Applied Energy 220 (2018) 814–828

Contents lists available at ScienceDirect

Applied Energy

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

Nonintrusive ultrasonic-based occupant identification for energy efficient smart building applications

Nacer Khalil*, Driss Benhaddou, Omprakash Gnawali, Jaspal Subhlok

University of Houston, 4800 Calhoun Road, Houston, TX 77004, United States

HIGHLIGHTS

- A novel method to identify occupants by sensing their shape and movement.
- This method is based on sensing height and width from ultrasonic ping sensors attached to the door.
- The system enables to identify up to 20 people with a 95% accuracy.
- The system does not require any training.

ARTICLE INFO

Keywords:
Occupant identification
Sensor networks
Smart buildings
Clustering
Machine learning

ABSTRACT

The ability to non-intrusively identify people will enable smart buildings to customize the environment to meet occupants' comfort level while saving energy. Occupant identification can help in energy savings effort in a building because we can retrieve each occupant's temperature preference profile and choose the temperature that minimizes the total discomfort of a group in the building. To enable occupant identification in buildings, many methods used can be intrusive, such as using cameras or requiring the users to carry mobile gadgets or a smart phone. Non-intrusive techniques are gaining interest in smart building applications. In this paper, we present a non-intrusive ultrasonic based sensing technique to identify people by sensing their body shape and movement. The ultrasonic sensors are placed on the top and sides of doors to measure the height and width as the occupant walks through the instrumented doorway. Height and width and their related features can give a unique signature to occupants to identify them. In this study, the proposed system senses a stream of height and width data, recognizes the walking event when a person walks through the door, and extracts features that capture a person's movement as well as physical shape. These features are fed to a clustering algorithm that associates each occupant with a distinct cluster. The system was deployed for a total of three months. The results show that the proposed approach achieves 95% accuracy with 20 occupants suggesting the suitability of our approach in commercial building settings. In addition, the results show that using girth to distinguish between occupants is more successful than using height. We show that this system generalizes beyond our datasets and works for different populations of different physical distributions.

1. Introduction

Buildings energy consumption constitutes about 40% of the total energy consumed in the United States. There are efforts to develop buildings that are smart and energy efficient. Developing energy efficient buildings requires addressing challenges from different aspects including energy efficient building material, integrating renewable energies, and using more efficient Heating, Ventilation and Air Conditioning (HVAC) systems. The latter is a significant consumer of power in buildings, especially in warmer or colder areas. By having accurate estimates of occupancy or human activities inside buildings, one can develop smarter algorithms to better manage the HVAC yielding saved energy and increased occupants' comfort. Occupancy allows the first order optimization of HVAC use in a building. However, occupancy does not provide any information about the occupants' comfort levels. In fact, occupancy is mostly used to adjust the airflow, but the temperature of the air flowing to the various rooms is not a function of occupancy but rather constant and predefined by building managers. As a result, many occupants are uncomfortable in buildings because of the lack of customization and inability to tailor to the present occupant's thermal preferences. Thus, there is a case for identifying occupants in a building to customize the HVAC operation to meet the

* Corresponding author.
E-mail addresses: nkhalil@uh.edu (N. Khalil), dbenhad@central.uh.edu (D. Benhaddou), gnawali@cs.uh.edu (O. Gnawali), Subhlok@cs.uh.edu (J. Subhlok).

https://doi.org/10.1016/j.apenergy.2018.03.018
Received 25 August 2017; Received in revised form 2 March 2018; Accepted 15 March 2018
Available online 27 March 2018
0306-2619/

occupants preferences and avoiding the *too hot* or *too cold* scenarios, and hence energy waste.

Smart and energy efficient buildings need to tailor the climate to the occupants' preferences. This is important because not only can it make the buildings more comfortable but also save energy. In fact, studies have shown that we can achieve at least 30% energy savings by having an accurate estimate of occupancy Erickson et al. [1]. With accurate occupancy estimate, Erickson and Cerpa [2] show that we can adjust the HVAC airflow by feeding the occupied areas with just enough air given the number of occupants. However, even though we can achieve optimum airflow thus saving energy, there is no way for the occupants to provide feedback about their comfort given the set temperature. This leads to increased discomfort and missed energy saving opportunities because by having the occupants' comfort profile, we can set the temperature to accommodate as much people as possible. Current HVAC systems have predefined temperature settings low enough to accommodate every occupant's preference in the case of cooling and vice versa for heating. However, there are different temperature setting strategies that can be performed such as choosing the median temperature, mean, or various voting-based strategies as shown in Zhang et al. [3] or by taking into account thermal occupant comfort constraints as shown by Ghahramani et al. [4]. Most importantly, having a way to retrieve the occupants' temperature preference is crucial to making the buildings more comfortable, more adaptable and more energy efficient than by simply relying on occupancy sensing.

In this paper, we propose a non-intrusive ultrasonic based indoor occupant identification system that can be implemented in doorways that scale to accurately identifying 20 users which can be used in commercial applications such as buildings and nursing homes. The sensing technique measures the occupant's shape and movement which are used by a clustering algorithm to identify people. The proposed solution mounts three ultrasonic ping sensors at the top and the side of the door frame to measure occupants' height and girth. The sensors at the side of the door are used to compute the width of the person passing through it. Using the height and width time series, the system extracts a set of features to infer the occupant's body shape and movement. We identity occupants by clustering data from these features to uniquely characterize occupants. This makes our solution easy to use since it doesn't require to train the model because it's based on clustering. This solution can easily be integrated into doors and is cost effective since it uses off-the-shelf ultrasonic sensors.

Given that we rely on height and width which are weak biometrics, the system is subject to uncertainty as some people may have similar height and width measures. This paper studies the extent to which such similarities will affect the system's performance. Our findings show that clustering with the body shape and time spent walking through the door (thus movement) enables us to accurately identify people. The results show that indeed these parameters are key to differentiating between people thus achieve 95% accuracy for 20 people and 75% for 50 people.

We introduce the system design which includes sensing methods, data filtering techniques to minimize noise from the measurements, and clustering algorithms to identify occupants. We conducted two experiments: a room-scale experiment that lasted one month long involving 20 people in a classroom environment at the University of Houston and a building-scale experiment involving five door frames for a period of two months with over 170 participants.

1.1. Contributions

The contributions of this paper are:

- We propose and implement a system that identifies occupants using height and width measured using ultrasonic sensors mounted in doors. The system extract a set of useful features which are used by a clustering algorithm to identify people.
- We Investigate the impact of the computed features on the accuracy of the system. In particular, our results show that gait contributes more significantly to identification than height.
- We compare the accuracy of different combinations of features in identifying occupants. The results show that clustering with girth and time provide more accurate results than height and time.
- We investigate how the method scales for larger populations of up to 50 people.
- We investigate how would the system perform on different populations of different physical characteristics. We perform that by drawing samples that match a physical characteristics and evaluating the performance of our model on it.

Relative to our preliminary work in Khalil et al. [5], we deployed a new large scale testbed composed of five door frames for a period of two months with over 170 people participants. This was important to evaluate how consistent is the system performance when conducted with more and different participants and in different locations involving more door frames. We also studied the extent to which such a system would scale to different populations with different characteristics and how would the performance change as we scale the population size. We also explored the use of Spectral clustering on how it performs compared to DBSCAN. We evaluated the walking event performance. Our research demonstrates that it is possible to identify occupant with 95% accuracy for 20 people and 75% for 50 people.

2. Related work

2.1. HVAC control

HVAC is one of the main power consumers in buildings in the US. In fact, it consumes 50% of building energy consumption and 20% of the total energy consumption Pérez-Lombard et al. [6]. Thus, numerous research has been done to make these systems more energy efficient. A few researchers have explored the possibility of augmenting the HVAC control with occupancy information for the purpose of energy saving Conte et al. [7], Färber et al. [8], Erickson and Cerpa [2], Erickson et al. [1]. Also, Brandemuehl and Braun [9] showed that energy could be saved by adjusting ventilation to maximum occupancy, and Erickson et al. [10] showed that the majority of current HVAC systems assume maximum occupancy during normal hours and turn it off during the evening which is a source of inefficiency because rooms are not always at maximum occupancy. Therefore, an accurate occupancy estimate could save energy by adjusting the ventilation to the estimated occupancy rather than maximum occupancy.

However, having an accurate estimate of occupancy is not enough to save energy. In fact, Hoyt et al. [11] has shown that by increasing the set point temperature by 1 °C, we can save 7–15% energy consumed by the HVAC in the summer in three cities namely San Francisco, Miami and Phoenix. These savings are achievable because a large fraction of users feel that the temperature set points in their building is too extreme. In fact, in a study at the University of Southern California, Jazizadeh et al. [12] showed that 60% of the users felt cool to cold inside the building during the summers showing the potential to raise the temperature which would increase the users' comfort and save energy. So one can make the occupants more comfortable as well as save energy having users comfort preference.

2.2. Setting HVAC temperature

Having an estimate of occupancy is crucial to achieve proper ventilation, but understanding the thermal preference of users would help save even further by taking into consideration the occupants' preferences. In fact, Karjalainen and Koistinen [13] discuss that the lack of control as well as individual control in buildings increase occupants' discomfort in offices because the systems are planned without an understanding of the users which change over time. Erickson and Cerpa

[14] show that by bringing the users "in the loop", you can improve the users' comfort and save 10% over standard baseline strategies involving ASHRAE 55 guidelines Standard [15]. Zhang et al. [3] shows a method to set the temperature based on thermal comfort voting from the occupants.

2.3. Identification for HVAC control

To set the temperature in the room, understanding not only how many people are in the room but also who is in the room is important for increased comfort and savings. Ghahramani et al. [4] show that by taking into account numerous parameters including the occupants' comfort constraints, the temperature could be set by solving an optimization problem which tries to maximize the comfort and take into consideration energy savings. Therefore, for such systems having access to the occupants' identity in a given room is important and this is not trivial because occupants move to different rooms throughout the day and a good example would be university classroom buildings.

Identification and tracking people has captured interest from the research community in the last decades and different sensing methods has been proposed. These sensing technologies rely on sensing strong biometrics such as facial recognition Lanitis et al. [16], fingerprint Hrechak and McHugh [17], iris and hand geometry Tisse et al. [18], Pan et al. [19]. Other technologies make use of weak biometrics such as height Hnat et al. [20], Srinivasan et al. [21] and weight Jenkins and Ellis [22].

