Advanced Engineering Informatics 49 (2021) 101289

Contents lists available at ScienceDirect

Advanced Engineering Informatics

journal homepage: www.elsevier.com/locate/aei

# Using footstep-induced vibrations for occupant detection and recognition in buildings

Slah Drira [a,c,*], Sai G.S. Pai [a,c], Yves Reuland [b], Nils F.H. Olsen [a], Ian F.C. Smith [a,c]

[a] School of Architecture, Civil and Environmental Engineering (ENAC), Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland
[b] Department of Civil, Environmental and Geomatic Engineering, ETH Zurich, Switzerland
[c] Future Cities Laboratory, Singapore-ETH Centre, Singapore

ARTICLE INFO

*Keywords:*
Footstep-induced floor-vibrations
Event detection
Event-signal extraction
Event classification
Occupant counting
Occupant recognition
Support vector machine
Convolutional neural network

ABSTRACT

Occupant detection and recognition support functional goals such as security, healthcare, and energy management in buildings. Typical sensing approaches, such as smartphones and cameras, undermine the privacy of building occupants and inherently affect their behavior. To overcome these drawbacks, a non-intrusive technique using floor-vibration measurements, induced by human footsteps, is outlined. Detection of human-footstep impacts is an essential step to estimate the number of occupants, recognize their identities and provide an estimate of their probable locations. Detecting the presence of occupants on a floor is challenging due to ambient noise that may mask footstep-induced floor vibrations. Also, signals from multiple occupants walking simultaneously overlap, which may lead to inaccurate event separation. Signals corresponding to events, once extracted, can be used to identify the number of occupants and their locations. Spurious events such as door closing, chair dragging and falling objects may produce vibrations similar to footstep-impacts. Signals from such spurious events have to be discarded as outliers to prevent inaccurate interpretations of floor vibrations for occupant detection. Walking styles differ among occupants due to their anatomies, walking speed, shoe type, health and mood. Thus, footstep-impact vibrations from the same person may vary significantly, which adds uncertainty and complicates occupant recognition. In this paper, efficient strategies for event-detection and event-signal extraction have been described. These strategies are based on variations in standard deviations over time of measured signals (using a moving window) that have been filtered to contain only low-frequency components. Methods described in this paper for event detection and event-signal extraction perform better than existing threshold-based methods (fewer false positives and false negatives). Support vector machine classifiers are used successfully to distinguish footsteps from other events and to determine the number of occupants on a floor. Convolutional neural networks help recognize the identity of occupants using footstep-induced floor vibrations. The utility of these strategies for footstep-event detection, occupant counting, and recognition is validated successfully using two full-scale case studies.

## 1. Introduction

Identifying occupants inside buildings is an important step in the development of an automatized understanding of building-occupant information. Detection of building occupants involves counting their number, recognition of their identities and determination of their trajectories. Information regarding indoor occupants enables optimization of functionalities in buildings, such as security enhancement [1], healthcare [2,3], as well as space and energy management [4–6].

Prior studies involved the development of sensing technologies for occupancy detection and recognition, such as acoustic instrumentation [7,8], $CO_2$ sensors [9,10], smart-flooring systems [11,12], optical sensors [13–15], and radio-frequency devices [16–19]. For example, acoustic-based methods were found to be sensitive to ambient audible noise [7,8,20]. The major limitation of $CO_2$-based approaches was related to the slow spreading of $CO_2$ within an indoor space where air ventilation compromised the concentration of $CO_2$ inside buildings, leading to ambiguous interpretations of occupancy levels [20,21]. Smart flooring systems required highly instrumented floors (thousands of sensors) [11,12]. Such systems are not suitable for large full-scale applications.

In addition, passive infrared (PIR) sensors were used to detect

* Corresponding author.
*E-mail address:* slah.drira@gmail.com (S. Drira).

https://doi.org/10.1016/j.aei.2021.101289
Received 5 February 2020; Received in revised form 25 March 2021; Accepted 31 March 2021
Available online 13 May 2021

occupants inside buildings for efficient operation of lighting and heating, ventilation and air-conditioning systems [13,15]. However, PIR detection systems could only provide a binary status of occupancy; no attempts were made to recognize and locate occupants [6,20]. Cameras were also used to estimate the level of occupancy and to study the behavior of occupants [22,23]. Video-based pattern recognition was used to recognize occupant identities [24–27]. However, optical sensors required clear lines of sight and large angles of coverage for accurate occupancy detection [20,27,28].Radio-frequency identification devices (RFIDs) including portable sensors (smartphones) [16,29], embedded Wi-Fi [30–32] and Bluetooth beacons [16,33], have been used to determine occupancy levels inside buildings [19,34–36] and to lesser extend for occupant recognition [30]. However, RFIDs needed regular maintenance [20,30,36]. Such devices require clear spaces to detect occupants in buildings due to multi-path problems that are induced by structural and non-structural elements such as walls and furniture [20,37]. Also, the RFIDs devices were insufficient to recognize indoor occupants.

Video-recording devices and RFIDs undermine the privacy of indoor occupants due to their intrusive nature [20,38]. For instance, cameras in office environments influence the behavior of occupants and radio-frequency-based techniques require the occupants to carry permanently-connected devices [20]. Therefore, non-intrusive and reliable strategies for occupant detection such as structural-vibration sensors are preferred, since they preserve privacy. In this paper, occupant detection, counting and identity recognition are carried out using only footstep-induced floor vibrations.

Detecting the presence of occupants on a floor is challenging due to variations in rigidities of floor slabs and the presence of obstructions such as beams and walls. Moreover, the dispersive nature of floor slabs may result in footstep-impact events with low signal-to-noise ratios (SNR) and variations in footstep-impact signatures at various floor locations [39,40]. Moreover, spurious events such as door closing, chair dragging or dropping objects have been found to result in vibrations that have similarities with footstep-impact signal signatures [39]. In addition, overlapping signals from multiple occupants walking simultaneously on a floor complicate event-signal extraction and estimating the number of occupants. Also, walking gaits of occupants are affected by various sources including their anatomies, walking speed, shoe type, health and mood [41–43]. Since the same occupant may walk differently [44], various walking patterns from the same person induce variability between footstep-impact signatures. Thus, this leads to challenges in performing occupant recognition.

Events (from footsteps or other sources) have been detected as anomalies when vibration amplitudes exceed a previously defined baseline level of ambient vibrations [37,39,45]. However, event vibrations with low SNR may be hidden by ambient noise [37]. Several solutions have been proposed to overcome this limitation. For example, increasing the signal resolution using amplifiers has been proposed by Pan et al. [38]. However, hard footstep-impacts might lead to clipping signals (amplitudes of measured signals exceed sensor range). A denoising technique has been proposed by Clemente et al. [46] using discrete-wavelet transforms with high-level filtering to differentiate impact events from ambient vibrations. However, some event signals with low SNR are also filtered out. Also, unsupervised learning techniques using Gaussian mixture models have been proposed by Anchal et al. [47] to detect and extract footstep-event signals.

In order to distinguish footstep events from other spurious events, supervised learning techniques using one-class support vector machines (SVM) [48] were proposed [37,46,49], for which training data was limited to footstep events only (i.e. no spurious events). One-class SVM classification was trained using normalized power spectral density of detected event signals as features, which led to a 90% F1 score [37]. The F1 score defines the overall performance metric that reflects the ability of the classifier to distinguish between classes. Moreover, event classification was performed to separate footstep events from other impulses and ambient noise. Footstep and fall events were classified using one-class SVM based on 13 features in time and frequency domains [46]. Spurious events such as dropping objects, closing doors and drawers, hitting tables and jumping were included to estimate the performance of one-class SVM and compared with Gaussian process and *k*-nearest neighbors (KNN) classifiers. One-class SVM led to an F1 score of 92% for footstep classification compared with 38% for fall classification. However, type II error, that defines the rate of non-footstep events that are identified as footstep events, has been approximately 10%. Thus, training with only footstep events might miss-classify spurious impulses as footstep events.

The KNN classifier calculates the similarity between new data instances and each training data instance. Then, the class labels of the k most similar neighbors are used to predict the class of the new data instances [50–52]. Boosted tree (BT) classifier is based on an ensemble of decision trees to predict new data instances [53,54]. A boosting algorithm such as AdaBoost [55] has been applied to many deemed weak classifiers to achieve a final strong classifier [56].

Footstep-induced floor-vibration measurements were used to estimate the number of occupants on floor slabs [57–59]. Occupancy-level estimation of sub-areas of full-scale floors was performed based on changes in floor-vibration energy induced by footstep impacts [57]. Similarly, occupancy-levels were updated in real time by tracking single occupants in floor zones [58]. However, the proposed framework led to inaccurate results in presence of multiple occupants walking simultaneously on the floor slab. A KNN classifier was proposed to identify the number of occupants from overlapped vibration measurements of four participants [59]. Cross-correlation between consecutive footstep-event signals from the same sensor, cross-correlation of footstep-event signals between all sensors, footstep-event signal duration and footstep-event signal entropy were used as features. However, classification accuracy was low for multiple occupants walking simultaneously on the floor (accuracy of 67% for two occupants and 33% for three occupants).

Recognition of occupant identities using floor-vibration measurements was proposed using multi-class SVM classifiers using time-and-frequency domain features [40,41,46]. Recognition accuracy was improved from 63% to 83% using a hierarchical classifier [41] that took classification results from all succeeding footstep events of walking measurements. A modified SVM learning algorithm, that provided higher accuracy rates than traditional SVM on low-sized training-data sets, was proposed by Pan et al. [40] for identity recognition. However, seven succeeding footsteps were required to recognize occupant identities with high performance rates. In another study, two succeeding footstep events were used to train a multi-class SVM classifier to recognize accurately the identities of six participants [46]. However, the average F1 score (overall performance score) for all participants did not exceed 89%.

In this paper, strategies for efficient event-detection, event-signal extraction, estimating number of occupants within a building space (while walking on instrumented floors) and occupant recognition are presented. Footstep and non-footstep impacts generate waves that travel through the floor slab with frequency-dependent phase velocities [39]. Appropriate decomposition of raw signals at several frequency ranges helps in determining characteristics of impact events that help improve the proposed strategies compared with those in literature.

The paper starts with a description of the occupant detection and localization framework (Section 2). Event detection, event-signal extraction, event classification, occupant counting, and identity recognition methodologies are presented in Section 3. Two full-scale case-study descriptions for occupant detection, counting and recognition are presented in Section 4. Discussion is then provided in Section 5 followed by conclusions in Section 6.

## 2. Framework for occupant detection and localization

Once the presence of an occupant is detected, the localization

framework utilizes footstep-induced floor vibrations to infer possible locations of occupants. Floor slabs are regularly subjected to impact from many types of activities (walking, door closing, object drops, etc.) due to interactions between occupants and their indoor environment. This leads to ambiguities in the interpretation of measured vibrations.

Significant variability in footstep-event signals is observed for the same occupants walking along similar trajectories due to several factors [43]. Changes in walking speeds, health, mood and other characteristics alter gait patterns of occupants. This increases the challenges of occupant detection and interpretation of detected events.

Footstep impacts on typical floors are impulses, which generate Lamb waves that travel through the floor slab [39]. The propagation of Lamb waves, within a dispersive medium, results in shape changes of the floor responses as recorded at sensors. Such behavior leads to distortion in the time-of-arrival of measured footstep-impact signals, which complicate the interpretation of floor vibrations. Apart from the dispersive nature of typical floor slabs, magnitudes of footstep-induced floor vibrations can be affected by structural behavior and its boundary conditions. These signal magnitudes are further influenced by obstructions such as beams and walls. These ambiguities make occupant detection and localization a challenging task.

A framework for occupant detection and localization is shown in Fig. 1, which takes into account the presence of uncertainties from the aforementioned sources. This framework involves multiple steps that help reduce uncertainties in detection and localization of occupants (see Fig. 1) to improve accuracy and precision.

Detecting footstep events from vibration measurements is a key first step. Event detection and signal extraction require the understanding of structural characteristics such as the fundamental frequencies of the structure, which are estimated using ambient vibrations. Using structural information (frequency ranges that cover the first few vertical modes of the structure), event detection and subsequently event extraction are carried out. Event-detection and signal-extraction strategies are discussed in Section 3.1. Subsequently, a supervised learning classifier based on a support vector machine (SVM) [48] is used to distinguish between extracted footstep and non-footstep event signals (see Section 3.2).

Extracted footstep-event signals are then used to count the number of occupants on the floor using another SVM classifier (see Section 3.3). The number of occupants may then be used to select appropriate model simulations for occupant localization. Occupant recognition is performed when an occupant is detected on the floor. Inspired by pattern recognition using deep learning approaches [60,61], footstep-induced floor-vibration signatures are recognized using a convolutional neural network (CNN) classification [62,63] (see Section 3.4).

Occupant localization is based on combining information from measured footstep-event signals with physics-based models [45,64–66]. Error-domain model-falsification (EDMF) [67] is a model-based data interpretation approach that is well-suited to identify a population of possible locations of occupants, as shown by Drira et al. [45]. In this paper, the focus is on event detection, occupant counting and identity recognition, as illustrated in Fig. 1 within the dashed box.

