

Smart Home Occupant Identification via Sensor Fusion Across On-Object Devices

JUN HAN, SHIJIA PAN, MANAL KUMAR SINHA, HAE YOUNG NOH, PEI ZHANG, and PATRICK TAGUE, Carnegie Mellon University

Occupant identification proves crucial in many smart home applications such as automated home control and activity recognition. Previous solutions are limited in terms of deployment costs, identification accuracy, or usability. We propose *SenseTribute*, a novel occupant identification solution that makes use of existing and prevalent on-object sensors that are originally designed to monitor the status of objects to which they are attached. *SenseTribute* extracts richer information content from such on-object sensors and analyzes the data to accurately identify the person interacting with the objects. This approach is based on the physical phenomenon that different occupants interact with objects in different ways. Moreover, *SenseTribute* may not rely on users' true identities, so the approach works even without labeled training data. However, resolution of information from a single on-object sensor may not be sufficient to differentiate occupants, which may lead to errors in identification. To overcome this problem, *SenseTribute* operates over a sequence of events within a user activity, leveraging recent work on activity segmentation. We evaluate *SenseTribute* using real-world experiments by deploying sensors on five distinct objects in a kitchen and inviting participants to interact with the objects. We demonstrate that *SenseTribute* can correctly identify occupants in 96% of trials without labeled training data, while per-sensor identification yields only 74% accuracy even with training data.

CCS Concepts: • **Computer systems organization** → **Sensor networks**;

Additional Key Words and Phrases: Occupant identification, on-object sensing, sensor fusion

ACM Reference format:

Jun Han, Shijia Pan, Manal Kumar Sinha, Hae Young Noh, Pei Zhang, and Patrick Tague. 2018. Smart Home Occupant Identification via Sensor Fusion Across On-Object Devices. *ACM Trans. Sen. Netw.* 14, 3–4, Article 23 (December 2018), 22 pages.
<https://doi.org/10.1145/3218584>

1 INTRODUCTION

Occupant identification is fundamental in providing many value-added services for smart homes. Personalized home control such as comfort adjustments for lighting and HVAC proves to be

This work is an extension to the paper “SenseTribute: Smart Home Occupant Identification via Fusion Across On-Object Sensing Devices,” published in the Proceedings of ACM International Conference on Systems for Energy-Efficient Built Environments (BuildSys) 2017.

This work was supported in part by the National Science Foundation (under grants CNS-1645759, CNS-1149611 and CMMI-1653550), Intel, and Google.

Authors' addresses: J. Han, S. Pan, and M. K. Sinha, Carnegie Mellon University, Moffett Field, CA; emails: {junhan, shijapan, manalkus}@cmu.edu; H. Y. Noh, Carnegie Mellon University, Pittsburgh, PA; email: noh@cmu.edu; P. Zhang and P. Tague, Carnegie Mellon University, Moffett Field, CA; emails: {peizhang, tague}@cmu.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2018 Association for Computing Machinery.

1550-4859/2018/12-ART23 \$15.00

<https://doi.org/10.1145/3218584>

important for user convenience as well as energy and cost savings (Balaji et al. 2013; Conte et al. 2014; Ranjan et al. 2014). Furthermore, occupant identification supports activity recognition and/or occupant behavior analysis (Zeng et al. 2016).

Prior works investigate the use of body-worn sensors for occupant identification (Gafurov and Snekenes 2009; Gafurov et al. 2007; Mantyjarvi et al. 2005). Such solutions, however, are intrusive and are less practical because users are required to always carry or wear the sensors. To solve this problem, infrastructure-based solutions have also been explored. However, they make use of sensors that may invade privacy, such as cameras and microphones (Lin et al. 1994; Stillman et al. 1998). To overcome such problems, researchers also introduce solutions leveraging special purpose sensors such as infrared or vibration sensors (Hnat et al. 2012; Khalil et al. 2016; Pan et al. 2015, 2017). Because these solutions deploy the sensors specifically for occupant identification purposes, the solutions come at high hardware and installation costs. Researchers also explore existing infrastructure, such as WiFi, to help identify occupants (Zeng et al. 2016; Zhang et al. 2016). However, they make strong assumptions—requiring a user to walk in a straight line or to stay within a line-of-sight between transceivers—limiting their practicality.

Hence, to overcome the aforementioned limitations of prior work and provide a more practical and yet cost-effective solution, we ask the following question—instead of building and deploying specific sensors to provide a practical occupant identification solution, can we leverage sensing capabilities of existing IoT devices within a smart home? To answer this question, we observe an emerging trend in commercial *on-object sensing devices* (Cao Gadgets LLC 2017; Ecolink 2017; MetaSensor 2017; Notion 2017; Samsung 2017), which are detachable wireless sensor nodes that retrofit home objects such as doors, windows, drawers, and/or refrigerators, to monitor and report the object status over the home network. These devices are already prevalent and are projected to be more ubiquitous throughout smart homes (Deloitte 2015; Murphy 2015).

On-object sensing devices are typically equipped with accelerometers and/or gyroscopes to monitor object status (e.g., door opened or closed). However, we explore the possibility of repurposing these devices to provide more expressive data rather than just object status. Specifically, we find that the way a person interacts with an object is rather unique and can differentiate among people. For example, different family members tend to open a door or refrigerator in different manners, possibly due to different physical build, strength, and habit. We present *SenseTribute*, a novel occupant identification mechanism for smart home settings, which takes advantage of this physical phenomenon using representative features from accelerometer and gyroscope measurements to distinguish home occupants. *SenseTribute* enables **attribution of sensory** measurements to the originating user, hence, the name. Figure 1 depicts an example of repeated accelerometer measurements for two different users, highlighting the important capabilities to distinguish between users and match subsequent user readings.

SenseTribute utilizes supervised learning techniques to first train the model using collected bootstrapping data as training data, along with the corresponding training labels. Subsequently, upon collecting testing data, *SenseTribute* performs classifications using the trained model. While some application scenarios may ask the occupants to initially provide the training labels (e.g., names of persons associated with the training data) during a bootstrapping phase, such approach may be impractical in other scenarios due to usability problems. Hence, we design *SenseTribute* to be robust against this challenge, specifically, even in scenarios where the users do not provide the training labels. In such cases, *SenseTribute* is still able to identify the occupants, but with *pseudo-identifiers* instead of explicit identifiers such as names (e.g., *Persons A* and *B* rather than *Alice* and *Bob*). Pseudo-identifiers still support most of smart home applications such as the aforementioned personalized home control and identifying occupants of recognized activities, and may even be suitable for privacy-preserving applications. This is made possible because *SenseTribute* infers the

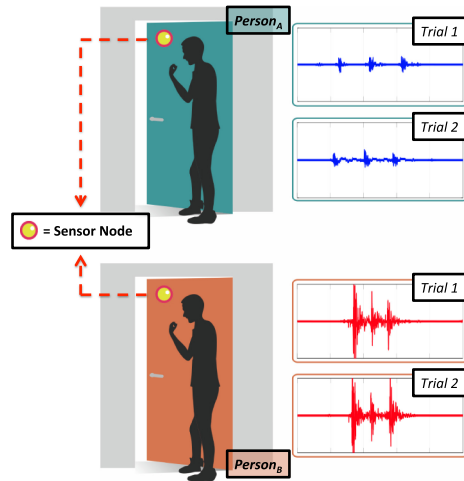


Fig. 1. *SenseTribute* utilizes the physical phenomenon that different people interact with objects at home (e.g., knocking or opening a door) differently such that resulting signals captured by on-object sensing devices are (1) sufficiently differentiable across different home occupants and (2) similar within different trials of the same occupant.

labels by utilizing unsupervised learning techniques to cluster the bootstrapping data into cluster identifiers. Subsequently, *SenseTribute* trains the model using the training data and the corresponding cluster identifiers as *quasi*-training labels. The *quasi*-training labels are labels that do not have information to map the cluster identifiers to occupant identities such as names.

Even with this classification approach, each on-object sensing device provides limited information content, yielding low identification accuracy. Performance degrades even further if training labels are not provided. In order to solve this challenge, we introduce *SenseTribute*'s *Ensemble Module* to amplify the information content across multiple on-object sensing devices, thereby boosting the accuracy of the overall occupant identification. The *Ensemble Module* relies on related research on *activity segmentation* (Helaoui et al. 2013; Kodeswaran et al. 2016; Riboni et al. 2016), which segments out a sequence of events belonging to a single activity segment performed by the same person, out of entire sensor data streams of multiple persons' events. For example, a *cooking breakfast* activity may consist of multiple sensor events performed by the same user such as *opening the refrigerator*, followed by *taking out a frying pan*, followed by *turning on the stove*.

We design and implement *SenseTribute* and evaluate its feasibility by conducting real-world empirical experiments with five distinct sensor pairs (accelerometer and gyroscope), each attached to five different objects—door, refrigerator, drawer, towel dispenser, and window. We invite five participants to perform a sequence of events that interact with these objects. We choose five participants, as this number is greater than an average of 3.14 persons per home in the United States (U.S. Census Bureau 2017). From our empirical analysis, *SenseTribute* is able to correctly identify occupants with 96% accuracy, even when the training labels are not provided, while the average accuracy from per-object identification yields 74%, even with training labels. Overall, we make the following contributions in this article.