2.4. Carried and wearable devices

Different proposed technologies have been proposed. Ranjan et al. [23] make use of RFID-based wearables, Subbu and Thomas [24] track users' Smartphones and Hnat et al. [20] make use of iBeacon technology. In RFID-based wearables, Ranjan et al. propose an RF Doormat which is an RF sensing system that can identify and track users' locations as they walk through doorways. Smartphones have been used as a mean to identify and track and localize users in buildings. These systems identify occupants with high accuracy but suffer from missing a user if she does not have the wearable.

2.5. Strong biometrics

Different systems have been proposed that use facial, fingerprint, iris and hand geometry and achieved high accuracy as have shown Lanitis et al. [16], Hrechak and McHugh [17], Tisse et al. [18], Pan et al. [19]. However, these methods raise privacy concerns and many require the users to interact with the system. Zhao et al. [25] show that vision-based person identification approaches for identification namely face recognition techniques has been explored for decades. In fact, vision-based system target different sets of applications ranging from entertainment and virtual reality to security and surveillance. Some vision-based identification methods rely on gait to identify occupants as have reported Kale et al. [26]. These systems are energy intensive, invasive and not easily deployed in environments such as buildings as they are privacy infringing.

Hrechak and McHugh [17], Tisse et al. [18] and Pan et al. [19] use other systems such as those using fingerprint, iris, hand and retina sensing require some level of engagement from the user to authenticate. This requirement is difficult to enforce and if users do not authenticate, then the identification process fails. These systems though highly accurate are privacy infringing and may not be applicable to large commercial buildings.

2.6. Weak biometrics

Numerous methods involve sensing weak biometrics. Methods that sense height, weight, footstep vibration and step force for occupant identification have been proposed by Hnat et al. [20], Srinivasan et al. [21], Pan et al. [27], Jenkins and Ellis [28], Liau et al. [29], Elrod and Shrader [30]. These sensing systems are non-intrusive and do not require effort from the occupant in the identification process. Among these systems, footstep vibration based systems have been proposed to detect occupants identity. Pan et al. [27] showed that their system achieves 96% accuracy for a population of 4–5 people. However, the system cannot operate when multiple occupants are present in the area because vibration signals would interfere with each other and it would be difficult to separate them.

Height as a weak biometric has been proposed in the literature using ultrasonic sensing by Hnat et al. [20], Srinivasan et al. [21]. In fact, Hnat et al. [20] introduce Doorjamb which is a system that uses height information, walking direction, and tracking information to identify users achieved a high accuracy rate within a population of 2–4 people. However, our system leverages height and width to sensors to identify up to 20 people with an accuracy of 95%.

Weight is also a good weak biometric and has been used in research to identify occupants Jenkins and Ellis [28], Liau et al. [29]. In fact, Jenkins and Ellis [28] showed that weight can be used as a weak biometric and it can achieve 80% accuracy for up to 15 people.

New systems emerged that directly identify occupants from sensed RF signals. In fact, with systems such as WiWho, Zeng et al. [45] proved that their system was able to identify occupants by observing the signature of the RF signal reflection intensity on the human body. However, many of these systems fail to identify more than 4 people with 90% and this quickly drops to 70% when the population is 6 people. A Similar system is proposed by Zhang et al. [31] that also uses WiFi signals to identify occupants. Table 1 compares some of the work in this area. These systems are non = intrusive, do not require user's engagement to authenticate but many of them fail to identify large populations which make them limited and difficult to use in large commercial buildings. (see Table 2).

3. System design

The system is designed to sense walkers' body shape and movement as they pass through the door. We design a door frame which we install three ultrasonic sensors (see Fig. 1a and c). One sensor is placed on top facing downwards referred to as UT (Ultrasonic Top), and other two sensors are placed on the sides horizontally at 1 meter high referred to as UR (Ultrasonic Right) and UL (Ultrasonic Left), which is at the right and left sides of the frame. The sensor-instrumented door is installed in room 219 at the Technology building at the University of Houston. The sensor-instrumented door sends the extracted features to the back end system to run the clustering algorithms and identify occupants.

The system includes four components: Sensing and calibration, event recognition, feature extraction, clustering and decision making. Fig. 2 depicts the process. The ultrasonic ping sensors compute the distance between the sensor and the closest object (in this case a person) and generate three streams of readings per walker. When a walking event is detected, the generated data is preprocessed and a set of features is computed and extracted. These feature instances are then used to build a clustering model for every user.

Table 1
Table showing a comparison of different non-intrusive identification methods.

Paper	Sensor	Accuracy (%)	population
Hnat et al. [20]	Ultrasonic	94	5
Pan et al. [27]	Geophone	96	5
Zeng et al. [45]	Wi-Fi RF	93	4
Jenkins and Ellis [28]	Pressure	80	15

Table 2
Table showing the needed hardware with their respective quantities and prices per unit per doorframe.

Hardware	Quantity	Cost(USD)
Arduino Uno	1	30
Parallax Ultrasonic Sensor	3	15
Raspberry PI 2 Model B	1	35

3.1. Sensor platform

We instrument each doorframe with ultrasonic ping sensors. This is composed of the following hardware:

- Arduino Uno
- Parallax Ultrasonic Ping sensor
- Raspberry PI 2 model B

The three ultrasonic sensors are connected to the Arduino which does the data acquisition and preprocessing. Data preprocessing consists of noise detection and correction as well as walking event detection. The walking event data is sent to the Raspberry PI through USB. The Arduino and Raspberry PI communicate over USB as opposed to low power Protocol such as Zigbee because the amount of data generated is much higher than what is supported by Zigbee. In fact, Zigbee's maximum data rate is 256Kbps which is not sufficient given the high sampling rate, and the transmission delay is larger than the sampling time making Zigbee a bottleneck. The feature extraction and the clustering algorithm can run on the Raspberry PI. However, during the experiment and data collection, We sent the extracted features to a server where the data is both stored and processed by the Machine Learning algorithm. We send the data to the server over WiFi and we use RabbitMQ as our message queue mechanism to transfer the data from the Raspberry PI to the server. The total cost of sensor hardware per doorframe is $80.

3.2. Sensing and calibration

Sensors need calibration as signal acquisition and sampling introduces errors in calculating the distance. For the sensor UT, the longest distance the ultrasonic pulse will travel is 4.2 m which is twice the height of the door frame as the ultrasound signal needs to go back and forth. Given the speed of sound of 341 m/s, we estimate that it will travel this distance (4.2 m) in 12 ms, which represents the maximum possible delay. For the sensors UL and UR which are separated by a distance of 1.2 meters, the maximum expected delay is 7 ms. The three sensors are sampled sequentially to avoid cross-talk between sensors.

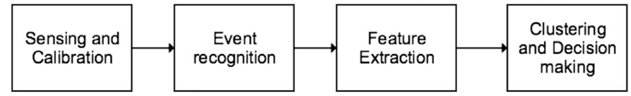


Fig. 2. Sequence of operations for occupant identification in a building.

The way we sense a walking event is as follows: we send a beam from one sensor and wait for the reflection to reach the sensor, we compare the time it took with the maximum allocated time for the sensor and set the node to sleep for the rest of the allotted time. The purpose of this is to keep the sampling rate at a stable fixed rate regardless of the walker's size. This also will prevent crosstalk between three ultrasonic sensors as it allocates enough time for the beam to be sent and reflected back to the sensor before operating the next sensor. In fact, synchronization between sensors is key to avoiding crosstalk and reducing noise in distance measurements.

The measured delay from UT is converted to height using the following formula:

$$d_{height}(t_{UT}) = d_{maxheight} - \frac{34.3}{2} t_{UT}$$

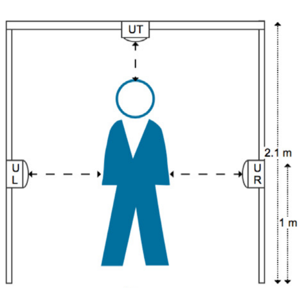
where 34.3 is the distance, in cm, traveled by sound every 1 ms. The variable t is the delay in ms. We divide by 2 because the measured delay represents the time for the pulse to go back and forth. $d_{maxheight}$ refers to the maximum distance separating the sensor and the ground in the case of UT. The maximum distance measured by UT, UL, and UR in our testbed are respectively 212 cm, 124 cm,124 cm.

The measured distances are then converted to the width of the person passing through the door using the following formula:

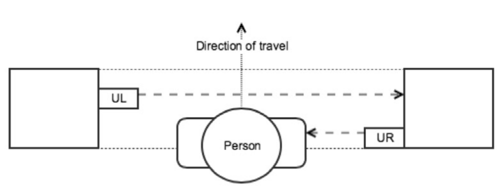
$$d_{width}(t) = d_{maxwidth} - \frac{34.3}{2} t_{UL} - \frac{34.3}{2} t_{UR}$$

When there is no one under the door frame, then the width formula returns a negative figure, more exactly $-d_{max_width}$ because both distances computed from UL and UR will be equal to $d_{maxwidth}$. To avoid this problem, we first check if $\frac{34.3}{2}t_{UL} = \frac{34.3}{2}t_{ULR} = d_{maxwidth}$, if true we return 0, and if not we compute the width using the formula.

Our first design uses periodic polling with an interval of 29 ms to simplify the implementation. Periodic polling is energy intensive and impacts the lifetime of the sensors, therefore is not the best way to operate in real world application. In a real world environment, we would add a motion sensor to optimize the operation to activate the sampling of the ultrasonic sensors when the motion sensor detects a person close to the door frame. Adding a motion sensor wouldn't significantly increase the overall cost of the system since off-the-shelf costs as low as $5 and wouldn't add significant complexity to the system. In addition, activated motion sensing is less energy intensive than running three ultrasonic sensors continuously.



(a) Figure showing the conceptual design of the sensor instrumented door frame.



(b) Top view of the doorframe showing direction detection as person walks through the door. UT has not been depicted in this figure to emphasize on the displacement between UL and UR

(c) Front view of the door frame with the ultrasonic sensors mounted.

Fig. 1. Figure showing the sensor instrumented doorframe schematic figures and a photo of our current testbed.

