ASCE

Step-Level Occupant Detection across Different Structures through Footstep-Induced Floor Vibration Using Model Transfer

Mostafa Mirshekari, S.M.ASCE[1]; Jonathon Fagert, P.E., S.M.ASCE[2]; Shijia Pan[3]; Pei Zhang[4]; and Hae Young Noh, A.M.ASCE[5]

Abstract: This paper presents a floor-vibration-based step-level occupant-detection approach that enables detection across different structures through model transfer. Detecting the occupants through detecting their footsteps (i.e., step-level occupant detection) is useful in various smart building applications such as senior/healthcare and energy management. Current sensing approaches (e.g., vision-based, pressure-based, radio frequency–based, and mobile-based) for step-level occupant detection are limited due to installation and maintenance requirements such as dense deployment and requiring the occupants to carry a device. To overcome these requirements, previous research used ambient structural vibration sensing for footstep modeling and step-level occupant detection together with supervised learning to train a footstep model to distinguish footsteps from nonfootsteps using a set of labeled data. However, floor-vibration-based footstep models are influenced by the structural properties, which may vary from structure to structure. Consequently, a footstep model in one structure does not accurately capture the responses in another structure, which leads to high detection errors and the costly need for acquiring labeled data in every structure. To address this challenge, the effect of the structure on the footstep-induced floor vibration responses is here characterized to develop a physics-driven model transfer approach that enables step-level occupant detection across structures. Specifically, the proposed model transfer approach projects the data into a feature space in which the structural effects are minimized. By minimizing the structure effect in this projected feature space, the footstep models mainly represent the differences in the excitation types and therefore are transferable across structures. To this end, it is analytically shown that the structural effects are correlated to the maximum-mean-discrepancy (MMD) distance between the source and target marginal data distributions. Therefore, to reduce the structural effect, the MMD between the distributions in the source and target structures is minimized. The robustness of the proposed approach was evaluated through field experiments in three types of structures. The evaluation consists of training a footstep model in a set of structures and testing it in a different structure. Across the three structures, the evaluation results show footstep detection F1 score of up to 99% for the proposed approach, corresponding to a 29-fold improvement compared to the baseline approach, which do not transfer the model. **DOI: 10.1061/(ASCE)EM.1943-7889.0001719.**

Introduction

Step-level occupant detection is important in various smart building applications, such as senior/healthcare and efficient energy management. Step-level information enables the estimation of temporal gait-related features, which are important for continuous monitoring and treatment of diseases such as dementia, chronic obstructive pulmonary disease, and muscular dystrophy (Oberg et al. 1993; Visser 1983). Some examples of these gait features are cadence, stride time, step time, and activity level. Specifically, cadence (or step frequency) is the rate at which a person walks; it is expressed in steps per minute and can be estimated by detecting and counting the number of footsteps in a given time. Stride time and step time are the time between one foot striking to the same foot striking or the opposite foot striking, respectively, and can be estimated by detecting the footsteps and finding the time between them. Finally, the number of footsteps in a given area represent the activity level. Additionally, existence of footsteps is a sign of the presence of occupants, which can be utilized for heating, ventilation, and air conditioning (HVAC) control to minimize energy consumption while maintaining occupants' comfort.

Some prior sensing approaches for occupant detection include vision-based (Snidaro et al. 2005; Muñoz-Salinas et al. 2007), pressure-based (Savio and Ludwig 2007; Murakita et al. 2004), radio frequency (RF)–based (Zetik et al. 2006; Li et al. 2012), acoustic-based (Uziel et al. 2013), and mobile-based (Wang et al. 2016; Saha et al. 2014) approaches. These approaches are potentially accurate in detecting occupants; however, installation and maintenance requirements limit their applications. Vision-based sensing is sensitive to visual occlusions such as furniture and columns (Snidaro et al. 2005; Muñoz-Salinas et al. 2007). Pressure-based and RF-based sensing require dense deployment (particularly if step-level detection is required) (Savio and Ludwig 2007; Murakita et al. 2004; Zetik et al. 2006; Li et al. 2012). Acoustic sensors are sensitive to various kind of environmental noise

[1]Dept. of Civil and Environmental Engineering, Carnegie Mellon Univ., Pittsburgh, PA 15213 (corresponding author). ORCID: https://orcid.org/0000-0001-7364-7284. Email: mmirshekari@cmu.edu

[2]Dept. of Civil and Environmental Engineering, Carnegie Mellon Univ., Pittsburgh, PA 15213.

[3]Professor, Dept. of Computer Science and Engineering, Univ. of California-Merced, Merced, CA 95343.

[4]Professor, Dept. of Electrical and Computer Engineering, Carnegie Mellon Univ., Moffett Field, CA 94035.

[5]Professor, Dept. of Civil and Environmental Engineering, Carnegie Mellon Univ., Pittsburgh, PA 15213.

Note. This manuscript was submitted on January 22, 2019; approved on July 3, 2019; published online on December 21, 2019. Discussion period open until May 21, 2020; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Engineering Mechanics***, © ASCE, ISSN 0733-9399.

(Uziel et al. 2013). Finally, wearable sensing require the occupants to wear or carry a device, which might not be possible in some applications (Wang et al. 2016; Saha et al. 2014).

To overcome these limitations, previous research has introduced floor-vibration-based sensing for step-level occupant detection (Mirshekari et al. 2017; Lam et al. 2016). Vibration-based sensing enables sparse sensing, does not require the occupant to carry or wear a device, and is not sensitive to visual occlusions. This approach detects the occupants through recognizing their footstep-induced vibrations. The main intuition is that the structural vibration responses caused by the footsteps are different from the ones caused by other types of excitations such as objects falling on the floor or a door closing. Due to this difference, a set of labeled vibration data from footsteps and nonfootsteps can be used to train a footstep model through supervised learning (Mirshekari et al. 2017; Lam et al. 2016).

However, the vibration data characteristics also depend on the underlying structure. Therefore, a footstep model trained in one structure is not accurate in other structures. In other words, the distribution of the vibration data is different across various structures and hence, the footstep model does not transfer well between them. Furthermore, acquiring labeled data requires a calibration phase in every structure and different locations even in the same structure, which is costly and difficult.

This paper presenst a step-level occupant-detection approach that is robust to structural changes across different structures using model transfer. To this end, the structural effects on its vibration responses to footsteps are characterized to develop a physics-driven model transfer approach. This characterization is used to project the data into a new feature space in which the structural effect is reduced. By lowering the structural effect, the footstep model mainly represents the excitation differences and is similar across structures (i.e., transfers between them). In other words, this model transfer through projection to a new feature space enables one to utilize the labeled data in a set of source structures to develop a footstep model for the target structures with no available labeled data. By requiring the labeled data only from a small number of structures, this approach significantly reduces the calibration effort.

The main steps of the proposed approach are (1) estimating a set of bases that project the source and target vibration data to a feature space with lower structural effect, (2) training a classifier using the projected labeled data in the source structure and utilizing it to label the samples in the target structure, and (3) utilizing the classification model confidence score for combining the labels from multiple sources. The main novelty of the presented work is in the model transfer approach, which can be combined with any classifier. For demonstration, support vector machines (SVM), which is a widely used classifier, is used. The proposed system is validated under different real-world scenarios in multiple buildings including a campus building and two eldercare facilities.

In summary, the core contributions of this paper are as follows:

- A step-level occupant-detection approach is presented, which is robust across different structures using footstep-induced floor vibration.
- Structural effects on the vibration responses are characterized analytically to develop a physics-driven model transfer approach that projects the data into a new feature space with reduced structural effect.
- Step-level occupant detection is evaluated through model transfer in various structures with different structural materials and characteristics.

The rest of the paper is organized as follows. First, related works and how this work is distinguished from them are discussed. Then, the physical intuition behind the proposed approach, including analytic physical characterization of model transfer, are discussed. Next, the approach and its modules are detailed. Finally, the evaluation procedure, including the experiments conducted and the analysis results, are described and conclusions are presented.

Related Work

In the recent years, footstep-induced floor vibration sensing has been used for many occupant-monitoring applications such as occupant localization (Mirshekari et al. 2016, 2018b; Poston et al. 2017b; Reuland et al. 2017), activity monitoring (Pan et al. 2018; Madarshahian et al. 2019; Fagert et al. 2017b), occupant identification (Pan et al. 2017c), and occupant balance estimation (Fagert et al. 2017a). This paper introduces a model transfer approach for step-level occupant detection using this sensing approach. Here, current floor-vibration-based occupant-detection approaches and their limitations are first discused. Then, the literature for model transfer (or transfer learning) is reviewed.

Conventional Learning Approaches for Floor-Vibration-Induced Occupant Detection

The current floor-vibration-based approaches for occupant detection use signal processing and machine learning approaches to distinguish footsteps from nonfootstep excitations. Some examples of these approaches include (1) statistics-based signal analysis [using features such as kurtosis (Koç and Yegin 2013; Succi et al. 2001), chi-squared (Mirshekari et al. 2018b; Poston et al. 2017a), and autocorrelation (Alyamkin and Eremenko 2011)], (2) matched filtering (Richman et al. 2001; Poston et al. 2017b), and (3) supervised learning [e.g., neural networks (Subramanian et al. 2010) and support vector machines (Lam et al. 2016)].

However, these approaches require calibration in every structure or noise level. For example, statistics-based approaches require a threshold which may change in various structures. Furthermore, the matched filtering and supervised learning approaches require labeled data in every new structure (as well as various locations in the same structure). Due to these calibration requirements, these approaches are difficult and expensive to implement in real applications.

Transfer Learning

Model transfer (or transfer learning) has been introduced in the machine learning literature to reduce the labeled data (i.e., calibration) requirements. The idea behind transfer learning for vibration-based occupant detection is to use the labeled vibration data in a specific structure (source) to train a footstep model in other structures (target) in which only unlabeled data are available. In the literature, this specific setting corresponds to transductive transfer learning (Pan and Yang 2010). Similar to model transfer, unsupervised learning approaches assume no labeled data in the structure. However, they do not utilize the labeled data in other structures (i.e., source structures), which results in lower model performance. Furthermore, semisupervised and supervised learning assume that at least some labeled data are available in the structure (Friedman et al. 2001). The main categories of transductive transfer learning are instance-based and feature-based approaches.

Instance-Based Transductive Transfer Learning

The main idea behind instance-based approaches is to assign higher weights to a subset of source instances that are more likely to happen in the target structure for model training (Dai et al. 2007;

Bickel et al. 2007; Sugiyama et al. 2008; Yu et al. 2018). Therefore, these approaches assume an overlap between the source and target data distribution in the original data space. In the proposed application, various structures have drastically different signal characteristics that typically do not overlap in the original data space, and hence instance-based transfer is not suitable.

Feature-Based Transductive Transfer Learning

The feature-based approaches aim to find a feature space (i.e., the projection of the original data) in which the distributions of data in the source and target share similarities (Blitzer et al. 2006; Ben-David et al. 2007; Ling et al. 2008; Pan et al. 2011; Mirshekari et al. 2018a). It is possible to find such a feature space even if in the original data space, the source and target distributions are different, and therefore, these approaches are more suitable for the present problem. To demonstrate this point, consider a set of input vibration responses $\mathbf{Y_S} \in \mathbb{R}^{n_s \times n_b}$ and $\mathbf{Y_T} \in \mathbb{R}^{n_t \times n_b}$ for the source and target structures, where n_s and n_t are the number of samples in the source and target structures and n_b is the length of each input vector. Assuming $n_s = n_t$ and to ensure similarity in the projected space, the aim is to find the solutions for $\mathbf{Y_S W} = \mathbf{Y_T W}$ where $\mathbf{W} \in \mathbb{R}^{n_b \times n_d}$ projects the source and target data into a n_d-dimensional feature space.

Here, without loss of generality, consider the simplest case of $n_d = 1$. Existence of solutions for this equation means that regardless of the differences in the original data space, the source and target distributions in the projected feature space share similarities. The equation can be rewritten as a homogeneous set of simultaneous equations $(\mathbf{Y_S} - \mathbf{Y_T})\mathbf{W} = 0$. This set of equations always has nontrivial solutions as long as there are more unknowns than equations ($n_b > n$) (Strang and Strang 2006). This condition is generally true in the proposed application because the length of the sample vectors (e.g., number of bins in frequency domain representation) are much higher than the number of samples. This proof shows that there exists a projected space in which the source and target distribution share similarity. However, this similarity does not ensure separable classes (e.g., footsteps and nonfootsteps) and high performance of the model in the projected feature space.

Furthermore, current feature-representation-based approaches only ensure that the marginal distribution of the data is similar across the source and target structures. In other words, there is no guarantee that in the projected feature space, the source structure features for footsteps are close to the target structure features for the footsteps and those for nonfootsteps are close to the nonfootsteps. Therefore, even after projection, the footstep models are not necessarily transferable across various structures. To overcome these limitations, the structural effect is reduced, which ensures that the footstep models in the projected feature space mainly represent the excitation characteristics. To this end, physical insights are utilized to (1) characterize the structural effect on the floor vibration response and its distribution, and (2) introduce a model transfer approach that reduces the structural effect to make the footstep models useful across various structures. By mainly representing the excitation effect, the features for footsteps in the source and target structure are close to each other, and similarly, the features for nonfootsteps are close in source and target structures. Hence, in the projected feature space, the footstep models are transferable across various structures.

Analytic Physical Characterization of Model Transfer for Occupant Detection

The proposed approach models the footstep floor vibrations to distinguish the signals caused by them from the ones caused by nonfootsteps. These floor vibrations depend on the underlying structure, and therefore, footstep models trained in one structure are significantly different from those trained in other structures. This section first describes the excitation mechanism and structural effects by showing a set of footstep and balldrop signals in two structures. Then, the structural effect on the vibration responses are characterized to transfer the models across various structures.

Structure-Dependent Excitation Mechanism

Intuitively, impulsive excitations cause deformations in the floor structure. The floor is then restored to its original position due to the elasticity of the structure. The repeated cycle of applied force and restoration force results in a deformation cycle and oscillations in the floor structure, which are commonly referred as vibrations. Various excitations result in different deformation patterns and hence vibration signals. Therefore, if one has labeled footstep and nonfootstep signals from a structure, one can train a classifier to distinguish them.

