in an instrumented building. In *Experimental Techniques, Rotating Machinery, and Acoustics, Volume 8*, pages 265–271. Springer, 2015.

[34] J. Song, Y.-F. Dong, X.-W. Yang, J.-H. Gu, and P.-P. Fan. Infrared passenger flow collection system based on rbf neural net. In *Machine Learning and Cybernetics, 2008 International Conference on*, volume 3, pages 1277–1281. IEEE, 2008.

[35] D. Sousa Nunes, P. Zhang, and J. Sa Silva. A survey on human-in-the-loop applications towards an internet of all. *Communications Surveys & Tutorials, IEEE*, 17(2):944–965, 2015.

[36] A. Subramanian, K. G. Mehrotra, C. K. Mohan, P. K. Varshney, and T. Damarla. Feature selection and occupancy classification using seismic sensors. In *Trends in Applied Intelligent Systems*, pages 605–614. Springer, 2010.

[37] Z. Sun, A. Purohit, K. Chen, S. Pan, T. Pering, and P. Zhang. Pandaa: physical arrangement detection of networked devices through ambient-sound awareness. In *Proceedings of the 13th international conference on Ubiquitous computing*, pages 425–434. ACM, 2011.

[38] L. Xia, C.-C. Chen, and J. Aggarwal. Human detection using depth information by kinect. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*, pages 15–22. IEEE, 2011.

[39] C. Xu, B. Firner, R. S. Moore, Y. Zhang, W. Trappe, R. Howard, F. Zhang, and N. An. 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*, pages 79–90. IEEE, 2013.

[40] C. Xu, B. Firner, Y. Zhang, R. Howard, J. Li, and X. Lin. Improving rf-based device-free passive localization in cluttered indoor environments through probabilistic classification methods. In *Proceedings of the 11th international conference on Information Processing in Sensor Networks*, pages 209–220. ACM, 2012.

[41] H. Xu, P. Lv, and L. Meng. A people counting system based on head-shoulder detection and tracking in surveillance video. In *Computer Design and Applications (ICCDA), 2010 International Conference on*, volume 1, pages V1–394. IEEE, 2010.

[42] H. Young Noh, K. Krishnan Nair, D. G. Lignos, and A. S. Kiremidjian. Use of wavelet-based damage-sensitive features for structural damage diagnosis using strong motion data. *Journal of Structural Engineering*, 137(10):1215–1228, 2011.

[43] C. Zeng and H. Ma. Robust head-shoulder detection by pca-based multilevel hog-lbp detector for people counting. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 2069–2072. IEEE, 2010.

[44] D. Zhang, Y. Liu, X. Guo, and L. M. Ni. Rass: A real-time, accurate, and scalable system for tracking transceiver-free objects. *Parallel and Distributed Systems, IEEE Transactions on*, 24(5):996–1008, 2013.

[45] D. Zhang and L. M. Ni. Dynamic clustering for tracking multiple transceiver-free objects. In *Pervasive Computing and Communications, 2009. PerCom 2009. IEEE International Conference on*, pages 1–8. IEEE, 2009.

[46] P. Zhang and M. Martonosi. Locale: Collaborative localization estimation for sparse mobile sensor networks. In *Information Processing in Sensor Networks, 2008. IPSN'08. International Conference on*, pages 195–206. IEEE, 2008.

[47] P. Zhang, Z. Sun, and A. Purohit. Coughloc: Location-aware indoor acoustic sensing for non-intrusive cough detection. In *International Workshop on Emerging Mobile Sensing Technologies, Systems, and Applications*, 2011.