3.3. Walking event recognition

A Walking event refers to the stream of {UT, UL, UR} readings. Every time a person walks through the door frame, we receive a stream of data and the number of readings varies between 35 and 40 depending on the speed of the person. Fig. 4a illustrates the data stream. The faster the person walks, the fewer the readings. This stream of data will contain noisy points and errors. These noisy points need to be corrected and recovered before processing the features.

Since our testbed is in periodic polling, we get continuous data stream from the ultrasonic sensors. To detect a walking event, we look at the height as a detection mechanism. When there is no one under UT, we expect the maximum value. Algorithm 1 shows how we extract the walking event from a stream of data.

It has been shown that the average height of people in the United States is 169 cm with a standard deviation of 7.5 cm Ogden et al. [32]. So a height interval of 3 standard deviations from the mean should statistically cover 99% of the walkers assuming height follows the Gaussian distribution. The lower end of the interval would be 146.5 cm and we chose 140 cm as a lower bound. The reason is that height measured by UT is not necessary the ground truth and many times it is lower because of how the person walks, especially if the walker is looking a bit downwards or holding a backpack or just using a smartphone. The walking event starts when the measured height is at least 140 cm and stops when the height is less than 140 cm by allowing at most 4 consecutive points that are out of this interval. This last condition is chosen to prevent erroneous readings from making the system think the walker is no longer at the door and gives the impression we have multiple events. We chose 4 consecutive experimentally because it yields the most accurate walking event.

Algorithm 1. Extract Walking Event

```
procedure EXTRACT WALKING EVENT
missed ← 0
minheight ← 140
maxmissed ← 4
queue ← FIFOQueue
walkingevent ← emptyFIFOQueue
    do
        reading ← Dequeueelementfromqueue
        if reading.height > min_height then
            Enqueue reading to walkingevent
            missed ← 0
        else if missed < max_missed then
            missed ← missed + 1
        else
            return walking event
        end if
    while queue not empty
end procedure
```

3.3.1. Person direction recognition

The width sensors UL and UR are displaced in parallel to the walking direction line. As the person is walking, one sensor is closer to her than the other and therefore the closest will be the one to detect her by returning a non-default (a default value means max width when no one is under the door) value indicating it came in contact with a person. Fig. 1b shows how sensors UL and UR are displaced and how this displacement helps detect the person's direction. This displacement helps not only with person direction recognition but also for more accurate width measurement as shown in Section 4.11.

The direction is set when setting up the door frame in the room. In our testbed, we positioned our door to have the sensor UR closer to the entrance, so if a person is entering the room, the first non-default reading will be from UR and when exiting the first non-default reading will be coming from UL. If a person intentionally rotates the door, then our system will start giving wrong directions, but in a real deployment, the sensors will be mounted on the door frame by drilling a hole into the door and such an intentional rotation would not be possible.

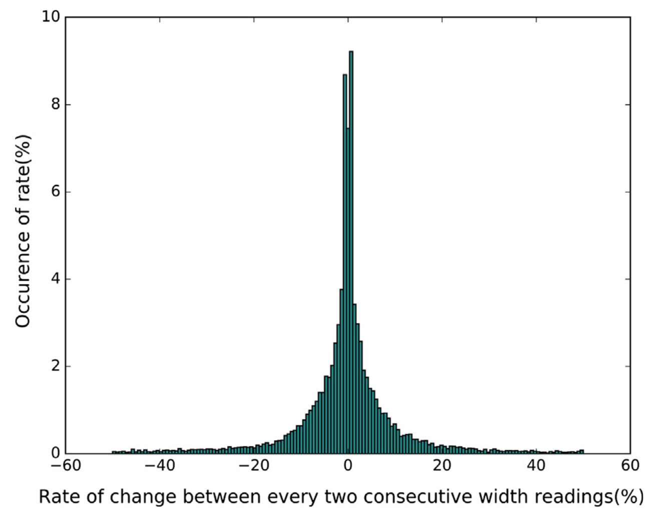


Fig. 3. Histogram of the distribution of adjacent width measurements rate of change.

3.4. Noise canceling and correction

Once the walking event is recognized and its respective sensory data stream is detected, we preprocess the data to filter noisy points before the event is further processed. We filter out the readings that are outside the interval [0, Max Height] and more than 30% difference between adjacent points. The reason is that we have observed that the width varies approximatively by 15% and the latter is higher when a person is carrying a purse. Fig. 3 is a Histogram that depicts the distribution of adjacent readings.

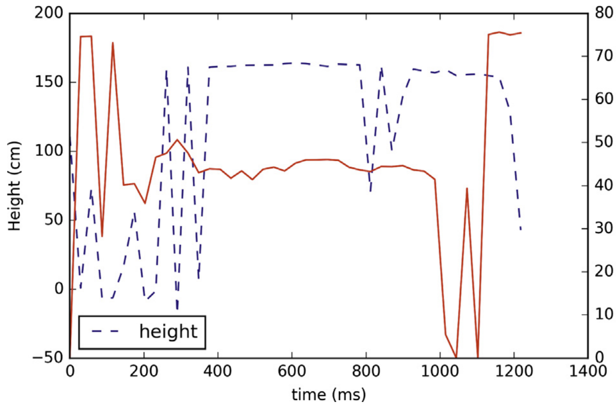
Once noisy points are identified, we need to either remove or recover them. Since the height measure uses only UT, then removing would not affect the overall height data. However, data from UL and UR are computed in pairs and removing one UR implies removing its equivalent UL or vice versa. For example, if we remove UL at t1 but not UR at t1, we will end up with more UR measures than UL measures and most importantly, we will end up with pairs that did not occur at the same time. Therefore, the best approach is to use linear interpolation to replace the noisy values. Therefore, after receiving the raw walking event data stream, we identify the noisy values and replace them using linear interpolation. The new stream of data is then used for feature extraction.

3.5. Main features

To detect occupants, we first use the preprocessed stream of data to extract a set of features that will be used to detect and identify users. It is illustrated in Fig. 4a and Table 4b. We experimented with several features including max, min, average, bounce, girth, and time under the door. Girth and time under door gave the best results for occupant identification.

3.5.1. Girth

Girth is a circumference measurement around a person's waist. In order to compute the girth, we use the stream of width data to create 2 point clouds where y is the instance number multiplied by the distance traveled each sampling interval and x is equal to $\frac{width}{2}$. We generate two



(a) Plot showing an instance of Height and Width sensory readings as a function of time when a person walked through the door.

Feature	Value
Average Height	156.1
Maximum Height	163.8
Average Width	39.5
Maximum Width	42.1
Girth	83.1
Hand-Waist Distance	13.2
Bounce	19.7

(b) Table showing list of extracted features from the walking event height and width readings.

Fig. 4. Figure showing an instance of a walking event height and width readings and the extracted Features.

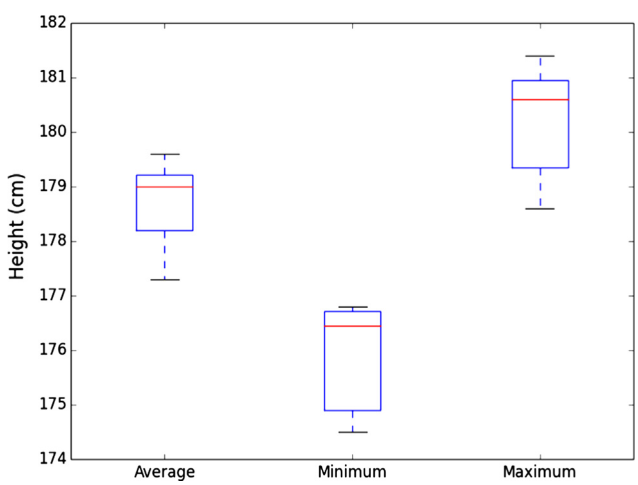


Fig. 5. Box plot showing variance in average, minimum and maximum height for 7 trials.

points, (x,y) and $(-x, y)$. We then construct the convex hull for all the point clouds and calculate its perimeter using Euclidean distance as the distance measure. Assuming an average speed of 5 km/h, a person will walk 3.6 cm every 29 ms. The pseudo-code for constructing the girth is presented in Algorithm 2.

Algorithm 2. Girth Calculation Algorithm

procedure COMPUTE GIRTH
$distancewalkedperiteration \leftarrow 3.6$
$edge \leftarrow 0$
$x_t \leftarrow \frac{width[t]}{2}$
$y_t \leftarrow iteration_t * distancewalkedperiteration$
$edge \leftarrow edge + \sqrt{(x_t - x_{t-1})^2 + (y_t - y_{t-1})^2}$
Return: $girth \leftarrow 2*$edge
end procedure

3.5.2. Time

Since everyone has a different walking speed, we measure time indirectly by counting the number of interval time the person spent under the door. Given the sampling rate, s, (s = 35 Hz in the testbed), we calculate the time by dividing the number of height measurements $H = \{h_1,...,h_n\}$ by s and width measurements $W = \{w_1,...,w_n\}$ by s. We select the result that gives the max time spent under the door, t, is therefore given by:

$t = \frac{1}{s} max(|H|,|W|)$ We take the maximum because noise may alter the length of the H or W and therefore select the longest since it was the least affected by noise.

3.5.3. Bounce

Bounce is a gait measure of how a person bounces as she walks. Some people tend to bounce more than others when walking. We capture bounce from the height measurement by subtracting the minimum from maximum height. Given height measurements $H = \{h_1,...,h_n\}$, we set $Bounce = max(H) - Min(H)$.

3.6. Other features

3.6.1. Maximum, minimum, and average height

From the stream of height measures, we compute the minimum, maximum and average height. To decide which of the three features is most appropriate for identification, we need to find which metrics has the least variance for the same person and ideally closer to the ground truth. To do that, we conducted a small experiment where the same person performs 7 walking events under the door. We compute for every event the minimum, maximum and average height. Fig. 5 is a box plot illustrating the data from the trials. Given a ground truth of 180 cm, maximum feature is the closest to the ground truth. However, we note that the average height is the feature with the least variance and this means it is more consistent for the same occupant. As consistency is important for identification, we choose average height.

3.6.2. Average width

Once a person passes through the door, we compute the person's width. The width measure is independent of the position of the occupant in the door i.e. if he is closer to one edge of the door as opposed to the other, the width measurement is still the same.