- We design *SenseTribute* to extract expressive data from on-object sensors and identify occupants in a smart home.
- We demonstrate how *SenseTribute* achieves high identification accuracy by combining observations across several sensors on different objects, even without labeled training data.

- We evaluate *SenseTribute* by conducting real-world experiments with participants interacting with different objects in a kitchen, each interfaced with a sensor node.

The remainder of this article is organized as follows. We present background information and related work in Section 2. We then present the details of *SenseTribute*'s design and implementation in Section 3, and its evaluation results in Section 4. Subsequently, we present a discussion and the conclusion in Sections 5 and 6, respectively.

2 BACKGROUND AND RELATED WORK

We first present *on-object sensing devices* and their prevalence. We then introduce *activity segmentation* often studied in the field of activity recognition, and how *SenseTribute* utilizes it. Furthermore, we describe related work on *occupant identification*.

2.1 On-Object Sensing Devices

On-object sensing devices are popular smart home gadgets that enable home owners to monitor the status of various objects—such as doors and drawers—by simply attaching the device to each object. An on-object device is commonly equipped with inertial sensors (e.g., accelerometer and/or gyroscope), which sense the movement of the object it is attached to. The sensor signals are then processed to output object status—such as door or drawer open/close—and reports the events to home owner's smartphone over the cloud. Companies such as Notion (Notion 2017) and Samsung SmartThings (Samsung 2017) are industry leaders, while there are many other commercial solutions from various vendors (Cao Gadgets LLC 2017; Ecolink 2017; MetaSensor 2017). These devices are projected to be more prevalent in smart homes in the near future (Deloitte 2015; Murphy 2015). We design *SenseTribute* to extract more expressive data than mere status of objects, namely, to infer the identities of occupants in a home. Hence, *SenseTribute* inherently eliminates the costly need to build and deploy specific sensing devices for occupant identification.

2.2 Activity Segmentation

Activity segmentation—an actively studied topic in the activity recognition field—segments out a sequence of events that are performed by a single occupant. However, this is a difficult problem because different events are performed by different persons that may be temporally overlapping within a single stream of sensor data. Hence, researchers make use of combinations of sensor patterns and temporal information to identify a sequence of *events* that constitute a single *activity segment* (Helaoui et al. 2013; Kodeswaran et al. 2016; Riboni et al. 2016). For example, consider *Person_A* *cooking breakfast*, and *Person_B* *watching TV* in the living room. The *cooking breakfast* activity segment may consist of a sequence of *events* such as: {*kitchen door opening*, *fridge door opening*, and *pasta drawer opening*}. On the other hand, *watching TV* activity segment may consist of a sequence of *events* such as: {*sitting down on sofa*, *taking out remote control*, *TV turning on*}. Each of the sequence of *events* belonging to the same *activity segment* are grouped together, even though there may be temporal overlaps between individual *events*. *Activity segmentation* is one of the important foundations when designing *SenseTribute*. Specifically, *Ensemble Module* exploits the above property that a sequence of events within an activity segment is performed by the same user, enabling *SenseTribute* to combine the confidence of a sequence of events (see Section 3.5).

2.3 Occupant Identification

Smart home occupant identification is an important problem. Personalized home control is gaining much attention such as user-specific comfort adjustments for lighting and HVAC for convenience as well as energy efficiency (Balaji et al. 2013; Conte et al. 2014; Ranjan et al. 2014). Due to

potentially significant cost-savings, this is a real-world problem that is heavily studied by appliance manufacturers as well. Furthermore, occupant identification supports many activity recognition applications. This is because understanding *who* is performing the recognized activity is a building block to associating activities to individual occupants, rather than just knowing that someone at home has performed the activity (Zeng et al. 2016) (e.g., splitting costs between roommates based on individual energy consumption or even simply providing feedback to which family member consumes the most energy).

Due to the importance of the occupant identification problem, prior works explore solutions by deploying infrastructure-based sensors. Researchers utilize ultrasonic-based doorway sensors to capture the movements and the physical characteristics such as height (Hnat et al. 2012) and/or weight (Khalil et al. 2016) of persons. Researchers also utilize structural vibration-based sensors to detect occupant's gait patterns (Pan et al. 2015, 2017). Occupants strike the floor with different gait patterns, inducing a unique structural vibration waveform. Similarly, researchers also exploit changes in body electric potential due to walking (Grosse-Puppendsahl et al. 2016). While these solutions are promising first steps, all of them utilize hardware that is specifically built and deployed to solve the occupant identification problem. This inevitably incurs high cost both in terms of hardware as well as deployment costs.

Prior work also explores solutions that use existing infrastructure such as Wi-Fi to utilize channel state information (CSI) induced by the occupant's walking pattern (Zeng et al. 2016; Zhang et al. 2016). While these solutions do not incur additional hardware or deployment costs, they face challenges in limited deployment practicality. This is because these solutions require the occupants to either (1) only walk in a straight line (Zeng et al. 2016) or (2) stay within the line of sight between WiFi transceivers (Zhang et al. 2016).

As opposed to the related work, *SenseTribute* inherently reduces the hardware and deployment cost because it *utilizes existing and prevalent on-object sensing devices* deployed by users, and simultaneously provides a more practical occupant identification by performing simple software modifications to extract information necessary to identify the occupants.

3 DESIGN AND IMPLEMENTATION

We now present *SenseTribute*'s design and implementation. We describe the details *SenseTribute*'s algorithm for both *known* and *unknown* training labels scenarios. We also explain how *SenseTribute* ensembles different objects to amplify the identification accuracy.

3.1 *SenseTribute* Overview

SenseTribute's goal is to identify the occupants by leveraging signals of on-object sensors utilizing supervised learning techniques. *SenseTribute* is divided into two phases—a *Bootstrapping Phase* and an *Identification Phase*. First, during the *Bootstrapping Phase*, *SenseTribute* trains a classification model from the collected sensor data (i.e., history data). Subsequently, in its *Identification Phase*, *SenseTribute* tests newly collected sensor data to finally identify the occupant.

In order to train the model for classification, the system requires training labels (i.e., ground truth occupant identity corresponding to the collected history data). However, it may be more practical for certain applications to not collect user provided training labels (e.g., to increase usability). We account for this problem and design *SenseTribute* to automatically adapt its training scheme based on the availability of user-provided labels.

We present the flow chart diagrams to depict the overall *SenseTribute* design as shown in Figure 2(a). Specifically, when the training labels are provided to the system by the users (i.e., *known labels* scenario), *SenseTribute* utilizes the traditional supervised learning techniques by taking as input for the *Training Module* the (1) training label and (2) data. For the training label,

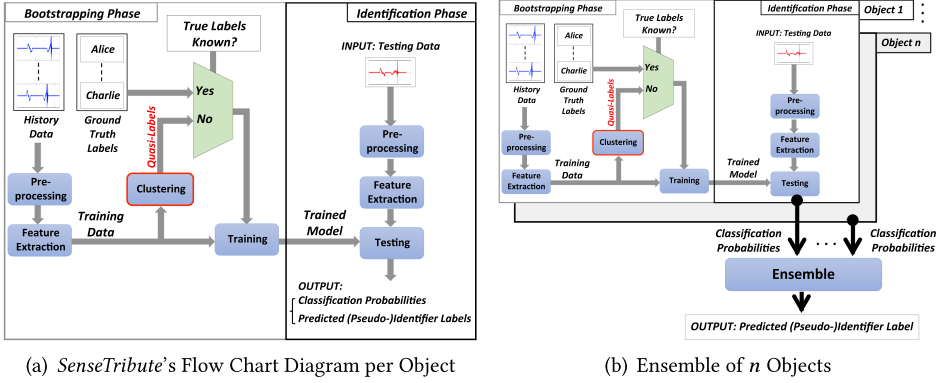


Fig. 2. (a) Flow-chart diagram of *SenseTribute* for an individual object. During the *Bootstrapping Phase*, collected *history data* and ground truth labels are used to train a model. If a ground truth label is unavailable, *SenseTribute* clusters history data to *infer* training labels. Subsequently, during *Identification Phase*, *SenseTribute* uses the trained model to predict occupant identity. (b) Subsequent to the identification phase in (a), *SenseTribute* further **ensembles** classification probabilities from n different objects and predicts occupant identity with higher accuracy.

SenseTribute utilizes the user-provided ground truth labels. For the training data, *SenseTribute* first processes the collected *history data* in the *Pre-processing Module* and then extracts necessary features in the *Feature Extraction Module*. Finally, at the end of the *Identification Phase*, the *Testing Module* outputs the *Predicted Occupant Label* along with the classification probabilities of all the potential classes.