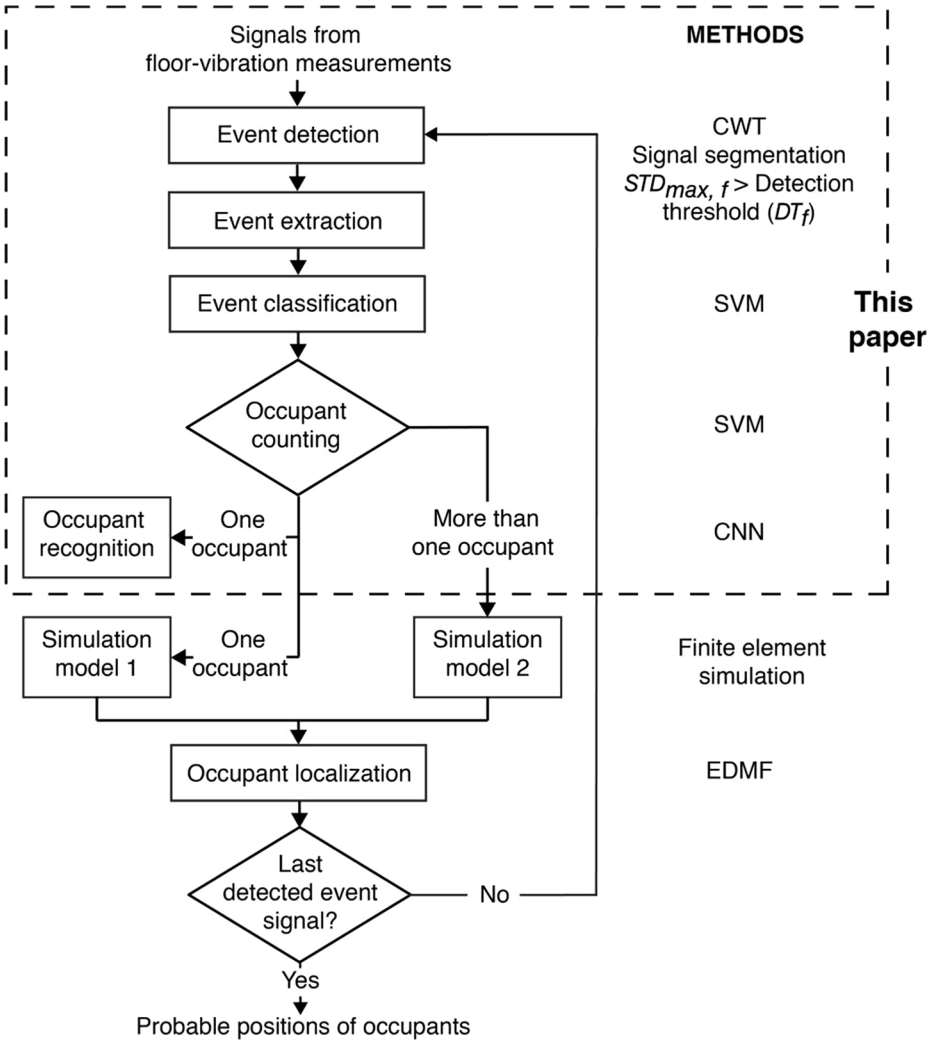


**Fig. 1.** Framework for occupant detection and localization. $STD_{max,f}$ is the maximum standard deviation from all sensors of a running window through a decomposed and reconstructed signal at frequency range $f$. $DT_f$ is detection threshold that corresponds to frequency range $f$. CWT is continuous wavelet transform, SVM is support vector machine and CNN is convolutional neural network. EDMF is error-domain model falsification. The focus of this paper are steps for occupant detection, classification and recognition (within dashed box).

## 3. Methodologies

The contribution of this paper is composed of four parts: event detection and signal extraction, event classification, occupant counting and identity recognition using floor-vibration measurements.

### 3.1. Event-detection and signal-extraction strategies

Floor-vibration measurements may include various activities including footstep-impact events (from one or multiple occupants) and other activities that result from interactions between occupants and the indoor environment (such as door closing, chair dragging and dropping objects). High levels of ambient noise may undermine detection of events that have low amplitudes and thus, a low SNR. Moreover, length of a signal that characterizes the vibrational response of the structure to an impact event (event-signal duration) depends on the type of event and on the walking gait of occupants. For example, the walking gait of occupants (and thus the impact on the floor) changes with their mood, shoe type and walking speed, resulting in variability in measured signals and even in event-signal durations. The first step of the framework involves detecting all possible events from recorded vibration measurements and subsequently, extracting the relevant event signals.

#### 3.1.1. Event-detection strategy

Event-detection strategy is intended to capture the occurrence times of possible events within the vibration measurement. Occurrence time of an event defines the signal segment that contains prominent magnitudes (from all sensors) of an event signal. Since footstep impacts generate non-stationary waves that travel through the floor slab, the event-detection strategy utilizes information from multiple frequency components of floor-vibration measurements [68]. In Fig. 2, the relevant steps involved in event-detection are outlined.

Event detection starts with signal processing, as illustrated by the upper dashed box in Fig. 2. The frequency range that contains the fundamental bending modes of the structure is assessed using ambient vibrations. Prominent peaks in the first singular values of the cross-power spectral density (CPSD) [69] help to delimit the range with most energy contribution. This frequency range is then divided into at least four equivalent and overlapping ranges to cover the fundamental vertical modes of the structure. The measured signal is then decomposed using continuous wavelet transform (CWT) [70] and reconstructed using inverse wavelet transform (IWT) at these frequency ranges.

Depending on the type of event, the frequency ranges that are most useful to differentiate between ambient vibrations and events may not be the same. Also, since occupants strike the floor differently, their

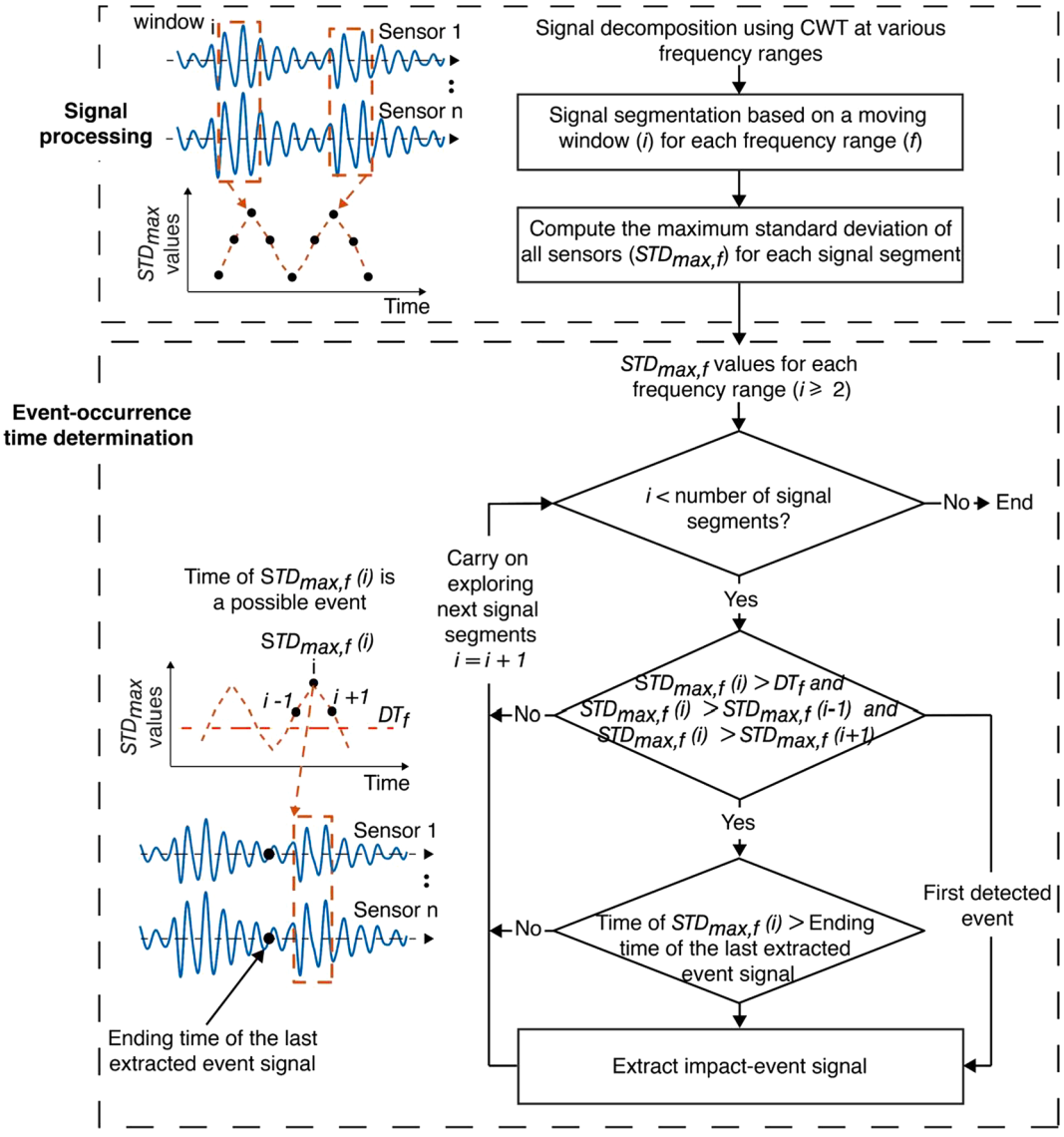


**Fig. 2.** Event detection is composed of a signal-processing part and an event-occurrence time determination part.

footstep impacts activate unequally several bending modes of the structure. Thus, careful selection of multiple frequency ranges has potential to enhance the detection of events characterized by low SNR that may be hidden by ambient noise.

The Morlet wavelet [71] is chosen as mother wavelet in CWT due to its shape similarity with the footstep impact signal [45]. Signal-decomposition using CWT is based on comparing the recorded vibration signal with varying scaling factors of the chosen mother wavelet. Thus, the signal is decomposed in the time domain. Subsequently, the reconstruction of the signal within a specific frequency range is carried out using IWT through combining information contained at the corresponding scales.

Experimental studies have revealed that walking frequency varies between 1.4 Hz and 2.5 Hz [42,72]. Thus, each decomposed signal is segmented into windows with a length 0.2 s moving with an increment of 0.1 s. The duration of the moving window corresponds to half of the minimum time between two footsteps (0.4 s).

The standard deviation of measured vibrations is taken as metric for event detection since it is correlated to the energy of the signal. Standard deviations calculated for data windows of the measured responses helps find abrupt variations in data due to peaks. The maximum standard deviation from all sensors ($STD_{max,f}$) is assessed for each segment of decomposed and reconstructed signals at frequency range $f$ (see signal processing part in Fig. 2). This operation is repeated for each frequency range.

Similarly, ambient vibration measurements are decomposed at the same frequency ranges, in order to establish detection thresholds ($DT_f$) for all frequency ranges. $DT_f$ is defined based on computing $STD_{max,f}$ values of segmented and decomposed ambient vibration signal at frequency range $f$. For each frequency range, the maximum value of the resulting $STD_{max,f}$ values is taken to be a $DT_f$.

$STD_{max,f}$ values assessed for each frequency range are then used to detect event signals within the captured vibration measurement (see event-occurrence time determination part in Fig. 2). A local maximum resulting from $STD_{max,f}$ values corresponds to a signal segment that contains prominent magnitudes (from all sensors) of an event signal. Thus, local maxima resulting from $STD_{max,f}$ values that exceed $DT_f$ over at least one decomposed signal indicate the occurrence times of possible events (see Fig. 2). Moreover, each local maximum has to be defined at least within an interval of 0.4 s, which defines the minimum time between two footsteps when an occupant walks at maximum speed of 2.5 Hz.

A signal segment that corresponds to an occurrence time of a possible event serves to extract the event signal at each sensor location (see Section 3.1.2). The objective of the signal extraction is to determine dynamically the starting and ending times of a detected event signal. The signal-extraction operation is incorporated within the event-detection strategy. Indeed, event-occurrence times resulting from local maxima from $STD_{max,f}$ values for each decomposed and reconstructed signal are not equivalent. Each decomposed and reconstructed signal may have different occurrence time pointing to a same event. Thus, in order to avoid extracting the same event signal, the event-occurrence time has to be greater than the ending time of the last extracted event signal (see event-occurrence time determination part in Fig. 2). This operation is investigated starting from the second detected event. Occupant detection strategy is operational until all signal segments of the vibration measurements are explored.

#### 3.1.2. Signal-extraction strategy

Event extraction dynamically ascertains starting and ending times of detected-event signals. It has been shown through empirical studies that a decomposed signal at a frequency range greater than the first natural frequency of the structure provides better event delimitation in the time domain [68]. Thus, $STD_{max,f}$ values of the decomposed signal at a frequency range greater than the first natural frequency of the structure are used as inputs for event extraction.

Event detection serves to define the occurrence time of an event (see Fig. 2). The captured occurrence time of an event corresponds to a signal segment that contains only the prominent magnitudes (from all sensors) of the event signal (see Fig. 2). A signal segment that corresponds to a detected event corresponds to a local maximum of $STD_{max,f}$ values. Accordingly, previous and succeeding signal segments from the local maximum contain information of the starting and ending times of a detected event signal. Therefore, starting and ending times of a detected event signal are determined using a backward and forward search in $STD_{max,f}$ values from the local maximum that defines the detected event.

In Fig. 3, signal segment $e$ corresponds to an occurrence time of a detected event and is used as a reference. Let $STD_{max,f}(e)$ be the maximum standard deviation from all sensors calculated for the reference signal segment $e$ (see Fig. 3). For a signal segment $i$, preceding the segment $e$, if the standard deviation $STD_{max,f}(i)$ is higher than $STD_{max,f}(e)$, then the event-starting time is in signal segment $i$. Otherwise, the current $STD_{max,f}(i)$ value becomes equal to the reference $STD_{max,f}(e)$ value and the comparison is repeated with its previous one ($e = i$ and $i = i - 1$). Similarly, for a signal segment $j$, succeeding the segment $e$, if the standard deviation $STD_{max,f}(j)$ is higher than $STD_{max,f}(e)$, then the event-ending time is in signal segment $j$. Otherwise, the current $STD_{max,f}(j)$ value becomes equal to the reference $STD_{max,f}(e)$ value and the comparison is repeated with its succeeding one ($e = j$ and $j = j + 1$). In order to avoid searching indefinitely for the starting or the ending time of a detect event, to the maximum duration of a footstep event is fixed as 0.7 s (derived from the minimum walking speed of 1.4 steps per second).

The signal-extraction operation is also capable of capturing the starting time of the detected event when this information is contained in the signal segment $e$ while $STD_{max,f}(i)$ is higher than $STD_{max,f}(e)$. This is due to the overlap during signal segmentation, which is equal to half of the running window (0.1 s in Section 3.1.1). This overlap also allows capture of the signal ending time event when the signal segment $j$ contains parts of the next event vibrations.

Subsequently, the sums of the absolute values of the amplitudes of the raw signal of all sensors are computed for the signals delimited by segments $i$ and $j$. A moving average using a Gaussian-weighted function [73] is applied to assess the trend of the resulting sums within the segments $i$ and $j$. This allows determination of the trend of the resulting sums. The weighted moving average uses a convolution of a moving window over data points with a weighting function (such as Gaussian-weighted function). Minimum values of the resulting trends (bounded by signal segments $i$ and $j$) define the starting and the ending times of a detected event. The detected event signal is finally extracted separately for all sensor locations.

This set of operations provides a method for dynamic selection of the starting and the ending time of impact events from vibration measurements, which is an improvement upon methods that apply a fixed window length [45,68]. Moreover, these strategies are suitable to extract event signals from overlapping signals resulting from multiple occupants walking together on a floor. Therefore, this leads to accurate delimitation of event signals for applications such as event classification, counting the number occupants and recognition of occupant identities.