3.6.3. Body-hand distance (WH)

This feature captures how close a person's hands are to her body as she walks. As the Walker swings her arms, the UR or UL sometimes measure the distance to the waist and sometimes to the arm. To compute this feature, we divide the measurements into 2 groups. The ones closer to by at most 10% of the minimum width and the others that are farther by at least 15%. We calculate the average of every group and return the difference. We do this for both sides using UL and UR and take the maximum. The reason for taking the maximum as opposed to the minimum or average is that you have people on one hand in the pocket.

3.7. Feature selection

Given the feature set generated by the system, choosing a subset of features has been extensively discussed in the literature. In this study, we use two methods to find the most successful subset of features to use to identify occupants. First, we evaluate how would a feature set composed of one feature perform and then evaluate pairs of features' accuracy. This approach is very similar to the sequential search methods Liu and Yu [33] where we start with one feature and add a second one to increase the goodness of our dataset.

In the second method, we perform feature selection using the Recursive Feature Elimination (RFE) algorithm Doak [34] whose goal is to find a subset of features that maximize accuracy and increase the robustness of the identification. Since we have 3 degrees of freedom, namely height, width and time, we derive one feature from every degree of freedom to minimize interdependence between the features. Principal Component Analysis (PCA) has been widely used as a dimensionality reduction method that leverages the variance to measure the importance of features. In fact, we use PCA to generate a new set of features that are a combination of the input features to maximize the variance Ghodsi [35]. Therefore, we first run RFE on the experimental dataset presented in Section 4. We found that the 3 most important features are girth, time and bounce. Then, we use PCA to create a new set for building our model. To validate the model, we search for the important features by forming a model for all possible feature pairs and evaluate each model's accuracy.

3.8. Occupant identification

We use clustering for occupant identification rather than rely on a supervised learning algorithm that requires training. Supervised learning would require users data for training the system and that could be cumbersome in a real-world building-scale deployment with dynamic and changing population. Hence, we rely on clustering and search for a set of features computed from the walking event that would cluster for the same user and do not cluster for different users. In other words, if a feature computed for a given person is consistent across different instances of walking trials, this feature can be used as a unique signature to identify the person. However, the feature should also meet an additional requirement: a feature computed for different people should be different. We process all sensor data stream and extract a set of features for the person. Since the system is unaware of the walker but rather tags her with a feature set, we treat this occupant identification problem as a clustering problem rather than classification, where every user will have her own cluster. In addition, the system does not need training to work and therefore is able to differentiate between people without prior information about the walkers. The way this method could help retrieve the occupant's profile to adjust the HVAC is to do the following; Have the person walk once through the door. The system will output the cluster id. Afterwards, we associate the cluster id with the profile, so whenever the person walks through any of the doors, the profile id could be retrieved.

Some of these features such as Girth and Bounce have been shown to be very consistent for the same user while other features did not meet this requirement. Other features such as Body-Hand distance did not vary much among different people and therefore are not be able to differentiate between different users. To address this issue, we decided to combine features into pairs and evaluate each pair to find the one that yields the highest identification accuracy.

There are many clustering algorithms in the literature designed with different goals in mind. We selected two algorithms, DBSCAN and Spectral clustering, which are different enough from each other. Experimenting with these two algorithms allows us to understand the performance achieved by these different clustering algorithms that were designed with different trade-offs in mind: DBSCAN is good for finding dense clusters but spectral clustering looks for connected graphs. In our preliminary work Khalil et al. [5], we explored the effectiveness of DBSCAN for clustering and showed promising result, achieving up to 95% accuracy with 20 people. We found two main weaknesses with DBSCAN::

- DBSCAN finds the number of clusters and the clusters at the same time which makes it sensitive to misidentification by wrongly estimating the number of clusters and also by finding the wrong cluster delimitations.
- DBSCAN assumes that most clusters have the same density which is not applicable to this problem as people can walk through the door at different frequencies depending on how often they come to the building. Therefore the clustering algorithm may fail in cases where the participation frequency is not homogeneous among Occupants.

While these weaknesses may not be important in some cases (largely uniform density distribution of population and events over time), we next explore Spectral clustering for scenarios in which these weaknesses may be critical (for example, when there is large variation in population and events and weather/behavior). In those cases, Spectral clustering would generally outperform DBSCAN. However, spectral clustering cannot make an accurate estimate of the number of clusters in the dataset. This is essential in our problem since we do not know at any time how many different people have walked through the door. To solve this problem, we refer to the paper by Zelnik-Manor and Perona [36]. The authors propose an improvement to Spectral clustering by self-tuning numerous parameters to improve the performance of the algorithm. Thus, these two algorithms represent unique trade-offs in design of our system.

To find the correct number of clusters, we use the method provided by Zelnik-Manor and Perona [36]. The main idea of spectral clustering is to find the connected regions rather than dense ones. To do that, the algorithm defines an affinity matrix and constructs a Laplacian Graph. Then an Eigenvector problem is solved. The idea behind finding a good number of clusters in the dataset is to sort the eigenvectors and analyze them by counting the number of eigenvectors with magnitude 1. We refer the reader to the paper by Zelnik-Manor and Perona [36] for further details. Once we have a good estimate of the number of clusters, we build our clustering model by making use of all the features defined.

4. Evaluation

We describe the setup used for evaluating the occupant identification system followed by evaluation results. The experiment protocol was approved by the University of Houston Committee for the Protection of Human Subjects.

4.1. Testbed

The sensing testbed is composed of five door frames. Fig. 1 shows one of the doorframes we built for this study. We connect an off-the-shelf Logitech C310 camera to each door to collect the ground truth.

The sensors are attached to a board that is attached to the frame as illustrated in Fig. 1c. However, this is not required by the system, it was designed for convenience but it would work similarly if the sensors were actually mounted inside the frame. The system is not susceptible to crosstalk because the sensor sampling is performed sequentially giving enough time for each sensor's signal to travel to the target and back. If installed in a wider door, this time parameters must be changed but could be computed given the dimensions of the door as shown in Section 3.2.

In order to compute the width of a person, we need both UL and UR readings to be at the same time in order to have accurate width

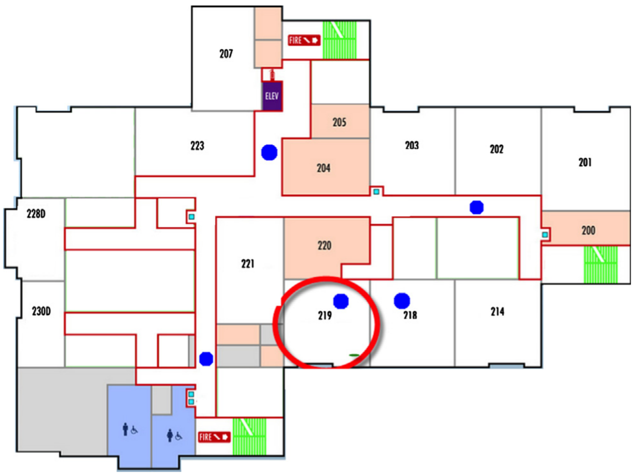


Fig. 6. Figure showing the floor plan for both Experiments. The circle shows the location of the doorframe of the first experiment and the blue dots show the location of the doorframes of the second large scale experiment. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

measure. To achieve this, we displace the sensors by 1.2 cm on the walking direction in order to account for the temporal difference between the consecutive samplings of UL and UR. This specific displacement distance (1.2 cm) is chosen to account for the temporal difference in sampling. Assuming the walker walks at an average speed of 5 km/h and having a sampling rate of 35 Hz, the Walker travels a distance of 1.2 cm every 8 ms. Also, the order to sampling the sensors is: $UT \rightarrow UL \rightarrow UR$ with a 8 ms time difference between UL and UR. Thus, this displacement is crucial because though UL and UR are sampled at different times, the width measurements of the walker though taken at different times are the same points as if the person was standing.

4.2. Single-door experiment

We conducted our experiment in lab room 219 (see Fig. 6) in Building T2 at the University of Houston for a month. We recruited students from one of the classes scheduled in the room as participants. Their age varies between 18 and 30 years old. We informed them about the purpose of the experiment and we asked them to walk naturally.

The door frame was at the entrance of the room and there was enough space for others to bypass it in case they don't want to participate. The camera was always recording. Whenever a person walks through the door, a walking event is recorded with the start and end time. Since we only keep records of when a person walks into the door frame, every minute, the video footage is processed and only the times when the walker(s) passed is extracted. We keep the video recording starting 3 s before the walking event start time. This extra time is added on purpose so that we can see the whole event when annotating the data.

Every time a person walks through the door, a stream of data {UT, UL, UR} as shown in Fig. 4a and video is captured. Each stream is then converted to the set of features extracted from the dataset. After one month which marked the end of the experiment, we annotated the data manually by looking at the video footages for every event and marked the data with the person that walked. This manual annotation was done twice by the same person and we believe it is 100% accurate.

The number of people that participated in this experiment is fifty-three. However, many of these participants passed through the door only once or twice. We discard data for those participants from our dataset as the clustering algorithm expects at least four points per cluster. We decided to take top 20 people in terms of the number of walking events. This group averaged 7.5 passes per person, with a maximum of 17 passes and a minimum of 4; Fig. 7 shows the distribution. Eleven participants were male. 9 were female. We did not measure the participants' true heights or width but it appeared to us that there was a fair distribution of body shapes.

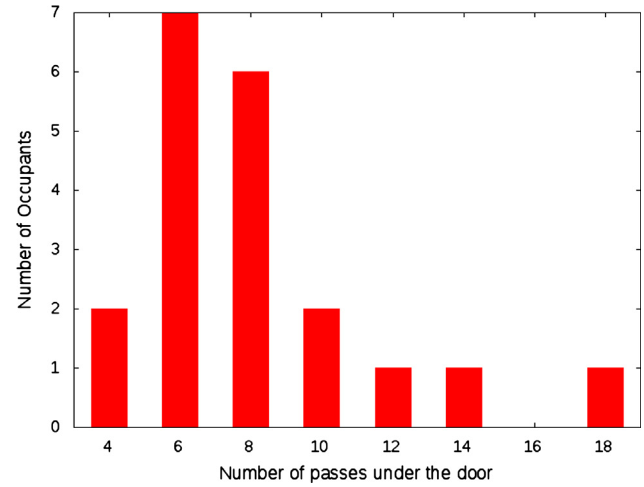


Fig. 7. Histogram showing the number of participants by number of passes under the door frame during the study.