However, *SenseTribute* is also capable of operating even when users do not provide the ground truth labels (i.e., *unknown labels* scenario) by utilizing a hybrid approach of unsupervised and supervised learning techniques. Similar to the *known labels* scenario, the *history data* are used to process and extract features. However, the features are now input to *Clustering Module*, which computes and outputs the clustered indices. We use these indices as *quasi-labels* that substitute the ground truth labels. *Quasi-labels* represent different clusters, or groups, corresponding to the history data. However, as opposed to the ground truth labels, *quasi-labels* (1) do not carry information to be directly mapped to specific occupant's explicit identities such as names and (2) are prone to some amount of error due to clustering. Finally, the *Testing Module* outputs the predicted *pseudo-identifier labels*, along with the classification probabilities of all the potential classes. Similar to *quasi-labels*, *pseudo-identifiers* do not carry information to be directly mapped to the specific occupant identities such as names, but are still valuable because they can be used to sufficiently distinguish different occupants (e.g., *Person_A* vs. *Person_B*). While clustering algorithms such as K-Means provide linear decision boundaries, we design *SenseTribute* using classification as the backbone framework for the simplicity of integrating both *known* and *unknown labels* scenarios.

Since the information content from a single object may not be sufficient to accurately identify the occupants, we introduce *SenseTribute*'s *Ensemble Module* subsequent to the *Identification Module* of each object, to “ensemble” the classification probabilities to arrive at a higher occupant identification accuracy. Figure 2(b) depicts the corresponding flowchart diagram.

3.2 Pre-processing and Feature Extraction

3.2.1 Pre-processing. Prior to extracting the features from the raw sensor data, we first perform noise reduction to increase the Signal-to-Noise Ratio (SNR) and the subsequent classification and clustering performance. We make use of *spectral subtraction* (Boll 1979), used in speech recogni-

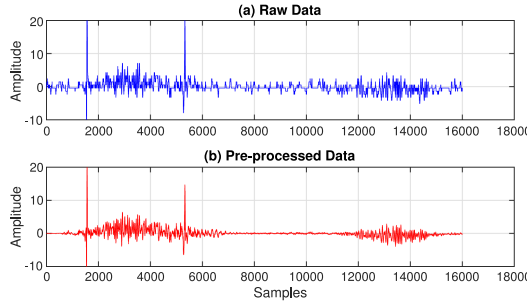


Fig. 3. We use spectral subtraction to increase the Signal-to-Noise Ratio (SNR). Figure 3(a) Depicts a raw time-series gyroscope signal; and (b) depicts the resulting spectral subtraction.

Table 1. Features Used in *SenseTribute*, Where Vectors, x_p and y_q are Time and Frequency Domain Representations of the Pre-processed Data, Respectively

Feature	Domain	Expression
RMS	time	$\sqrt{1/N \sum_{p=1}^N x_p^2}$
FFT_{RMS}	frequency	$\sqrt{1/M \sum_{q=1}^M y_q^2}$
$Peak2RMS$	time	$\max(x) / \sqrt{1/N \sum_{p=1}^N x_p ^2}$
$Energy$	time	$\sum_{p=1}^N \log(x_p^2)$
SMA	time	$1/N \sum_{p=1}^N x_p $
FFT_{max}	frequency	$\max(y_q)$
$Mean$	time	$\frac{1}{N} \sum_{p=1}^N x_p$
$Median$	time	$median(x_p)$

N and M are the number of elements in x and y , respectively.

tion to remove background noise, because the ambient noise is similar to inherent sensor noise. Spectral subtraction performs the operation $S(\omega) = Y(\omega) - N(\omega)$, where $Y(\omega)$, $S(\omega)$, and $N(\omega)$ are the frequency-domain spectra of the noisy sensor reading, desired signal, and noise, respectively. We estimate the noise spectrum $N(\omega)$ by sampling the ambient noise, which can be performed by sensor nodes, in practice, prior to the *Pre-processing Module*. Figure 3 depicts an example of single-axis gyroscope signal corresponding to opening and closing a drawer.

3.2.2 Feature Extraction. *SenseTribute* then performs feature extraction on the pre-processed signal. We extract features from both time and frequency domains as characteristics of the induced signal. We list the features used in this work in Table 1. Vectors x_p and y_q are time and frequency domain representations of the data, respectively, and N and M are the number of elements in x and y , respectively. The Root Mean Square (RMS) (in time or frequency domains as RMS and FFT_{RMS} , respectively) reflects the variation within a signal segment, a relatively widely used feature that effectively describes the signal. The peak-to-RMS ratio of time domain signal, $Peak2RMS$, measures more detailed signal distribution in addition to RMS . For example, a person with thicker finger bones knocking on the door may trigger an impulse signal with a sharp waveform, which may lead to a higher $Peak2RMS$ value. We also compute the log energy entropy (Coifman and Wickerhauser 1992; Mathworks 2017), $Energy$, which measures the signal distribution. Signal magnitude area (Bersch et al. 2014; Bouten et al. 1997), SMA , measures the average of the signal

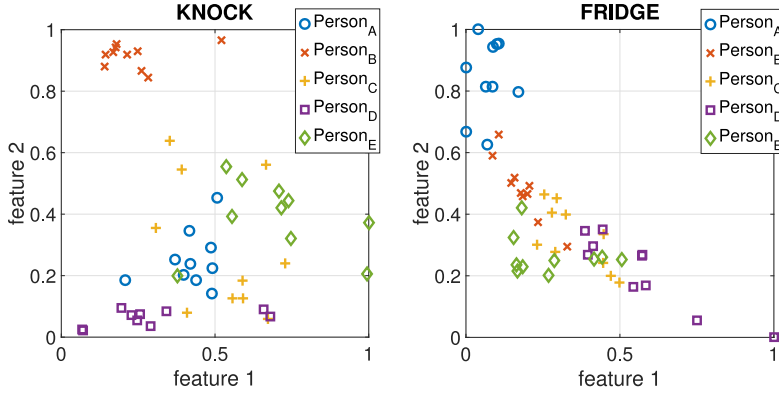


Fig. 4. We plot feature pairs for *knock* and *fridge door open/close event types* for comparison. *Knock* plot depicts sufficient separation of features across *Persons A, B, D, and E*, while *Person C* has a large overlapping area. *Fridge* plot depicts sufficient separation for *Persons A, B, and E*, while *Persons C and D* have large overlapping areas.

amplitude. The maximum value of frequency domain signal, FFT_{max} , provides the peak amplitude of y_q . Finally, we use the common statistical *mean* and *median* of x_p as measurements of central tendency.

Furthermore, we compare the feature distributions of different occupants by plotting feature pairs. Figure 4 depicts two examples of feature pairs (*Peak2RMS* vs. *SMA*) from two distinct sensors on a door (capturing knocking events) and refrigerator (capturing refrigerator opening and closing events). Each marker represents a feature comparison per occupant (i.e., *Person_A* to *Person_E*). We make the following two observations. First, we observe that within each sensor, the feature pair provides information to distinguish different occupants in a fairly sufficient manner. For example, for a knocking event, *Person_A*, *Person_B*, *Person_D*, and *Person_E* exhibit sufficient separation, while *Person_C* exhibits a large overlapping area with other occupants. Second, we also observe that across the two events from different objects, different feature pairs contribute to separating the occupants. For example, the feature pairs performed well in distinguishing *Person_D* for knocking on a door but poorly for opening and closing a refrigerator.

3.3 Known Labels Scenario

In the application scenario where the user provides the ground truth labels for the training label, we leverage supervised learning techniques to perform occupant identification. We implement *SenseTribute*'s classification modules (i.e., *Training* and *Testing Modules*) with Support Vector Machines (SVM) (Bennett and Campbell 2000) using a Radial Basis Function (RBF) kernel. We choose SVM because it requires a relatively small amount of training data to achieve high classification accuracy, compared to other classification methods such as neural networks. We implement the modules using publicly available LIBSVM (Chang and Lin 2011). We use multi-class SVM classification to classify n occupants in smart home settings, where $n \geq 2$. First, the *Training Module* takes as input the aforementioned feature vector of the training data and the training label to compute the trained model. This module concludes the end of the *Bootstrapping Phase*.

Second, the *Testing Module* in the *Identification Phase* takes as input the trained model and the feature vector of the testing data. This module performs the SVM classification to output the following: (1) classification probabilities, $Pr[O = o_i]$, of all possible classes (i.e., occupants), o_1, \dots, o_n ; and (2) the final predicted label, which is the occupant, o_i , that yields the highest $Pr[O = o_i]$.

3.4 Unknown Labels Scenario

When the user does not provide any training labels, we leverage a hybrid approach of supervised and unsupervised learning techniques to perform occupant identification. The **unknown** and **known labels** scenarios are equivalent in computing the feature vector. However, it differs in that the system no longer has the given training labels to be input to the classification modules. Hence, we *infer* the training labels using the unsupervised learning techniques.

Specifically, we implement the *Clustering Module* with K-Means clustering (Kanungo et al. 2002; Lloyd 1982), which takes a feature vector from the history data and the number of cluster groups K . We assume that K , i.e., the number of occupants in a home, is known (see Section 5.2). The K-Means clustering algorithm groups each of the input observations into K clusters with the smallest distance to the corresponding computed centroid. This module outputs the clustered indices, which will be subsequently used as the training label in the *Training Module*. We note that the clustered indices are *quasi-labels*, which do not map directly to occupants' explicit identifiers (e.g., occupants' names such as *Amy* vs. *Bob*). However, *quasi-labels* provide adequate information to identify occupants to their *pseudo-identifiers* (e.g., *Person_A* vs. *Person_B*) at the end of the *Identification Phase*.