However, the vibration signals are also affected by the the underlying structure and its characteristics. To illustrate this point, Fig. 1 shows sample balldrop and footstep signals from two structures. In each structure, the footstep and balldrop are different in shape, which enables training the footstep model. However, the differences between the footstep and the balldrop are not consistent in these two structures. Therefore, the model trained using the labeled signals from one of the structures is not applicable to the other structure. To better understand this problem and the proposed solution, the rest of this section first analytically describes the effect of the structure and the excitation type on the vibration signals. Then, it is shown that the proposed model transfer approach minimizes the effect of the structural differences on the vibration signals. This minimization enables the same footstep model from the source structure to be used for footstep prediction in the target structure.

Structural Effect Formulation

To transfer the models across structures, the structural effects on the vibration responses are first formulated through the convolution theorem assuming an linear time-invariant (LTI) system (Oppenheim et al. 1997). Specifically, in frequency domain, a specific sample (e.g., ith) can be described

$$\mathbf{Y_i} = \mathbf{H_i X_i} \tag{1}$$

where $\mathbf{Y_i} \in \mathbb{R}^{n_b \times 1}$ is the vibration frequency representation, in which n_b is the number of frequency bins and $\mathbb{R}$ is the set of real numbers; $\mathbf{X_i} \in \mathbb{R}^{n_b \times 1}$ is the input force spectrum; and $\mathbf{H_i} \in \mathbb{R}^{n_b \times n_b}$ is the frequency response function (FRF) of the structure, which can be described as the following diagonal matrix:

$$\mathbf{H_i} = \begin{pmatrix} h_1 & & & \\ & h_2 & & \\ & & \ddots & \\ & & & h_{n_b} \end{pmatrix} \tag{2}$$

Minimizing the Structural Effect

The vibration responses depend on the structure and the excitation, as shown in Eq. (1). Therefore, by minimizing the structural effect, the trained model mainly represents the excitation effect and transfers across structures. To this end, the rest of this section shows that the structural effect is correlated to the maximum-mean-discrepancy (MMD) distance between the source and target structure in a given kernel space. This insight implies that by minimizing

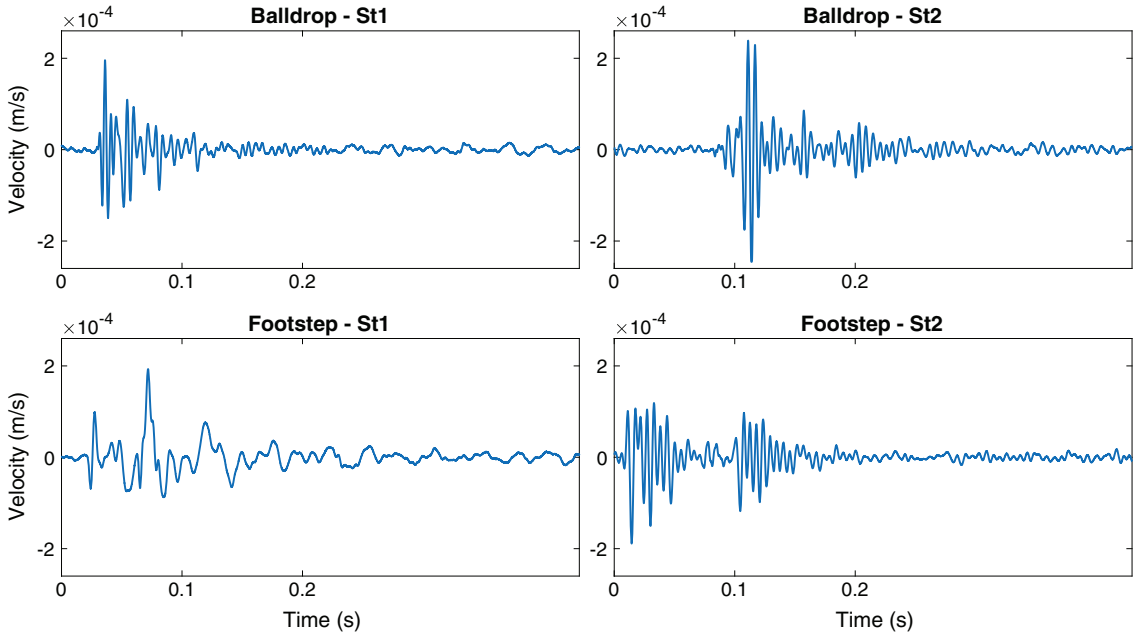

Fig. 1. Sample signals of balldrop and footstep impulses in two structures. Differences between the footstep and the balldrop are not consistent in these two structures. Therefore, the model trained in one structure is not suitable in the other structure.

the MMD distance between the source and target distributions, the objective of minimizing the structural effect can be achieved, which enables transferring models across the structures.

To show the aforementioned correlation, the MMD distance is analytically described with respect to the structural effects. MMD is a nonparametric distribution distance metric that does not require an intermediate density estimate and hence does not require distribution-type assumption (Smola et al. 2007; Gretton et al. 2012). Given the data sets $\mathbf{S} = \{\mathbf{Y_i^s}\}$ and $\mathbf{T} = \{\mathbf{Y_i^t}\}$ from the source and target structures, respectively, the empirical estimate of the MMD can be described as follows (Pan et al. 2008):

$$\text{MMD}(\mathbf{S}, \mathbf{T}, \phi) = \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} \phi(\mathbf{Y_i^s}) - \frac{1}{n_t} \sum_{i=1}^{n_t} \phi(\mathbf{Y_i^t}) \right\|_{\mathcal{H}} \tag{3}$$

where ϕ = kernel-induced feature map; n_s and n_t = number of samples in the source and target structures; $\mathbf{Y_i^s}$ and $\mathbf{Y_i^t}$ = ith samples in the source and target structures; and $\| \cdot \|_{\mathcal{H}}$ = norm in the reproducing kernel Hilbert space (RKHS). RKHS is a Hilbert space with two properties: (1) the feature map of every point is in the feature space, and (2) it has reproducing property, meaning that the values of the functions can be evaluated through an inner product (Gretton 2013). The kernel matrix can be defined as $K = [\phi(\mathbf{Y_i})^T \quad \phi(\mathbf{Y_j})]$ to rewrite the Eq. (3) through the kernel trick as follows (Pan et al. 2011):

$$\text{MMD}(\mathbf{S}, \mathbf{T}, \mathbf{K}) = tr(\mathbf{KL}) \tag{4}$$

where $\mathbf{K} \in \mathbb{R}^{(n_s+n_t)\times(n_s+n_t)}$ is the kernel matrix for the data set consisting of data from both the target and source structures. The submatrices forming the kernel matrix are shown in Eq. (5)

$$\mathbf{K} = \begin{bmatrix} \mathbf{K_{S,S}} & \mathbf{K_{S,T}} \\ \mathbf{K_{T,S}} & \mathbf{K_{T,T}} \end{bmatrix} \tag{5}$$

where $\mathbf{K_{S,S}} \in \mathbb{R}^{n_s\times n_s}$ and $\mathbf{K_{T,T}} \in \mathbb{R}^{n_t\times n_t}$ are the kernel matrices between the source samples and target samples, respectively; $\mathbf{K_{S,T}} \in \mathbb{R}^{n_s\times n_t}$ and $\mathbf{K_{T,S}} \in \mathbb{R}^{n_t\times n_s}$ are the kernels across the samples in source and target structures; and $\mathbf{L} \in \mathbb{R}^{(n_s+n_t)\times(n_s+n_t)}$ is a matrix of coefficients, which is found by

$$\mathbf{L_{ij}} = \begin{cases} \dfrac{1}{n_s^2} & \mathbf{Y_i}, \mathbf{Y_j} \in \mathbf{S} \\ \dfrac{1}{n_t^2} & \mathbf{Y_i}, \mathbf{Y_j} \in \mathbf{T} \\ -\dfrac{1}{n_s n_t} & \text{otherwise} \end{cases} \tag{6}$$

Knowing that the trace of a product of two matrices can be rewritten as the sum of their elementwise product, Eqs. (5) and (6) can be used to rewrite Eq. (4)

$$\text{MMD}(\mathbf{S}, \mathbf{T}, \mathbf{K}) = \frac{1}{n_s^2}\mathbf{K_{S,S}} + \frac{1}{n_t^2}\mathbf{K_{T,T}} - \frac{2}{n_t n_s}\mathbf{K_{T,S}} \tag{7}$$

The distance defined by Eq. (7) is correlated to the structural effects. To show this, the kernel matrix is first described in terms of the input force spectrum and frequency response function. For better understanding, start with the simpler case of assuming a linear kernel for which $\phi(\mathbf{Y_i}) = \mathbf{Y_i}$. In this case, the MMD distance from Eq. (3) is equivalent to the Euclidean distance between the mean of the source and target samples in the original data space. Using the expression in Eq. (1), the linear kernel matrix for a set of n samples can be found using

$$\mathbf{K} = \mathbf{X^T H^T H X} \tag{8}$$

where $\mathbf{H} \in \mathbb{R}^{n_b\times(n_b\times n)}$ and $\mathbf{X} \in \mathbb{R}^{(n_b\times n)\times n}$ are matrices containing the FRF and input force spectrum for all the samples and can be described

$$\mathbf{H} = (\mathbf{H_1} \quad \mathbf{H_2} \quad \dots \quad \mathbf{H_n}) \tag{9}$$

$$\mathbf{X} = \begin{pmatrix} \mathbf{X_1} & & & \\ & \mathbf{X_2} & & \\ & & \ddots & \\ & & & \mathbf{X_n} \end{pmatrix} \tag{10}$$

The next step is to rewrite Eq. (7) in terms of the structural and input responses using the kernel matrix in Eq. (8). Assuming that the input matrix is similar for the source and target structures (i.e., $\mathbf{X_T} = \mathbf{X_S} = \mathbf{X}$) and using the kernel description from Eq. (8), one can rewrite Eq. (7) as follows:

$$\text{MMD}(\mathbf{S}, \mathbf{T}) = \frac{1}{n_s^2}\mathbf{X^T H_S^T H_S X} - \frac{2}{n_t n_s}\mathbf{X^T H_S^T H_T X} + \frac{1}{n_t^2}\mathbf{X^T H_T^T H_T X} \tag{11}$$

Eq. (11) can be rewritten in the following format:

$$\text{MMD}(\mathbf{S}, \mathbf{T}) = \mathbf{X^T}\left(\frac{1}{n_s^2}\mathbf{H_S^T H_S} - \frac{2}{n_t n_s}\mathbf{H_S^T H_T} + \frac{1}{n_t^2}\mathbf{H_T^T H_T}\right)\mathbf{X} \tag{12}$$

Eq. (12) represents the distance between the source and the target structure data distributions. The term $\mathbf{H_S^T H_T}$ shows the cross-similarity between the structure FRFs. According to this equation, the MMD between the source and target distributions is negatively correlated to the cross-similarity of the structures. In other words, assuming that the norms of $\mathbf{H_S}$ and $\mathbf{H_T}$ are fixed, a lower distance corresponds to higher cross-similarity between the structures and lower structural effect.

To generalize this derivation to other kernels, the more general description of the kernel, $\mathbf{K} = [\,\phi(\mathbf{Y_i})^T \quad \phi(\mathbf{Y_j})\,]$, is used. In this case, Eq. (8) can be rewritten

$$\mathbf{K} = \phi_M(\mathbf{HX})^T\phi_M(\mathbf{HX}) \tag{13}$$

where ϕ_M = function that maps each column of the matrix using the ϕ mapping. Using this definition of the kernel matrix, Eq. (11) can be rewritten

$$\text{MMD}(\mathbf{S}, \mathbf{T}, \phi) = \frac{1}{n_s^2}\phi_M(\mathbf{H_S X})^T\phi_M(\mathbf{H_S X}) - \frac{2}{n_t n_s}\phi_M(\mathbf{H_S X})^T\phi_M(\mathbf{H_T X}) + \frac{1}{n_t^2}\phi_M(\mathbf{H_T X})^T\phi_M(\mathbf{H_T X}) \tag{14}$$

By defining $\phi_{MX}(\mathbf{H}) = \phi_M(\mathbf{XH})$, Eq. (14) can be rewritten in the following format:

$$\text{MMD}(\mathbf{S}, \mathbf{T}, \phi) = \frac{1}{n_s^2}\phi_{MX}(\mathbf{H_S})^T\phi_{MX}(\mathbf{H_S}) - \frac{2}{n_t n_s}\phi_{MX}(\mathbf{H_S})^T\phi_{MX}(\mathbf{H_T}) + \frac{1}{n_t^2}\phi_{MX}(\mathbf{H_T})^T\phi_{MX}(\mathbf{H_T}) \tag{15}$$

In Eq. (15), $\phi_{MX}(\mathbf{H_S})^T\phi_{MX}(\mathbf{H_T})$ = cross-similarity of a projected version of the structure FRFs. Eq. (15) shows that the MMD between the source and target structure is negatively correlated to the projected version of the cross-similarity of the structures, assuming that the norms of $\phi_{MX}(\mathbf{H_T})$ and $\phi_{MX}(\mathbf{H_S})$ are fixed. Therefore, to reduce the structural effects, the proposed model transfer projects the data into a feature space in which the MMD between the source and target data distributions is minimized.

However, just minimizing the MMD might not be satisfactory because (1) it might result in a trivial solution for which there is zero distance between the distributions, and all the data points are projected to the origin, and (2) in these derivations, it is assumed that the data are noise-free, which is not the case in real applications. In these cases, one might end up in a projected feature space in which the footstep and nonfootsteps are mixed and not distinguishable. When introducing the proposed method, other objectives will be discussed, which will be combined with distribution distance minimization to overcome these limitations.

Footstep Models before and after Projection

Due to the reduced structural effect, in the projected feature space, the trained models mainly represent the excitation effects and hence transfer well between the structures. In other words, in the projected feature space, the footsteps and nonfootsteps from the source and target structures are close to other footsteps and nonfootsteps, respectively. Figs. 2(a and b) show the data from a target and source structure before and after projection. Before projecting the data, the distance between the data distributions caused by the structural effect is significant. Therefore, the footsteps models in the target source structures are not similar. In this figure, the Frequency features A and B are the log amplitudes of the fast Fourier transform (FFT) at 42 and 260 Hz. These specific frequencies are chosen for better illustration of the model transfer intuition. Furthermore, the log transform of the fft amplitudes is used to reduce the right-skewness of the data and improve the model training. After the projection, this structural effect and the distance between the distributions are reduced. Thus, the model trained in the source structure well represents the data in the target structure.