4.3. Building-scale experiment

We performed a set of experiments involving more participants, more door frames and over a more extended period of time. We deployed five door frames on the second floor of the Technology-2 building at the University of Houston for two months. Fig. 6 show the locations of the five door frames as shown in blue dots. People were invited to participate in the experiment by simply passing through the doors and encouraged to walk as naturally as possible. There was no direct communication with the participants with the exception of banners inviting them to participate. Most of the participants were students, faculty, and staff. The experiment lasted for two months and data of over 13000 walks through the door were collected during this period. We discovered that almost 4000 events were false positives (when a person does not walk through the door and the system thinks the person did). Fortunately, it is quite easy to identify the false positives where we can observe them having no more than 2 samples per event. Also, after annotating the data, we found that over 215 participants walked more than once. All the data from the people who participated once or twice were discarded as our model cannot predict anything from a few occurrences of a person with the exception of the system understanding that this person was never seen before. Although we did not compute an exact gender mix among the participants, manual inspection of a small subset of ground truth video showed a good mix of participants from both genders.

Given the large number of walks through the door, it is not feasibly to manually tag the data from each walk by watching the video. We use automated annotation face identification technology developed by the computer vision research community. We used Facenet Schroff et al. [37] which provides an algorithm that consumes an image composed of a face and provides the end user with a vector of 256 dimensions using Deep Learning. This vector is a mapping of the Face in the Euclidean space and similar faces have a Euclidean distance of less than 1. This Algorithm is well established in the field and was tested under numerous well-known face recognition datasets achieving an accuracy of over 99.63% in the Labeled Faces in the Wild Huang et al. [38] dataset that is composed of 13,000 different faces. An implementation of Facenet is provided using Openface Amos et al. [39].

As a person walks towards the door, we extract 30–100 images containing the face, depending on the door position and how far was she initially. We collected a total of over 700,000 face images. Once, these images are collected, we clustered them such as each cluster refers to one person making sure that similar images of the same person are no farther than 1 Euclidean distance from another in the 256-dimensional feature space.

We convert every image to a 256-dim vector and note the ID of the walking event. We then search for the number of clusters in the dataset using Gaussian Mixture Model (GMM) estimation using Expectation-Maximization (EM). The parameter we vary is the distance between different images, which is initially set to 1. For each walking event, we look for the percentage of the largest cluster of the set of captured images (30–100) and compute their average. We then vary distance value parameter passed to the GMM model until we find local maxima.

For sanity check, we manually checked three images in every cluster and checked if they are of the same person. We checked 30 events. We did not find any errors in the annotation. Thus, the automatic association process is robust. We believe this annotation could scale to a larger deployment at the multi-building level or more.

4.4. Evaluation metric

Since we model our system using an unsupervised method, training data is not required. Evaluation metrics such as Purity Manning et al. [40] have been proposed in the literature. In fact, Purity is calculated as the ratio of the count of the most frequent label as a total number of labels in a particular cluster. However, this metric is unsuitable to evaluate our system because knowing how pure our clusters do not indicate how well our algorithm is able to identify occupants. We decide to evaluate it by dividing the dataset into a training and a testing dataset: 1/3 of the data for training and the remaining 2/3 for testing.

The issue with clustering is that it can create the correct number of clusters but may end up having different people in one cluster thus creating impure clusters. That said, we need to first find the feature pair that is able to generate the correct number of clusters and for those pairs, we then associate a cluster label with the class with the majority class. For instance, if a cluster is composed of instances of different classes, then we label the cluster with the most frequently observed class in the cluster. Even if some clusters may be less pure, then this would affect the overall accuracy at the testing phase. Therefore the most successful feature pair would yield the purest clusters and highest accuracy.

Ideally, we should expect two instances of the same class to belong to the same cluster. In other words, we would like two different walking events of the same person to belong to the same cluster. All the instances of possible True/False Positives/Negatives are illustrated in Table 3: We define accuracy as:

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

.

4.5. Clustering with pairs of features

We combined pairs of features in DBSCAN and for every pair using the training set and test the accuracy of every model based on a pair of features. Table 4 illustrates the result. The pair (girth, time) achieved the highest accuracy. Clustering with Girth achieved the highest accuracy in single feature clustering, so pairing it with another feature seems to increase accuracy. However, time is not accurate, but combining it with girth is most accurate because they do not commit the same mistake and people with close girth values appear to have different time values and vice versa. We should note that the pair (Bounce, Average Height) performs better than (Average Height, Average Width) which proves that it is better to use Bounce over Average Height.

In Fig. 8, the confusion matrix of clustering with the pair (Girth, Time) is illustrated. The darkness of the color indicates the percentage of the trace of the i^{th} person (P_i,$i = 0. .19$)which was recognized as the $i = j^{th}$ person (P_j,$j = 0. .19$) Using the confusion matrix, we can observe how each person is identified and misidentified. The confusion matrix shows that most of the occupants are correctly identified all the time with the exception of person 2 and 9 which seem to confused with 2 other occupants. We also observe some misidentification for 14 and 16.

Table 3
Different outcomes from clustering.

	Same cluster	Different cluster
Same Person	True Positive	False Negative
Different Person	False Positive	True Negative

Table 4
Accuracy (in %) achieved by clustering with feature pairs constructed from the features in row and column.

	Height	Width	Bounce	Time	Girth	WH
Height	84.3	89.5	89.5	90.5	93.2	86.4
Width		87.6	90.5	91.0	93.7	87.2
Bounce			88.1	87.6	94.7	89.4
Time				82.6	95.4	85.2
Girth					89.5	90.3
WH						76.9

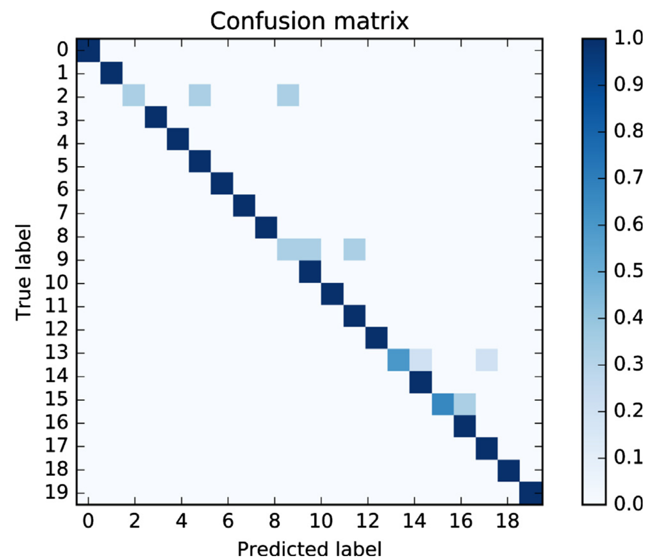


Fig. 8. Confusion matrix showing the identification and misidentification of the occupants and how which occupants was identified as another.

4.6. Model scaling with increasing number of occupants

To evaluate the performance of the technique as a function of the number of users, we calculated the accuracy of the technique in different population sizes. Fig. 9 shows the accuracy as a function of the number of occupants. As expected, the accuracy decreases as the population size increases. We can observe that using the pair (Girth, time), we are able to differentiate between people with an accuracy of 97% in the case of 5 people. Our system achieves a slightly higher accuracy for the same number of people compared to systems such as Doorjamb Hnat et al. [20] which achieves 93% accuracy and Pan et al. footstep induced identification system which achieves 96.5% accuracy Pan et al. [27]. Fig. 9 shows the plot of accuracy as a function of the number of occupants for 2 clustering models. We show how the accuracy of clustering with (Girth, Time) changes as the number of occupants increases. We also compare it to clustering with Average Height. We

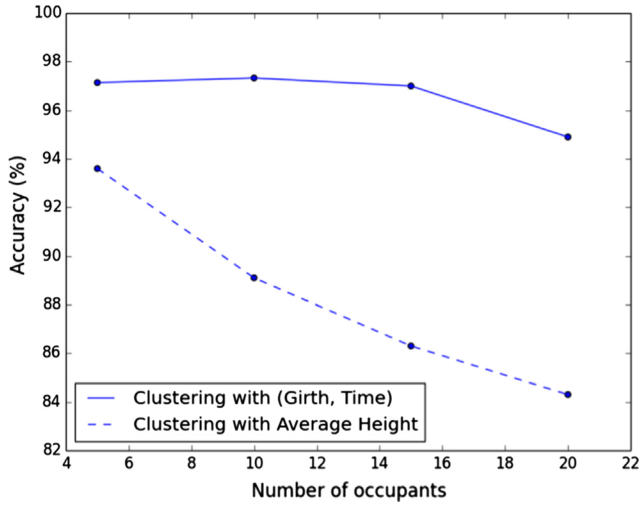


Fig. 9. Plot showing identification accuracy of clustering with (Girth, Time) and clustering with Average Height as a function of the number of occupants.

observe that clustering with the (Girth, Time) not only achieves higher accuracy but does maintain higher accuracy compared to clustering with average height when we increase the number of occupants.

4.7. Evaluating spectral clustering

We built the Spectral clustering model using 2/3 of the data as in the experiment with DBSCAN. We evaluated the accuracy on rest of the data. The model achieved an accuracy of 94.6% which is slightly lower than clustering with DBSCAN where we achieved 95.5%. Though it appears that DBSCAN outperforms the spectral clustering, we note that the dataset is not large enough to show the weaknesses of DBSCAN mainly its inability to cope with varying densities throughout different clusters.

4.8. Comparing spectral clustering and DBSCAN with increasing number of occupants

In this section, we compare the performance of DBSCAN and Spectral clustering using the new dataset as we vary the number of occupants. We select subsets from the dataset with populations of 5, 10, 15, 20, 25, 30 up to 50 with increments of 5. We evaluate how the performance of clustering changes as the number of occupants increase. First, we select subsets from the larger dataset such that we match the population size we target. To draw a subset, we do the following:

1. Arbitrarily select a person from the dataset and select all her walks
2. Repeat operation until we meet the subset population size.