3.5 Ensemble Module

Each object's identification accuracy (output from Figure 2(a)) is limited because each object has either low resolution of information, or the same occupant may occasionally interact with the object in a slightly different manner. Furthermore, for the case of the *unknown labels*, accumulated errors from clustering contributes to lower per-object identification accuracy. Hence, we design the *Ensemble Module* to amplify the occupant identification accuracy. *SenseTribute* ensembles identification probabilities of individual objects, as depicted in the flow chart diagram in Figure 2(b). Specifically, this module takes as input the resulting classification probabilities, $Pr[O = o_i]$, where $i = 1, \dots, n$ (indicating n potential classes, i.e., n occupants), from each of the *Testing Modules* belonging to m different sensors each interfaced with different objects, defined as S_j , where $j = 1, \dots, m$. Subsequently, this module outputs the final predicted occupant identity, o^* , which has an amplified identification accuracy, which we evaluate in Section 4.

To implement the *ensemble algorithm*, we formulate this problem as the conditional probability depicted in Equation (1):

$$o^* = \arg \max_{o_i} Pr[O = o_i \mid S_1, \dots, S_m], \quad (1)$$

This finds the most likely occupant o^* , given sensor data S_1, \dots, S_m . We assume independence across each sensor, S_j , and use Bayes' theorem to rewrite this formulation as shown in Equation (2):

$$o^* = \arg \max_{o_i} \prod_{j=1}^m Pr[O = o_i \mid S_j], \quad (2)$$

where each of the probabilities, $Pr[O = o_i \mid S_j]$, is equivalent to the output probabilities, $Pr[O = o_i]$, of each sensor.

4 EVALUATION

In this section, we first present the experiment setup and evaluate *SenseTribute*'s performance.

4.1 Experiment Setup

4.1.1 Apparatus. We conduct our experiment by facilitating five objects in a kitchen, each with a sensor node. The objects include—*door*, *fridge door*, *drawer*, *towel dispenser*, and *window*. Each

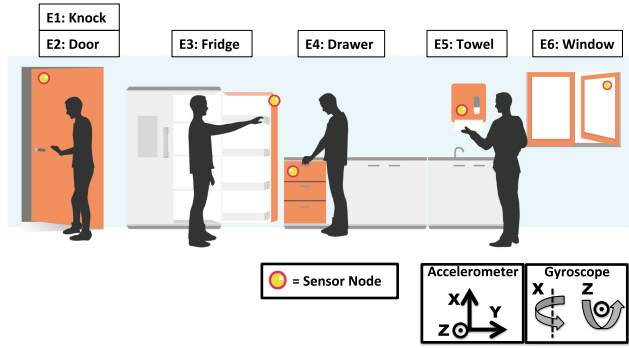


Fig. 5. We depict experiment setup conducted in a kitchen. Participants are asked to perform prescribed events.

sensor node is comprised of an Arduino Uno (Arduino 2017) interfaced with ADXL335 tri-axis accelerometer (Analog Devices 2017) and LPY503AL dual-axis gyroscope (STMicroelectronics 2017), sampling each axis at 5KHz. These sensors are attached to the objects so that the accelerometer’s z -axis is perpendicular to the object’s surface, and the gyroscope’s x -axis revolves around an imaginary line perpendicular to the floor, as depicted in Figure 5. The rest of this evaluation only considers using the two axes, and we discuss practical considerations later in Section 5.3.

4.1.2 Data Collection. We invited five participants, which is higher than the average number of people per family, which is 3.14 persons (U.S. Census Bureau 2017). We asked each participant to perform a predefined activity of operating the aforementioned objects—i.e., opening/closing *door*, *fridge*, *cabinet drawer*, and *window*, and pulling a towel from the *towel* dispenser. We now refer to *event type* as the *event type*–*object* pair (e.g., *door* represents opening/closing door). We performed the study after obtaining approval from the Institutional Review Board (IRB) and conducted the experiment in compliance to the IRB’s recommendations.

4.2 Known Labels Scenario

We evaluated *SenseTribute* when the system was given the training labels, performing an SVM classification as described in Section 3.3. We reported the classification accuracy by varying the number of occupants from $i = 2, \dots, 5$, where each variation is an average of all possible combinations, $\binom{5}{i}$ (e.g., 3 *occupants* case is an average of $\binom{5}{3} = 10$ instances). Each instance of combination is an average result of a 10-fold cross validation (i.e., Leave-One-Out) as we have ten trials per occupant. Figure 6(a) depicts the results for all six *event types*. We observe that as the number of occupants increases, the classification accuracy decreases for each of the *event types*. This is intuitive as introducing more *classes* (i.e., occupants) to the classifier introduces more room for error. The average of all six *event types* with five occupants yields 74%, as reported in Section 1.

We further note that different *event types* result in different accuracy, due to certain objects being more distinctive. We observe that objects that provide relatively consistent interaction yielded better classification accuracy. For example, *knocking on door* and *opening and closing a drawer* leads to more information to sufficiently distinguish occupants, while dispensing a *towel* did not produce sufficient information on its own.

To understand the reason for such low accuracy for certain *event types* such as dispensing a *towel event type*, we also report the classification accuracy per occupant (*Person A* through *E*), per *event type* (Figure 6(b)). As depicted from this figure, the accuracy across a *towel event* is generally low for certain persons such as *Persons D* and *E*. To further understand the reason, we analyze the

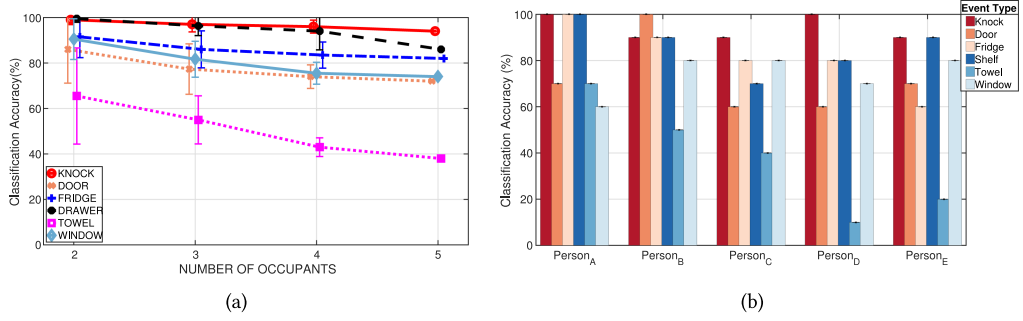


Fig. 6. (a) Figure depicts classification accuracy by varying number of occupants of each *event* type (for the *known labels* scenario). Each data point is an average accuracy of all combinations within each number of occupants. As the number of occupants increases, classification accuracy decreases. The average accuracy of different *event types* for five occupants case yields 74%. (b) Figure depicts classification accuracy due to different occupants. Certain pair of *event types* yield contradicting accuracy across different occupants. *SenseTribute* takes advantage of such phenomenon to amplify the final accuracy in its *Ensemble Module*.

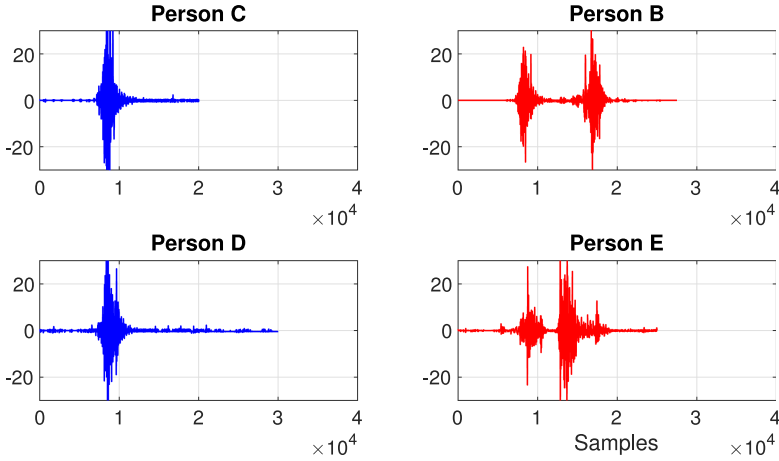


Fig. 7. This figure illustrates raw accelerometer signals of *Persons B, C, D, and E*, when dispensing towels. The signals depict high similarities between *Persons C and D*, as well as *Persons B and E*. This can possibly be explained when considering a similar physical build between the occupant pairs.

confusion matrix for the five-person scenario depicted in Table 3. From the matrix, we note that *Person D* is often misclassified with *Person C* and vice versa (as highlighted), while *Person E* is often misclassified as *Person B* and vice versa (as highlighted). To further analyze the misclassifications, we present the raw accelerometer data of the four occupants in Figure 7. From visual inspections, signals from *Persons C and D* are similar together, while those of *Persons B and E* are also similar to each other. This phenomena can possibly be explained when considering similar physical build between *Persons C and D*, and *Persons B and E* as described in Table 2. The similarities between the physical build is correlated to the similarities of the raw signals when dispensing towels. However, in general the *towel event type* proves to yield poor classification accuracy for all five occupants. This can be understood when noticing that the towels often got ripped during dispensing in multiple trials, and, consequently, yielded different interactions, even within the same subject.