Model Transfer for Step-Level Occupant Detection across Structures

This section describes the floor-vibration-based approach for step-level occupant detection across different structures through model transfer. The proposed approach has two main modules: (1) impulse-detection module, which measures the structural vibration and distinguishes possible footstep-induced vibration signals from the background noise, and (2) structure-informed model transfer module, which utilizes the data from the source structure to train a footstep model in the target structure. An overview of the approach is presented in Fig. 3.

Impulse Detection

The impulse-detection module measures the ambient structural vibration of the target structure and detects the impulsive vibration events. To this end, the vertical ambient structural vibrations were first measured using a geophone sensor placed on the floor. Geophones are low-cost and low-distortion sensors that convert the velocity of the floor vibration to voltage (I/O Sensor Nederland bv 2006). Then, the vibration signals are amplified using an op-amp to improve their resolution (Pan et al. 2017b). Depending on the floor structure, the proposed system has an effective sensing range of 20 m in diameter.

The ambient vibration signal consists of impulsive vibration events (such as footsteps, object falls, and door closings) and stationary background noise (such as machinery or measurement noise). It was observed that the impulsive events have higher

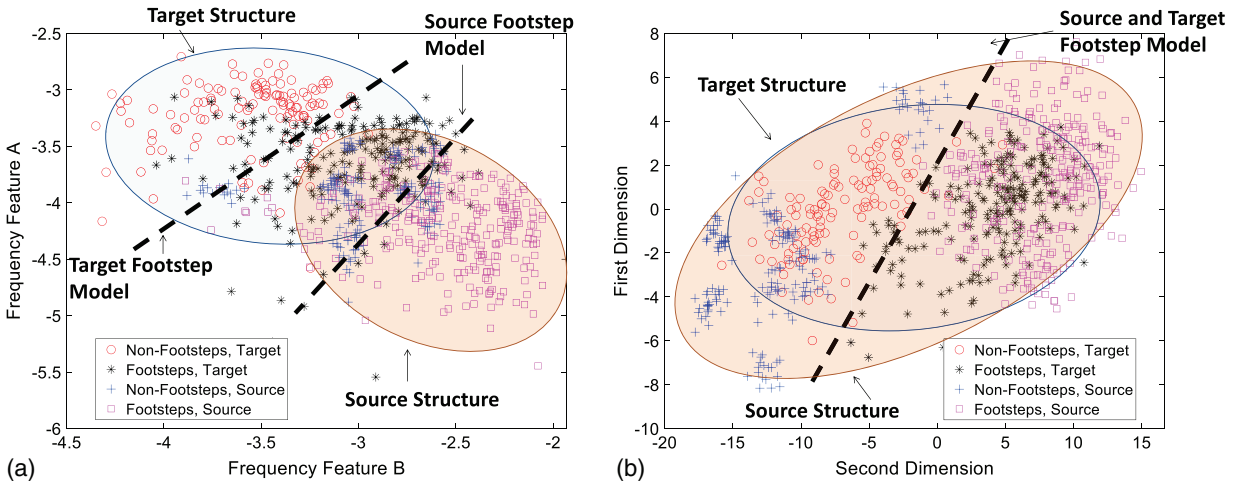

Fig. 2. Structural effect (a) before; and (b) after projection. Part (a) shows that before projection the distance between the distributions caused by the structural effects results in major differences in the footstep model. However, part (b) shows the distributions after projection where the footstep models are aligned.

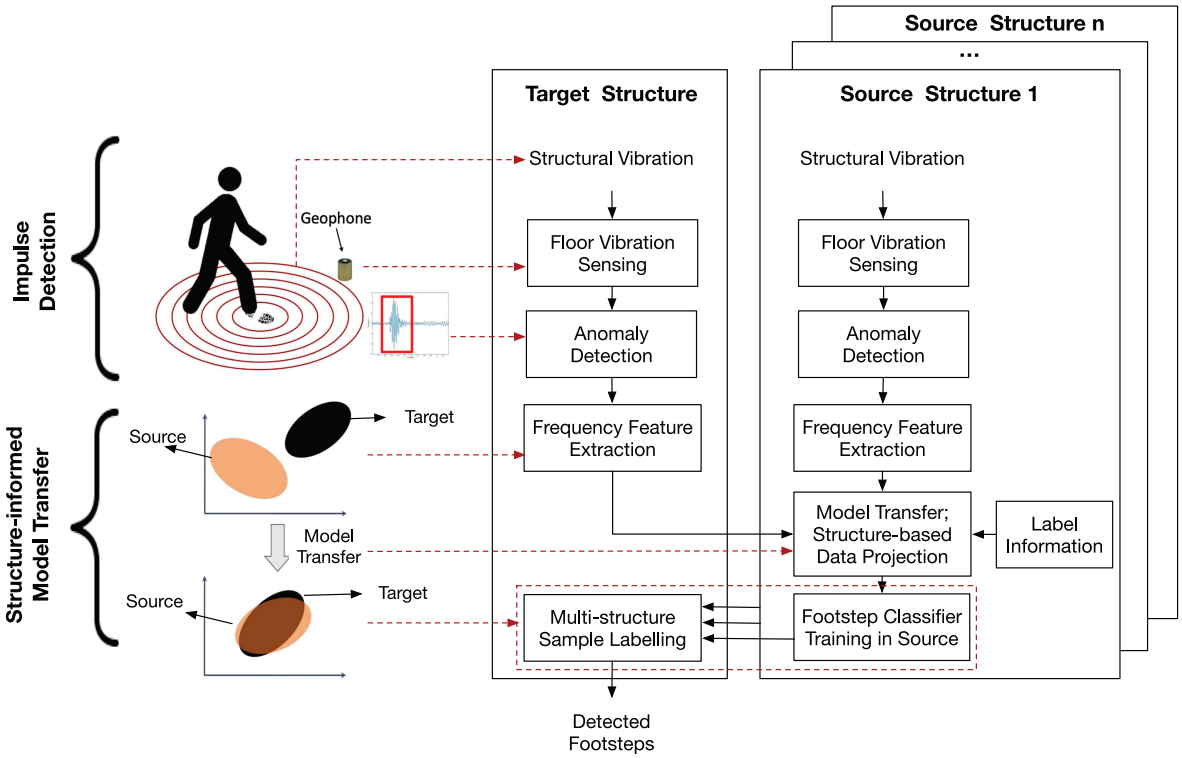

Fig. 3. System overview. The proposed approach consists of impulse detection and structure-informed model transfer. The left-hand side conceptually shows the steps of the proposed approach. Gray arrows relate the conceptual figures to the different steps shown in the flow chart. The solid arrows in the flowchart show the relationship between different steps.

variance than the stationary noise, as shown in Fig. 4. Therefore, impulsive events are separated through a variance-based detection algorithm (Mirshekari et al. 2018b). This approach considers a sliding window for the signal and compares the variance of the signal window σ_w^2 with the variance of noise σ_n^2 using a hypothesis test. The noise variance is estimated based on the part of the signal that does not contain an impulsive excitation. The hypothesis test evaluates the null hypothesis H_0:$\sigma_w^2 = \sigma_n^2$ (i.e., the signal is noise) against the alternative hypothesis H_1:$\sigma_w^2 = \sigma_n^2$ (i.e., the signal is an impulsive event).

This hypothesis test is a chi-squared test because the variance of the ambient noise signal follows a scaled chi-squared distribution. Therefore, first, the chi-squared statistics for each signal window, χ_w^2, are found and compared with χ_α^2, which is the statistic value corresponding to the significance levels of α. The significance level is the probability of having samples with χ^2-statistics higher than

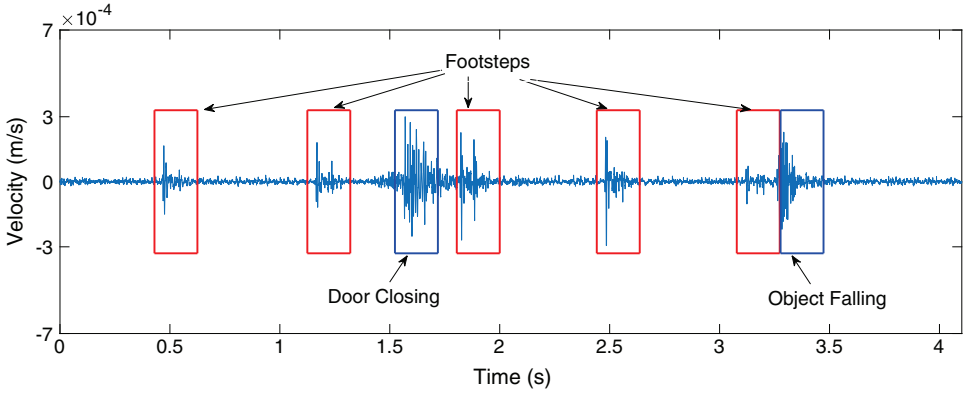

Fig. 4. Example of impulse detection showing that the detected impulses can be footsteps or other impulsive excitations such as a door closing.

χ^2_α given the null hypothesis. Then, the null hypothesis is rejected if $\chi^2_w \geq \chi^2_\alpha$ (Baron 2013). Mirshekari et al. (2018b) have given more details on this impulse-detection approach.

An example of the vibration signal and impulse-detection results are presented in Fig. 4. In this figure, the output of the impulse-detection module is shown with the light gray boxes. However, these separated impulses include both footsteps and other impulsive vibration events. The next modules distinguish the footsteps from other impulsive excitations.

Structure-Informed Model Transfer

The main objective of the structure-informed model transfer module is to use the labeled data in the source structures to develop a footstep model in the target structure. This model distinguishes the footsteps from the other impulsive excitations and therefore enables step-level occupant detection. This objective is achieved by (1) extracting the frequency features, (2) projecting the features from the source and target structures to a new feature space in which the structural effect is reduced, (3) training a footstep model using the labeled source data in the projected feature space, and (4) predicting the sample labels in the target structure by combining the predictions of the models from different source structures.

Frequency Feature Extraction

The proposed model transfer approach takes the frequency representation of the vibration signals as the original data features. As shown in Eq. (1), in the frequency domain, the excitation and structure effect can be simply separated using matrix multiplication, which enables the formulation of the relationship between the MMD and the structural effect. Therefore, the features for model training are developed using the frequency-domain representation, which is more suitable for the present problem compared with time-domain and time-frequency-domain representations.

However, directly using the frequency representation estimated through fast Fourier transform (FFT) has two limitations. First, the overall amplitude of the measured signals shows how much the signal is attenuated, which is location-dependent and does not represent the excitation type. Therefore, before estimating the FFT of the signal, the signals are normalized to ensure the same energy for all the signals by multiplying the signal by constant$/\sqrt{\sum s_i^2}$ in which s_i is the ith element of the signal. Second, the frequency data distribution is right-skewed, which is caused by the fact that the values of signal amplitude in each frequency are positive. This right-skewness results in a data distribution that is less similar to Gaussian distribution and reduces the classifier training performance. To reduce this right-skewness and improve the classifier training, log transform is performed for the frequency data by finding the logarithm of the signal amplitude in each frequency (Changyong et al. 2014).

Structure-Based Data Projection

The objective of structure-based data projection is to find a set of basis that projects the data into a feature space in which structural effects are minimized while keeping the footsteps and nonfootsteps separable. Due to reduced structural effects in this projected feature space, the footstep models trained in various structures are similar and successfully transfer between the structures. In real-life applications, minimizing only the structural effects might not be enough for model transfer because of the noise in the data and the possibility of ending up with a trivial solution. In these situations, even though the projected feature space has lower structural effect, the footsteps and nonfootsteps are not separable, and this results in low model performance. The following sections further discuss the structural effect minimization objective and then address the aforementioned limitations by adding additional terms to the objective function using the semisupervised transfer component analysis (SSTCA) (Pan et al. 2011) model transfer framework.

Minimizing the Structural Effect. To minimize the structural effect, the MMD across the source and target structures is minimized. Eqs. (12) and (15) have shown that the MMD between the source and target data distributions is correlated to the structural effect in the kernel space. MMD is a nonparametric distribution distance metric that does not make distributional assumptions about the data. Hence, it is more suitable to the present problem compared with alternative approaches [e.g., Kullback–Leibler (KL) divergence] because the distribution of the vibration data is unknown (Gretton et al. 2007). This MMD can be defined as $tr(\mathbf{KL})$, where $\mathbf{K}$ is the kernel matrix for the data set consisting of the data from the target and source structures and L is a coefficient matrix. To find a feature space with lowest structural effect, instead of assuming a fixed kernel, one can solve for the kernel matrix $\min_{\mathbf{K}\geq 0} tr(\mathbf{KL})$ through semidefinite programming (SDP) (Pan et al. 2008). However, there are three limitations using this formulation:

- SDP is computationally expensive.
- For each new unseen test sample in the target structure, a new kernel matrix needs to be computed, which adds to the computational cost.
- To reduce the dimension of the final projected feature space (and consider a subspace with lower structural effects), additional dimensionality reduction approach [such as principal component analysis (PCA)] is necessary (Pan et al. 2011).

To overcome these limitations, an alternative representation of the objective function is formed based on the SSTCA model

transfer approach (Pan et al. 2011). Assuming positive definite (and hence invertible) kernel, one can decompose $\mathbf{K}$ and rewrite the objective function

$$\min_{\mathbf{K}} tr((\mathbf{K}\mathbf{K}^{-1/2})(\mathbf{K}^{-1/2}\mathbf{K})\mathbf{L}) \tag{16}$$

This decomposition is generally known as empirical kernel map (Schölkopf et al. 1998). Decompose $\mathbf{K}\mathbf{K}^{-1/2} = \bar{\mathbf{K}}\bar{\mathbf{K}}^{-1/2}\tilde{\mathbf{W}}$ where $\bar{\mathbf{K}} \in \mathbb{R}^{(n_s+n_t)\times(n_s+n_t)}$ is a fixed kernel matrix [e.g., linear or radial basis function (RBF), among others] and $\tilde{\mathbf{W}} \in \mathbb{R}^{(n_s+n_t)\times(n_s+n_t)}$ is a weight matrix. This decomposition holds because $\mathbf{K}$, $\bar{\mathbf{K}}$, $\mathbf{K}^{-1/2}$, and $\bar{\mathbf{K}}^{-1/2}$ are positive definite and invertible. Similarly, decompose $\mathbf{K}^{-1/2}\mathbf{K} = \tilde{\mathbf{W}}^{\mathbf{T}}\bar{\mathbf{K}}^{-1/2}\bar{\mathbf{K}}$. Based on these decompositions, the objective function is rewritten

$$\min_{\tilde{\mathbf{W}}} tr((\bar{\mathbf{K}}\bar{\mathbf{K}}^{-1/2})\tilde{\mathbf{W}}\tilde{\mathbf{W}}^{\mathbf{T}}(\bar{\mathbf{K}}^{-1/2}\bar{\mathbf{K}})\mathbf{L}) \tag{17}$$

Next, simplify the equations by defining $\mathbf{W} = \bar{\mathbf{K}}^{-1/2}\tilde{\mathbf{W}}$, which results in

$$\min_{\mathbf{W}} tr((\bar{\mathbf{K}}\mathbf{W}\mathbf{W}^{\mathbf{T}}\bar{\mathbf{K}})\mathbf{L}) \tag{18}$$

Finally, the cyclic property of the trace can be used to rewrite the equation

$$\min_{\mathbf{W}} tr(\mathbf{W}^{\mathbf{T}}\bar{\mathbf{K}}\mathbf{L}\bar{\mathbf{K}}\mathbf{W}) \tag{19}$$

This new representation addresses the limitations of the previous objective function. First, this representation enables finding a close-form solution, which is computationally inexpensive because it does not require solving the SDP, as will be described subsequently. Second, for new samples, the corresponding kernel values can be computed and added to the kernel matrix and there is no need to resolve the optimization problem. Finally, the dimensions of $\mathbf{W}$ can be defined as $\mathbb{R}^{(n_s+n_t)\times m}$ to project the kernel data into a m-dimensional feature space. Therefore, additional dimensionality-reduction approaches might not be necessary.