Then we divide each subset into training and testing set as done previously and evaluate the performance of both models with regards to the size. In Fig. 10, we plot the model accuracy of both models as we vary the population size. We observe that the spectral clustering performs better than DBSCAN at larger scale.

4.9. Statistical significance of achieved model accuracy

We have shown that we achieved 95% accuracy in clustering in our dataset. We also used 3-fold cross validation to show that this result generalizes. To further make sure that the results are statistically sound and generalizable, we conducted further evaluation on a bigger and different data set that we generated using the statistical bootstrapping technique. It consists of random sampling method with data replacement. In this test, the goal is to generate subset data different than the

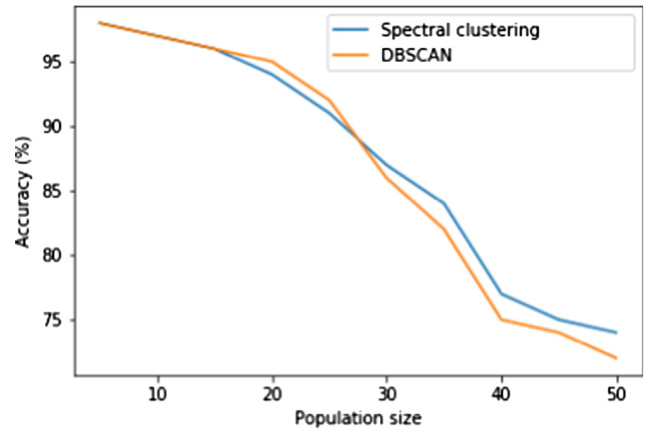


Fig. 10. Plot comparing Spectral Clustering and DBSCAN with three features as we increase the population from 5 to 50 people.

original dataset and compute achieved accuracy. Each subset is used to evaluate the model's performance. The process goes as follows:

1. Arbitrarily draw with replacement a subset from the original dataset
2. Divide subset into training and testing with respective proportions of 2/3 and 1/3
3. Measure Accuracy of model trained on the training set and tested on the testing set
4. Repeat the process for 1000 times.

To determine the statistical significance of the 95% accuracy on our dataset and its generalization, we created 1000 subsets with replacement. For each subset, we calculated the model's accuracy using the same method described in Section 4.4. We plotted the histogram of the distribution of the model's accuracies computed over the 1000 subsets. From Fig. 11, we measured the mean and standard deviation which are respectively 94.5% and 1.4%. This means that our accuracy of 95% is statistically sound and achievable.

4.10. Accuracy for different population height-width distribution

We have shown that our system achieves 95% accuracy for a group of up to 20 people. We have also shown that we would expect such accuracy to degrade as the population size increases. However, the model's accuracy is not only affected by the population size but also by the height-width distribution in the dataset.

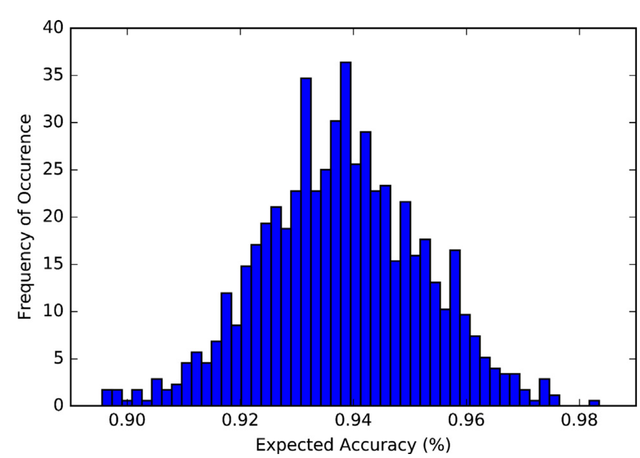


Fig. 11. Histogram showing the expected accuracy distribution of the model. This accuracy distribution is generated using the statistical bootstrapping method.

We run a simulation where we select subsets of the occupants from our large scale deployment to match a target height, width spread to evaluate how our model performs for different groups and population. Some countries have more height, width variance than others and therefore the expected model's performance may differ.

To run this simulation, we define a height and width spread interval. This spread refers to the variance in a population. We set the height and width intervals to be respectively [0,15 cm] and [0,5 cm]. This defines a 2-dimensional spread space. We then define points by incrementally moving with a step of 0.5 cm. Therefore we select all (height, width) points between (0,0) and (15,5) by incrementally moving with a step of 0.5 cm. Each point in this space is referred to as a spread target. For every spread target, we simulate 10 different populations whose variance is the spread target. To do that, we perform the following operation.

1. Choose target spread $s = [s_h, s_w]$
2. Arbitrarily select a person from the population. This person x would at the center of the dataset.
3. Define a Bivariate Normal Distribution $X \sim N(\mu, \Sigma)$ where $\mu = [x_h, x_w]^T$ and $\Sigma = s$ which is the target spread.
4. Arbitrarily select another person y with replacement. Calculate average height and width from all walks of user we call (y_h, y_w).
5. Calculate probability of point being generated by the Normal distribution $p = P(X = y)$
6. We duplicate the data belonging to person y by a factor of $p*1000$
7. repeat for another 18 people.

The reason we duplicate the data is to be able to match the target spread as explained in Step 5. All the retrieved people's data is duplicated to match the target spread. The extent to which they are duplicated is defined by the probability that the person belongs to the defined distribution. Therefore, the farther the person's height-width from the mean of the distribution, the smaller is the duplicate factor. With this process, even though we select people of various height and width arbitrarily, we are able to match our target spread by varying the duplication factor according to the probability of being observed in such spread distribution.

In total, we generated ten populations per spread and we have 30 spread points. We chose ten populations instead of simply one because the mean of the bivariate distribution is dependent upon the first arbitrarily sampled person. So if the first person chosen is very tall or very short, then the other randomly sampled people would more likely be farther on average and therefore would be duplicated with a smaller factor and therefore the instances of peoples' walks would be disproportionately distributed. To mitigate this issue, we simulate ten datasets per target spread instead of one. This will mitigate the risk of sampling the person with "anomalous" physical shapes. There is a risk to chose the same person as the first for different distributions but since our population of 170 is large enough, this would rarely happen.

Each simulated dataset is divided into two subsets: training and testing set of sizes 2/3 and 1/3 respectively. We calculate the accuracy for every dataset which are then grouped by spread target. For every spread target, we compute the mean and standard deviation of the accuracy over the datasets associated with it.

In Fig. 12, we plot a heat map of the mean accuracy as a function of height and weight spread. Note that the original model based on DBSCAN did not make use of height to construct the model. However, we observe from Fig. 12 that as the height spread increases, which indicates that the model gives more weight to the height as the spread increases.

In Fig. 13, we plot a heat-map of the standard deviation of accuracy as a function of height and weight spread for a certain population spread. We observe that as the accuracy increases, the variance decreases. Even though the figure depicts that the variance is higher towards the middle, it is an absolute value and the lower spread varies as much but since the absolute accuracy is low then it varies less. So we can conclude that the model's accuracy becomes more stable as the spread increases.

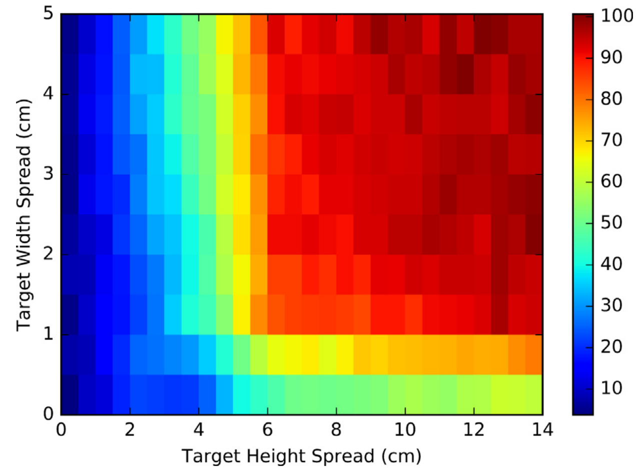


Fig. 12. Heat map showing the expected accuracy of the model as a function of the height and width spread as observed in the dataset.

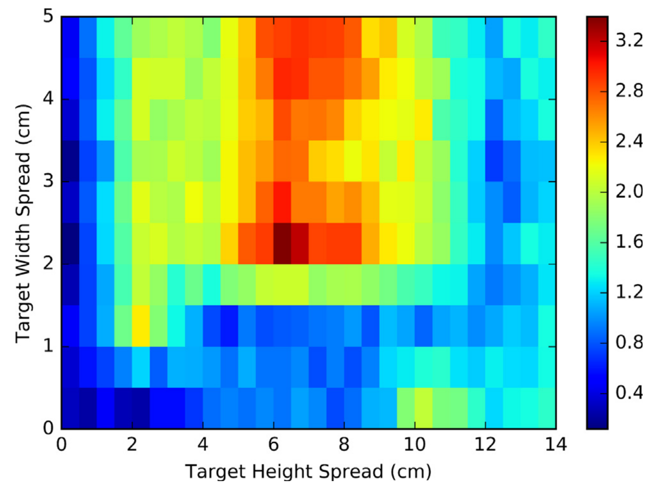


Fig. 13. Heat map showing the expected standard deviation of the expected accuracy of the model as a function of the height and width spread as observed in the dataset.

4.11. Width sensors positioning evaluation

The objective of this section is to evaluate if displacing the sensors UL and UR (as depicted in Fig. 1b) or aligning them for measuring the Walker's width is useful. We conducted an experiment where one person performed 6 passes through the doorway with the sensor aligned and repeated the same procedure having the sensors displaced.

With the width sensors were aligned, the occupant performed 6 passes under the door with the hands raised. The hands are raised to not bias the width measurements. The same operation is repeated but with the sensors, UL and UR displaced by 1.2 cm as explained. The same occupant walked for 6 times. Fig. 14 illustrates the result of the experiment.

We can see that displacing the sensors improves the accuracy of width measurement. However, if we install the sensors in an aligned fashion, we get more variation in width measurements. In addition, the average width for all 6 passes is 40.25 cm and 38.50 cm respectively for the displaced sensors and aligned whereas the ground truth is 40 cm.