Table 2. Table Presents Demographics of Five Participants

	Gender	Height	Weight	Age
Person A	Female	1.59m	51kg	28
Person B	Male	1.75m	76kg	33
Person C	Male	1.65m	45kg	27
Person D	Female	1.65m	50kg	30
Person E	Female	1.85m	95kg	26

Table 3. Table Depicts a Confusion Matrix of *Towel Event Type* for the Five Occupant Scenario from Figure 6(a)

		Predicted as:					Classification Accuracy
		<i>Person A</i>	<i>Person B</i>	<i>Person C</i>	<i>Person D</i>	<i>Person E</i>	
Ground Truth	<i>Person A</i>	7	0	2	1	0	7/10
	<i>Person B</i>	0	5	1	1	3	5/10
	<i>Person C</i>	1	0	4	4	1	4/10
	<i>Person D</i>	2	1	5	1	1	1/10
	<i>Person E</i>	2	4	1	1	2	2/10

As highlighted, *Persons C* and *D* are often misclassified to each other. Similarly, *Persons B* and *E* are also misclassified to each other. This can be explained when considering similar physical build between the occupant pairs.

Table 4. The table Depicts a Confusion Matrix of *Door Event Type* for the Five Occupant Scenario from Figure 6(a)

		Predicted as:					Classification Accuracy
		<i>Person A</i>	<i>Person B</i>	<i>Person C</i>	<i>Person D</i>	<i>Person E</i>	
Ground Truth	<i>Person A</i>	7	0	1	1	1	7/10
	<i>Person B</i>	0	10	0	0	0	10/10
	<i>Person C</i>	1	0	6	1	1	6/10
	<i>Person D</i>	2	0	2	6	0	6/10
	<i>Person E</i>	1	0	0	2	7	7/10

Average classification accuracy for *Person B* is highest. Similarly, all other occupants did not misclassify to *Person B*. This can be explained by *Person B*'s unique motion of opening and closing the door. *SenseTribute* eventually benefits from such unique *occupant-object* pair when performing ensemble across different objects.

Furthermore, from Figure 6(b), we note another interesting point about classification accuracy across different occupants. Even though the average classification accuracy was relatively high for certain *event types* (as depicted in Figure 6(a)), different occupants have varying results. For example, while *door* yielded relatively high accuracy, we find that only *Person B*'s result is particularly good compared to other occupants. To understand this phenomenon better, we plot the raw accelerometer signals of *Persons A, B, and C* in Figure 8. From visual inspection, it is sufficient to differentiate *Person B*'s signal (colored in red) from the rest, while *Person A* and *Person B* yielded similar signals (colored in blue). This is also depicted in Table 4 as a confusion matrix. Notice that *Person B* is uniquely and correctly classified to himself, while the accuracy is lowered for other occupants. Furthermore, other occupants were never incorrectly classified to *Person B* (as highlighted).¹ These results demonstrate how different people interact with certain objects

¹We further present confusion matrices of other *event types* in Appendix A.

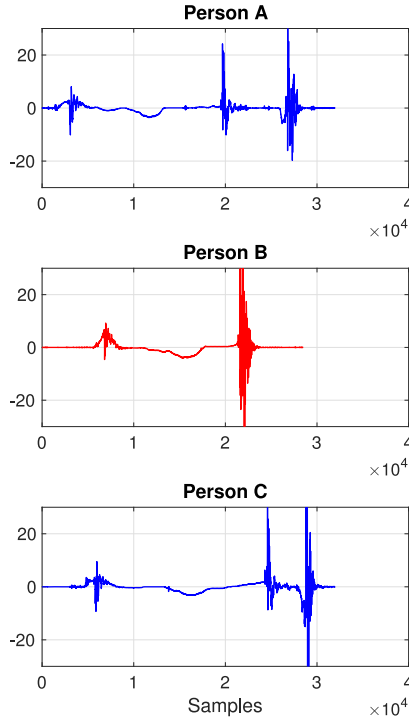


Fig. 8. The figure illustrates raw accelerometer signals of *Persons A, B, and C* when opening and closing a door. The signals depict high similarities between *Persons A and C*, while *Person B* is relatively unique. Such uniqueness yields high classification accuracy for *Person B* with the *door event type*.

uniquely, but interact rather similarly with other objects. This is one of the reasons why *SenseTribute* benefits from its *Ensemble Module* to fuse across classification probabilities from different objects. In addition, for certain *event types*, the classification accuracy is flipped across a pair of occupants on two different *event types*. For example, *Knock* and *Door event types* yield relatively high and low classification accuracy for *Person A*, respectively. However, the two *event types* conversely yield relatively low and high accuracy for *Person B*, respectively.

4.3 Unknown Labels Scenario

We now evaluate *SenseTribute* when the training labels are not provided by the user. As presented in Section 3.4, *SenseTribute* utilizes a hybrid approach of unsupervised and supervised learning—i.e., using clustering results as *quasi-labels*, to replace the unknown training labels. To provide a comprehensive view of how clustering accuracy affects the classification accuracy, we set clustering accuracy artificially from 25% to 100%, for each *event type*, as depicted in Figure 9. For example, a clustering accuracy of 50% indicates that half of the training labels selected at random are made incorrect on purpose. We repeat this process a thousand times and report the average for each data point in this figure. We show the results of the five occupants' cases as an example. This figure illustrates that as the clustering accuracy increases, the corresponding classification accuracy also increases (with 100% corresponding to known labels).

We now evaluate the performance of the *Clustering Module*. The clustering accuracy is computed as *Rand Index* (Rand 1971), which is defined as Equation (3):

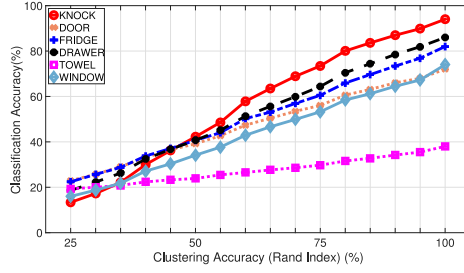


Fig. 9. The figure depicts how varying clustering accuracy affects classification accuracy by artificially setting clustering accuracy from 25% to 100% for each *event type* (for five occupants case). The classification accuracy increases as the clustering accuracy increases for all *event types*.

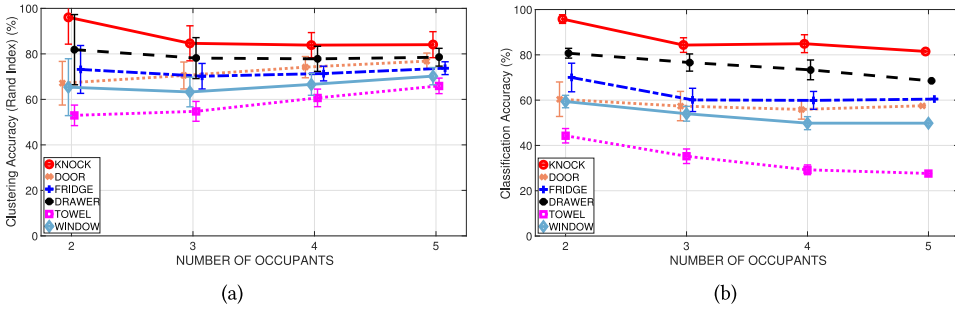


Fig. 10. (a) The figure depicts clustering accuracy when varying the number of occupants for each *event type* (for *unknown labels* scenario). *Knock* and *Drawer* yield decreasing accuracy while other *event types* yield increasing accuracy, as the number of occupants increases. (b) The figure depicts the classification accuracy when varying the number of occupants for each *event type*, for the *unknown labels* scenario. As the number of occupants increases, the corresponding accuracy also decreases.

$$\text{Clustering Accuracy (Rand Index)} = \frac{TP + TN}{TP + TN + FP + FN}, \quad (3)$$

where TP, TN, FP, and FN, are True Positive and Negative, and False Positive and Negative, respectively. Figure 10(a) depicts the clustering accuracy (i.e., Rand Index), when we vary the number of occupants $i = 2, \dots, 5$. Each of the data points is an average of all possible combinations of i occupants, $\binom{5}{i}$. Furthermore, we report the average of a thousand iterations for all instances. We note that the clustering accuracy decreases as the number of occupants increases for *Knock* and *Drawer* event types. However, the rest of the *event types* yield results that have increasing clustering accuracy as the number of occupants increases. This is because *Knock* and *Drawer*, which yield high classification accuracy for the *Known Labels* scenario, have features that are sufficiently differentiable, while the rest of the *event types* do not follow this trend. Hence, during the clustering of two occupants' cases, the two centroids may be very close to each other, yielding low clustering accuracy. However, when the number of occupants increases, more centroids are introduced, yielding higher clustering accuracy.