### 3.2. Event-classification strategy

Floor-vibration measurements contain footstep-impact events that are often affected by ambient activities due to the interaction of occupants with the indoor environment (such as door closing, chair dragging or dropping objects) as well as external activities (traffic, wind, etc.). Thus, a supervised learning technique based on binary-SVM is used to distinguish between footstep and non-footstep event signals. It has been shown that SVM classifiers provide good performance with small training sets with respect to feature numbers compared with neural-

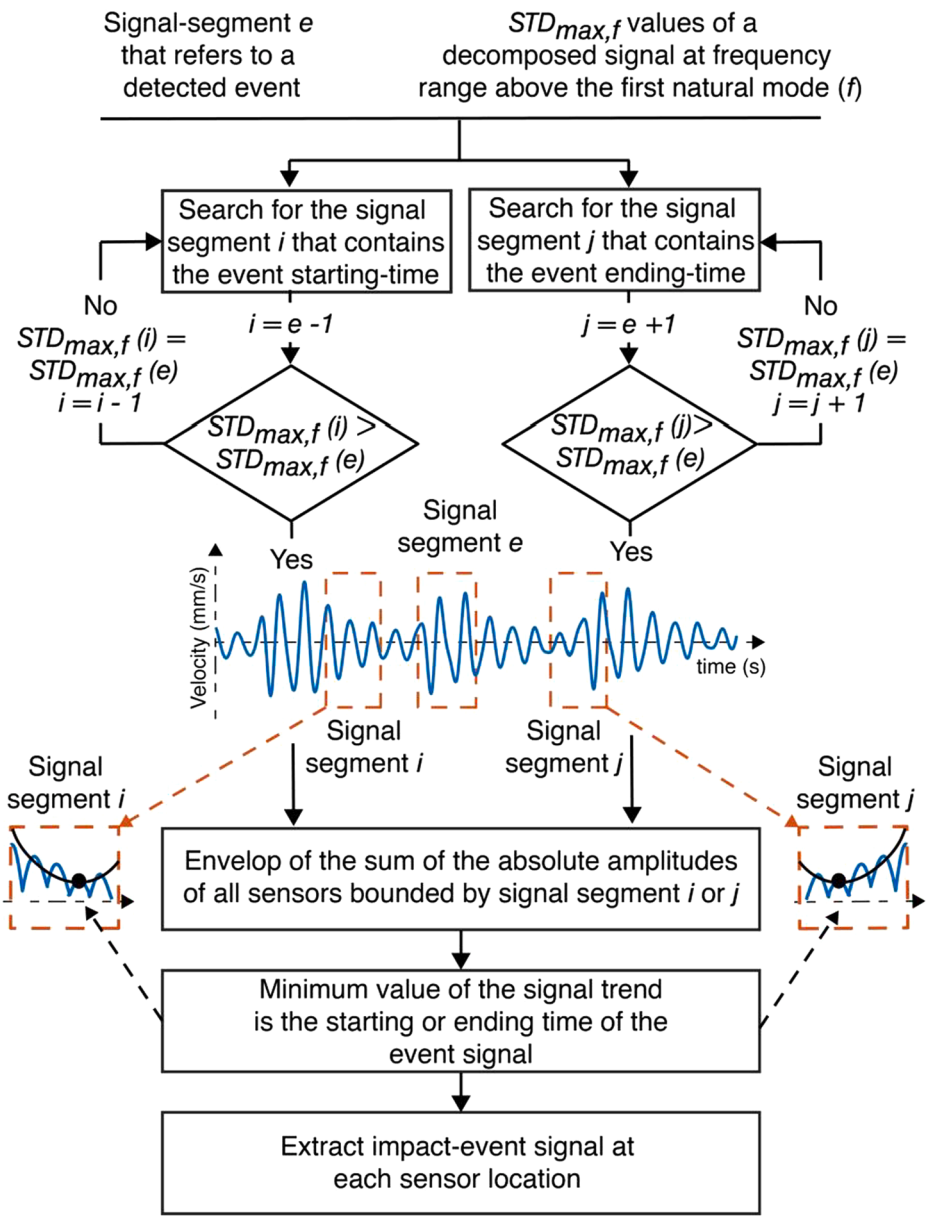


**Fig. 3.** Event extraction strategy is used to identify dynamically the starting and the ending times of a detected event signal.

network-based methods [40,63]. A binary-SVM classifier is trained with footstep and non-footstep events to improve accuracy and prevent miss-classification of spurious impulses as footstep events when compared with one-class SVM.

Feature extraction from raw event signals is a crucial step to perform classification-based methods. Since impact-event signals are influenced by the natural vibration modes of structures, features are assessed in time and frequency domains to effectively differentiate between footstep and non-footstep events [46]. Frequency-domain metrics that are used include the frequency value that corresponds to the maximum of the first singular values of CPSD of all sensors ($FSV_{max}$) and the centroid of first singular values of CPSD ($C_{CPSD}$). Time-domain metrics include standard deviation ($\sigma$), maximum difference in amplitudes ($\Delta_{amp}$), root-mean-square (RMS), kurtosis ($Kr$) and median ($Md$) of the entire event signal. These time-domain metrics are assessed for decomposed and reconstructed event signals at various frequency ranges using CWT and IWT. The frequency band that vibration sensors can provide is divided into equivalent ranges with an overlap to cover signal components from low-to-high frequency ranges. The size of each frequency range and overlap is determined based on engineering judgment through several tests. The maximum values of all sensors as well as the average values of all sensors are recorded for all time-domain metrics.

Time-domain metrics assessed at specific frequency ranges that maximize the discrepancy between footstep and non-footstep event classes are selected as features for classification. Selection of appropriate frequency range for each time-domain metric is carried out using null-hypothesis based on the Kolmogorov-Smirnov test [74]. For a given time-domain metric that is assessed for decomposed signals at a certain frequency range, the null-hypothesis test is rejected when footstep and non-footstep data are from different distributions with a predefined level of confidence (typically 5%). Otherwise, the two data populations are defined to be from the same distribution. Null-hypothesis based on Kolmogorov-Smirnov test is used to estimate the discrepancy level between footstep and non-footstep populations for each time-domain metric and for each frequency range. Therefore, for each time-domain metric, the frequency range that has the highest discrepancy level between footstep and non-footstep populations is selected.

Time-domain metrics may be correlated, which leads to redundant

information in the training process. Thus, correlation coefficients are assessed based on the Pearson linear-correlation method [75] between every two time-domain metrics. Highly correlated metrics (above 90%) are disregarded.

#### 3.2.1. Support vector Machine (SVM)

SVM classifier infers the optimal decision boundary that maximizes the distance between data sets and the separating hyperplane [48,76,77]. For example, for two-class training data set, an SVM learning starts with transforming the input data into a higher dimensional space by means of a kernel function such as Polynomial, Gaussian and Radial basis functions. Then, an optimal separating hyperplane is constructed between the two classes in the higher-dimensional space by minimizing an objective function. The choice of a kernel function is not subjected to any rules. One kernel function may provide better classification performance than another over given the initial data set. Thus, testing of several kernel functions is recommended.

### 3.3. Occupant-counting strategy

The occupant-counting strategy is intended to determine the number of occupants walking together on the floor slab using a multi-class SVM classification. A multi-class training set contains features from vibration signals from one and multiple people walking together on the floor slab. In real-life applications, multiple people walk regularly at the same time on the same floor-slab. The resulting floor-vibration measurements include a superposition of the structural responses from multiple occupants walking with their own speed on their respective trajectory. Thus, footstep impacts of multiple people may be: 1) fully synchronized; 2) off-synchronized, leading to overlapping signals; and 3) staggered, leading to non-overlapping signals [59].

Footstep-induced floor-vibrations from a single or multiple occupants (synchronized, off-synchronized or staggered footstep impacts) are altered by the structure and depend on their footstep-impact locations. Despite the influence of the structure on the vibration measurements, floor vibrations at a sensor locations present higher amplitudes when the footstep impact is in close distance. Therefore, a cross-correlation between event signals at all sensor locations from the same footstep-event has the potential to infer the number of occupants on the floor.

Cross-correlation coefficients between event signals at all sensor locations from each footstep event are used as features to count the number of occupants on the floor using an SVM classifier. These coefficients are computed as the pair-wise correlation of velocity amplitudes of each footstep event captured at each sensor location. The cross-correlation coefficient matrix is calculated based on the Pearson linear-correlation method [75].

In addition, most of the footstep impacts from multiple people walking on the same floor area are off-synchronized, as shown by [78]. This results in overlapping floor responses, gathering the contribution of each occupant. Since the standard deviation ($\sigma$) values and the power spectral density (PSD) of event signals are correlated to the impact force induced by footsteps at a sensor [65], $\sigma$ values and maximum PSD of overlapping signals induced by multiple occupants have higher magnitudes than those from single occupants. Thus, apart from cross-correlation coefficients between event signals measured at multiple sensors, including $\sigma$ values of event signals recorded at each sensor location and maximum CPSD of all sensors as features have the potential to increase the classification performance to determine the number of occupants walking on the floor.

### 3.4. Occupant-recognition strategy

The objective of occupant recognition is to ascertain the identity of occupants based on the floor vibrations. Thus, non-intrusive occupant recognition can be performed. While this may not be acceptable in care-home and office contexts due to privacy concerns, security contexts, such as banks and computer areas may benefit. Convolutional neural network (CNN) classification is applied to the vibrations generated by occupants walking on a floor slab (see Section 2). CNN has emerged as a powerful supervised learning approach for pattern recognition [60,63,79]. Inspired by pattern recognition, footstep-impact signatures of multiple people, captured by several sensors, are used as inputs to train the CNN classification. Occupant recognition is tested using three CNN classification models and compared with a shallow NN (traditional NN) classifier.

#### 3.4.1. Convolutional neural network (CNN)

CNNs are artificial neural networks (NNs) that use convolution instead of general matrix multiplication in the presence of multilayer perceptrons and fully connected networks [63]. Fully connected networks have each neuron in one layer connected to all neurons in the next layer. Fully connected networks in traditional NN involve the interaction between each input unit with each output unit, which makes them prone to overfitting. Overfitting happens when a learning model captures the small variation along with the underlying pattern in data. CNNs are more efficient in terms of memory and computational-time requirements compared with traditional NNs. This is because CNNs have sparse interactions between input and output units through picking important features from input data using kernel-based filters.

A convolutional network layer accommodates three-step processing [63]. The input instance is first subjected to several convolutions using a kernel-based filter (size of the kernel has to be smaller than the size of the input instance) providing a set of linear activations (kernel-based filtering results). Subsequently, a nonlinear activation function such as Rectified Linear Unit (ReLU) [80] is applied to each linear activation. Finally, a pooling operation, which is a down-sampling strategy in CNN such as max-pooling or average-pooling is used to reduce the dimensionality of the convolutional filter output. Applying successive convolutional layers within a CNN architecture allows extracting high-level features from the input instance and leads to a better understanding of the data set.

For classification purposes, the outputs of the convolutional layers are flattened, leading to down-sampling of the convolutional outputs into a one-dimensional feature vector [81]. The flattened layer is then connected to regular NN dense layers that end with a class label layer.

#### 3.4.2. Convolutional layers

The CNN classifier for occupant-recognition strategy is defined by two-dimensional convolutional architecture. Convolutional layers constitute the skeleton of the CNN. Footstep-event signals from all sensors are connected to two successive convolutional layers (i.e. filters) using a window (i.e. kernel). Resulting convolutional outputs are evaluated using a ReLU activation function. Also, a max-pooling operation is used to reduce the dimensionality of each convolutional output layer. Finally, a flattened layer is used to convert the output data into a one-dimensional layer.

The flattened output is then fed to a feed-forward neural network and back-propagation that are applied to every neuron for each iteration of training. Over a series of epochs (i.e. iterations), the model can distinguish dominating features within input patterns and classify them using the Softmax classification technique. The Softmax function produces probability-like predictions for each class (occupant identity) [82–84]. The choice behind the window size of each layer is inferred based on repetitive tests and trials. No fixed rules are available to define these window sizes.

#### 3.4.3. CNN model #1

CNN model #1, as illustrated in Fig. 4, is trained based on separate footstep events as input patterns. These input patterns are processed through convolutional layers. A cross-entropy loss function [84] (i.e. error function or objective function that tends to minimize the

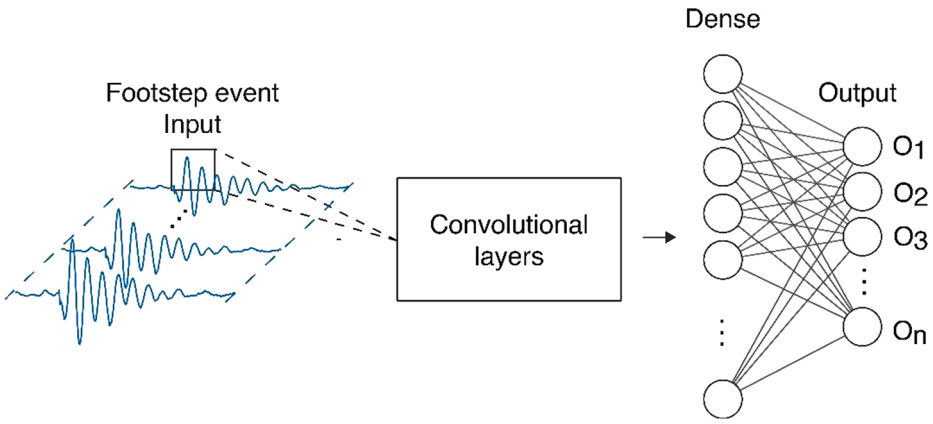


**Fig. 4.** CNN model #1 contains a dense layer that is connected to the convolutional layers. ReLU activation function is used for the dense layer. The dense layer is connected to an output layer that assimilates the identity of occupants ($O_n$) using Softmax activation function.

classification error) is used to evaluate the performance of the classification model in every epoch in order to determine the best classifier parameters.

#### 3.4.4. CNN model #2

CNN model #2 is characterized by rearranging the data set into couples of two succeeding footstep events as shown in Fig. 5. Each footstep event of each couple is processed in parallel through the convolutional layers. Each flatten layer is then connected to a dense layer using ReLU activation functions. The two dense layers are subsequently concatenated. The resulting layer is finally connected to the main output layer using Softmax activation function.

#### 3.4.5. CNN model #3

CNN model #3 is defined by the same architecture as the CNN model #2 (see Fig. 5). In addition to CNN model #2, the training process of the learning algorithm of the CNN model #3 is performed through minimizing a multi-objective function. The multi-objective function includes weighted losses from convolutional outputs #1 and #2, that correspond to the succeeding footstep events as well as the main output loss (see Fig. 5). Weights of 0.5 are attributed to convolutional outputs #1 and #2 and 1.0 to the main output.

## 4. Application to full-scale floor-slabs

Two full-scale structures, which are used for validation, are described in Sections 4.1 and 4.2. Occupant activities within these structures are measured and methodologies described in Section 3 have been applied for event detection, occupant counting and occupant recognition.