However, only minimizing the distance between the distributions is not enough for successful model transfer. First, the derivation in Eqs. (3)–(15) is based on the assumption that the data are noise-free. Existence of noise in real life applications might result in noise-governing low-variance projected features spaces. Second, it results in the trivial solution for which all the data points are projected to the origin (i.e., zero distance between the distributions). In these cases, one ends up with a feature space in which the footsteps and nonfootsteps are mixed and not separable. To avoid these challenges, additional terms are added to the objective function.

Specifically, to avoid the challenges regarding the trivial solution and noisy data, a term is added to preserve the distance pattern of the samples (i.e., to ensure that the neighbor samples in the original data space are close to each other after projection). Furthermore, to improve the classification accuracy, another term is added to utilize the source label data by maximizing the dependence between the labels and projected data. The following sections discuss these additional terms, the final objective function, and the close-form solution to the objective function.

Maintaining the Data Variation. In the cases of trivial solution and noisy data, one might end up in projected spaces in which data have low variance. In these low-variance spaces, even though the distance between the distributions is small, the footsteps and nonfootsteps are mixed and inseparable, which in turn results in low classification accuracy. Therefore, the second term in SSTCA aims to maintain the variation and the distance pattern of the data through locality preservation. The intuition is that if there are two samples, $\mathbf{Y_i}$ and $\mathbf{Y_j}$, which are neighbors in the original data space, they should be close to each other after projection. To ensure locality preservation, first, the data are considered as a graph with affinity of $m_{ij} = \exp(-d_{ij}^2/2\sigma^2)$. This affinity is then used to form $\mathcal{N} = (\mathbf{Y_i}, \mathbf{Y_j})$, which is the set of sample pairs that are k-nearest neighbors of each other. Then, one can minimize

$$\sum_{i,j\in\mathbf{N}} m_{ij}\|[\mathbf{W}^{\mathbf{T}}\bar{\mathbf{K}}]_i - [\mathbf{W}^{\mathbf{T}}\bar{\mathbf{K}}]_j\|^2 \tag{20}$$

where $\|[\mathbf{W}^{\mathbf{T}}\bar{\mathbf{K}}]_i - [\mathbf{W}^{\mathbf{T}}\bar{\mathbf{K}}]_j\|^2$ = distance between samples $\mathbf{Y_i}$ and $\mathbf{Y_j}$ the projected feature space, respectively. Eq. (20) can be rewritten

$$tr(\mathbf{W}^{\mathbf{T}}\bar{\mathbf{K}}\mathcal{L}\bar{\mathbf{K}}\mathbf{W}) \tag{21}$$

where $\mathcal{L} = \mathbf{D} - \mathbf{M}$, where $\mathbf{M} = [m_{ij}]$ and $\mathbf{D}$ is a diagonal matrix with entries $d_{ii} = \sum_{j=1}^{n} m_{ij}$.

Using the Label Information. To improve the classification accuracy, the third objective aims to take advantage of the labeled data in the source structure. To this end, SSTCA maximizes the dependence of the projected data and the labels using the Hilbert-Schmidt independence criterion (HSIC) which is a nonparametric approach for estimating the dependence between two sets (Gretton et al. 2005). Specifically, SSTCA considering the HSIC between the original data set containing all of the data in source an target, D, and the label set, L, is

$$\text{HSIC}(\mathbf{D}, \mathbf{L}) = \frac{1}{(n-1)^2} tr(\mathbf{C}\bar{\mathbf{K}}\mathbf{C}\mathbf{K}_{\mathbf{L}}) \tag{22}$$

where $\mathbf{K_L}$ is a kernel matrix representing the label dependence; $\mathbf{C}$ is a centering matrix defined as $\mathbf{C} = \mathbf{I} - (1/n_s + n_t)\mathbb{1}\mathbb{1}^{\mathbf{T}}$; and n = number of all the samples in the source and target structure and is equal to $n_s + n_t$. Furthermore, $\mathbf{K_L}$ can be described

$$\tilde{\mathbf{K}}_{\mathbf{L}} = \gamma\mathbf{K}_{\mathbf{l}} + (1-\gamma)\mathbf{K}_{\nu} \tag{23}$$

where $[\mathbf{K_l}]_{ij}$ = kernel value between the ith and jth sample labels in the labeled source data; $\mathbf{K}_{\nu} = \mathbf{I}$; and γ = trade-off parameter. It has been shown empirically that $\gamma = 0.5$ works well on all the data sets (Pan et al. 2011). Replacing the kernel matrix definition from Eq. (19) in Eq. (22) and removing the constant coefficient results in

$$\text{HSIC}(\mathbf{D}, \mathbf{L}) = \mathbf{W}^{\mathbf{T}}\bar{\mathbf{K}}\mathbf{C}\tilde{\mathbf{K}}_{\mathbf{L}}\mathbf{C}\bar{\mathbf{K}}\mathbf{W} \tag{24}$$

Updating the Objective Function and Finding a Solution. Finally, the three objectives shown in Eqs. (19), (21), and (24) are combined to form the SSTCA optimization problem (Pan et al. 2011) which is

$$\begin{aligned} \underset{\mathbf{W}}{\text{minimize}} \quad & tr(\mathbf{W}^{\mathbf{T}}\bar{\mathbf{K}}\mathbf{L}\bar{\mathbf{K}}\mathbf{W}) + \mu tr(\mathbf{W}^{\mathbf{T}}\mathbf{W}) + \frac{\lambda}{n^2} tr(\mathbf{W}^{\mathbf{T}}\bar{\mathbf{K}}\mathcal{L}\bar{\mathbf{K}}\mathbf{W}) \\ \text{subject to} \quad & \mathbf{W}^{\mathbf{T}}\bar{\mathbf{K}}\mathbf{C}\tilde{\mathbf{K}}_{\mathbf{L}}\mathbf{C}\bar{\mathbf{K}}\mathbf{W} = \mathbf{I} \end{aligned} \tag{25}$$

where $\lambda \geq 0$ = trade-off parameter; $n^2 = (n_s + n_t)^2$ is a normalization term in which n_s and n_t are the number of samples in the source and target structures; and $tr(\mathbf{W}^{\mathbf{T}}\mathbf{W})$ = regularization term, which aims to control the complexity of $\mathbf{W}$ and avoid overfitting. The solutions of this problem are the eigenvectors of $(\bar{\mathbf{K}}(\mathbf{L} + \lambda\mathcal{L})\bar{\mathbf{K}} + \mu\mathbf{I})^{-1}\bar{\mathbf{K}}\mathbf{C}\tilde{\mathbf{K}}_{\mathbf{L}}\mathbf{C}\bar{\mathbf{K}}$ (Pan et al. 2011). The steps of structure-informed data projection are summarized in Algorithm 1. These eigenvectors are the dimension components that can be used for projecting the data. Furthermore, the order of the corresponding eigenvalues shows how well the data projected using the dimension components satisfy the objective function. Therefore, to project the data into a m-dimensional feature space, the m eigenvectors with leading eigenvalues are chosen for data projection. The number of

the projected data dimensions m affects the performance of the model transfer and is evaluated in the "Evaluation Metric and Model Parameters" section.

Algorithm 1. Structure-Informed Data Projection

$\mathbf{Y} \leftarrow [\mathbf{Y^s}; \mathbf{Y^t}]$
$n_s \leftarrow size(\mathbf{Y^s}, 1)$; $n_t \leftarrow size(\mathbf{Y^t}, 1)$
Estimate $\bar{\mathbf{K}}(\mathbf{Y})$ *(e.g., for linear kernel* $\bar{\mathbf{K}}(\mathbf{Y}) = \mathbf{YY^T}$*)*
for $i, j \leq n_s, n_t$ **do**
  if $\mathbf{Y_i}, \mathbf{Y_j} \in \mathbf{S}$ **then**
    $\mathbf{L_{ij}} \leftarrow \frac{1}{n_s^2}$
  else if $\mathbf{Y_i}, \mathbf{Y_j} \in \mathbf{T}$ **then**
    $\mathbf{L_{ij}} \leftarrow \frac{1}{n_t^2}$
  else
    $\mathbf{L_{ij}} \leftarrow -\frac{1}{n_s n_t}$
  end if
end for
$\mathbf{M} = [m_{ij}] \leftarrow \exp(-d_{ij}^2/2\sigma^2)$ ▷ For k-nearest neighbors
$\mathbf{D} = [d_{ii}] \leftarrow \sum_{j=1}^{n} m_{ij}$ ▷ $\mathbf{D}$ is Diagonal
$\mathcal{L} \leftarrow \mathbf{D} - \mathbf{M}$ ▷ *Laplacian Matrix*
Estimate the label kernel $[\mathbf{K_l}]_{ij}$.
$\mathbf{K}_\nu \leftarrow \mathbf{I}$
$\tilde{\mathbf{K}}_\mathbf{L} \leftarrow \gamma \mathbf{K_l} + (1-\gamma)\mathbf{K}_\nu$ ▷ *Label Dependence Matrix*
$\mathbf{C} \leftarrow \mathbf{I} - (1/n_s + n_t)\mathbf{1}\mathbf{1^T}$
$eig((\bar{\mathbf{K}}(\mathbf{L} + \lambda\mathcal{L})\bar{\mathbf{K}} + \mu\mathbf{I})^{-1}\bar{\mathbf{K}}\mathbf{C}\tilde{\mathbf{K}}_\mathbf{L}\mathbf{C}\bar{\mathbf{K}})$
return the first m eigenvectors

Footstep Classifier Training in Source. To distinguish the footstep-induced vibrations from vibration caused by nonfootsteps, binary classifiers are used. Without loss of generality, SVM, a common classification approach that does not make distributional assumption about the data, is used. Instead, SVM maximizes the distance (i.e., margin) of the sample points to the decision boundary (Friedman et al. 2001). After finding the suitable $\mathbf{W}$, the source and target data aare projected by finding the $\mathbf{W}\bar{\mathbf{K}}$. In this projected space, the structural effects on the vibration data is minimized, and the source and target data distributions are similar. Therefore, the SVM classifier is trained on the labeled projected data from the source structure or structures and is used for predicting the target structure data labels. The output of the SVM classifier for each vibration sample is a score whose value and sign represent the distance of the sample to the decision boundary and whether the sample is a footstep or a nonfootstep, respectively.

Multistructure Sample Labeling. When there are more than one source structures available, each one of them is used for data projection and developing SVM classifiers for the target structure. To improve the accuracy and robustness of footstep classification (and occupant detection), the scores are combined across these classifiers. To this end, the SVM scores for target samples from these classifiers from all the source structures are added. The sign and magnitude of the final score represents the predicted label (footstep or nonfootstep) and the prediction confidence.

Occupant-Detection Evaluation

To evaluate the performance of the proposed model transfer for step-level occupant detection, a set of experiments were conducted with two human participants and in real-world structures. First, the experimental setup is introducecd and the metrics used for evaluation are defined. For evaluation, the step-level occupant-detection results are presented and analyzed based on the whole approach (as shown in Fig. 3), which combines the results of multiple source structures. In addition, the model transfer part is analyzed without combining the results for multiple sources to show that transferring the model improves the classification performance regardless of the choice of the source structure. To this end, the model transfer performance is evaluated for single-source structure cases. Finally, the approach is also evaluated for sensitivity to the number of available data in the target structure and in the source structures, as well as sensitivity to the dimension of the projected feature space.

Experimental Setup in Three Buildings

To explore the effect of using various types of structure as the source and target structure, experiments were conducted in three types of structures, as shown in Fig. 5. These three structures, located in Pittsburgh, Pennsylania, include (1) a carpeted wooden floor in the third floor of the Baptist senior care, (2) a carpeted concrete hallway on the ground level of the Porter Hall, which is a campus building at Carnegie Mellon University, and (3) a carpeted metal deck floor on the second floor of the Vincentian senior care facility. The first observed natural frequency of these structures are 16.02, 23.83, and 14.84 Hz, respectively. These natural frequency estimations were obtained using the basic frequency domain (BFD) or peak-picking approach (Brincker et al. 2000), where the first observed natural frequency is identified as the first peak in the Fourier transform of ambient vibration data. The varying natural frequency of the three structures reinforces the theory that their vibration responses will vary and justifies the use of the proposed model transfer approach.

Furthermore, to evaluate how accurate the proposed approach is in distinguishing footsteps from other impulsive excitations, data were collected on footsteps as well as other impulsive excitations, such as door closings, dropping objects (ball and keychain) on the floor, and hammer striking in the experiments. Specifically, 220, 100, and 290 footsteps and 130, 70, and 190 nonfootsteps in Baptist, Porter, and Vincentian, respectively, were included. The impact locations are within 5 m around the sensor. Fig. 6 presents the experimental setup used for all the structures.