Finally, we evaluate the classification accuracy of *SenseTribute's Unknown Labels* scenario (i.e., output of *Testing Module*). We compute the classification accuracy in a similar manner to the aforementioned Figure 9, namely, purposely degrading the correctness of the training label. Only this time, we take the actual empirical results of clustering accuracy from Figure 10(a) instead of the

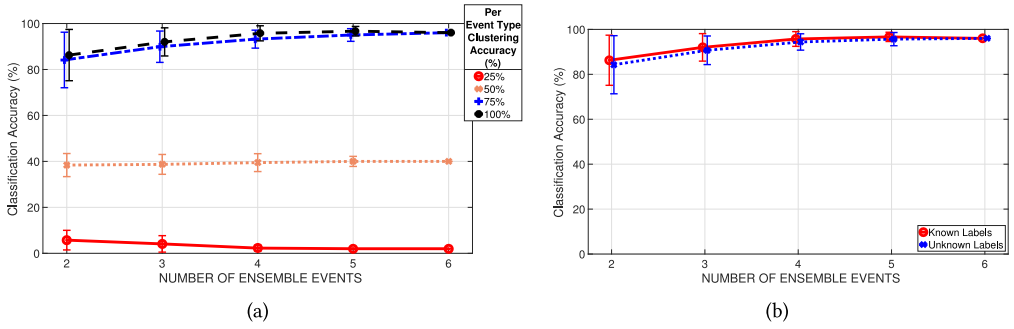


Fig. 11. (a) The figure depicts how (1) the number of *ensemble event types* and (2) the availability of training labels affect classification accuracy. We artificially assign equal clustering accuracy per *event type*. As the number of *ensemble event types* increases, accuracy increases, except for the 25% case. Also, lower per *event type* clustering accuracy yields lower classification accuracy due to more mislabeled training data. (b) This figure depicts increasing classification accuracy as we ensemble more *event types*, for both *Known* and *Unknown Labels* scenarios. We observe high classification accuracy even if the training labels are not known.

artificial numbers. We apply this strategy rather than directly applying the output of the clustering indices as the training label. This is because the *Clustering Module* outputs clustered indices, which is at times difficult to map to corresponding ground truth labels. However, this is necessary when computing the final classification accuracy for evaluation purposes. While improving the clustering algorithm would certainly help to solve this issue, we concentrate on evaluating the effects of the clustering accuracy on the classification accuracy. Figure 10(b) depicts the effect of the classification accuracy as we vary the number of occupants, where each data point, again, depicts an average of all possible $\binom{5}{i}$ combinations, and each combination is an average of 10-fold cross validation (i.e., Leave-One-Out). We make two interesting observations. First, similar to Figure 6(a) of the *Known Labels* scenario, this figure depicts an intuitive trend of decreasing classification accuracy as the number of occupants increases. This trend exists even for the *event types* that have increasing clustering accuracy with the number of occupants from Figure 10(a). This is because the effect of increasing the number of SVM classes outweighs the effect of correct labels. Second, we also observe that classification accuracy is relatively lowered compared to Figure 6(a) of the *Known Labels* scenario due to the incorrect labels.

4.4 Ensemble Classification Accuracy

We now evaluate *SenseTribute's Ensemble Module* for both the *known* and *unknown labels* scenarios. To provide a comprehensive view of how the number of *ensemble event types* and availability of training labels affect the classification accuracy, we present Figure 11(a). We report the classification accuracy when varying the number of events to ensemble from $j = 2, \dots, 6$, where each variation is an average of all $\binom{6}{j}$ combinations. Again, each combination is an average of 10-fold cross validation (i.e., Leave-One-Out). We artificially assign equal clustering accuracy per *event type*, again, by artificially degrading the correctness of the training label accordingly. We degrade different training labels at random and repeat this process for a thousand times to report an average value. Each of the line plots depict a different clustering accuracy—25%, 50%, 75%, and 100%—assigned per *event type*. The 100% clustering accuracy line graph represents the *known labels* scenario. We observe the trend of increasing classification accuracy as we ensemble more *event types*. This is intuitive as we have more information content to amplify the confidence of occupant identification. The 25% per *event type* curve does not follow this trend, however, due to

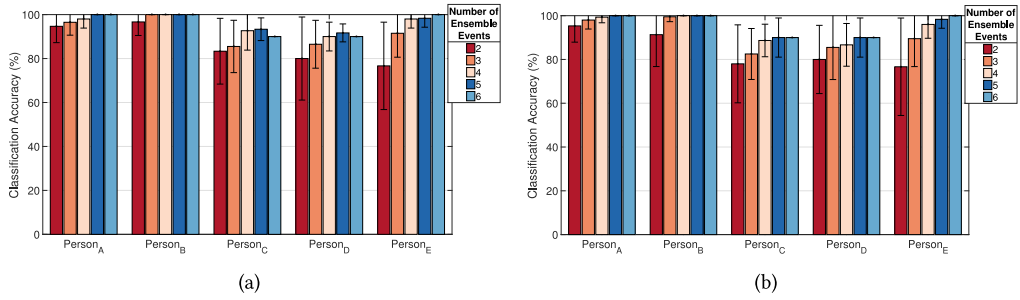


Fig. 12. (a) The figure depicts classification accuracy of ensemble of *event types* for different occupants when labels are *known*. As the number of ensemble *event types* increases, the accuracy also increases. (b) This figure depicts classification accuracy of ensemble of *event types* for different occupants when labels are *unknown*. As the number of ensemble *event types* increases, the accuracy also increases.

the fact that most of the training labels are incorrect, which would actually hurt the performance as the number of *event types* increases.

Noting the effects of number of *event types* and availability of training labels on classification accuracy, we now evaluate the performance of ensemble for both *known* and *unknown labels* scenarios, as depicted in Figure 11(b). From these two plots, we make the following two observations. (1) We observe that the classification accuracy increases as we ensemble more *event types* for both *known* and *unknown labels* scenarios. For example, we observe for the *unknown labels* scenario an increase from 84% to 96%. This is intuitive, and, in fact, one of the main contributions of *SenseTribute*, as increasing information content ultimately amplifies the accuracy of occupant identification. (2) We observe only a small difference in the resulting classification accuracy between the *known* and *unknown labels* scenarios. We further observe that the difference reduces as we ensemble more *event types*. This important observation means that *SenseTribute* provides a practical solution that *does not require users to provide manual labels* with no significant impact on occupant identification.

We also present the classification accuracy per occupant for different numbers of *event types*. Again, we report the average over all combinations. Figure 12(a) and (b) depict the corresponding results for *known* and *unknown labels* scenarios, respectively. For both figures, we observe the similar trend that as we ensemble more *event types*, we achieve higher classification accuracy. This trend is more evident with *Persons C, D, and E*, as these occupants depict low average classification accuracy when we ensemble two events, compared to *Persons A and B*. This can be explained by revisiting Figure 6(b). Recall that the accuracy for the dispensing *towel event type* were extremely low for these three occupants (as explained from Table 3 and Figure 7). However, the low accuracy of one particular *event type* gets compensated by performing ensemble with other *event types*, hence, demonstrating how *SenseTribute* benefits from the ensemble process.

5 DISCUSSION

In this section, we further discuss practical considerations and directions for further study with respect to activity segmentation, unsupervised learning techniques, sensor calibration, and correlation with prior work.

5.1 Additional Contextual Information

We highlight two additional contextual pieces of information that may potentially be helpful for *SenseTribute*, namely *order* and *time* of events. In this work, we design *SenseTribute* to perform

occupant identification based on the results of activity segmentation, which provides a sequence of events that are performed by a single person. In Section 4, we evaluate scenarios where the *order* of events (in an activity segment) are the same across different participants. However, in practice, there is a high probability that the *order* may vary. For example, when making a bowl of cereal, $Person_A$ may take out a bowl from the cabinet, milk from the fridge, and cereal from the cupboard, while $Person_B$ may perform the same activity in an opposite order. In addition, different occupants may conduct the same activity at different *times* of the day. For example, $Person_A$ usually makes cereal around 8 a.m., while $Person_B$ does the same at 10 a.m. Taking the above two observations into account, we hint at the possibility of a hybrid approach of solving both the activity segmentation and occupant identification problem simultaneously. This hybrid approach would potentially increase the performance with the additional contextual information. Furthermore, the hybrid approach may even increase the identification accuracy despite inconsistencies in different interactions by the same user over time, or similar interactions by different users.

5.2 Unsupervised Learning

Recall that when the training labels are not provided by the user, *SenseTribute* utilizes clustering to *infer* the quasi-training labels. We evaluate our results by clustering the history data during the bootstrapping phase. When *SenseTribute* is deployed in practice, we can utilize online learning techniques (Charikar et al. 1997; Choromanska and Monteleoni 2011; Liang and Klein 2009) to improve the results of clustering. This is because, over time, the clustering accuracy would increase as the system collects more data, ultimately leading to potentially higher identification accuracy.

Furthermore, in our evaluation, we assume the knowledge of “K” (i.e., number of occupants) in the K-means clustering algorithm. We make such assumptions because it is practical to have such prior knowledge of how many people live at home. Granted, we note that if guests are introduced to smart home, it may lead to less accurate results. In practice, however, there are clustering methods to estimate the optimal “K,” such as the Elbow method (Ketchen and Shook 1996). Also, there are other clustering algorithms that do not require the number of clusters (Ester et al. 1996). However, we leave this study for future work.