### 4.1. Description of Case Study 1

Detection of occupants, signal extraction and occupant counting in Sections 3.1 and 3.3 are tested on the full-scale floor-slab [66], shown in Fig. 6. The full-scale multi-story building is located in Singapore (approximate floor area is 950 $m^2$). The test area is approximately 600 $m^2$. The floor is a continuous reinforced-concrete slab. Ten concrete columns as well as several reinforced-concrete walls support the floor slab as shown in Fig. 6. The concrete slab is 25 cm thick and is covered by a linoleum finishing. Uni-directional reinforced-concrete beams connect the slab with the concrete columns (see section A-A in Fig. 6). Several masonry and plasterboard walls are used for separation purposes.

The floor is instrumented with eight low-cost uni-directional vibration sensors (Geophones SM-24 by I/O Sensor Nederland) to measure vertical velocity-response of the slab (one sensor per ~ 75 $m^2$). These sensors (25.4 mm diameter – 32 mm height) are low-distortion

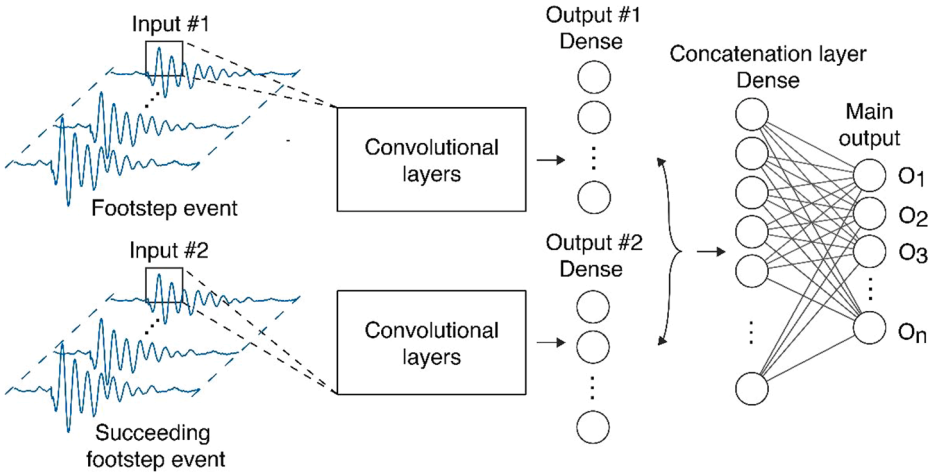


**Fig. 5.** CNN models #2 and #3 are trained with a data set that contains couples of succeeding footstep events. Each footstep event of each couple is processed in parallel through the convolutional layers. Each flatten layer is connected to a dense layer using ReLU activation function. The two dense layers are concatenated. The resulting layer is connected to the main output layer that assimilates the identity of occupants ($O_n$) using Softmax activation function.

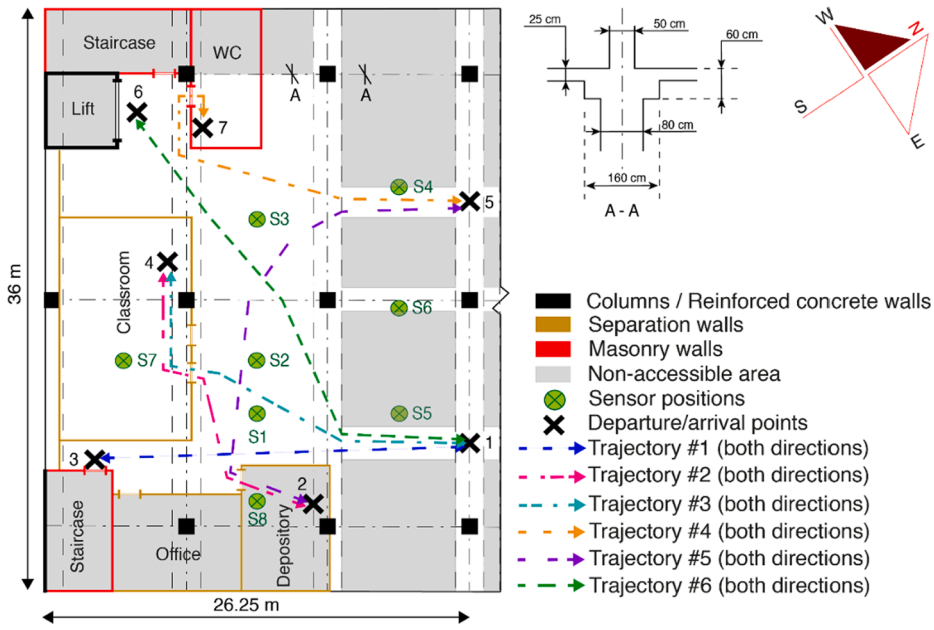


**Fig. 6.** Case Study 1. Detection of occupants, signal extraction and occupant counting are tested on a full-scale concrete slab (~950 $m^2$). Bi-directional trajectories of single occupants walking along six trajectories are used for testing. The same trajectories are used for two occupants walking simultaneously (see Table 1).

geophones (less than 0.1%) with a bandwidth up to 240 Hz. Sensors are placed at locations where the fundamental bending modes are assumed to have largest amplitudes. These locations are assumed to have high SNR signals. All sensors are wired to an acquisition unit (NI USB-6003) that is used to capture the vertical vibration measurements with a sampling rate of 1000 Hz.

Measurements are recorded for three occupants walking on the floor slab individually and following six trajectories (both directions; back and forth) as illustrated in Fig. 6. All occupants walk while wearing various types of shoes (with hard, intermediate or soft soles). The three occupants weigh between 75 and 93 kg. Moreover, measurements are recorded for two occupants walking simultaneously following eight trajectory configurations (both directions; back and forth), as explained in Table 1.

Single occupants walking along six trajectories (both directions) (see Fig. 6) as well as eight trajectory configurations of two occupants walking together (see Table 1) are used for testing the strategies described in Sections 3.1 and 3.3. Walks along these trajectories are repeated several times. During these walks, the occupant moves with self-selected step length and speed. The walking speed (in terms of steps per second) is estimated to be between 1.5 Hz and 1.8 Hz using measurements. During the walking tests, the participants count their steps.

The average velocity is removed prior to data analysis in order to withdraw the spurious offset from the non-calibrated sensors. Based on ambient vibration measurements, the modes of the structure with most energy contribution to vertical bending have frequencies between 5 and 30 Hz. The fundamental bending mode of the structure is contained within the frequency range of 9–11 Hz. Floor vibrations are contaminated by electrical devices (such as fans) that run at a fixed frequency of 50 Hz. Thus, a Butterworth stop-band filter [85] is used to remove the frequency band between 49 and 51 Hz. The goal of filtering is to enhance the SNR of the footstep-event signals.

**Table 1**
Trajectory combinations for two occupants walking simultaneously on the floor of Case Study 1. FS is footstep, T is trajectory and X is departure point (see Fig. 6).

| Configuration | Occupant/Trajectory |
|---|---|
| 1 | O1: T6 from X6 - O3: T3 from X1 |
| 2 | O1: T2 from X2 - O3: T1 from X1 |
| 3 | O1: T3 from X4 - O3: T3 from X1 |
| 4 | O1: T6 from X5 (After 4 FSs) - O3: T6 from X5 |
| 5 | O1: T6 from X5 (After 6 FSs) - O3: T6 from X5 |
| 6 | O2: T6 from X5 (After 4 FSs) - O3: T6 from X5 |
| 7 | O1: T3 from X1 (After 4 FSs) - O2: T1 from X3 |
| 8 | O2: T3 from X1 (After 4 FSs) - O1: T3 from X1 |

### 4.2. Description of case study 2

The utility of strategies explained in Sections 3.1, 3.2 and 3.4 for occupant detection, signal extraction, event classification and occupant recognition is further demonstrated using a second full-scale case study. This case study involves a reinforced concrete slab (approximately 100 $m^2$) supported by steel beams, as shown in Fig. 7. The multi-story building is located in Switzerland. The slab is 20 cm thick covered by a linoleum finishing. The steel frame is composed of five H-beams in the north, west and east ends and 12 I-beams. Six steel columns support the part of the structure that has been instrumented for occupant localization. A non-structural wall made of plasterboard is above the structure on the east end. The lower half of the west end of the slab is connected to prefabricated structural walls made of reinforced concrete. The remaining parts of the slab are joined to masonry walls.

The vertical vibrations are measured with the same sensors and acquisition unit as for Case Study 1. Sensor locations were chosen to cover the two-thirds of the space (a sensor per ~ 10 $m^2$), as shown in Fig. 7. Processing the ambient vibration measurements, the modes with most energy contribution to vertical bending are delimited by the frequency range of 15–40 Hz (see Section 3.1). The fundamental bending modes of the floor slab is contained within the frequency range of 15–18 Hz.

Measurements are carried out for five people walking individually (without fixing speed and impact locations) along a fixed trajectory (see Fig. 7) to train and test the event classification strategy. Measurements are taken for occupants walking along the trajectory without fixing the precise footstep-impact locations, without fixing the number of steps and without fixing the walking speed (steps per second). All measurements are repeated several times. Occupants are estimated to weigh between 60 and 90 kg. In addition, vibrations are recorded for other activities, such as book-dropping, chair-dragging, hand and mug

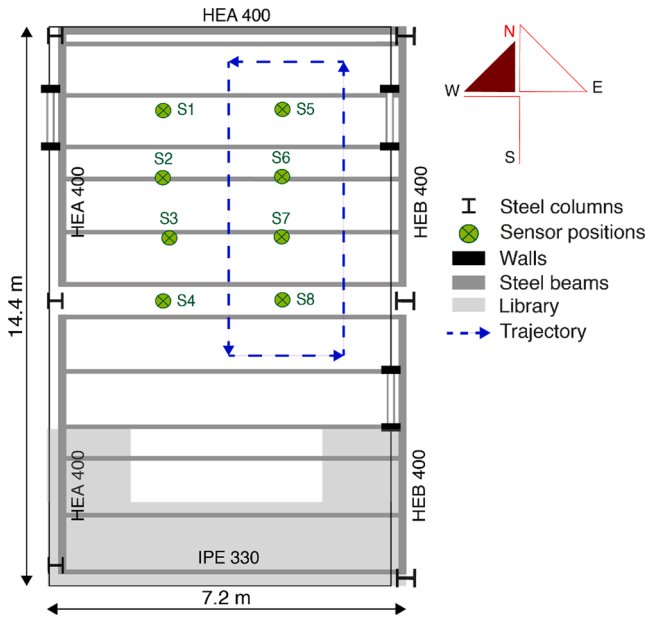


**Fig. 7.** Case Study 2. Occupant detection, event classification and occupant recognition are tested on a full-scale concrete slab supported by multiple steel beams (~100 $m^2$). Measurements have been carried out from seven people walking individually along a fixed trajectory. Book-dropping, chair-dragging, hand and mug impacts on a table, opening/closing-door events and jumping events have been recorded.

impacts on a table as well as opening and closing of doors. These non-footstep events are typical activities that occur in office environments. A second data set that contains vibration measurement induced by chair-dragging, opening/closing-door, footstep events from two other occupants (not involved in the training data) as well as jumping events are used for testing the classifier. This second data set is not used in the training phase.

Additional measurements have been carried out for five occupants walking individually along a fixed trajectory (see Fig. 7) multiple times to test the occupant-recognition strategy. Measurements refer to measurements of walking occupants with fixed footstep-impact locations and fixed walking speeds. Measurements are recorded for footsteps at 28 fixed locations separated by 75 cm (step length). Each occupant repeats the walks with two types of shoes (hard-and-soft soled shoes) and five walking speeds. Walking speeds are 1.4 Hz; 1.6 Hz; 1.8 Hz; 2 Hz and 2.2 Hz. Measurements along the trajectory being tested are repeated on average 15 times for each occupant, wearing a particular shoe type and walking at a controlled speed. For each of the five occupants, wearing two types of shoes and walking at 5-speed levels, an average of 150 measurement iterations are carried out.

### 4.3. Event-detection results

For event detection, information from low-frequency components of floor-vibration measurements is incorporated to determine the occurrence time of possible events (see Section 3.1.1). The frequency band that contains the modes with most energy contribution to vertical bending of each structure (see Sections 4.1 and 4.2) is divided into equivalent ranges of 10 Hz with an overlap of 5 Hz (see Section 3.1). Thus, vibration measurements from Case Study 1 are decomposed into frequency ranges of 5–15 Hz, 10–20 Hz, 15–25 Hz and 20–30 Hz. Vibration measurements from Case Study 2 are decomposed at frequency ranges of 15–25 Hz, 20–30 Hz, 25–35 Hz and 30–40 Hz. These decompositions help focus on the frequency components that are influenced by impact events.

Example of velocity time series from one sensor, for an occupant walking on the slabs of case studies 1 and 2 are shown in Figure 8 a1 and a2. These measurements are used to illustrate the application of the proposed event-detection operation for both case studies. In Figure 8 a1 and a2, dashed lines represent three standard deviations of the ambient noise. Data points in Figure 8, b1 and b2, are $STD_{max,r}$ values corresponding to data windows with fixed length and fixed increments.

Data points in Figure 8, c1 and d1, are $STD_{max,f}$ values that are assessed over the segmented and decomposed vibration signal from Case Study 1. In this example, vibration signals that are decomposed and reconstructed at frequency ranges of 5–15 Hz and 20–30 Hz are only presented in Figure 8, c1 and d1. Data points in Figure 8, c2 and d2, are $STD_{max,f}$ values that are assessed over the segmented and decomposed vibration signal from Case Study 2. The vibration signals shown in Figure 8, c2 and d2, are decomposed and reconstructed at frequency ranges of 15–25 Hz and 30–40 Hz. These frequency ranges are selected for illustration reasons.

Dashed lines in Figure 8, c1, c2, d1 and d2, are $DT_f$ values that are assessed at their corresponding frequency ranges (see Section 3.1.1). Triangular pointers in Fig. 8 represent local maxima that indicate the occurrence time of detected impact events (see Fig. 2).

$STD_{max,r}$ values resulting from a moving window over the non-processed signal from Case Study 1 do not lead to the detection of the first four impact events since $STD_{max,r}$ values are below the threshold of three standard deviations of the ambient noise (see arrows in Figure 8, b1). Thus, events with low SNR signals cannot be detected, which leads to inaccurate event detection. Since vibration measurements are influenced by the structure, $STD_{max,f}$ values of the decomposed and reconstructed signal at a frequency range that contains the first bending mode of the structure (see Figure 8, c1 and c2) are similar to those of the non-processed signal.