Geophone sensors were used to measure the structural vibration. Geophones are low-cost and low-distortion sensors that measure the vertical velocity of floor vibration (I/O Sensor Nederland bv 2006). Fig. 7 shows an example of the sensing node. LMV358 operational amplifiers (Texas Instruments 2014) were used to amplify the geophone vibration measurements by the orders of 200× –2,000× to improve the resolution of the signals while reducing the signal clipping. The geophone sensor has a frequency range of 10–240 Hz and sensitivity of 28.8 V/m/s (I/O Sensor Nederland bv 2006). The experiments used a 25-kHz sampling frequency for data collection. The reason for choosing this high sampling frequency is that the detected footsteps will be used with other applications such as occupant localization, which requires high temporal resolution, as discussed by Mirshekari et al. (2018b). After amplification, which varies depending on the structure type and footstep strike energy, the effective sensing range of this system for footstep detection is up to 20 m in diameter. Amplified signals are then digitized and transferred to a server using a 24-bit analog/digital (A/D) converter.

Evaluation Metric and Model Parameters

As the performance metric, the F1 score was used, which has been commonly used for evaluating classification algorithms. F1 score is

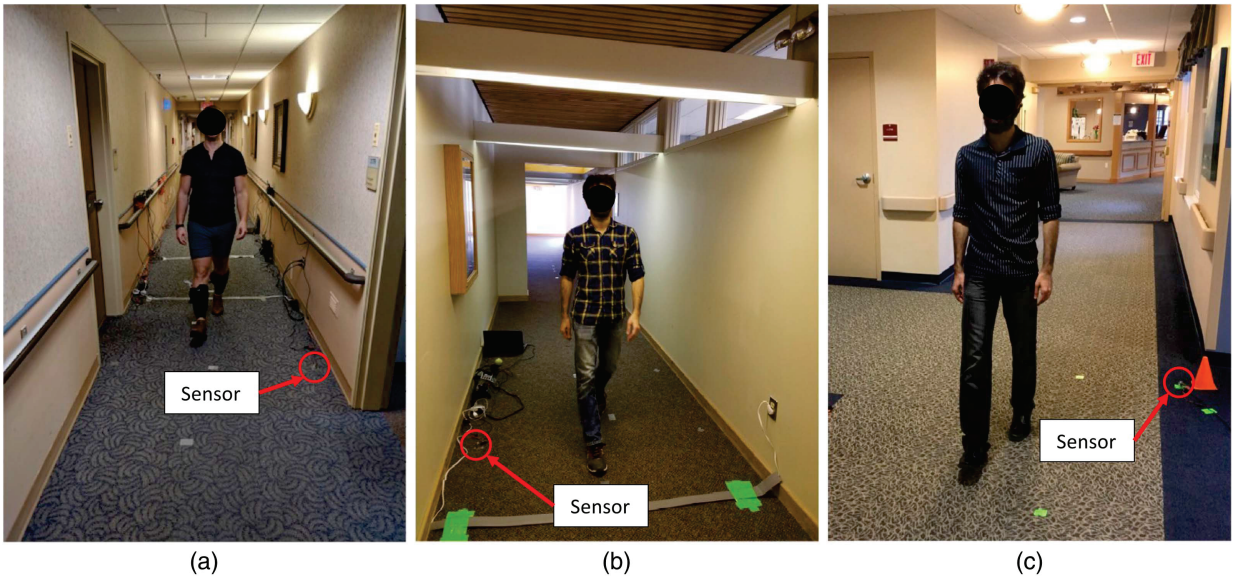

Fig. 5. Experiment locations: (a) Baptist senior care facility; (b) Carnegie Mellon University (CMU) Porter Hall A18; and (c) Vincentian senior care facility.

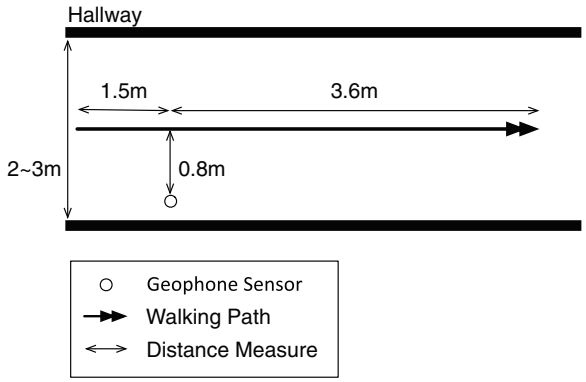

Fig. 6. Experimental setup: sensor configuration and footstep trace utilized in all the structures.

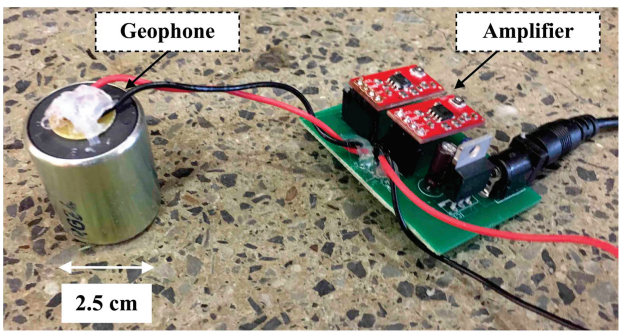

Fig. 7. Sensing node. The geophone measures the vertical velocity of vibrations. These measurements are amplified and transferred to a server for further analysis.

the harmonic mean of the precision and recall rate and can be found through the following equation (Sasaki et al. 2007):

$$\text{Recall} = \frac{\text{TruePositives}}{\text{TruePositives} + \text{FalseNegatives}} \tag{26}$$

$$\text{Precision} = \frac{\text{TruePositives}}{\text{TruePositives} + \text{FalsePositives}} \tag{27}$$

$$F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{28}$$

where TruePositives = number of correctly detected footsteps; FalsePositives = number of nonfootstep mistakenly detected as footsteps; and FalseNegatives = number of missed footstep events. The F1 score is estimated using the k-fold cross-validation approach using $k = 10$. Using the fitcsvm MATLAB version R2018b function (MATLAB 2018), the training of the SVM model for 300 samples takes approximately 0.009 s using a MacBook pro with 8 gigabytes RAM and 2.7 GHz Intel Core i5 processor. With respect to the SVM parameters, the regularization term C (or the box constraint) is set as 1. A simple linear kernel is used to ensure that the SVM model does not overfit to the data in the source structure. Further, the software divides the input data by an appropriate scale factor, which is estimated using a heuristic procedure, before applying the kernel (MATLAB 2018).

Overall Footstep Classification Robustness

This section evaluates the accuracy and robustness of the overall footstep classification approach. To this end, the robustness of the proposed approach is discussed and compared with two baseline approaches in three structures. The two baseline approaches utilize the time-domain (TD) and frequency-domain (FD) representations of the signal. Both TD-based and FD-based baseline approaches first train an SVM classifier in one structure and then test the model in a different structure (i.e., without model transfer). Then, each sample in the target structure is labeled by combining the SVM scores from multiple sources.

Fig. 8(a) shows the comparison of the F1 score for each structure as the target. For the results of each structure, the other two structures are the source structures. For example, the Baptist results represent the case with Baptist as the target structure and Porter and Vincentian as the source structures. Based on these results, in the Baptist nursing home, the proposed approach has resulted in 0.96 F1-score, which is equivalent to a 9.25× and 7.5× reduction in

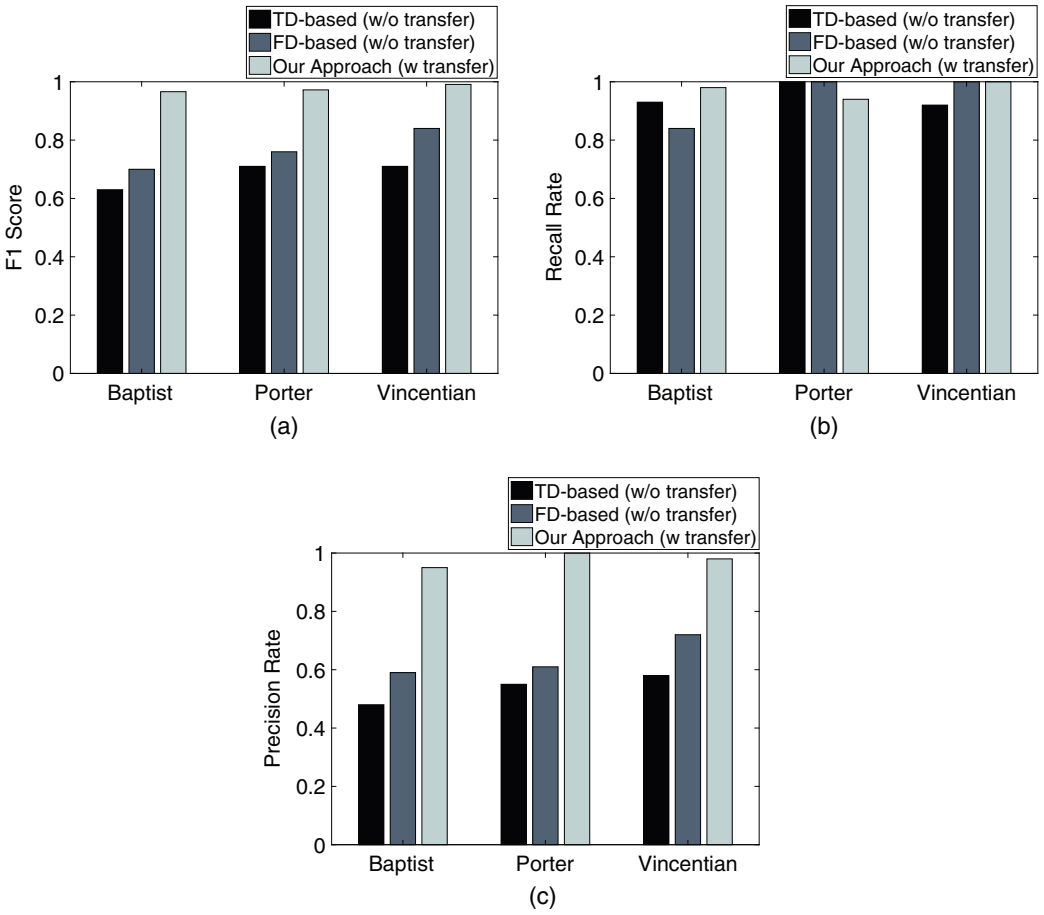

Fig. 8. Evaluation results in three structures: (a) F1-score; (b) recall rate; and (c) precision rate are compared using the proposed transfer-based approach with two methods that utilize the time-domain and frequency-domain signals in the source without transfer to predict the sample labels in the target structure.

classification error over the TD-based and FD-based approaches, which resulted in 0.63 and 0.7 F1 score, respectively. Similarly, in Porter Hall, the proposed approach results in 0.97 F1-score compared with 0.71 and 0.76 F1-scores using the TD-based and FD-based approach, which is equivalent to a 9.7× and 8× reduction in error, respectively. Finally, in Vincentian nursing home, the proposed approach results in a F1-score of 0.99 compared with the F1-scores of 0.71 and 0.84 using the TD-based and FD-based approaches, equivalent to 29× and 16× reductions in error, respectively.

These higher improvements in Vincentian are due to larger separation between the footsteps and nonfootsteps in this structure. To explore this effect, a metric was defined as the mean of $\mu_f - \mu_{nf}$ where μ_f and μ_{nf} are the mean of the fft of footstep and nonfootstep samples, respectively. This separation metric is 0.029 in Vincentian compared with 0.019 in Baptist and 0.018 in Porter. The models trained in the structures with smaller separation metric contain more information about the lower-confidence samples, which are closer to the decision surface compared to the models trained in structures with larger separation metric. Therefore, the models trained in Baptist and Porter, which have smaller separation metric, perform well in the Vincentian, which has larger separation metric. This effect will be described in more detail in the next section, which discusses model transfer with one source structure.

To provide more detail, Figs. 8(b and c) show the evaluation results for recall and precision rate, respectively. These figures show that the baseline approaches generally have higher recall rate than precision rate. Specifically, the recall rate is between 0.92 and 1 for TD-based and between 0.84 and 1 for FD-based approaches. In comparison, the precision rate is between 0.48 and 0.58 for the TD-based and between 0.59 and 0.72 for FD-based approach. High recall rate and low precision rate implies that the baseline approaches cause a large number of false positives and detect nonfootsteps as footsteps. This high false positive rate is caused by the higher variation in the footstep-induced excitations, which results in distribution imbalance between the footstep and nonfootsteps. In this case, the model will be biased toward classifying the samples as footsteps because the higher-variance footstep class account for more of the data space than the lower-variance nonfootstep class.

In comparison, the proposed approach results in recall rate of 0.98, 0.94, and 1 and precision rate of 0.95, 1, and 0.98 for Baptist, Porter, and Vincentian, respectively. Consistent improvement in the F1-score in all the three structures shows that the proposed approach is more robust to the changes in the structure. Furthermore, the combined high recall and high precision rates using the proposed approach shows that it is more robust to the footstep variations.

Model Transfer Evaluation for Different Source Structures

By reducing structural effects through ensuring a similar data distribution in the target and source structures, the structure-informed model transfer enables footstep modeling and classification in the target structure with no labeled data. This section focuses on the model transfer performance by comparing the results of the proposed approach with the TD-based and FD-based baseline

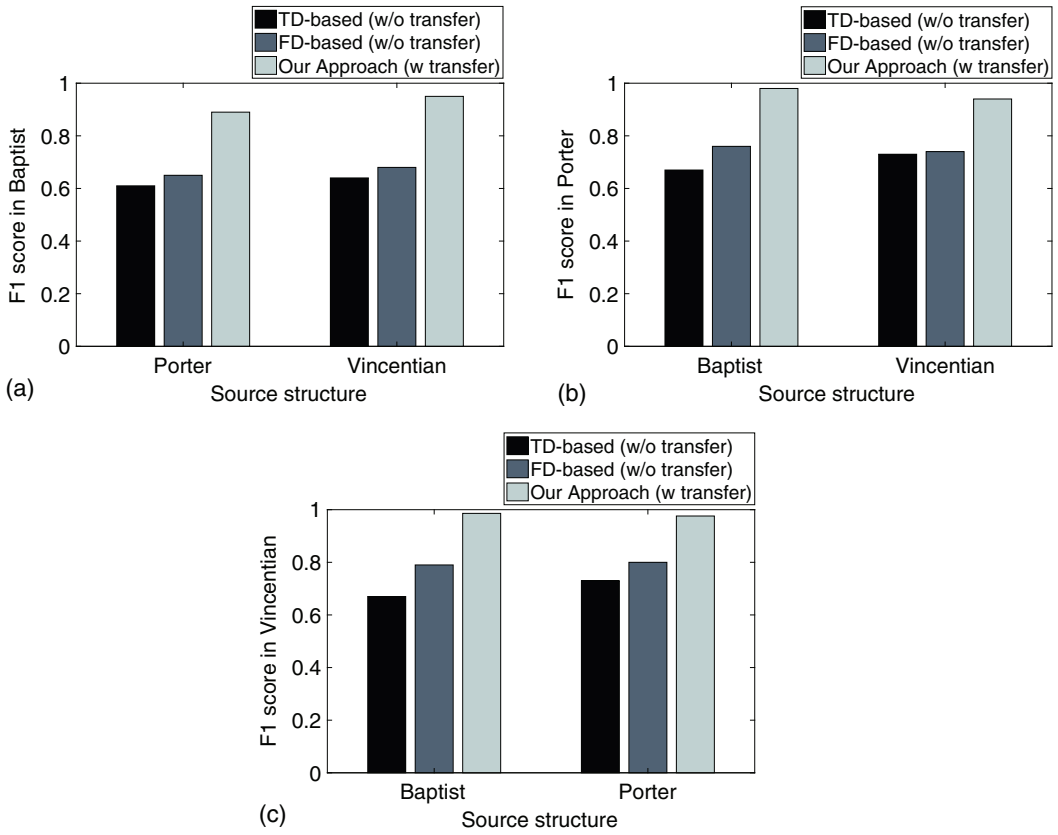

Fig. 9. F1 score using one structure as the source: (a) Baptist building; (b) Porter Hall; and (c) Vincentian building. F1-scores for the proposed transfer-based approach are compared with the baseline approaches when different source targets are utilized.

approaches when there is only one source structure. Figs. 9(a–c) shows the results in Baptist, Porter, and Vincentian, respectively.