5.3 Sensor Calibration

Recall from our evaluation that we deploy sensors on different objects with consistent orientation of accelerometers and gyroscopes as presented in Section 4.1. While we conducted the experiment as a proof-of-concept, in practice, we cannot assume such deployment. Hence, the system would need a simple but important calibration phase to identify the axes that have relatively richer information content. *SenseTribute* may benefit from the calibration phase, as identifying a specific set of features and axes per object and/or event type would ultimately increase the identification performance. For example, Figure 13 depicts the signals from x -, y -, and z -axes of the accelerometer when a participant opens and closes a fridge door. It is evident that the z -axis contains richer information content as opposed to x and y -axes. This is because of the way that the accelerometer is attached to the fridge. Recall from Figure 5 in Section 4.1.1 that the z -axis is perpendicular, while the x - and y -axes are parallel to the surface of the fridge door. Hence, when the fridge door opens and closes, the Z -axis will capture more information content. Thus, the calibration phase would analyze the information content of the three axes to find that the x and y -axes do not contain additional information than the z -axis in identifying the occupants. Specifically, the analysis could entail simple time or frequency domain analysis to determine the axis that contains sufficient information for *SenseTribute*.

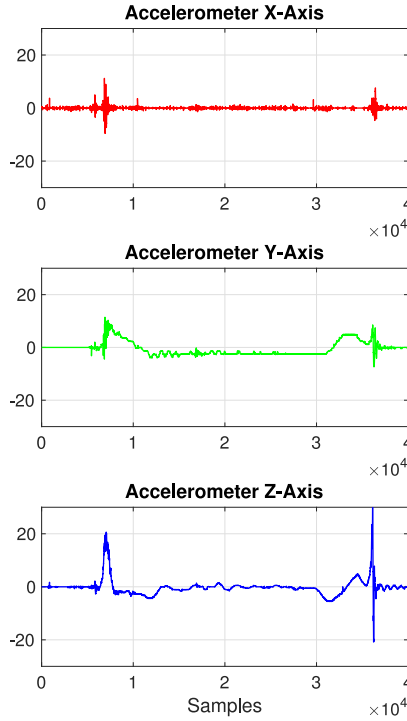


Fig. 13. This figure illustrates the raw accelerometer signals of x -, y -, and z -axes when a participant opens and closes a fridge door. Due to the orientation of the accelerometer installed on the fridge door during the experiment (as described in Section 4.1.1), the z -axis captures rich information content of the fridge door opening and closing motion, while the other two axes do not add additional information when classifying occupants.

5.4 Correlating with Prior Work

In this work, we design *SenseTribute* to specifically leverage existing on-object sensing devices to eliminate the need for deployment of sensors specifically for home occupant identification. However, we envision that in the near future, smart homes will be equipped with more built-in sensors. This is an opportunity for us as well because *SenseTribute* can take advantage of more pervasive sensors that are already deployed to potentially further increase the identification accuracy. For example, vibration sensors such as geophones and accelerometers are emerging as common solutions for structural health monitoring to monitor conditions of structural buildings (Kim et al. 2007; Ulusoy et al. 2012). Given the availability of such vibration sensors at home, we could take advantage by extracting information of home occupants by differentiating induced footsteps similar to prior work (Pan et al. 2014, 2015). Such information could then extend *SenseTribute* by performing ensemble together with the existing on-object sensors. Footstep is especially attractive for the following two reasons. First, footstep information could lead to more accurate activity segmentation as people often walk in between different activities (e.g., open a door, walk to a fridge, then open the fridge). Second, footstep signals can also compensate signals from on-object sensors due to different sources of error. While a person's walking pattern is known to be relatively consistent, identifying a person via gait is difficult due to other factors such as different types of shoes or location to the sensor. On the contrary, events sensed by on-object

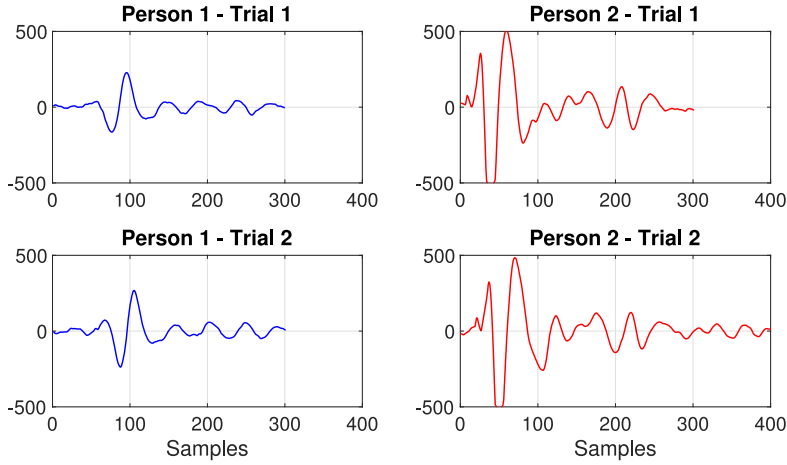


Fig. 14. This figure illustrates raw signals of footsteps from two participants captured by a geophone. The signals depict a high correlation between trials from the same participant. However, the signals are sufficiently differentiable across the two participants due to differing gait movements.

sensors usually do not have errors due to locations but have higher variations in the way that occupants interact with the object (e.g., door opening).

Envisioning the deployment of vibration sensors in a home, we conducted a preliminary experiment to test the feasibility of differentiating different occupants. Figure 14 depicts footstep signals collected by a vibration sensor (specifically a geophone) sampled at 1KHz. In particular, two participants were each asked to walk in a straight line for two separate trials. As depicted, the signals are similar between the two trials of the same person, while the signals can be sufficiently differentiated across the two participants.

6 CONCLUSION

We present *SenseTribute*, a smart home occupant identification system that leverages existing and prevalent *on-object sensing devices* equipped with inertial sensors, which are traditionally designed to monitor the status of objects such as doors. *SenseTribute* re-purposes these devices and exploits machine learning techniques to provide a low-cost, non-intrusive, and practical occupant identification system in a smart home with high accuracy, even when training labels are unavailable. Furthermore, *SenseTribute* combines information from multiple sensors on different objects to amplify the identification accuracy. We evaluate *SenseTribute* using real-world experiments with five on-object sensors deployed on distinct objects. Compared to an average of 74% per-object classification when the training labels are known, *SenseTribute* achieves an accuracy of 96% even when the labels are unknown.

APPENDIX

A EVALUATION DETAILS

We now provide details of some of the evaluations. In particular, we present tables that describe the confusion matrices of classification of Persons A to E interacting with different objects for the five occupant scenario depicted in Figure 6(a) in Section 4.2.

Table 5. Table Depicts a Confusion Matrix of *Knock Event Type* for the Five Occupant Scenario from Figure 6(a)

		Predicted as:					
		Person A	Person B	Person C	Person D	Person E	Classification Accuracy
Ground Truth	Person A	10	0	0	0	0	10/10
	Person B	0	9	0	0	1	9/10
	Person C	0	0	9	0	1	9/10
	Person D	0	0	0	10	0	10/10
	Person E	1	0	0	0	9	9/10

Table 6. Table Depicts a Confusion Matrix of *Fridge Event Type* for the Five Occupant Scenario from Figure 6(a)

		Predicted as:					
		Person A	Person B	Person C	Person D	Person E	Classification Accuracy
Ground Truth	Person A	10	0	0	0	0	10/10
	Person B	1	9	0	0	0	9/10
	Person C	0	0	8	0	2	8/10
	Person D	0	0	0	8	2	8/10
	Person E	0	0	2	2	6	6/10

Table 7. Table Depicts a Confusion Matrix of *Drawer Event Type* for the Five Occupant Scenario from Figure 6(a)

		Predicted as:					
		Person A	Person B	Person C	Person D	Person E	Classification Accuracy
Ground Truth	Person A	10	0	0	0	0	10/10
	Person B	0	9	0	1	0	9/10
	Person C	0	0	7	2	1	7/10
	Person D	0	1	1	8	0	8/10
	Person E	0	0	1	0	9	9/10

Table 8. Table Depicts a Confusion Matrix of *Window Event Type* for the Five Occupant Scenario from Figure 6(a)

		Predicted as:					
		Person A	Person B	Person C	Person D	Person E	Classification Accuracy
Ground Truth	Person A	6	0	1	2	1	6/10
	Person B	1	8	0	1	0	8/10
	Person C	2	0	8	0	0	8/10
	Person D	1	2	0	7	0	7/10
	Person E	2	0	0	0	8	8/10

ACKNOWLEDGMENTS

The views and conclusions contained here are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either express or implied, of CMU, NSF, or the U.S. Government or any of its agencies.