$STD_{max,f}$ values that are assessed over the decomposed signal at a frequency range that contains the first bending mode of the structure, as shown by arrows in Figure 8, c1, are insufficient to detect the first three footstep events. This is due to $STD_{max,f}$ values that are below the $DT_f$. However, these events are differentiable from ambient vibrations at higher frequency ranges, as illustrated in Figure 8, d1. Moreover, $STD_{max,f}$ values that are assessed over the decomposed signal at a frequency range that contains the first bending mode of the structure from Case Study 2 do not guarantee the detection of the seventh and the tenth impact events (see circles in Figure 8, c2). However, $STD_{max,f}$ values that correspond to the seventh and the tenth impact events of the decomposed signal at higher frequency range (see Figure 8, d2) significantly exceed the $DT_f$.

$STD_{max,f}$ values of the decomposed signal segments at a high-frequency range of 30–40 Hz do not lead to the detection of the third impact event (see arrow in Figure 8, d2), whereas $STD_{max,f}$ values of the decomposed signal segments at the first bending mode of the structure lead to accurate detection of this event, as illustrated in Figure 8, c2. Detection is successful if any frequency range indicates that the threshold has been exceeded. Thus, combining information from decomposed signals at several frequency ranges helps the event detection strategy reduce false negatives (undetected events), leading to accurate event detection.

Event detection is successfully tested to ascertain more than 24,000 footstep and non-footstep events on both case studies. For example, out of 2605 footstep events for Case Study 1, less than 1% of the events are not detected and less than 1% of the detected events are incorrect (see Table 2). Similarly, undetected events and incorrectly detected events are less than 1% of the number of footstep events (1854 footstep impacts) for Case Study 2 (see Table 2). Incorrect detection is due to the presence of additional local maxima in $STD_{max,f}$ values or incorrect signal extraction. These additional local maxima in $STD_{max,f}$ values result in additional vibrations in the dynamic response of event signals. These additional vibrations may be caused by ambient activities or reflected waves from boundary conditions. Incorrect signal extraction

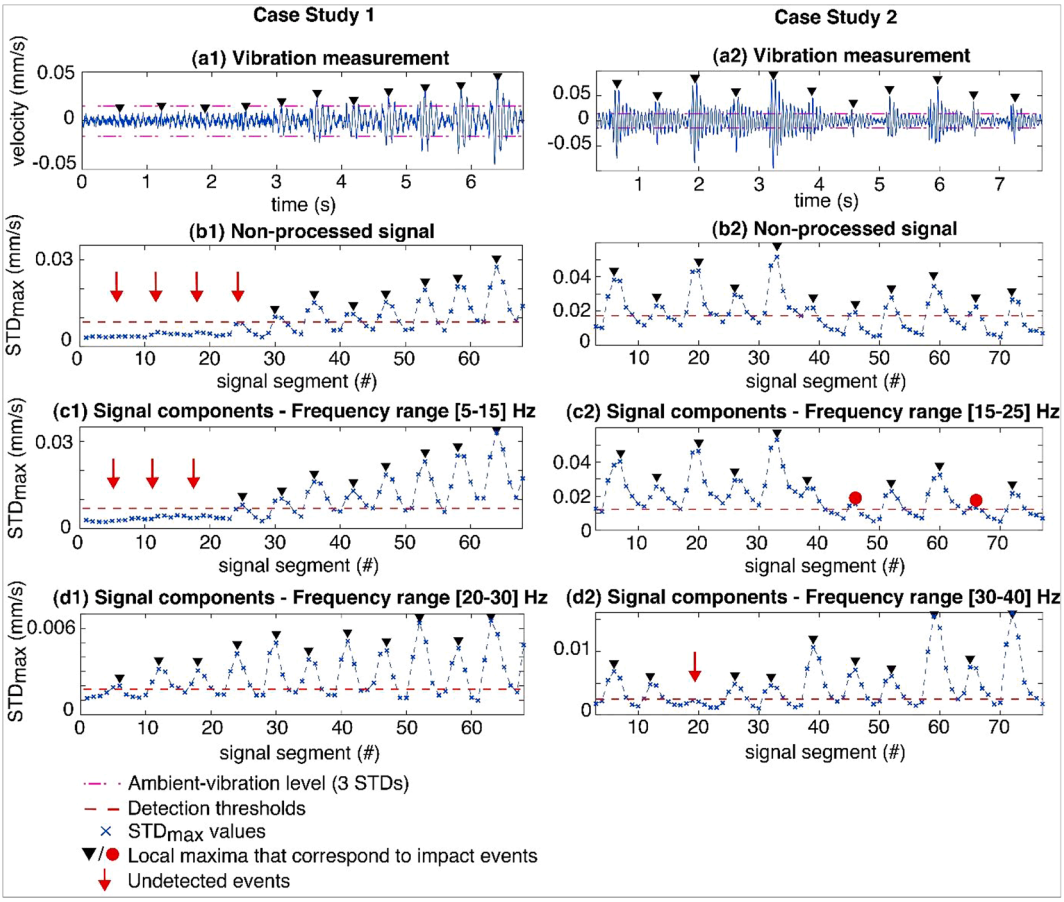


**Fig. 8.** (a1 to d1) example of event-detection results from vibrations generated by a walking occupant captured at one sensor location from Case Study 1 (see Fig. 6). (a2 to d2) example of event-detection results from vibrations generated by a walking occupant captured at one sensor location from Case Study 2 (see Fig. 7).

**Table 2**
Number of undetected and incorrectly detected events for each trajectory of single occupants walking on the floors of Case Studies 1 and 2 (see Figs. 6 and 7).

| | Trajectory | Occupant | Average number of events per test | Total number of events | Undetected events | Incorrect detection |
|---|---|---|---|---|---|---|
| Case Study 1 | 1 | O1 | 31 | 250 | 0 | 1 |
| | | O2 | 28 | 168 | 2 | 0 |
| | | O3 | 32 | 186 | 0 | 0 |
| | 2 | O1 | 24 | 240 | 0 | 0 |
| | 3 | O1 | 33 | 264 | 1 | 0 |
| | | O2 | 31 | 124 | 0 | 0 |
| | 4 | O1 | 32 | 384 | 1 | 1 |
| | | O3 | 30 | 120 | 2 | 0 |
| | 5 | O1 | 41 | 246 | 0 | 0 |
| | | O2 | 39 | 155 | 0 | 0 |
| | 6 | O1 | 40 | 318 | 1 | 0 |
| | | O2 | 37 | 150 | 1 | 0 |
| Case Study 2 | – | O1 | 28 | 526 | 0 | 0 |
| | | O2 | 28 | 379 | 1 | 1 |
| | | O3 | 28 | 449 | 0 | 2 |
| | | O4 | 28 | 242 | 0 | 1 |
| | | O5 | 28 | 258 | 1 | 6 |

results in incorrect determination of the ending time of detected event signal (see Section 3.1.2). As each trajectory is composed of several footsteps, this detection accuracy of at least 99% is deemed sufficient. Moreover, no more than two false negatives or false positives are present within a single trajectory measurement of a person walking for both case studies. Therefore, accurate event detection is achieved through combining information from multiple frequency components of measurement vibrations.

Despite the accuracy of occupant detection when involving $STD_{max,f}$ values of floor vibrations at multiple frequency components in these cases, the event-detection strategy may show limitations when applied to other building structures that are characterized by assemblies of prefabricated elements and the presence of thick and highly dissipating floor finishing materials.

### 4.4. Signal-extraction results

The goal of event extraction is to determine the starting and ending times (no fixed duration of events) of detected event signals (see Fig. 3 in Section 3.1.2). In Fig. 9, an illustration of the steps involved in

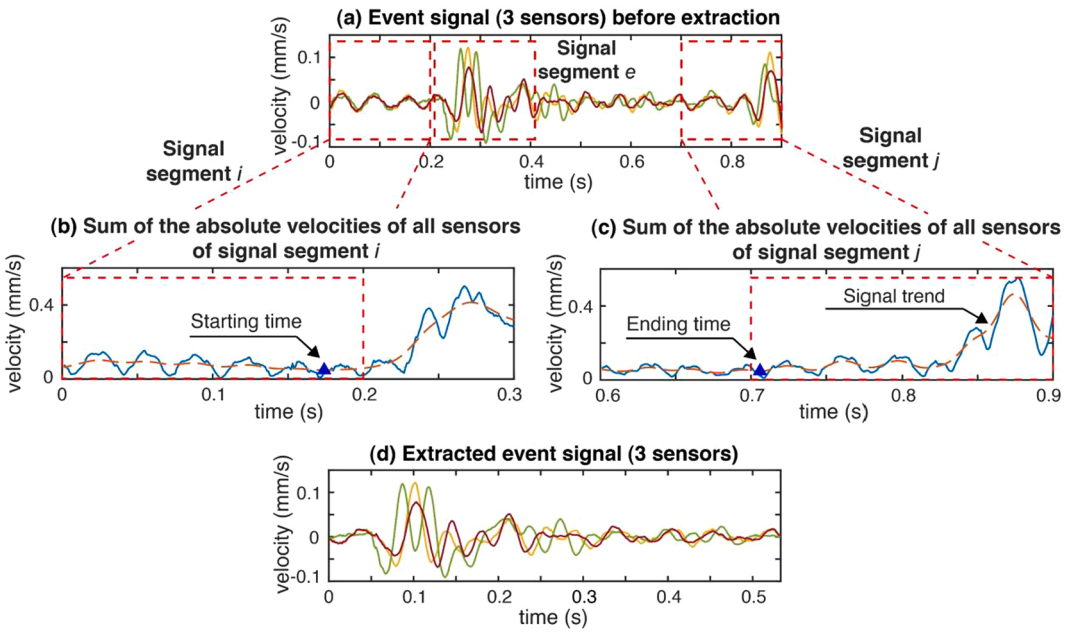


**Fig. 9.** (a) an example of a detected event signal (at three sensor locations) resulting from the determination of signal segments *i* and *j* (see Fig. 3) that contain the starting and ending times of the signal. (b and c) the sum of the absolute velocities of all sensors of signal segments *i* and *j* as well as the resulting signal trends. Minimum values of the signal trends define the starting and the ending times of the detected event signal. (d) the extracted signal at three sensor locations.

delimitating the signal of a detected footstep event from vibration measurements for Case Study 1 is shown. Signal segment *e* (see Fig. 9, a) that corresponds to the ascertained event-occurrence time, is used to perform backward and forward searches in order to determine the starting and the ending time of the event signal, as explained in Section 3.1.2 and Fig. 3.

$STD_{max,f}$ values of the segmented signal that is decomposed and recomposed at a frequency range of 20–30 Hz are used. This frequency range is greater than the first natural frequency of the structure (see Section 4.2). Signal segment *e* is used to perform backward and forward searches based $STD_{max,f}$ values (see Section 3.1.2). The resulting signal segment *i* contains the starting time of the detected event signal, as shown in Fig. 9, b. The resulting signal segment *j* contains the ending time of the detected event signal, as shown in Fig. 9, c.

In Fig. 9, b and c present the sum of the absolute velocities of the non-processed signal of all sensors, bounded by signal segments *i* and *j*, as

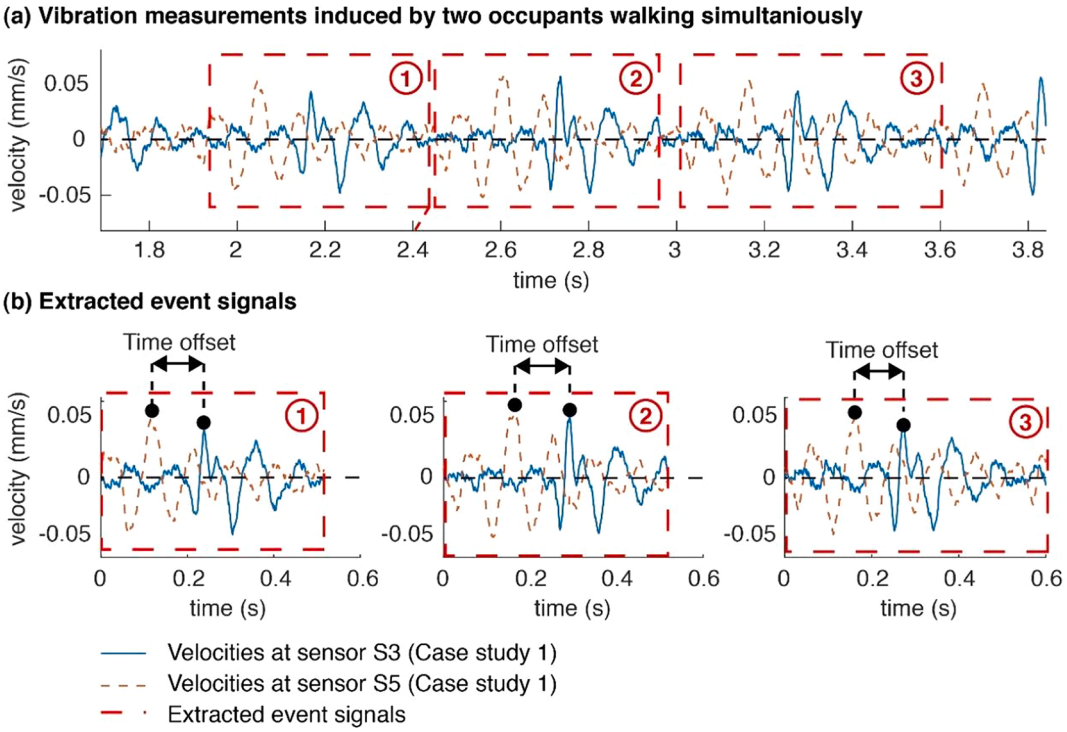


**Fig. 10.** (a) vibration measurements from two occupants walking simultaneously on the floor of case study 1 (see Fig. 6) following the first trajectory configuration (see Table1). (b) the extracted event signals as well as time offsets within the overlapping signals.

well as their corresponding signal trends. Signal trends are assessed based on applying a weighted moving average on the resulting absolute cumulative signals (see Section 3.1.2). Minimum values of the resulting trends define the starting and the ending times of a detected event signal. Finally, the detected event signal is extracted for all sensors as illustrated in Fig. 9, d.

Signal extraction is carried out along with the event-detection operation (see Figs. 2 and 3). Signal extraction is successfully tested to extract footstep and non-footstep event signals from continuous vibration measurements (see Table 2).