Fig. 9(a) shows the F1-scores for the Baptist as the target structure. It can be seen that when the source structure is Vincentian, the proposed approach achieves a F1-score of 0.95 compared with 0.61 and 0.65 using the TD-based and FD-based approaches, respectively, which is a 7.8× and 7× improvement in error. On the other hand, when the source structure is Porter, the proposed approach results in F1-score of 0.89 compared with 0.64 and 0.68 using the TD-based and FD-based approaches, equivalent to a 3.3× and 2.9× improvement in error.

Fig. 9(b) shows the F1-scores for the case with Porter as the target structure. The proposed approach achieves a F1-score of 0.98 compared with 0.67 and 0.76 using the TD-based and FD-based approaches, which is equivalent to a 16.5× and 12× improvement in error when the source structure is Baptist. On the other hand, when the source structure is Vincentian, the proposed approach results in F1-score of 0.94 compared with 0.73 and 0.74 using the TD-based and FD-based approaches, which correspond to a 4.5× and 4.3× improvement in error, respectively.

Fig. 9(c) shows the F1-scores for the case with Vincentian as the target structure. In this case, when the source structure is Baptist, the proposed approach results in a F1-score of 0.99 compared with 0.67 and 0.79 using the TD-based and FD-based approaches, which is equivalent to a 33× and 21× improvement in error. On the other hand, when the source structure is Porter, the proposed approach results in F1-score of 0.98 compared with 0.73 and 0.8 using the TD-based and FD-based approaches, which is equivalent to a 13.5× and 10× improvement in error, respectively.

The first observation in these results is that the model performance is higher when the Vincentian is the source structure compared with when it is the target structure. As discussed in the previous section, the reason behind this effect is the larger separation between the footsteps and nonfootsteps in Vincentian. Specifically, the separation metric is 0.029 in Vincentian compared with 0.019 in Baptist and 0.018 in Porter. Models with a smaller separation metric contain more information about the low-confidence samples, which are closer to the decision surface. Therefore, the models trained in Baptist and Porter perform well in Vincentian (which is shown by F1-score of 0.99 and 0.98, respectively). However, the models trained in Vincentian have lower performance in Baptist and Porter (shown by F1-scores of 0.95 and 0.94, respectively).

The second observation is that using the Baptist as the source and Porter as the target outperforms using Porter as the source and Baptist as the target. This is true even though the separation metric is similar in Baptist and Porter. The reason behind this observation is that the heterogeneity of the wooden floor in Baptist results in higher variance in the data distribution. To explore this effect, a metric was defined as the mean of the σ/μ where σ and μ are the standard deviation and the mean of the frequency representation (FFT) of the samples in each structure. The values of this metric are 1.29, 1.06, and 1.05 in Baptist, Porter, and Vincentian, respectively. The higher value of the metric show the higher data variance and heterogeneity of the data from Baptist. Therefore, the model trained in Baptist is more informative than the model trained in Porter and works well in Porter (F1-score of 0.98), whereas the model trained in Porter has lower performance in Baptist (F1-score of 0.89).

Using more than one source structure improves the robustness of the results. In other words, the combined source structure case outperforms the source structure with lower model performance; however, the improvement over the source structure with higher

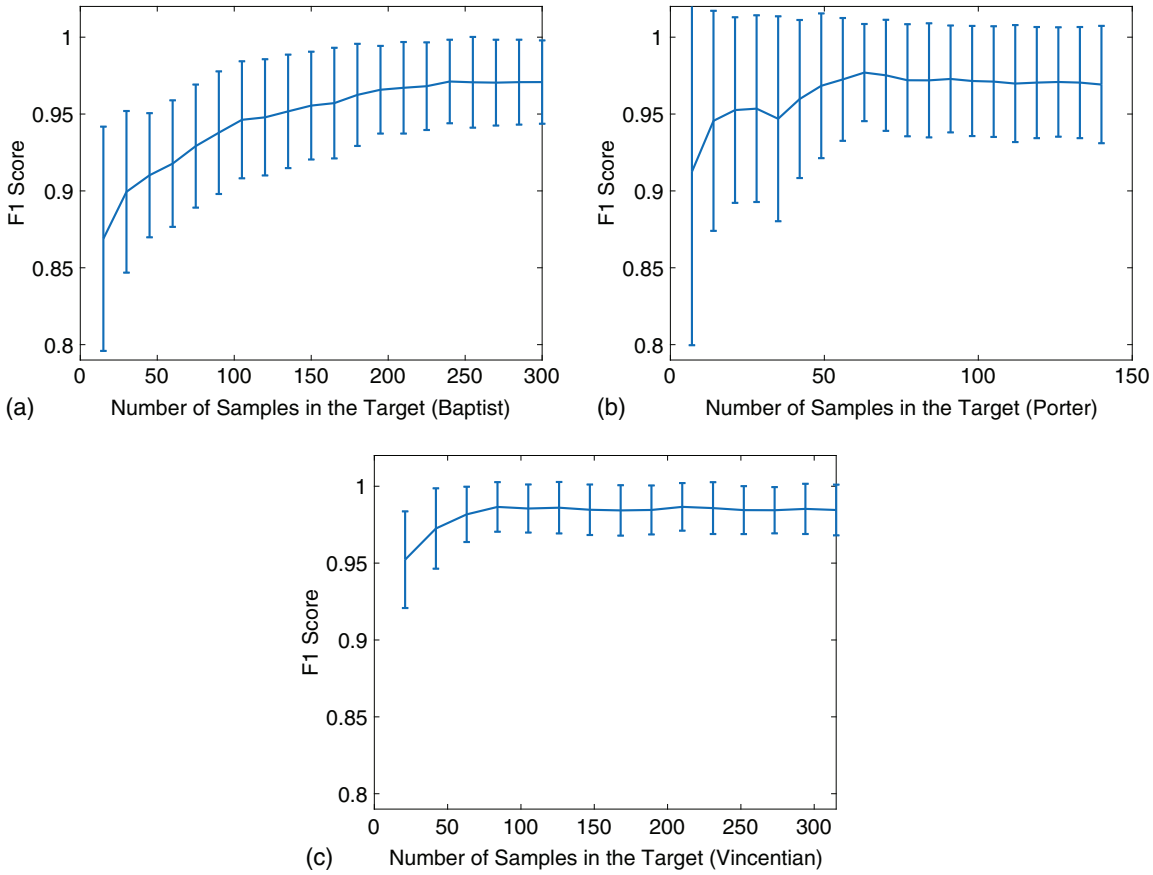

Fig. 10. Sensitivity of F1-score to the amount of data in the target structures for cases with (a) Baptist building; (b) Porter Hall; and (c) Vincentian building as the target structure, respectively. As expected, in general, increasing the amount of target data results in better performance (i.e., higher F1-score) and lower standard deviation of F1-scores.

performance is not substantial. Specifically, for Baptist, Vincentian, and Porter, the F1-scores are 0.96, 0.99, and 0.97 when using two source structures, 0.89, 0.98, and 0.94 when using the source structure with lower performance, which is equivalent to a 2.7×, 2×, and 2× improvement, respectively. However, the F1-scores when using the source structure with the higher performance are 0.95, and 0.99, and 0.98 for Baptist, Vincentian, and Porter which are equivalent to a 1.25×, 0, and −1.5× change in the error, respectively.

In summary, although model transfer performance depends on the source structure, the proposed model transfer approach results in a 0.89–0.99 F1-score and 2.9 × –33× improvement in the error compared with the baseline approach in different cases of source and target structures. Therefore, it successfully transfers the footstep model between the structures.

Sensitivity to the Amount of Target Data

The amount of the available data in the target structure potentially affects the performance of the model transfer approach. To evaluate this effect, 10% of the target data were kept as the test data. Then, among the rest of the target data, varying amounts of the remaining target data (unlabeled) were selected and used with the source data (labeled) for training. The trained model is then used for labeling the test samples and finding the F1-scores.

Figs. 10(a–c) show the results of this evaluation. As expected, having a data set with more target data results in higher F1-scores in the target structure because it better represents the target structure data distribution in model transfer. Specifically, in Baptist, Porter, and Vincentian, the F1-score increased from 0.87 and 0.91, and 0.95 to 0.97, 0.97, and 0.99, respectively. The number of samples necessary to reach to the maximum accuracy is 240, 55, and 85 samples in Baptist, Porter, and Vincentian, respectively. The reason behind the need for more target data in Baptist is the higher variance of the target data distribution caused by higher structural heterogeneity in Baptist, as discussed in the previous section. Higher structural heterogeneity results in different characteristics for the impulse signals in different locations of the structure. Therefore, more samples are necessary to represent the data distribution in the Baptist location.

Furthermore, having more target data decreases the variance of the estimated F1-scores and hence increases the robustness of the footstep model in the target structure. These variances for Baptist, Porter, and Vincentian are reduced from 0.07, 0.11, and 0.03 to 0.03, 0.04, and 0.016, respectively. These evaluation results can be used for determining the amount of target data necessary for different applications.

Sensitivity to Amount of Source Data

One of the factors affecting the performance of the model transfer approach is the amount of the data available in the source structure. To evaluate this effect, a varying amount of the source data were kept and the F1-scores in the target data were estimated. Furthermore, to find the variance of the F1-score, the analysis was repeated for 10 subsets of the source data. Figs. 11(a–c) show the results of evaluation. For the sake of comparison, these figures also present the same evaluation for the TD-based and FD-based baseline approaches.

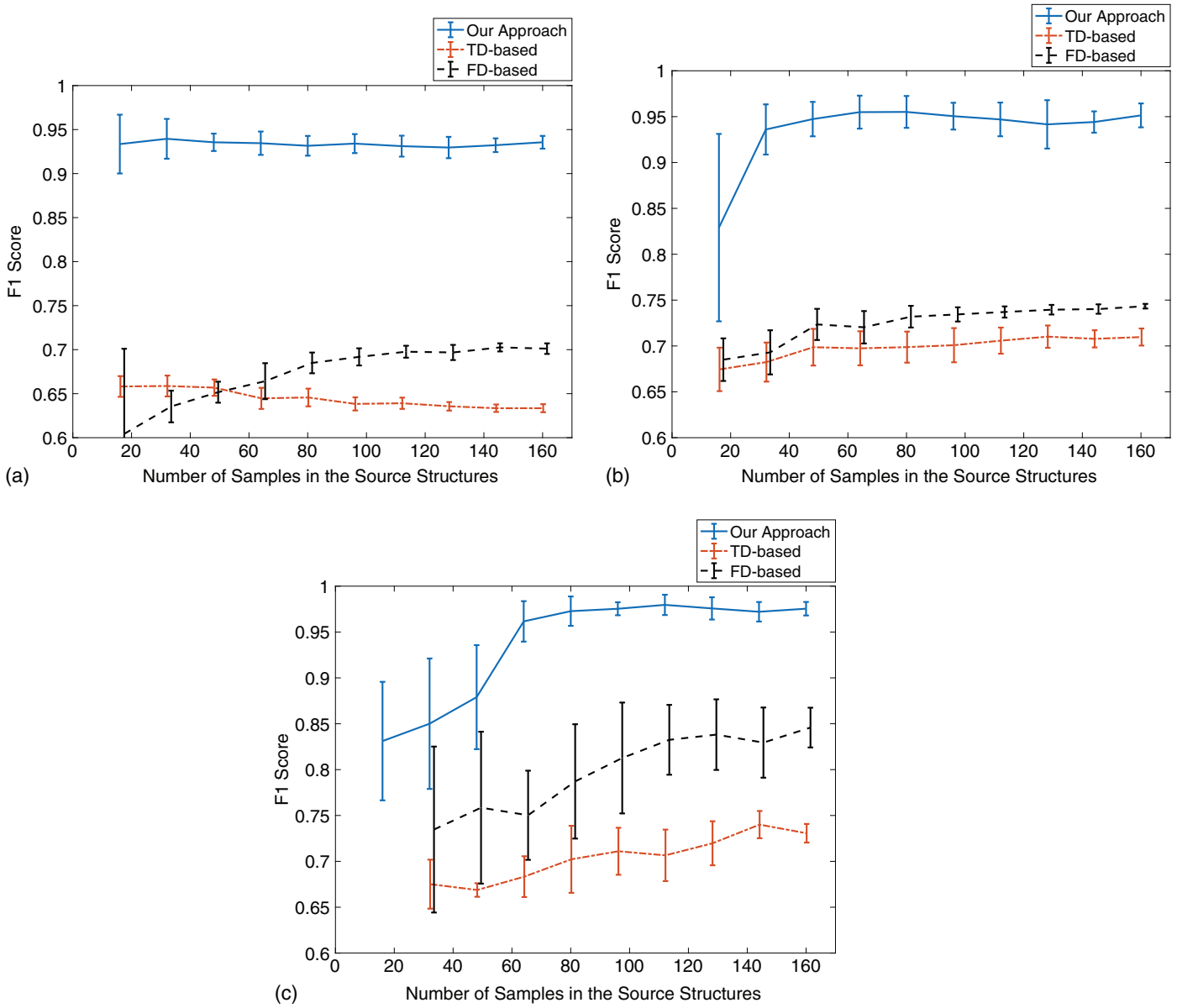

Fig. 11. Sensitivity of F1-score to the amount of data in the source structures for cases with (a) Baptist building; (b) Porter Hall; and (c) Vincentian building as the target structure, respectively. As expected, in general, increasing the amount of source data results in better performance (i.e., higher F1-score) and lower standard deviation of F1-scores.