REFERENCES

- Analog Devices 2017. *ADXL335 Datasheet*. Retrieved from <http://www.analog.com/media/en/technical-documentation/data-sheets/ADXL335.pdf>.
- Arduino 2017. *Arduino/Genuino UNO*. Retrieved from <https://www.arduino.cc/en/Main/arduinoBoardUno>.
- Bharathan Balaji, Jian Xu, Anthony Nwokafor, Rajesh Gupta, and Yuvraj Agarwal. 2013. Sentinel: Occupancy based HVAC actuation using existing WiFi infrastructure within commercial buildings. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems (SenSys'13)*. ACM, New York, Article 17, 14 pages. DOI: <https://doi.org/10.1145/2517351.2517370>
- Kristin P. Bennett and Colin Campbell. 2000. Support vector machines: Hype or hallelujah? *SIGKDD Explor. Newsl.* (2000).
- Sebastian D. Bersch, Djamel Azzi, Rinat Khusainov, Ifeyinwa E. Achumba, and Jana Ries. 2014. Sensor data acquisition and processing parameters for human activity classification. Retrieved from <http://www.mdpi.com/1424-8220/14/3/4239>.
- S. Boll. 1979. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 27, 2 (1979), 113–120. DOI: <https://doi.org/10.1109/TASSP.1979.1163209>
- C. V. C. Bouten, K. T. M. Koekkoek, M. Verduin, R. Kodde, and J. D. Janssen. 1997. A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. *IEEE Transactions on Biomedical Engineering* 44, 3 (1997), 136–147. DOI: <https://doi.org/10.1109/10.554760>
- Cao Gadgets LLC. 2017. *Wireless Sensor Tag System: Monitor Everything from the Internet*. Retrieved from <http://www.wirelesstag.net>.
- Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2, 3, Article 27 (2011), 27. DOI: <https://doi.org/10.1145/1961189.1961199>
- Moses Charikar, Chandra Chekuri, Tomás Feder, and Rajeev Motwani. 1997. Incremental clustering and dynamic information retrieval. In *ACM Symposium on Theory of Computing (STOC'97)*.
- Anna Choromanska and Claire Monteleoni. 2011. Online clustering with experts. In *Proceedings of the 2011 International Conference on On-line Trading of Exploration and Exploitation 2—Volume 26 (OTEAE'11)*.
- R. R. Coifman and M. V. Wickerhauser. 1992. Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory* 38, 2 (March 1992), 713–718. DOI: <https://doi.org/10.1109/18.119732>
- Giorgio Conte, Massimo De Marchi, Alessandro A. Nacci, Vincenzo Rana, and Donatella Sciuto. 2014. BlueSentinel: A first approach using iBeacon for an energy efficient occupancy detection system (*ACM BuildSys'14*). ACM.
- Deloitte. 2015. *The Digital Predictions—The Deloitte Consumer Review*.
- Ecolink. 2017. *Ecolink Z-Wave Garage Door Tilt Sensor*. Retrieved from <http://discoverecolink.com/product/tiltzwave2-eco/>.
- Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. 1996. A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise (*KDD'96*).
- Davrondzhon Gafurov and Einar Snekkenes. 2009. Gait recognition using wearable motion recording sensors. *EURASIP J. Adv. Signal Process* 2009, Article 7 (2009), 16. DOI: <https://doi.org/10.1155/2009/415817>
- D. Gafurov, E. Snekkenes, and P. Bours. 2007. Gait authentication and identification using wearable accelerometer sensor. In *2007 IEEE Workshop on Automatic Identification Advanced Technologies*. 220–225. DOI: <https://doi.org/10.1109/AUTOID.2007.380623>
- Tobias Grosse-Puppenthal, Xavier Dellangnol, Christian Hatzfeld, Biying Fu, Mario Kupnik, Arjan Kuijper, Matthias R. Hastall, James Scott, and Marco Gruteser. 2016. Platypus: Indoor localization and identification through sensing of electric potential changes in human bodies (*ACM MobiSys'16*).
- Rim Helaoui, Daniele Riboni, and Heiner Stuckenschmidt. 2013. A probabilistic ontological framework for the recognition of multilevel human activities (*ACM UbiComp'13*).
- Timothy W. Hnat, Erin Griffiths, Ray Dawson, and Kamin Whitehouse. 2012. Doorjamb: Unobtrusive room-level tracking of people in homes using doorway sensors (*SenSys'12*).
- T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu. 2002. An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24, 7 (2002), 881–892. DOI: <https://doi.org/10.1109/TPAMI.2002.1017616>
- David J. Ketchen and Christopher L. Shook. 1996. The application of cluster analysis in strategic management research: An analysis and critique. *Strategic Management Journal* 17, 6 (1996), 441–458.
- Nacer Khalil, Driss Benhaddou, Omprakash Gnawali, and Jaspal Subhlok. 2016. Nonintrusive occupant identification by sensing body shape and movement (*ACM BuildSys'16*).

- Sukun Kim, S. Pakzad, D. Culler, J. Demmel, G. Fennes, S. Glaser, and M. Turon. 2007. Health monitoring of civil infrastructures using wireless sensor networks. In *Information Processing in Sensor Networks (IPSN'07)*. ACM, New York, NY, 254–263. DOI : <https://doi.org/10.1145/1236360.1236395>
- P. Kodeswaran, R. Kokku, M. Mallick, and S. Sen. 2016. Demultiplexing activities of daily living in IoT enabled smarthomes. In *IEEE INFOCOM 2016*.
- Percy Liang and Dan Klein. 2009. Online EM for unsupervised models. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL'09)*.
- Qiguang Lin, Ea-Ee Jan, and James Flanagan. 1994. Microphone arrays and speaker identification. *IEEE Transactions on Speech and Audio Processing* 2, 4 (1994), 622–629. DOI : <https://doi.org/10.1109/89.326620>
- S. Lloyd. 1982. Least squares quantization in PCM. *IEEE Transactions on Information Theory* 28, 2 (1982), 129–137. DOI : <https://doi.org/10.1109/TIT.1982.1056489>
- J. Mantyjarvi, M. Lindholm, E. Vildjiounaite, S. M. Makela, and H. A. Ailisto. 2005. Identifying users of portable devices from gait pattern with accelerometers. In *ICASSP 2005*.
- Mathworks. 2017. *Entropy (Wavelet Package)*. Retrieved from <https://www.mathworks.com/help/wavelet/ref/wentropy.html>.
- MetaSensor 2017. *MetaSensor: Meet Sensor-1*. Retrieved from <https://www.metasensor.com/>.
- Samantha Murphy. 2015. *Samsung: By 2020, All of Our Products Will be Connected to the Web*. Retrieved from http://mashable.com/2015/01/05/samsung-internet-of-things/#c3_K6GzxJ65N.
- Notion 2017. *Notion: Home Awareness, Simplified*. Retrieved from <http://getnotion.com/>.
- Shijia Pan, Amelie Bonde, Jie Jing, Lin Zhang, Pei Zhang, and Hae Young Noh. 2014. BOES: Building occupancy estimation system using sparse ambient vibration monitoring. In *Proc. SPIE 9061*. DOI : <https://doi.org/10.1117/12.2046510>
- Shijia Pan, Ningning Wang, Yuqiu Qian, Irem Velibeyoglu, Hae Young Noh, and Pei Zhang. 2015. Indoor person identification through footstep induced structural vibration (*ACM HotMobile'15*).
- Shijia Pan, Tong Yu, Mostafa Mirshekari, Jonathon Fagert, Amelie Bonde, Ole J. Mengshoel, Hae Young Noh, and Pei Zhang. 2017. FootprintID: Indoor pedestrian identification through ambient structural vibration sensing. *ACM Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3, Article 89 (2017), 31. DOI : <https://doi.org/10.1145/3130954>
- William M. Rand. 1971. Objective criteria for the evaluation of clustering methods. *J. Amer. Statist. Assoc.* 66, 336 (1971), 846–850. DOI : <https://doi.org/10.2307/2284239>
- Juhi Ranjan, Erin Griffiths, and Kamin Whitehouse. 2014. Discerning electrical and water usage by individuals in homes (*ACM BuildSys'14*).
- Daniele Riboni, Timo Sztyler, Gabriele Civitarese, and Heiner Stuckenschmidt. 2016. Unsupervised recognition of interleaved activities of daily living through ontological and probabilistic reasoning (*ACM UbiComp'16*).
- Samsung 2017. *Samsung SmartThings Multipurpose Sensor*. Retrieved from <https://support.smartthings.com/hc/en-us/articles/205382174-Samsung-SmartThings-Multipurpose-Sensor>.
- Scott T. Stillman, Rawesak Tanawongsuwan, and Irfan A. Essa. 1998. *A System for Tracking and Recognizing Multiple People with Multiple Cameras*. Technical Report. Georgia Institute of Technology.
- STMicroelectronics 2017. *LPY503AL Datasheet*. Retrieved from <http://www.st.com/web/en/resource/technical/document/datasheet/CD00237199.pdf>.
- Hasan S. Ulusoy, Erol Kalkan, Jon Peter B. Fletcher, Paul Friberg, W. K. Leith, and Krishna Banga. 2012. Design and implementation of a structural health monitoring and alerting system for hospital buildings in the United States. In *World Conference on Earthquake Engineering*.
- U.S. Census Bureau 2017. *Average Number of People per Household in the United States from 1960 to 2016*. Retrieved from <https://www.statista.com/statistics/183648/average-size-of-households-in-the-us/>.
- Yunze Zeng, Parth H. Pathak, and Prasant Mohapatra. 2016. WiWho: Wifi-based person identification in smart spaces (*ACM/IEEE IPSN'16*).
- J. Zhang, B. Wei, W. Hu, and S. S. Kanhere. 2016. WiFi-ID: Human identification using WiFi signal. In *2016 International Conference on Distributed Computing in Sensor Systems (DCOSS)*.

Received January 2018; revised April 2018; accepted May 2018