The reason why the aligned sensors generate more variation is that since we sample sequentially, after sampling from UR, the Walker's

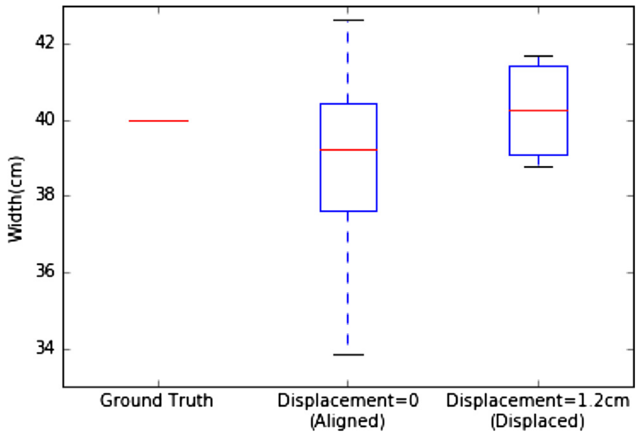


Fig. 14. Box plot comparing width measurements when the sensors UL and UR are aligned Vs Displaced.

position has changed by the time sample with UL. Therefore, both readings do not refer to the true width and depending on how fast the person walks or the direction (for example if he gets closer to one side as he walks), the width measurement will vary more.

4.12. Robustness to walking angle

Girth is the most successful feature in differentiating between occupants as is shown in Table 4. One of the strengths of girth as a feature is its ability to not change the direction of the walker. In fact, all width measure (minimum, maximum and average) suffer from the direction of the walker. The measures change drastically if a walker walks straight towards the door or at an angle relative to the door. However, girth does not appear to suffer from the direction because it represents a circumference of the person's waist and therefore is insensitive to the angle of walking when we make the measurements. We conducted an experiment where one Walker walks through the door at the angles relative to the door of 0,45 degrees and 90 degrees. For each angle, the Walker passed six times. For each pass, we computed the Walker's girth having the sensors UL and UR displaced. Fig. 15 shows a box plot of the girth computed for every pass and every angle. We observe that the mean girth does not vary much regardless of the angle at which you walk. Also, most of the girth measures fall within less than 1 cm away from the mean. We conclude that the girth is not sensitive to the angle at which the person walks which make it a practical feature to differentiate between walkers.

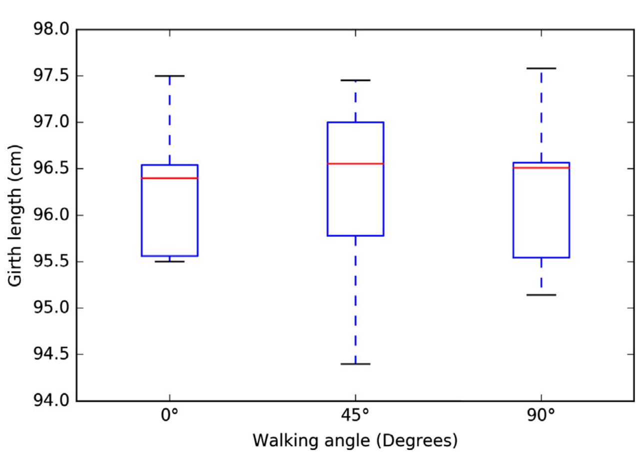


Fig. 15. Box Plot showing Girth sensing distribution for three walking multiple angles relative to the door.

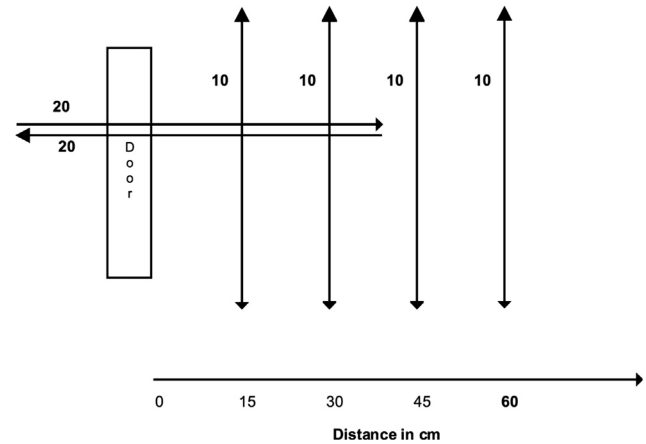


Fig. 16. Figure showing the experimental setup of walking scenarios to determine the accuracy of walk-through event detection. The arrows shows the directions of the walks and the number of times each direction was taken as noted by a number next to each arrow.

4.13. Accuracy of walk-through detection

To evaluate the performance of detecting walking events, we conduct an experiment in which we have a subject walk through the door 40 times and along side of the door 40 times as follows. First, we have the subject walk 20 times through the door in one direction and 20 times through the same door in the opposite direction to emulate a scenario of walking into a room and out of the room. Then, we have the same subject walk along the door (not through) ten times each at a distance of 15 cm, 30 cm, 45 cm, and 60 cm from the door to emulate a scenario in which the subject passes through a hallway without entering a room but possibly triggering the ultrasonic sensors and generating false positives. Fig. 16 shows the experimental setup and the different walking scenarios.

We achieved a precision of 100% and a recall of 81.6% in this experiment. There is a significant difference between precision and recall. In fact, Table 5 shows the individual performance of each scenario depicted in Fig. 16. Note Table 5, we measure the correct detection and in the case of walking along the door, a correct detection is when the walker did not walk through the door and door did not trigger the event. There is a strong justification for such results. Our system samples at 35 Hz, therefore it is almost impossible to walk through the door without affecting the measured distance triggering the walking event. Secondly, if we walk close enough to the door but not through it, the ultrasonic wave will reach the walker and be reflected back to the receiver triggering the start of the walking event. However, these false positives are easily detected because when UR and UL measuring the person's width, it is less than 20% of the width of the door, which is rare when a walker walks through the door. Typical width observed for a real walk-through is in the range 26%-52%. Thus, we can use a simple filter to eliminate these false positives.

Table 5
Table showing the accuracy of correct walk detection performance when walking through the door and along the door at different distances as illustrated in Fig. 16.

Walk	Accuracy of Correct detection (%)
Through the door	100
Along the door: 15 cm	30
Along the door: 30 cm	60
Along the door: 45 cm	100
Along the door: 60 cm	100

4.14. Accuracy of walking direction

In an experiment with 30 walk-throughs (15 times each direction), we found that the system can determine the direction of walk-through at 90% accuracy for walking speed faster than 5 km/h and 100% accuracy for slower walk-throughs.

4.15. Evaluating energy savings from identification

The result of occupant Identification can be used to retrieve people's temperature preference profiles as they walk through the door and set the temperature accordingly. In current commercial buildings, the temperature is set cool enough to accomodate most people but that leads to many people feeling too hot or too cold.

A better approach is to extract the people's preference and set the temperature using one of the following strategies for the population currently in the building:

- Set the temperature to the mean temperature
- Set the temperature to the median temperature

There are other strategies discussed in the literature for setting the temperature for a group of people but for the purpose of evaluating how setting the temperature according to the automatically retrieved temperature preference of the occupants can make the room more energy efficient, we will consider only the mean and the median. We note that in most buildings in the United States, the temperature is usually set to around 22 °C following the ASHRAE standard for thermal comfort setting ASHRAE [42].

In this study, we focus on the percentage of people who feel comfortable at various temperatures (as shown in Fig. 17) and assume that for that portion of people, that is their ideal temperature setting. We rely on an experiment performed by Kim et al. [41] in a school in Australia. They collected over 3356 samples where For every temperature interval, we obtain the number of people who feel comfortable at that temperature interval and assign the temperature preference to be that preference of all those people. This makes up a population of people appreciating different temperatures. In our paper, we consider different subsets of these people of populations of up to 20 people sitting in the same room with their individual temperature preference. We would like to set the temperature for the room using the two strategies discussed. Given the proportions of people of each temperature setting, we arbitrarily select 20 each with from one of the temperature setting groups while respecting the proportions of each temperature. Then for each group we measure the median and mean temperature which we would set in the room. We redo this operation for 1000 times to measure the distribution of temperatures that would be set in the room.

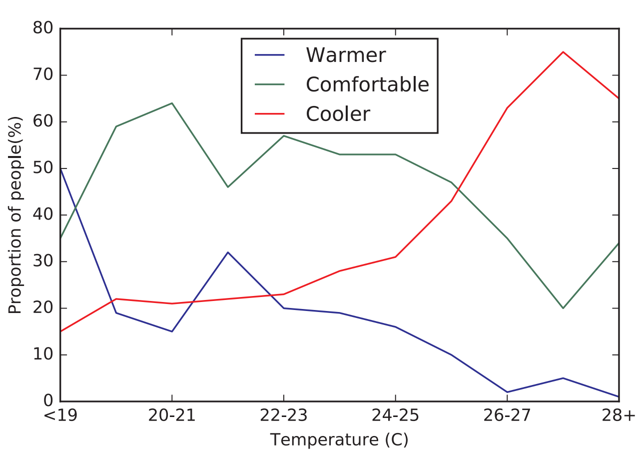


Fig. 17. Plot showing the change of percentage of people feeling cold, warm or like the current temperature as we vary the temperature in the room. This data is based on study performed by Kim et al. [41].

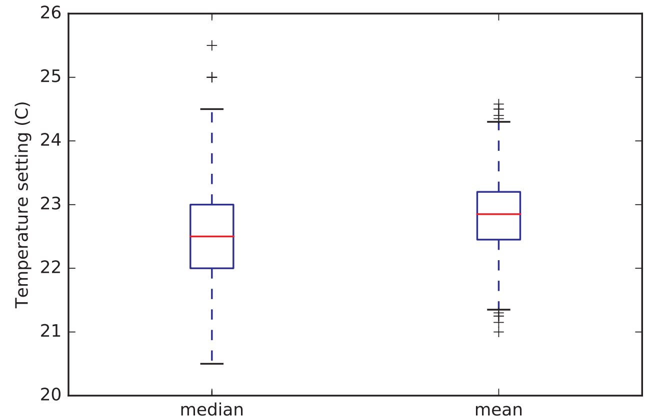


Fig. 18. Boxplot showing the median and mean temperature distribution when choosing 1000 subgroups each of size 20 people of different temperature preferences.