Signal extraction is successfully tested on measurements conducted by two people walking simultaneously. In Fig. 10, an example of event signals extracted from overlapping signals for two occupants walking simultaneously on the floor of Case Study 1 (see Fig. 6) is shown. The two occupants followed the first trajectory configuration, as defined in Table 1. In Fig. 10, vibration measurements at sensor locations S3 and S5 (see Fig. 6) are illustrated. Extracted event signals are shown in Fig. 10, b. The signal-extraction strategy involves capturing the time offsets within the overlapping footstep event signals (see Fig. 10, b). The starting and the ending times of each extracted event signal cover accurately amplitudes that are contributed by footstep impacts of each occupant. Vibration measurements of several occupants walking together show that footstep events overlap to varying offsets (see Fig. 10, b).

### 4.5. Event-Classification results

Subsequent to event detection, the next step involves differentiating between footsteps and non-footstep events using the extracted vibration signals (see Section 2). A binary-SVM learning approach is used to differentiate footstep events from spurious (non-footstep) events. Binary-SVM classifier performance is compared with k-nearest neighbors (KNN) [51] and boosted tree (BT) [53] classifiers.

Feature selection is important in order to ensure good classification performance. Several metrics are assessed in time and frequency domains as explained in Section 3.2. Frequency domain metrics are $FSV_{max}$ and $C_{CPSD}$ (see Section 3.2) matrices. Time-domain metrics are maximum and average $\sigma$, $\Delta_{amp}$, RMS, $Kr$ and $Md$ of event signals (see Section 3.2). These metrics are calculated for all sensors at various frequency ranges. Event signals are decomposed using CWT and reconstructed at equivalent frequency intervals of 20 Hz with an overlap of 10 Hz (see Section 3.2). This covers the frequency band (10 Hz to 240 Hz) of vibration sensors used to instrument the floor slab (see Section 4.1).

Time-domain metrics at specific frequency ranges that maximize the discrepancy between footstep and non-footstep event classes are selected as features for classification using null-hypothesis tests, as explained in Section 3.2. Metrics such as $\Delta_{amp}$, RMS and $Md$ are correlated with the $\sigma$ and therefore excluded from the training process to avoid overfitting. The features that are found to be useful in separating footsteps from other events are presented in Table 3. These features are average standard deviation ($\sigma_{mean}$) of all sensors of event signals at frequency ranges of 10–30 Hz and 70–90 Hz, maximum kurtosis ($Kr_{max}$) of all sensors of event signals at a frequency range of 50–70 Hz, average kurtosis ($Kr_{mean}$) of all sensors of event signals at a frequency range of 90–110 Hz as well as $FSV_{max}$ and $C_{CPSD}$. Classification performances using the selected frequency ranges for time-domain features (see Table 3) are compared with classifiers trained using data from raw-event signals.

**Table 3**
Metrics that maximize the discrepancy between footstep and non-footstep event classes are selected as features for classification. Null-hypothesis tests of each time-domain metric (assessed at various frequency ranges) have been conducted to select the frequency ranges that best differentiates footsteps from other events (see Section 3.2).

| Feature | Frequency range |
|---|---|
| $\sigma_{mean}$ | [10 –30] Hz |
| $\sigma_{mean}$ | [70–90] Hz |
| $Kr_{max}$ | [50 –70] Hz |
| $Kr_{mean}$ | [90–110] Hz |
| $C_{CPSD}$ | All frequencies |
| $FSV_{max}$ | All frequencies |

A dataset composed of measurements from Case Study 2 that correspond to five people walking separately along a trajectory (see Fig. 7) multiple times is used to train and test the classifiers. Since the adopted strategy is based on a binary-SVM classifier, non-footstep events including book-dropping, chair-dragging, hand and mug impacts on a table as well as opening and closing of doors have been measured and used to train the event classification learning. The data set is composed of 1853 footstep events and 390 non-footstep events. The data set is randomly split into 75% for training and 25% for validation. These ratios for training and validation are chosen based on the available number of measurements. A second data set that contains vibration measurement induced by chair-dragging, opening and closing of doors as well as footsteps and jumps of two additional occupants, which have not been involved in the training data, are used to test the classifier performance. The second data set, measured on another day on the same slab, includes 169 footstep events and 42 non-footstep events.

The performance scores (accuracy, precision, recall and F1) of the SVM classifier, which is compared with KNN and BT classifiers are presented in Table 4. The features used for training KNN and BT classifiers are presented in Table 3. A Gaussian kernel is used to train the binary-SVM classifier since it provides better performance compared with other kernels.

Accuracy is the number of correct predictions divided by the total number of predictions. Accuracy score is calculated using Eq. 1. In Eq. 1, $TP$ is true positives that present correct classification of footstep events, $TN$ is true negatives that present correct classification of non-footstep events, $FP$ is false positives that present incorrect classification of non-footstep events as footsteps, $FN$ is false negatives that present incorrect classification of footstep events as non-footstep events. Precision is the proportion of the number of correct classifications of footstep events ($TP$) from all events classified as footstep events ($TP + FP$). Precision score is calculated using Eq. 2. Recall is the proportion of the number of correct classifications of footstep events ($TP$) from all footstep events. Recall score is calculated using Eq. 3. F1score is the overall performance metric that reflects the classifier model's ability to distinguish between footstep and non-footstep events. F1 score is calculated using Eq. 4. Type I and II errors are used for further comparison between classification approaches. Type I error is the rate of footstep events that are identified as non-footstep events. Type II error is the rate of non-footstep events that are identified as footstep events.

$$Accuracy = \frac{TN + TP}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1 = 2\frac{Precision^{*}Recall}{Precision + Recall} \tag{4}$$

Binary-SVM classifier, trained using raw and decomposed signals at selected frequency ranges, provides similar prediction performance as KNN and BT classifiers. For example, the overall performance as defined by F1 score (see Eq.4) exceeds 95% in most cases. Predictions obtained using a test dataset show that binary-SVM classifiers based on decomposed and raw signals perform better than KNN and BT classifiers. Thus,

**Table 4**
Event-classification performances of the binary-SVM classifier based on Gaussian kernel compared with KNN and BT classifiers for Case Study 2 (see Fig. 7). The first data set is split into 75% for training and 25% for validation, which is reported. The second data set is used to estimate prediction performances. Classification performance of models that are trained using raw and decomposed signals at selected frequency ranges using the null-hypothesis tests is included for comparison.

| | | SVM | | KNN | | Boosted tree | |
|---|---|---|---|---|---|---|---|
| | | **Validation test** | **Second data test** | **Validation test** | **Second data test** | **Validation test** | **Second data test** |
| **Decomposed signals (CWT)** | Accuracy (%) | 97.5 | 98.6 | 97.3 | 91.9 | 98.2 | 93.8 |
| | Precision (%) | 98.2 | 100 | 97.8 | 93.2 | 98.0 | 95.9 |
| | Recall (%) | 98.7 | 98.2 | 98.9 | 97.1 | 99.8 | 96.5 |
| | F1 (%) | 98.5 | 99.1 | 98.4 | 95.1 | 98.9 | 96.2 |
| | **Error I (%)** | **1.3** | **1.8** | **1.1** | **3** | **0.2** | **3.6** |
| | **Error II (%)** | **8.1** | **0** | **10.2** | **28.6** | **9.2** | **16.7** |
| **Raw signals** | Accuracy (%) | 95.0 | 92.9 | 94.7 | 85.3 | 97.3 | 85.3 |
| | Precision (%) | 95.8 | 98.1 | 95.9 | 90.1 | 97.4 | 91.0 |
| | Recall (%) | 98.3 | 92.9 | 97.6 | 91.7 | 99.3 | 90.5 |
| | F1 (%) | 97.0 | 95.4 | 96.8 | 91.0 | 98.4 | 91.0 |
| | **Error I (%)** | **1.7** | **7.1** | **2.4** | **8.3** | **0.7** | **9.5** |
| | **Error II (%)** | **20** | **7.2** | **19** | **40.5** | **12** | **35.7** |

binary-SVM classifier provides better generalization than KNN and BT classifiers.

Event classification using binary-SVM is less sensitive to type I and type II errors compared with KNN and BT classifiers. Also, using raw signals for training increases type II errors for all tested classifiers. Thus, feature selection using null-hypothesis tests at various frequency ranges helps improve classifier performance. Therefore, event-classification performance is enhanced by selecting appropriate frequency components of vibration measurements. Although the data sets that are used to train and validate the binary-SVM result in good event classification, vibration measurements from other floor configurations, other spurious sources, and footstep events from multiple occupants may affect the classification performance.

### 4.6. Occupant-counting results

Occupant counting, as explained in Section 3.3, has been tested with footstep-induced vibration measurements from several participants walking on the floor of Case Study 1 (see Fig. 6). All occupants have been either walking alone (single occupants) or together (two occupants), following multiple trajectories (see Table 1).

The data set for occupant counting includes 2597 footstep events from single occupants and 1452 events from two people walking together. The data set is randomly split into 75% for training and 25% for validation based on the available number of measurements (see Section 4.5). Counting the number of occupants using the binary-SVM classifier is performed on processed footstep-event signals. Measured signals are processed using a stop-band filter to enhance their SNR regarding noise from electrical devices (see Section 4.1).

Two feature classes are used to train the learning algorithm. The feature class #1 includes only cross-correlation coefficients between signals between sensors while the feature class #2 includes the standard deviation (σ) of events signals recorded at each sensor and the maximum CPSD along with the cross-correlation coefficients. Classification performance based on the two feature classes is illustrated in Table 5 along with a comparison between SVM, KNN and BT classifiers.

Accuracy, precision, recall and F1 scores, calculated using Eq. 1 to Eq. 4 are presented in Table 5. These metrics help to assess the performance scores for each classifier using raw and decomposed footstep-event signals. Also, type I error that defines the rate of one occupant classified as two occupants and type II error that defines the rate of two occupants classified as one occupant are used for further comparison between classification approaches. Binary-SVM classifier to determine the number of occupants is trained based on several kernels such as linear kernel, Gaussian kernel, third-and-fourth degree polynomial kernels. Third-degree polynomial kernel provides best classification performance.

**Table 5**
Classification performance, based on validation test (25% of data set), to distinguish between one and two occupants on the floor of Case Study 1 (see Fig. 6).

| | | SVM | KNN | Boosted tree |
|---|---|---|---|---|
| **Feature class #1** | Accuracy (%) | 89.6 | 82.4 | 76 |
| | Precision (%) | 90.4 | 82.7 | 76.6 |
| | Recall (%) | 93.5 | 91.1 | 89.3 |
| | F1 (%) | 91.9 | 86.7 | 82.4 |
| | **Error I (%)** | **6.5** | **8.9** | **10.7** |
| | **Error II (%)** | **17.1** | **32.5** | **46.7** |
| **Feature class #2** | Accuracy (%) | 94.7 | 91.9 | 89.1 |
| | Precision (%) | 94.4 | 90.6 | 89.5 |
| | Recall (%) | 97.5 | 97.5 | 94 |
| | F1 (%) | 95.9 | 93.9 | 91.7 |
| | **Error I (%)** | **2.5** | **2.5** | **6.0** |
| | **Error II (%)** | **10.4** | **18.1** | **19.7** |

Based on validation test (25% of data set), the binary-SVM classifier is able to differentiate between the presence of either one or two occupants with performance scores exceeding 90% when using both feature classes on processed event signals for training. Binary-SVM classifier provides better performance scores (average increase of 7%) than KNN and BT classifiers. For example, when training the SVM with only cross-correlation coefficients, the overall prediction performance defined by F1 score is equal to 92% for SVM and less than 90% for KNN and BT classifiers (see Feature class #1 in Table 5).

Incorporating $\sigma$ values and maximum CPSD along with the cross-correlation coefficients (see Feature class #2 in Table 5) for training increases the classification performance scores with an average of 5% for SVM, 9% for KNN and 13% for BT. This improvement varies between 4% and 17% for all performance scores of all classifiers compared with using only cross-correlation coefficients as features (see Feature class #1 in Table 5).

Using the Feature class #2 for training, fewer type II errors (two occupants are classified as one occupant) are produced using SVM (10%) compared with KNN and BT classifiers (18% and 20%). Therefore, using cross-correlations between footstep signals recorded at various sensors as features provides efficient classifier in distinguishing between one and two occupants. Incorporating $\sigma$ values and maximum CPSD along with the cross-correlation coefficients enhances the performances of the occupant-counting classifier.

Since the floor responses are governed by the structural behavior of the slab, distances between impact and sensor locations significantly affect the cross-correlation coefficients between sensor recordings from the same footstep event. Thus, sensor configurations that do not systematically cover the entire floor space may affect the efficiency of

classification.

### 4.7. Occupant recognition results

Occupant recognition is performed using CNN classification as explained in Section 3.4. TensorFlow, an open-source library for machine-learning applications provided by Google [86], is used for classification. The CNN classification is trained with footstep-induced floor vibrations. The occupant-recognition strategy, explained in Section 3.4, is tested on Case Study 2 (see Fig. 7) to recognize the occupant walking on the floor slab, among five participants for which data is available for training. Controlled vibration measurements have been repeatedly conducted with five single occupants. Each occupant has walked with hard-and-soft soled shoes and at five walking-speed levels from slow to fast walking, leading to a total of 19526 footstep events (see Section 4.2).

The length of extracted footstep-event signals varies between 0.4 s and 0.7 s (see Section 3.1.1, which corresponds to 400–700 data points given the measurement-sampling rate (see Section 4.2). All footstep-event signals captured at eight sensors (see Fig. 7) are resampled into an equivalent length of 200 points to reduce the computational cost during training process without losing signal information. A finite impulse response (FIR) filter based on least-squared design [87] is used for resampling. Occupant recognition has been tested using three CNN classification models (see Section 3.4) and compared with a shallow NN classifier.

Footstep-event signals from all sensors (size of 8x200) are connected to two successive convolutional layers using window sizes of (3x50) and (3x20). In addition, a max-pooling operation is used on each convolutional output layer using a pool size of (1x3) and (2x3).

CNN model #1 is trained with separate footstep events as input patterns (see Fig. 4). The flatten layer resulting from convolutional layers is connected to a dense layer of 100 neurons using a ReLU activation function. Using the Softmax activation function, the dense layer ends with a class label-layer output. The output layer is composed of five neurons that assimilate the identity of five participants. The data set (19526 events) is split into 80% for training and 20% for validation test based on the available number of measurements (see Section 4.5). 30 epochs are used to train the deep learning classifier.

In CNN models #2 and #3 input data are rearranged into couples of two succeeding footstep events (see Fig. 5). Thus, the data set includes 18,775 footstep-event couples. Each flatten layer is connected to a dense layer of 50 neurons using ReLU activation functions. The two dense layers are subsequently concatenated. The resulting layer is finally connected to the main output layer using Softmax activation function.