Having a data set with more source data generally results in higher F1-scores in the target structure because it better represents the source structure data distribution. In Vincentian, the F1-score is increased from 0.83 from cases with 15 samples to 0.98 for cases with 150 samples, respectively. In Porter, the F1-score is increased from 0.83 from cases with 15 samples to 0.95 for cases with 150 samples, respectively. As discussed in the previous section, Vincentian and Porter, which are the source structures for Baptist results, are less heterogeneous. Therefore, high accuracy is achieved even with lower number of source samples for Baptist.

Furthermore, having more source data decreases the variance of the estimated F1-scores and hence increases the robustness of the footstep model in the target structure. These variances for Baptist, Porter, and Vincentian are reduced from 0.03, 0.1, and 0.065 for cases with 15 source samples to 0.007, 0.01, and 0.007 for cases with 150 samples. These results show that using 50 or more samples from the source structure results in the F1-score greater than 0.9.

Sensitivity to the Projected Feature Space Dimension

An important factor in the model transfer performance is the number of dimensions in the projected feature space. The dimension components for model transfer approach are ordered with respect to the objective function. Therefore, the effect of the number of dimensions on model transfer is a trade-off between better satisfaction of the objective function and higher model flexibility. On the one hand, as the number of dimensions is increased, components are added for which the objective function is less satisfied, and this might decrease the model performance. On the other hand, using more dimensions can potentially result in higher model flexibility and performance.

The model performance was evaluated through finding the F1-score in the target structure using different numbers of dimensions. Figs. 12(a–c) show the results of this evaluation. The general trend of F1-scores in all the three structures shows an initial increase, which is caused by higher model flexibility using

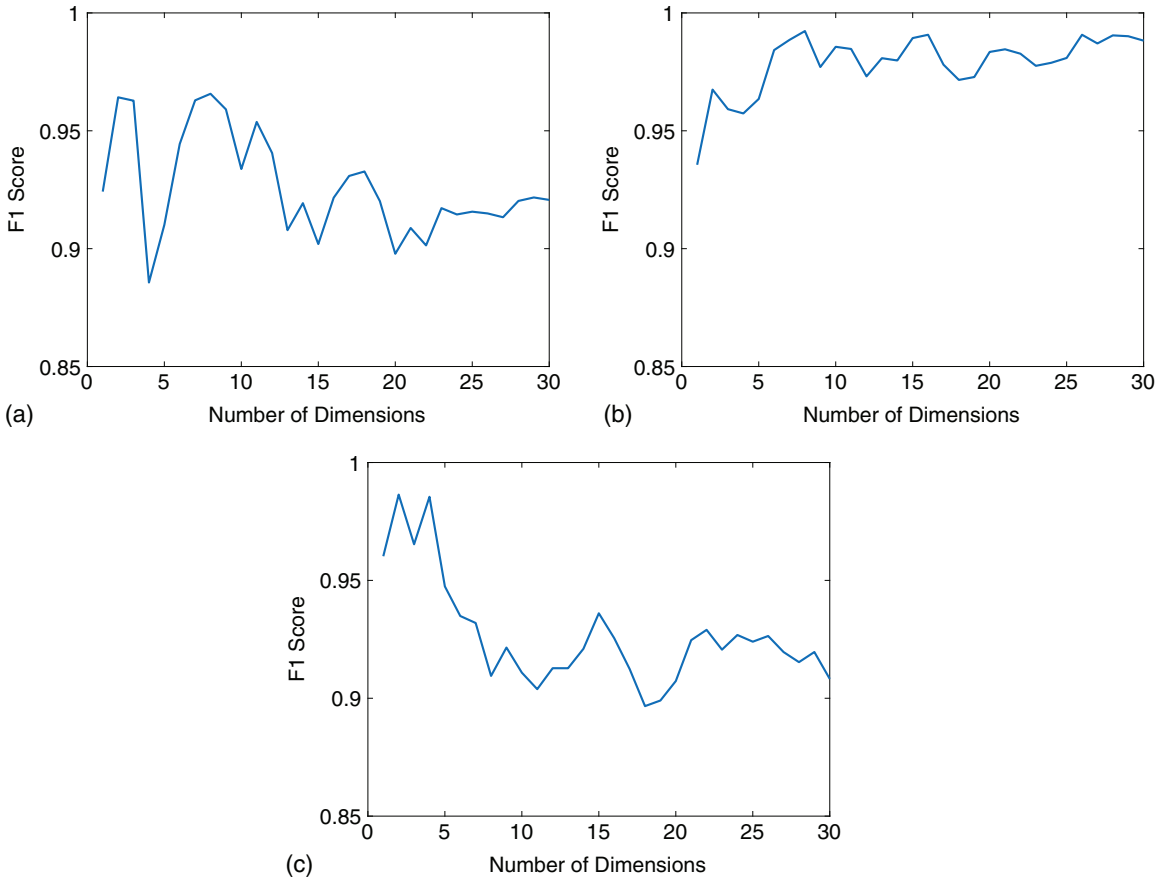

Fig. 12. Sensitivity of F1-score to the number of dimensions for cases with (a) Baptist building; (b) Porter Hall; and (c) Vincentian building as the target structure, respectively.

more dimensions. However, after the initial increase, the F1-score for Vincentian and Baptist decrease, whereas in Porter Hall, the F1-score does not show significant change.

To understand this trend, the cumulative normalized eigenvalues are depicted in Fig. 13. As discussed when describing Eq. (25), these eigenvalues show how well the data projected using a specific dimension's components satisfy the objective function. To make the eigenvalues comparable across the structures, they were normalized by dividing them by the summation of the eigenvalues for each structure. The range of the normalized eigenvalues is between 0 and 1, where higher values show higher contribution of the dimension component. This figure shows that the first few dimensions have more contribution compared with Baptist and Vincentian. For example, four initial dimensions count for 0.82 of the total value in Porter. The corresponding values for the same number of dimensions in Baptist and Vincentian are 0.76 and 0.68. The higher contribution of the initial components with lower noise results in a model that is less affected by the later noisier dimensions. Therefore, in Porter, the performance of the model does not decrease by adding more dimensions. Furthermore, the largest decrease is happening in Vincentian, where the initial dimensions show the lowest contribution.

Furthermore, for Baptist, the trend is mostly decreasing after the initial part. However, there is an outlier with low F1-score for the case of four dimensions, which means that the fourth basis projects the data into a feature space in which the footsteps and nonfootsteps are not separable. This effect is caused by higher noise in the wooden floor, which results in some errors in the model transfer. In this paper, to ensure good performance in all the structures, the first two dimensions were chosen for model transfer.

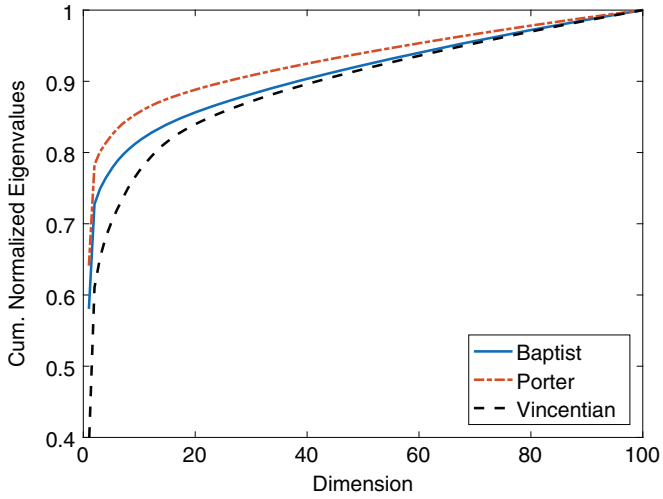

Fig. 13. Dimension contribution through normalized eigenvalue.

Sensitivity to the Signal-to-Noise-Ratio

The existence of environmental and measurement noise can affect the model performance. To evaluate this factor, the relationship between the signal-to-noise ratio (SNR) and the model F1-score was studied. Specifically, this relationship for the proposed model transfer approach was compared with the ones using the TD-based and FD-based approaches. For each impulsive event, SNR values are computed as the ratio of the summed squared magnitude of the

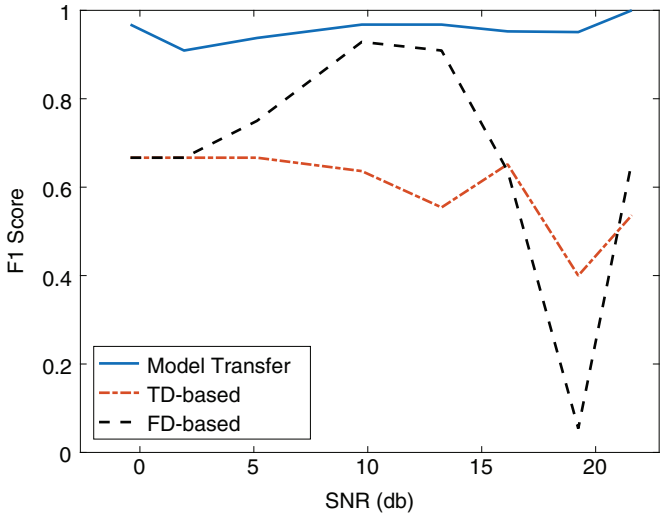

Fig. 14. Sensitivity to the SNR. The proposed model transfer approach has higher F1-score compared to the baseline approaches in various SNR ranges.

event signal to that of the noise (of the same length) and are described in decibels. Higher SNR values indicate high level of event signal and low level of noise, and vice versa. The noise consists of ambient vibration measurements, which does not include impulsive excitation and footstep events.

Fig. 14 shows the F1-score for 10 different levels of SNR. The results show that the proposed approach outperforms other approaches in all SNR ranges, with F1-scores between 0.91 and 1.00. In comparison, the baseline TD-based and FD-based approaches result in F1-score ranges of [0.4, 0.63] and [0.05, 0.92], respectively. Consistent improvement in the performance using the proposed approach compared to the baseline approaches shows that the proposed approach is more robust to the changes in the SNR.

Discussion of Applicability of Vibration-Based Sensing

This paper has introduced a vibration-based step-level occupant-detection approach that is robust across various structures. Fig. 15 shows an example deployment in elder monitoring. In this deployment, multiple sensors are used to enable further evaluation. However, in practice, only one geophone sensor is needed to measure and detect the footsteps in each sensing range. As discussed previously, the sensing range of these sensors are up to 20 m in common structure types, including steel, concrete, and wooden structures. In real-life applications, the exact range may vary by a few meters depending on the floor type and footstep forces. Further, the commercial scale price of these nodes is estimated to be USD 30, including processing units and sensors (Pan et al. 2017a).

Conclusion

This paper has presented step-level occupant detection that enables detection across different buildings using footstep-induced floor vibrations. Step-level occupant detection is important in various smart building applications such as senior/healthcare and energy management. To ensure successful transfer across various structures, the approach first projects the original frequency-based features into a feature space in which the structural effects are minimized. It has been shown that reducing structural effects can be achieved through minimizing the MMD between the source and target data distributions. Then, a footstep model was trained in this projected feature space to predict the samples in the target structure. The proposed approach was evaluated in three structures and through comparing with time-domain-based and frequency-domain-based approaches, which do not transfer the data. The proposed approach results in up to 29× and 16× improvement compared to the baseline TD-based and FD-based approaches, respectively. By providing a method to transfer models between different structures, the proposed system can significantly reduce labor requirement for occupant sensing in future smart buildings.

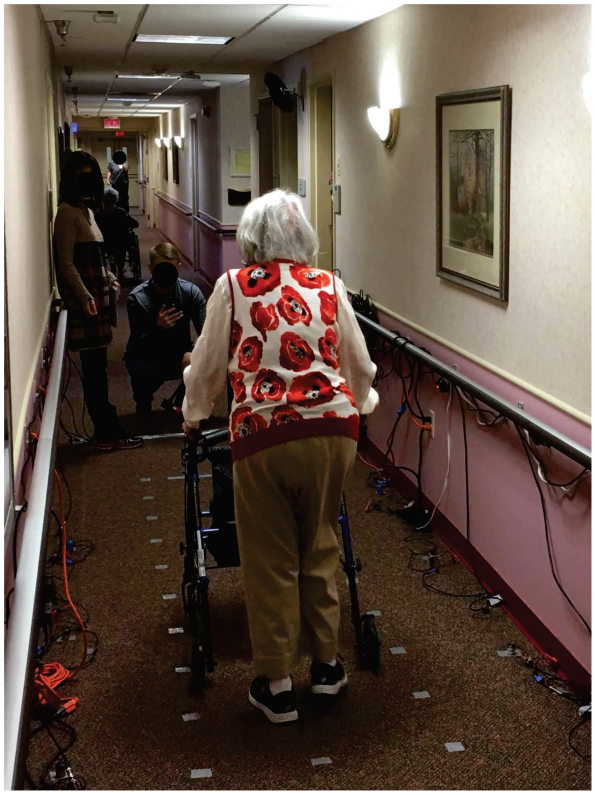

Fig. 15. Example real-life implementation for elder monitoring.

Acknowledgments

This research was partially supported by NSF (CNS-1149611 and CMMI-1653550), Google, Intel, and Highmark. The authors would also like to acknowledge Vincentian and Baptist Nursing Homes for providing deployment sites to conduct experiments and collect data.

Notation

The following symbols are used in this paper:

C = centering matrix;

$\mathbf{H}$ = frequency response function (FRF) of the structure for all the samples;

$\mathbf{H_i}$ = diagonalized frequency response function (FRF) of the structure for the ith sample;

$\mathbf{K}$ = general kernel matrix across the source and target structures;

$\bar{\mathbf{K}}$ = fixed kernel matrix across the source and target structures (e.g., linear kernel);
$\tilde{K}_L$ = label kernel matrix for labeled source data;
$\mathbf{L}$ = matrix of coefficients;
$\mathcal{L}$ = graph Laplacian matrix;
n_b = number of frequency bins;
n_s, n_t = number of samples in the source and target structures;
$\mathbf{S}$, $\mathbf{T}$ = source and target data set (distribution);
$\mathbf{W}$ = feature projection matrix;
$\mathbf{X_i}$ = input force spectrum for the ith sample;
$\mathbf{X}$ = input force spectrum for all the samples;
$\mathbf{Y_i^t}$, $\mathbf{Y_i^s}$ = ith samples in the target and source structures;
$\mathbf{Y_i}$ = vibration frequency representation for the ith sample;
σ_w^2 = variance of the signal window;
σ_n^2 = variance of noise;
ϕ = kernel-induced feature map;
χ_w^2 = chi-squared statistics for each signal window;
χ_α^2 = chi-squared statistics for significance level of α; and
λ, μ, γ = trade-off parameters.