In this application, the shallow NN architecture is composed of three hidden dense layers that connect the input layer using a ReLU activation function. The first dense layer (fully connected layer) contains 1600 neurons that correspond to the size of the captured footstep event at eight sensors (8x200). The second and the third layers contain 800 and 200 neurons respectively. These dense layers end with a label-layer output, using the Softmax activation function. The data set includes 19,526 footstep events with an approximately equivalent contribution from each occupant. The data set is split into 80% for training and 20% for validation.

Occupant-recognition performances of the three CNN models are compared with a shallow NN classifier, as presented in Table 6. Classification using shallow NN results in low-performance scores. For example, classification accuracy, calculated according to Eq. 1, is less than 60% for all five participants. Also, the precision of the shallow NN classifier (see Eq. 2) varies between 44% and 78%. Recall score, as defined by Eq. 3, varies between 48% and 71%. Classification using CNN model #1 (see Fig. 4) results in an improvement of 22% in recognition accuracy and recall scores as well as 21% in precision compared with shallow NN.

Training CNN with separate footstep events provides recognition performances (accuracy, precision and recall scores) of 80% for all occupants. Moreover, reinforcing the training data of CNN classification (CNN model #2 in Fig. 5) with two succeeding footstep-events leads to better recognition scores than using isolated footstep events. For example, CNN model #2 has an accuracy of approximately 90% and precision of 91% for all participants. However, precision of identity recognition of occupant #4 (see O4 in Table 6) is less than 80% compared with other occupants. Also, recall score of identity recognition of occupant #5 (see O5 in Table 6) using CNN model #2 is less than 75%.

**Table 6**
Occupant-recognition performances of five participants, based on validation test (20% of data set). Occupant-recognition classification has been performed using a shallow NN classifier and three CNN models (see Section 3.4).

| | Occupant | Shallow NN | CNN model #1 | CNN model #2 | CNN model #3 |
|---|---|---|---|---|---|
| **Accuracy (%)** | **Average** | **57.7** | **79.7** | **89.6** | **96.3** |
| **Precision (%)** | O1 | 58.2 | 78.4 | 91.9 | 98.7 |
| | O2 | 65.1 | 76.4 | 90.0 | 92.4 |
| | O3 | 77.8 | 89.7 | 94.7 | 98.6 |
| | O4 | 51.3 | 80.5 | 79.6 | 95.6 |
| | O5 | 44.4 | 74.1 | 96.3 | 96.7 |
| | **Average** | **59.4** | **79.8** | **90.5** | **96.4** |
| **Recall (%)** | O1 | 48.4 | 76.6 | 86.2 | 91.1 |
| | O2 | 51.0 | 80.5 | 92.8 | 97.7 |
| | O3 | 70.7 | 88.1 | 96.7 | 99.2 |
| | O4 | 54.9 | 75.5 | 96.5 | 96.9 |
| | O5 | 64.8 | 78.3 | 74.1 | 96.4 |
| | **Average** | **57.9** | **79.8** | **89.3** | **96.3** |

In order to enhance recognition scores for all occupants, CNN model #3 involves a multi-objective function that includes weighted losses from convolutional outputs #1 and #2 as well as from the main output layer (see Section 3.4). Thus, CNN model #3 provides good accuracy (96%) for the recognition of all occupants compared with approximately 90% with CNN model #2. Also, CNN model #3 yields good precision and recall scores of more than 95% for occupants O3, O4 and O5.

Therefore, feeding the CNN classifier with footstep signals captured at various sensors leads to an accurate occupant recognition. Moreover, training CNN classifier with two succeeding footstep events improves occupant-recognition performances compared with the use of isolated footstep events. CNN classification performances are enhanced by taking into account the errors from each convolutional output layer of each footstep event as well as from the main output layer during the training process.

Recognition using CNN on succeeding footstep events from five occupants results in high classification performance. Footstep-event signatures that are used for training and validation are from occupants that were walking individually. Generalization towards recognition of multiple occupants walking together may require separation of superimposed signals in order to maintain applicability of the occupant-recognition strategy.

## 5. Summary of results and discussion

Occupant detection, counting and recognition strategies have been applied to vibration measurements of two full-scale case studies. Vibration sensors provide non-intrusive occupant detection. Sensors are used to cover areas of up to 600 $m^2$. Once trained, processing the measurements for validation is not computationally expensive (near-real time). Thus, the proposed strategies have potential to be used in practice.

Accurate event detection is achieved through assessment of $STD_{max,f}$ values over segmented and decomposed vibration measurements at multiple frequency ranges. These frequency ranges cover the fundamental vertical modes of the structure. The frequency band with the

most energy contribution is delimited by the prominent peaks in the singular values of the CPSD derived from ambient vibrations.

Classification of events as footsteps and non-footstep events is carried out using a binary-SVM. Time-and-frequency domain features are used for training. Time-domains features are assessed for event signals decomposed and reconstructed at low-and-high frequency ranges. These ranges cover the frequency band that sensors provide. Null-hypothesis tests are used to select the most informative frequency ranges of time-domain features in order to improve the efficiency of the event classification methodology. However, the event-classification strategy has been tested using only individual footstep events from a single occupant walking on the floor slab. Thus, testing the strategy for overlapping footstep-event signals from multiple occupants walking simultaneously may be required for practical applications. Such testing could follow the same methodology presented in this paper.

Differentiation between one and two occupants has been achieved using a binary-SVM classifier. Cross-correlation coefficients between event signals at all sensor locations from each footstep event are used to train the SVM classifier. Using cross-correlations between footstep signals recorded at various sensors as features provides an efficient classifier in distinguishing between one and two occupants. Incorporating $\sigma$ values and maximum CPSD along with the cross-correlation coefficients enhances the performances of the occupant-counting classifier. However, counting the number of occupants is currently limited to one or two walking occupants. Testing for determining the number of individuals when more than two occupants walk together is future work.

In addition, correlation coefficients between sensor signals from the same footstep event are influenced by the distance between the impact and sensor locations. Sensor configuration that does not cover the floor space may lead to inefficient classification. Thus, careful placement of sensors over the floor slab is necessary to guarantee good performance of classifiers and algorithms for occupant detection and localization. Strategies for optimal sensor placement are part of future work. A solution to optimize the sensor configuration could be carried out using the joint entropy of footstep impacts [88].

Training the CNN classifier with two succeeding footstep signatures captured at various sensors improves occupant recognition. Occupant recognition has been enhanced by taking into account errors from convolutional output layers of each footstep event and from the main output layer during the training process. However, the CNN training process is computationally expensive (~one minute per iteration with Intel Core i7-4770 CPU). Also, occupant recognition has been tested using measurements from five participants walking individually on fixed impact locations and speed levels (see Section 4.1).

Evaluation of the recognition strategy using measurements from more than five occupants walking with self-selected step lengths and speed levels would be required for comprehensive occupant recognition. Further testing with realistic scenarios involving multiple occupants walking simultaneously would help recognize each occupant more accurately. Including measurements from occupants that are not involved in the training for occupant recognition would also be useful to evaluate the capability of the system to recognize strangers. In order to ensure high recognition performance, video-based monitoring could be employed during commissioning. The evaluation of occupant recognition on other full-scale floor slabs would further assess generality.

Structures having multiple assemblies such as wooden buildings and prefabricated elements as well as the presence of thick and energy-dissipating floor finishing materials may limit the applicability of the proposed strategies. Since the vast majority of multi-story structures have continuous concrete slabs, the results presented in this study give a good estimate of the performance that can be expected in most cases. In addition, due to the uniqueness of structural behavior amongst floor slabs, applications of the classification strategies to other floor slabs would require modal analysis for commissioning and re-training the learning algorithms with appropriate vibration measurements. Strategies for occupant detection and counting have been successfully validated on other full-scale case studies and with other measurement scenarios [89]. Although there is potential for transferability, further validation of the strategies for occupancy detection and recognition on other case studies is necessary.

A classification-performance analysis to determine the appropriate ratio of data classes for training and testing is needed. Although, the accuracy of the strategies for occupancy detection and recognition exceeds 95%, type I and type II errors remain greater than 5% for some cases as shown in Tables 4 and 5. In order to ensure a high classification performance, iterative transductive learning algorithms that update the labeled data set model with unlabeled testing results based on either physical insights may be useful [40].

Use of occupancy detection and recognition may be reused on several floors of multi-story-buildings when they have similar floor-slab configuration. This would allow use of the same data for all floors for training the learning classifiers. However, the dynamic response regarding footstep impacts may vary between floors. Thus, studies of the dynamic variability in floor responses are needed for multi-story-building applications.

Detected footstep-event signals, knowledge of number of occupants on the floors and occupant identities are valuable information for occupant localization. Using a model-based approach, localization of walking occupant may then perform separately for each captured footstep-event signal. Identifying the possible locations of consecutive footstep event enables the tracking of walking occupants. The number of occupants, as well as their identities, are essential to generate appropriate footstep-impact simulations for model-based occupant localization.

## 6. Conclusions

Occupant detection, counting and recognition strategies have been developed and evaluated using two full-scale case studies. The following conclusions are drawn:

- Combining information from multiple frequency components of measured vibrations improves the accuracy of event detection.
- Selection of appropriate frequency components for training enhances the performance of classifiers that are developed to distinguish between footstep and non-footstep events.
- Using cross-correlations between event signals measured at multiple sensor locations as features improve the performance of the classifier to distinguish between the presence of either one or two occupants.
- Incorporating information from multiple footsteps and knowledge of errors from convolutional output layers during the training process improves the performance of CNN classifiers for occupant recognition.

## Funding

This work was funded by the Applied Computing and Mechanics Laboratory (IMAC) EPFL and the Singapore-ETH Center (SEC) under contract no. FI 370074011-370074016.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

The authors acknowledge the participants that have been involved in vibration measurements: Amal Trabelsi, Marco Proverbio, Gennaro Senatore, Kazuki Hayashi and Arka Reksowardojo. The authors also acknowledge Tectus Dreamlab Pte Ltd and BBR Holdings Singapore for

access and resources provided for full-scale evaluations of strategies that are described in this paper.

## References


[1] B. Song, H. Choi, H.S. Lee, Surveillance tracking system using passive infrared motion sensors in wireless sensor network, in: 2008 Int. Conf. Inf. Netw., 2008: pp. 1–5.

[2] W.P.L. Cully, S.L. Cotton, W.G. Scanlon, J.B. McQuiston, Localization algorithm performance in ultra low power active RFID based patient tracking, in: 2011 IEEE 22nd Int. Symp. Pers. Indoor Mob. Radio Commun., 2011: pp. 2158–2162.

[3] W.P.L. Cully, S.L. Cotton, W.G. Scanlon, Empirical performance of RSSI-based Monte Carlo localisation for active RFID patient tracking systems, Int. J. Wirel. Inf. Networks. 19 (2012) 173–184.

[4] G. Diraco, A. Leone, P. Siciliano, People occupancy detection and profiling with 3D depth sensors for building energy management, Energy Build. 92 (2015) 246–266.

[5] C.M. Stoppel, F. Leite, Integrating probabilistic methods for describing occupant presence with building energy simulation models, Energy Build. 68 (2014) 99–107.

[6] W. Shen, G. Newsham, B. Gunay, Leveraging existing occupancy-related data for optimal control of commercial office buildings: a review, Adv. Eng. Informatics. 33 (2017) 230–242.

[7] D.T. Alpert, M. Allen, Acoustic gait recognition on a staircase, in: 2010 World Autom. Congr., 2010: pp. 1–6.

[8] J.T. Geiger, M. Kneißl, B.W. Schuller, G. Rigoll, Acoustic gait-based person identification using hidden Markov models, in: Proc. 2014 Work. Mapp. Personal. Trait. Chall. Work., 2014: pp. 25–30.

[9] L.M. Candanedo, V. Feldheim, Accurate occupancy detection of an office room from light, temperature, humidity and CO2 measurements using statistical learning models, Energy Build. 112 (2016) 28–39.

[10] C. Jiang, M.K. Masood, Y.C. Soh, H. Li, Indoor occupancy estimation from carbon dioxide concentration, Energy Build. 131 (2016) 132–141.

[11] R. Serra, P. Di Croce, R. Peres, D. Knittel, Human step detection from a piezoelectric polymer floor sensor using normalization algorithms, in: SENSORS, 2014 IEEE, 2014: pp. 1169–1172.

[12] R. Serra, D. Knittel, P. Di Croce, R. Peres, Activity recognition with smart polymer floor sensor: Application to human footstep recognition, IEEE Sens. J. 16 (2016) 5757–5775.

[13] V.L. Erickson, S. Achleitner, A.E. Cerpa, POEM: Power-efficient occupancy-based energy management system, in: Proc. 12th Int. Conf. Inf. Process. Sens. Networks, Philadelphia, Pennsylvania, USA, 2013: pp. 203–216.

[14] P. Henry, M. Krainin, E. Herbst, X. Ren, D. Fox, RGB-D mapping: Using Kinect-style depth cameras for dense 3D modeling of indoor environments, Int. J. Rob. Res. 31 (2012) 647–663.

[15] J. Lu, T. Sookoor, V. Srinivasan, G. Gao, B. Holben, J. Stankovic, E. Field, K. Whitehouse, The smart thermostat: using occupancy sensors to save energy in homes, in: Proc. 8th ACM Conf. Embed. Networked Sens. Syst., Zürich, Switzerland, 2010: pp. 211–224.

[16] P. Lazik, N. Rajagopal, O. Shih, B. Sinopoli, A. Rowe, ALPS: A bluetooth and ultrasound platform for mapping and localization, in: Proc. 13th ACM Conf. Embed. Networked Sens. Syst., Seoul, South Korea, 2015: pp. 73–84.

[17] J.T. Biehl, M. Cooper, G. Filby, S. Kratz, Loco: a ready-to-deploy framework for efficient room localization using wi-fi, in: Proc. 2014 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput., 2014: pp. 183–187.

[18] W. Wang, J. Chen, T. Hong, Occupancy prediction through machine learning and data fusion of environmental sensing and Wi-Fi sensing in buildings, Autom. Constr. 94 (2018) 233–243.

[19] N. Li, B. Becerik-Gerber, Performance-based evaluation of RFID-based indoor location sensing solutions for the built environment, Adv. Eng. Informatics. 25 (2011) 535–546.

[20] Z.D. Tekler, R. Low, B. Gunay, R.K. Andersen, L. Blessing, A scalable Bluetooth Low Energy approach to identify occupancy patterns and profiles in office spaces, Build. Environ. 171 (2020), 106681.

[21] K. Weekly, N. Bekiaris-Liberis, M. Jin, A.M. Bayen, Modeling and estimation of the humans' effect on the CO 2 dynamics inside a conference room, IEEE Trans. Control Syst. Technol. 23 (2015) 1770–1781.