References

Alyamkin, S., and S. Eremenko. 2011. "Pedestrian detection algorithms based on an analysis of the autocorrelation function of a seismic signal." *Optoelectronics Instrum. Data Process.* 47 (2): 124–129. https://doi.org/10.3103/S8756699011020038.

Baron, M. 2013. *Probability and statistics for computer scientists*. Boca Raton, FL: CRC Press.

Ben-David, S., J. Blitzer, K. Crammer, and F. Pereira. 2007. "Analysis of representations for domain adaptation." In *Proc., 19th Int. Conf. on Neural Information Processing Systems*, 137–144. Cambridge, MA: MIT Press.

Bickel, S., M. Brückner, and T. Scheffer. 2007. "Discriminative learning for differing training and test distributions." In *Proc., 24th Int. Conf. on Machine Learning*, 81–88. New York: Association for Computing Machinery.

Blitzer, J., R. McDonald, and F. Pereira. 2006. "Domain adaptation with structural correspondence learning." In *Proc., 2006 Conf. Empirical Methods in Natural Language Processing*, 120–128. Stroudsburg, PA: Association for Computational Linguistics.

Brincker, R., L. Zhang, and P. Andersen. 2000. "Modal identification from ambient responses using frequency domain decomposition." In *Proc., 18th Int. Modal Analysis Conf.* Aalborg, Denmark: Univ. of Aalborg.

Changyong, F., W. Hongyue, L. Naiji, C. Tian, H. Hua, and L. Ying. 2014. "Log-transformation and its implications for data analysis." *Shanghai Arc. Psychiatry* 26 (2): 105.

Dai, W., G.-R. Xue, Q. Yang, and Y. Yu. 2007. "Transferring naive bayes classifiers for text classification." In Vol. 1 of *Proc., 22nd National Conf. on Artificial Intelligence*, 540–545. Palo Alto, CA: AAAI Press.

Fagert, J., M. Mirshekari, S. Pan, P. Zhang, and H. Y. Noh. 2017a. "Characterizing left-right gait balance using footstep-induced structural vibrations." In Vol. 10168 of *Proc., Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*. Bellingham, WA: International Society for Optics and Photonics.

Fagert, J., M. Mirshekari, S. Pan, P. Zhang, and H. Y. Noh. 2017b. "Monitoring hand-washing practices using structural vibrations." In *Proc., Structural Health Monitoring 2017*. Lancaster, PA: DEStech Publications.

Friedman, J., T. Hastie, and R. Tibshirani. 2001. Vol. 1of *The elements of statistical learning*. New York: Springer.

Gretton, A. 2013. "Introduction to RKHS, and some simple kernel algorithms." *Adv. Top. Mach. Learn. Lect. Conducted Univ. Coll. London* 16.

Gretton, A., K. M. Borgwardt, M. Rasch, B. Schölkopf, and A. J. Smola. 2007. "A kernel method for the two-sample-problem." In *Advances in neural information processing systems*, 513–520. London: MIT Press.

Gretton, A., K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola. 2012. "A kernel two-sample test." *J. Mach. Learn. Res.* 13 (1): 723–773.

Gretton, A., O. Bousquet, A. Smola, and B. Schölkopf. 2005. "Measuring statistical dependence with hilbert-schmidt norms." In *Proc., Int. Conf. on Algorithmic Learning Theory*, 63–77. New York: Springer.

I/O Sensor Nederland bv. 2006. *SM-24 geophone element*. P/N 1004117. Drunen, Netherlands: I/O Sensor Nederland bv.

Koç, G., and K. Yegin. 2013. "Footstep and vehicle detection using slow and quick adaptive thresholds algorithm." *Int. J. Distrib. Sens. Netw.* 9 (10): 783604. https://doi.org/10.1155/2013/783604.

Lam, M., M. Mirshekari, S. Pan, P. Zhang, and H. Y. Noh. 2016. "Robust occupant detection through step-induced floor vibration by incorporating structural characteristics." In Vol. 4 of *Proc., Dynamics of Coupled Structures*, 357–367. New York: Springer.

Li, Y., X. Jing, H. Lv, and J. Wang. 2012. "Analysis of characteristics of two close stationary human targets detected by impulse radio UWB radar." *Prog. Electromagnet. Res.* 126: 429–447. https://doi.org/10.2528/PIER12011908.

Ling, X., W. Dai, G.-R. Xue, Q. Yang, and Y. Yu. 2008. "Spectral domain-transfer learning." In *Proc., 14th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 488–496. New York: Association for Computing Machinery.

Madarshahian, R., J. M. Caicedo, and N. Haerens. 2019. "Human activity benchmark classification using multilayer artificial neural network." In Vol. 2 of *Proc., Dynamics of Civil Structures*, 207–210. New York: Springer.

MATLAB. 2018. *Matlab statistics and machine learning toolbox*. Natick, MA: MathWorks.

Mirshekari, M., J. Fagert, A. Bonde, P. Zhang, and H. Y. Noh. 2018a. "Human gait monitoring using footstep-induced floor vibrations across different structures." In *Proc., ACM Int. Joint Conf. and 2018 Int. Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, 1382–1391. New York: Association for Computing Machinery.

Mirshekari, M., S. Pan, J. Fagert, E. M. Schooler, P. Zhang, and H. Y. Noh. 2018b. "Occupant localization using footstep-induced structural vibration." *Mech. Syst. Sig. Process.* 112 (Nov): 77–97. https://doi.org/10.1016/j.ymssp.2018.04.026.

Mirshekari, M., S. Pan, P. Zhang, and H. Y. Noh. 2016. "Characterizing wave propagation to improve indoor step-level person localization using floor vibration." In Vol. 9803 of *Proc., Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*. Bellingham, WA: International Society for Optics and Photonics.

Mirshekari, M., P. Zhang, and H. Y. Noh. 2017. "Calibration-free footstep frequency estimation using structural vibration." In Vol. 2 of *Proc., Dynamics of Civil Structures*, 287–289. New York: Springer.

Muñoz-Salinas, R., E. Aguirre, and M. Garca-Silvente. 2007. "People detection and tracking using stereo vision and color." *Image Vision Comput.* 25 (6): 995–1007. https://doi.org/10.1016/j.imavis.2006.07.012.

Murakita, T., T. Ikeda, and H. Ishiguro. 2004. "Human tracking using floor sensors based on the markov chain monte carlo method." In Vol. 4 of *Proc., 17th Int. Conf. on Pattern Recognition*, 917–920. New York: IEEE.

Oberg, T., A. Karsznia, and K. Oberg. 1993. "Basic gait parameters: Reference data for normal subjects, 10-79 years of age." *J. Rehabil. Res. Dev.* 30 (2): 210.

Oppenheim, A. V., A. S. Willsky, and S. H. Nawab. 1997. *Signals and systems*. 2nd ed. Upper Saddle River, NJ: Prentice-Hall.

Pan, S., M. Mirshekari, J. Fagert, C. G. Ramirez, A. J. Chung, C. C. Hu, J. P. Shen, P. Zhang, and H. Y. Noh. 2018. "Characterizing human activity induced impulse and slip-pulse excitations through structural vibration." *J. Sound Vib.* 414 (Feb): 61–80. https://doi.org/10.1016/j.jsv.2017.10.034.

Pan, S., C. G. Ramirez, M. Mirshekari, J. Fagert, A. J. Chung, C. C. Hu, J. P. Shen, H. Y. Noh, and P. Zhang. 2017a. "Surfacevibe: Vibration-based tap and swipe tracking on ubiquitous surfaces." In *Proc., 16th ACM/IEEE Int. Conf. on Information Processing in Sensor Networks*, 197–208. New York: IEEE.

Pan, S., S. Xu, M. Mirshekari, P. Zhang, and H. Y. Noh. 2017b. "Collaboratively adaptive vibration sensing system for high-fidelity monitoring of structural responses induced by pedestrians." *Front. Built Environ.* 3: 28. https://doi.org/10.3389/fbuil.2017.00028.

Pan, S., T. Yu, M. Mirshekari, J. Fagert, A. Bonde, O. J. Mengshoel, H. Y. Noh, and P Zhang. 2017c. "Footprintid: Indoor pedestrian identification through ambient structural vibration sensing." *Proc., ACM Interactive, Mobile, Wearable and Ubiquitous Technol.* 1 (3): 89. https://doi.org/10.1145/3130954.

Pan, S. J., J. T. Kwok, and Q. Yang. 2008. "Transfer learning via dimensionality reduction." In *Proc., 23rd AAAI Conf. on Artificial Intelligence*, 677–682. Palo Alto, CA: AAAI Press.

Pan, S. J., I. W. Tsang, J. T. Kwok, and Q. Yang. 2011. "Domain adaptation via transfer component analysis." *IEEE Trans. Neural Networks* 22 (2): 199–210. https://doi.org/10.1109/TNN.2010.2091281.

Pan, S. J., and Q. Yang. 2010. "A survey on transfer learning." *IEEE Trans. Knowl. Data Eng.* 22 (10): 1345–1359. https://doi.org/10.1109/TKDE.2009.191.

Poston, J. D., R. M. Buehrer, and P. A. Tarazaga. 2017a. "A framework for occupancy tracking in a building via structural dynamics sensing of footstep vibrations." *Front. Built Environ.* 3: 65. https://doi.org/10.3389/fbuil.2017.00065.

Poston, J. D., R. M. Buehrer, and P. A. Tarazaga. 2017b. "Indoor footstep localization from structural dynamics instrumentation." *Mech. Syst. Sig. Process.* 88 (May): 224–239. https://doi.org/10.1016/j.ymssp.2016.11.023.

Reuland, Y., S. G. Pai, S. Drira, and I. F. Smith. 2017. "Vibration-based occupant detection using a multiple-model approach." In Vol. 2 of *Dynamics of civil structures*, 49–56. New York: Springer.

Richman, M. S., D. S. Deadrick, R. J. Nation, and S. Whitney. 2001. "Personnel tracking using seismic sensors." In *Aerospace/defense sensing, simulation, and controls*, 14–21. Bellingham, WA: International Society for Optics and Photonics.

Saha, M., S. Thakur, A. Singh, and Y. Agarwal. 2014. "Energylens: Combining smartphones with electricity meter for accurate activity detection and user annotation." In *Proc., 5th Int. Conf. on Future Energy Systems*, 289–300. New York: Association for Computing Machinery.

Sasaki, Y. 2007. "The truth of the f-measure." *Teach Tutor Mater.* 1 (5): 1–5.

Savio, D., and T. Ludwig. 2007. "Smart carpet: A footstep tracking interface." In Vol. 2 of *Proc., 21st Int. Conf. on Advanced Information Networking and Applications Workshops*, 754–760. New York: IEEE.

Schölkopf, B., A. Smola, and K.-R. Müller. 1998. "Nonlinear component analysis as a kernel eigenvalue problem." *Neural Comput.* 10 (5): 1299–1319. https://doi.org/10.1162/089976698300017467.

Smola, A., A. Gretton, L. Song, and B. Schölkopf. 2007. "A hilbert space embedding for distributions." In *Proc., Int. Conf. on Algorithmic Learning Theory*, 13–31. New York: Springer.

Snidaro, L., C. Micheloni, and C. Chiavedale. 2005. "Video security for ambient intelligence." *IEEE Trans. Syst. Man Cybern. Part A Syst. Humans* 35 (1): 133–144. https://doi.org/10.1109/TSMCA.2004.838478.

Strang, G., and S. Strang. 2006. *Linear algebra and its applications*. Belmont, CA: Thomson Brooks/Cole.

Subramanian, A., K. G. Mehrotra, C. K. Mohan, P. K. Varshney, and T. Damarla. 2010. "Feature selection and occupancy classification using seismic sensors." In *Trends in applied intelligent systems*, 605–614. New York: Springer.

Succi, G. P., D. Clapp, R. Gampert, and G. Prado. 2001. "Footstep detection and tracking." In *Aerospace/defense sensing, simulation, and controls*, 22–29. Bellingham, WA: International Society for Optics and Photonics.

Sugiyama, M., S. Nakajima, H. Kashima, P. V. Buenau, and M. Kawanabe. 2008. "Direct importance estimation with model selection and its application to covariate shift adaptation." In *Proc., 20th Int. Conf. on Neural Information Processing Systems*, 1433–1440. New York: Association for Computing Machinery.

Texas Instruments. 2014. *OpAmp breakout*. Dallas: Texas Instruments.

Uziel, S., T. Elste, W. Kattanek, D. Hollosi, S. Gerlach, and S. Goetze. 2013. "Networked embedded acoustic processing system for smart building applications." In *Proc., 2013 Conf. on Design and Architectures for Signal and Image Processing*, 349–350. New York: IEEE.

Visser, H. 1983. "Gait and balance in senile dementia of Alzheimer's type." *Age Ageing* 12 (4): 296–301. https://doi.org/10.1093/ageing/12.4.296.

Wang, C., X. Wang, Z. Long, J. Yuan, Y. Qian, and J. Li. 2016. "Estimation of temporal gait parameters using a wearable microphone-sensor-based system." *Sensors* 16 (12): 2167. https://doi.org/10.3390/s16122167.

Yu, T., S. Pan, S. Xu, X. Chen, M. Mirshekari, J. Fagert, P. Zhang, H. Y. Noh, and O. J. Mengshoel. 2018. "ILPC: Iterative learning using physical constraints in real-world sensing data." In *Proc., SmartIoT 2018 Workshop: AI Enhanced IoT Data Processing for Intelligent Applications*, 120. Palo Alto, CA: AAAI Press.

Zetik, R., S. Crabbe, J. Krajnak, P. Peyerl, J. Sachs, and R. Thomä. 2006. "Detection and localization of persons behind obstacles using m-sequence through-the-wall radar." In Vol. 6201 of *Proc., Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense V*, 62010I. Bellingham, WA: International Society for Optics and Photonics.