[22] A. Kamthe, L. Jiang, M. Dudys, A. Cerpa, Scopes: Smart cameras object position estimation system, in: Eur. Conf. Wirel. Sens. Networks, Cork, Ireland, 2009: pp. 279–295.

[23] A. Bamis, D. Lymberopoulos, T. Teixeira, A. Savvides, The BehaviorScope framework for enabling ambient assisted living, Pers. Ubiquitous Comput. 14 (2010) 473–487.

[24] K.S. Gautam, S.K. Thangavel, Video analytics-based intelligent surveillance system for smart buildings, Soft Comput. 23 (2019) 2813–2837.

[25] S. Budi, K. Hyoungseop, T.J. Kooi, I. Seiji, Real time tracking and identification of moving persons by using a camera in outdoor environment, (2009).

[26] L. Wang, T. Tan, H. Ning, W. Hu, Silhouette analysis-based gait recognition for human identification, IEEE Trans. Pattern Anal. Mach. Intell. 25 (2003) 1505–1518.

[27] W.-S. Zheng, S. Gong, T. Xiang, Person re-identification by probabilistic relative distance comparison, in: CVPR 2011, 2011: pp. 649–656.

[28] S. Narayana, R.V. Prasad, V.S. Rao, T. V Prabhakar, S.S. Kowshik, M.S. Iyer, PIR sensors: Characterization and novel localization technique, in: Proc. 14th Int. Conf. Inf. Process. Sens. Networks, Seattle, Washington, 2015: pp. 142–153.

[29] G. Fierro, O. Rehmane, A. Krioukov, D. Culler, Zone-level occupancy counting with existing infrastructure, in: Proc. Fourth ACM Work. Embed. Sens. Syst. Energy-Efficiency Build., Toronto, Ontario, Canada, 2012: pp. 205–206.

[30] Y. Zeng, P.H. Pathak, P. Mohapatra, WiWho: wifi-based person identification in smart spaces, in: Proc. 15th Int. Conf. Inf. Process. Sens. Networks, 2016: p. 4.

[31] Q. Pu, S. Gupta, S. Gollakota, S. Patel, Whole-home gesture recognition using wireless signals, in: Proc. 19th Annu. Int. Conf. Mob. Comput. Netw., 2013: pp. 27–38.

[32] H. Lee, C.R. Ahn, N. Choi, Fine-grained occupant activity monitoring with Wi-Fi channel state information: Practical implementation of multiple receiver settings, Adv. Eng. Informatics. 46 (2020), 101147.

[33] S. Feldmann, K. Kyamakya, A. Zapater, Z. Lue, An indoor bluetooth-based positioning system: Concept, implementation and experimental evaluation., in: Int. Conf. Wirel. Networks, 2003.

[34] T. Alhmiedat, G. Samara, A.O.A. Salem, An Indoor Fingerprinting Localization Approach for ZigBee Wireless Sensor Networks, Eur. J. Sci. Res. ISSN 1450-216X / 1450-202X. 105(2) (2013) 190–202. https://arxiv.org/abs/1308.1809 (accessed October 5, 2018).

[35] A. Purohit, Z. Sun, S. Pan, P. Zhang, SugarTrail: Indoor navigation in retail environments without surveys and maps, in: Sensor, Mesh Ad Hoc Commun. Networks (SECON), 2013 10th Annu. IEEE Commun. Soc. Conf., New Orleans, LA, USA, 2013: pp. 300–308.

[36] C. Xu, B. Firner, R.S. Moore, Y. Zhang, W. Trappe, R. Howard, F. Zhang, N. An, SCPL: indoor device-free multi-subject counting and localization using radio signal strength, in: Proc. 12th Int. Conf. Inf. Process. Sens. Networks, Philadelphia, PA, USA, 2013: pp. 79–90.

[37] M. Lam, M. Mirshekari, S. Pan, P. Zhang, H.Y. Noh, Robust occupant detection through step-induced floor vibration by incorporating structural characteristics, in: Dyn. Coupled Struct. Vol. 4, Springer, 2016: pp. 357–367.

[38] S. Pan, S. Xu, M. Mirshekari, P. Zhang, H.Y. Noh, Collaboratively adaptive vibration sensing system for high-fidelity monitoring of structural responses induced by pedestrians, Front. Built Environ. 3 (2017) 28.

[39] M. Mirshekari, S. Pan, J. Fagert, E.M. Schooler, P. Zhang, H.Y. Noh, Occupant localization using footstep-induced structural vibration, Mech. Syst. Signal Process. 112 (2018) 77–97.

[40] S. Pan, T. Yu, M. Mirshekari, J. Fagert, A. Bonde, O.J. Mengshoel, H.Y. Noh, P. Zhang, https://doi.org/10.1145/3130954, in: Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol., ACM, 2017: pp. 1–31. https://doi.org/10.1145/3130954.

[41] S. Pan, N. Wang, Y. Qian, I. Velibeyoglu, H.Y. Noh, P. Zhang, Indoor person identification through footstep induced structural vibration, in: Proc. 16th Int. Work. Mob. Comput. Syst. Appl., Santa Fe, New Mexico, USA, 2015: pp. 81–86.

[42] V. Racic, A. Pavic, J.M.W. Brownjohn, Experimental identification and analytical modelling of human walking forces: literature review, J. Sound Vib. 326 (2009) 1–49.

[43] S. Drira, S.G.S. Pai, I.F.C. Smith, Uncertainties in structural behavior for model-based occupant localization using floor vibrations, Front. Built Environ. 7 (2021) 13.

[44] J.R. Gage, P.A. Deluca, T.S. Renshaw, Gait analysis: principles and applications, JBJS. 77 (1995) 1607–1623.

[45] S. Drira, Y. Reuland, S.G.S. Pai, H.Y. Noh, I.F.C. Smith, Model-Based Occupant Tracking Using Slab-Vibration Measurements, Front. Built Environ. 5 (2019) 63, https://doi.org/10.3389/fbuil.2019.00063.

[46] J. Clemente, F. Li, M. Valero, W. Song, Smart seismic sensing for indoor fall detection, location and notification, IEEE J. Biomed. Heal. Informatics. (2019).

[47] S. Anchal, B. Mukhopadhyay, S. Kar, UREDT: Unsupervised learning based Real-Time footfall event detection technique in seismic signal, IEEE Sensors Lett. 2 (2017) 1–4.

[48] R.K. Begg, M. Palaniswami, B. Owen, Support vector machines for automated gait classification, IEEE Trans. Biomed. Eng. 52 (2005) 828–838.

[49] M. Mirshekari, J. Fagert, A. Bonde, P. Zhang, H.Y. Noh, Human Gait Monitoring Using Footstep-Induced Floor Vibrations Across Different Structures, in: Proc. 2018 ACM Int. Jt. Conf. 2018 Int. Symp. Pervasive Ubiquitous Comput. Wearable Comput., 2018: pp. 1382–1391.

[50] Y. Liao, V.R. Vemuri, Use of k-nearest neighbor classifier for intrusion detection, Comput. Secur. 21 (2002) 439–448.

[51] C.-L. Liu, C.-H. Lee, P.-M. Lin, A fall detection system using k-nearest neighbor classifier, Expert Syst. Appl. 37 (2010) 7174–7181.

[52] S. Tan, An effective refinement strategy for KNN text classifier, Expert Syst. Appl. 30 (2006) 290–298.

[53] B. Wu, R. Nevatia, Cluster boosted tree classifier for multi-view, multi-pose object detection, in: 2007 IEEE 11th Int. Conf. Comput. Vis., 2007: pp. 1–8.

[54] E.-J. Ong, R. Bowden, A boosted classifier tree for hand shape detection, in: Sixth IEEE Int. Conf. Autom. Face Gesture Recognition, 2004. Proceedings., 2004: pp. 889–894.

[55] Y. Freund, R.E. Schapire, et al., Experiments with a new boosting algorithm, Icml (1996) 148–156.

[56] B.P. Roe, H.-J. Yang, J. Zhu, Y. Liu, I. Stancu, G. McGregor, Boosted decision trees as an alternative to artificial neural networks for particle identification, Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip. 543 (2005) 577–584.

[57] Y. Zhang, S. Pan, J. Fagert, M. Mirshekari, H.Y. Noh, P. Zhang, L. Zhang, Occupant Activity Level Estimation Using Floor Vibration, in: Proc. 2018 ACM Int. Jt. Conf. 2018 Int. Symp. Pervasive Ubiquitous Comput. Wearable Comput., 2018: pp. 1355–1363.

[58] J.D. Poston, R.M. Buehrer, P.A. Tarazaga, A framework for occupancy tracking in a building via structural dynamics sensing of footstep vibrations, Front. Built Environ. 3 (2017) 65.
[59] S. Pan, M. Mirshekari, P. Zhang, H.Y. Noh, Occupant traffic estimation through structural vibration sensing, in: Sensors Smart Struct. Technol. Civil, Mech. Aerosp. Syst. 2016, Las Vegas, Nevada, USA, 2016: p. 980306.
[60] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–444.
[61] J. Schmidhuber, Deep learning in neural networks: An overview, Neural Networks. 61 (2015) 85–117.
[62] S. Lawrence, C.L. Giles, A.C. Tsoi, A.D. Back, Face recognition: a convolutional neural-network approach, IEEE Trans. Neural Networks. 8 (1997) 98–113.
[63] I. Goodfellow, Y. Bengio, A. Courville, Deep learning, MIT Press, 2016.
[64] S. Drira, Y. Reuland, I.F.C. Smith, Occupant tracking using model-based data interpretation of structural vibrations, in: 9th Int. Conf. Struct. Heal. Monit. Intell. Infrastruct., St. Louis, MO, USA, 2019.
[65] S.G.S. Pai, Y. Reuland, S. Drira, I.F.C. Smith, Is there a relationship between footstep-impact locations and measured signal characteristics?, in: 1st ACM Int. Work. Device-Free Hum. Sens., New York, USA, 2019.
[66] S. Drira, Y. Reuland, I.F.C. Smith, Model-based interpretation of floor vibrations for indoor occupant tracking, in: 26th Int. Work. Intell. Comput. Eng., Leuven Belgium, 2019.
[67] J.-A. Goulet, I.F.C. Smith, Structural identification with systematic errors and unknown uncertainty dependencies, Comput. Struct. 128 (2013) 251–258.
[68] S. Drira, Y. Reuland, N.F.H. Olsen, S.G.S. Pai, I.F.C. Smith, Occupant-detection strategy using footstep-induced floor vibrations, in: Proc. 1st ACM Int. Work. Device-Free Hum. Sens., ACM, New York, NY, USA, 2019: pp. 31–34. https://doi.org/10.1145/3360773.3360881.
[69] K. Kanazawa, K. Hirata, Parametric estimation of the cross-power spectral density, J. Sound Vib. 282 (2005) 1–35.
[70] M.S. Ford, The illustrated wavelet transform handbook: introductory theory and applications in science, Health Phys. 84 (2003) 667–668.
[71] J. Lin, L. Qu, Feature extraction based on Morlet wavelet and its application for mechanical fault diagnosis, J. Sound Vib. 234 (2000) 135–148.
[72] S. Živanović, A. Pavic, P. Reynolds, Vibration serviceability of footbridges under human-induced excitation: a literature review, J. Sound Vib. 279 (2005) 1–74.
[73] M.H.F. Wilkinson, Gaussian-weighted moving-window robust automatic threshold selection, in: Int. Conf. Comput. Anal. Images Patterns, 2003: pp. 369–376.
[74] G. Zhang, X. Wang, Y.-C. Liang, J. Liu, Fast and robust spectrum sensing via Kolmogorov-Smirnov test, IEEE Trans. Commun. 58 (2010) 3410–3416.
[75] J. Hauke, T. Kossowski, Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data, Quaest. Geogr. 30 (2011) 87–93.
[76] C.-W. Hsu, C.-J. Lin, A comparison of methods for multiclass support vector machines, IEEE Trans. Neural Networks. 13 (2002) 415–425.
[77] V.N. Valdimir, N. Vapnik, The nature of statistical learning theory, (1995).
[78] L. Shi, M. Mirshekari, J. Fagert, Y. Chi, H.Y. Noh, P. Zhang, S. Pan, Device-free Multiple People Localization through Floor Vibration, in, Proc. 1st ACM Int. Work. Device-Free Hum. Sens. (2019) 57–61.
[79] W. Fang, L. Ding, B. Zhong, P.E.D. Love, H. Luo, Automated detection of workers and heavy equipment on construction sites: A convolutional neural network approach, Adv. Eng. Informatics. 37 (2018) 139–149.
[80] G.E. Dahl, T.N. Sainath, G.E. Hinton, Improving deep neural networks for LVCSR using rectified linear units and dropout, in: 2013 IEEE Int. Conf. Acoust. Speech Signal Process., 2013: pp. 8609–8613.
[81] D.C. Ciresan, U. Meier, J. Masci, L.M. Gambardella, J. Schmidhuber, Flexible, high performance convolutional neural networks for image classification, in: Twenty-Second Int. Jt. Conf. Artif. Intell., 2011.
[82] A. Ashiquzzaman, A.K. Tushar, Handwritten Arabic numeral recognition using deep learning neural networks, in: 2017 IEEE Int. Conf. Imaging, Vis. Pattern Recognit., 2017: pp. 1–4.
[83] S.S. Roy, A. Mallik, R. Gulati, M.S. Obaidat, P.V. Krishna, A deep learning based artificial neural network approach for intrusion detection, in: Int. Conf. Math. Comput., 2017: pp. 44–53.
[84] S. Sudholt, G.A. Fink, PHOCNet: A deep convolutional neural network for word spotting in handwritten documents, in: 2016 15th Int. Conf. Front. Handwrit. Recognit., 2016: pp. 277–282.
[85] I.W. Selesnick, C.S. Burrus, Generalized digital Butterworth filter design, IEEE Trans. Signal Process. 46 (1998) 1688–1694.
[86] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, others, Tensorflow: A system for large-scale machine learning, in: 12th ${$USENIX$}$ Symp. Oper. Syst. Des. Implement. (${$OSDI$}$ 16), 2016: pp. 265–283.
[87] M. Okuda, M. Ikehara, S. Takahashi, Fast and stable least-squares approach for the design of linear phase FIR filters, IEEE Trans. Signal Process. 46 (1998) 1485–1493.
[88] N.J. Bertola, A. Costa, I.F.C. Smith, Strategy to validate sensor-placement methodologies in the context of sparse measurement in complex urban systems, IEEE Sens. J. 20 (2020) 5501–5509.
[89] S. Drira, Occupancy detection and tracking in buildings using floor-vibration signals, École Polytechnique Fédérale de Lausanne - EPFL, Thesis n° 8289, 2020.