Clearly these strategies will not make everyone happy and there are better ones discussed in literature, but the user communicates with the room as he walks through the door using our identification as opposed to current commercial buildings where there is either no user input in temperature selection and setting (controlled by the central plant) or there are techniques such as Thermovote Erickson and Cerpa [14] requires active intervention from the user on a mobile device or control in the building. Compared to these techniques, we do not require the users to carry any device or push buttons on the physical controls in the room while we are able to customize the room temperature automatically depending on the occupant population in the room. However, in all 1000 trials the preferred temperature setpoint for the group of occupants is around 23 °C on average for both the mean and median strategy as shown in the boxplots in Fig. 18. Thus, our system would keep the temperature in the rooms at least 1 °C warmer in the summer compared to the standard baseline of 22 °C as recommended by ASHRAE [42]. Results from Hoyt et al. [43] and Ghahramani et al. [44] suggest HVAC-related energy savings of 7–15% and 4–10% when we keep the rooms warmer by 1 °C.

5. Discussions

We discuss a set of challenges that we face in real world deployment of the door frame. We also discuss ways to tackle these issues.

5.1. Multiple entries

In this study, we assume that only one occupant passes through the door at a time. The features extracted assume that there is only one person. If more than one person walks through the door simultaneously, the whole stream will be seen as one person with an unusual width and time. To deal with multiple simultaneous entries per door, our system needs to disaggregate the data into multiple walkers. In the case of large doors that are designed for multiple entries, we can extend them by adding an extra UT to capture the second walker.

5.2. Higher number of users

Previous methods have scaled their system up to 5–6 people whereas we were able to accurately identify up to 20 people. However, as the number of users increases past 20, the identification accuracy would decrease because the similarity between different features tends to become more probable. However, Height was not used for identification and therefore could be used to push further the number of people by grouping them by height and considering them as different groups.

5.3. Impact of belongings

A person carrying a backpack or a woman holding a purse will be reflected in the data and drive the overall identification accuracy down. However, the bias arising from extra objects follows a pattern and can be removed. For instance, a purse can be detected by noting a higher body-hand distance on the hand carrying the purse compared to the other hand. In the case of a backpack, the height sensor will show a unique pattern showing a person carrying a backpack. These cases could be solved individually by identifying them first and pruning the data from such a bias.

5.4. Impact of walking pattern

If a person walks faster or slower, the data stream length will be impacted because she will spend more/less time under the door. However, in our studies we make two observations: (1) speed among the same person rarely varies outside of the average mean time ± sampling time, (2) it was observed that only three participants had different speeds. Sometimes the features from one subject created multiple (in our data up to three) clusters corresponding to different walking speeds and patterns for that subject. However, each cluster always was associated with a single subject. Thus there was never an ambiguity in mapping from a cluster to an individual. We believe this result holds in general but we have not performed experiments in other settings to confirm.

5.5. People with disability

Our current deployment does not account for people with special disability. For instance, people with wheelchairs would appear as having the same width and height and may fail to distinguish between them. Moreover, People with crutches usually walk slower. However, we may be able to detect such cases by observing a more square shaped girth rather than a regular oval shaped one.

5.6. Location of door frame

Having the door frame at a hall will force people to walk differently than having the door frame right by a corner where people have to turn. We guess that identification will be consistent for the same door. But taking the same clusters generated for one door and using them in another door may not yield the same identification accuracy observed at the previous door even if the participants are the same.

5.7. Low power sensing

In the current setup, the sensors sample continuously and independently of whether there is an occupant. However, this would pose a problem in a real-world deployment because the current setup is energy inefficient. We suggest adding a passive motion detector and only when a person is detected, we activate the ultrasonic sampling. This would make the system much more energy efficient because only when a person approaches the door that we start sampling.

5.8. Multi-people door

In this work, we have deployed our system in single person doors. In other words, we assumed the door is able to fit no more than one walker at a time. Though the vast majority of doors only allow one person at a time, most buildings contain a few wide doors that allow more than one person to walk through it. Deploying our solution in such door will not work out of the box and we would want to consider more sensors and at different placements to achieve that. Also, there are other door types that we didn't consider such as circular doors which turn and let go more people at once. Even though our system would not fit these various door types, we believe that the vast majority of doors in buildings and homes are single-person doors and our system is designed for such doors.

5.9. Adding more ultrasonic sensors

Adding more sensors will likely improve the performance of the system, but the significance of the improvement is questionable. Having these three sensors, we are able to capture three dimensions: height, width and time. Adding more width sensors at a different placement will probably be in a way correlated with the original ones (and will capture only marginally different aspects of height, width, and time we already capture) whereas the top and side sensors are uncorrelated which makes them useful. A more rewarding approach may be to use a completely different type of sensor to capture aspects of motion ultrasonic sensors are not able to capture.

6. Conclusions

Occupant identification plays an important role in smart building management systems. In particular, non-intrusive identification is expected to enable many new applications in which buildings can be customized to the occupants preferences making the buildings more comfortable and energy-efficient. Non-intrusiveness will enable buildings to actively identify occupants without engaging them with a device.

This paper developed and implemented a non-intrusive and device-free ultrasonic-based doorway system that identifies occupants in commercial building applications. The door is composed of ultrasonic sensing components that collect data about the height and width of a passing by an occupant, then the system computes body shape including height, width, and girth as well as movement. The sensors generate a stream of data whenever an occupant passes under the doorway and the features are extracted from each stream. Using the feature sets, the doorway system uses a clustering machine learning algorithm to determine the identity of the occupant. The developed system is able to identify a person among a population of 20 people with an accuracy of 95%.

The system is subject to uncertainty that is due to the fact that some people may exhibit similar features and the clustering algorithm may confuse some of them. The paper investigated the combination of different features and measured their impact on identification. The combination that was able to distinguish most between people is girth and walking speed, with an accuracy of 95%. The results also showed that bounce is more accurate in clustering occupants. However, bounce did not contribute significantly to increasing the accuracy when included as a third feature for clustering. However, in the initial experiment, height was not a strong contributing factor to the model. On the other hand, Spectral clustering was able to make use of all the features and we observed that height seems to become more important as the height spread in the population increases. We evaluate the statistical significance of such performance and how it would generalize in different populations of different height-width distributions. We finally evaluate the energy savings potential of identification and how it can increase thermal comfort by raising the temperature in the summer while saving up to 15% of energy.

References

[1] Erickson VL, Achleitner S, Cerpa AE. POEM: power-efficient occupancy-based energy management system. Proceedings of the 12th international conference on Information processing in sensor networks. ACM; 2013. p. 203–16.
[2] Erickson VL, Cerpa AE. Occupancy based demand response HVAC control strategy. Proceedings of the 2nd ACM workshop on embedded sensing systems for energy-efficiency in building. ACM; 2010. p. 7–12.
[3] Zhang L, Lam AH-y, Wang D. Strategy-proof thermal comfort voting in buildings. Proceedings of the 1st ACM conference on embedded systems for energy-efficient

buildings. ACM; 2014. p. 160–3.
[4] Ghahramani A, Jazizadeh F, Becerik-Gerber B. A knowledge based approach for selecting energy-aware and comfort-driven HVAC temperature set points. Energy Build 2014;85:536–48.
[5] Khalil N, Benhaddou D, Gnawali O, Subhlok J. Nonintrusive occupant identification by sensing body shape and movement. Proceedings of the 3rd ACM international conference on systems for energy-efficient built environments. ACM; 2016. p. 1–10.
[6] Pérez-Lombard L, Ortiz J, Pout C. A review on buildings energy consumption information. Energy Build 2008;40(3):394–8.
[7] Conte G, De Marchi M, Nacci AA, Rana V, Sciuto D. BlueSentinel: a first approach using iBeacon for an energy efficient occupancy detection system. In: BuildSys'14; 2014.
[8] Färber I, Günnemann S, Kriegel HP, Kröger P, Müller E, Schubert E, et al. On using class-labels in evaluation of clusterings. In: MultiClust: 1st international workshop on discovering, summarizing and using multiple clusterings held in conjunction with KDD, 1; 2010.
[9] Brandemuehl MJ, Braun JE. The impact of demand-controlled and economizer ventilation strategies on energy use in buildings. ASHRAE Trans 1999;105:39.
[10] Erickson VL, Carreira-Perpiñán MÁ, Cerpa AE. OBSERVE: occupancy-based system for efficient reduction of HVAC energy. 2011 10th International conference on information processing in sensor networks (IPSN). IEEE; 2011. p. 258–69.
[11] Hoyt T, Lee KH, Zhang H, Arens E, Webster T. Energy savings from extended air temperature setpoints and reductions in room air mixing. In: International conference on environmental ergonomics 2009; 2005.
[12] Jazizadeh F, Ghahramani A, Becerik-Gerber B, Kichkaylo T, Orosz M. User-led decentralized thermal comfort driven HVAC operations for improved efficiency in office buildings. Energy Build 2014;70:398–410.
[13] Karjalainen S, Koistinen O. User problems with individual temperature control in offices. Build Environ 2007;42(8):2880–7.
[14] Erickson VL, Cerpa AE. Thermovote: participatory sensing for efficient building hvac conditioning. Proceedings of the fourth ACM workshop on embedded sensing systems for energy-efficiency in buildings. ACM; 2012. p. 9–16.
[15] A. Standard, Standard 55-2010:Thermal Environmental Conditions for Human Occupancy; ASHRAE, Atlanta USA.
[16] Lanitis A, Taylor CJ, Cootes TF. Automatic face identification system using flexible appearance models. Image Vis Comput 1995;13(5):393–401.
[17] Hrechak AK, McHugh JA. Automated fingerprint recognition using structural matching. Pattern Recogn 1990;23(8):893–904.
[18] Tisse Cl, Martin L, Torres L, Robert M et al. Person identification technique using human iris recognition. In: Proc vision interface; 2002. p. 294–9.
[19] Pan S, Chen A, Zhang P. Securitas: user identification through rgb-nir camera pair on mobile devices. In: SPSM'13; 2013.
[20] Hnat TW, Griffiths E, Dawson R, Whitehouse K. Doorjamb: unobtrusive room-level tracking of people in homes using doorway sensors. In: SenSys'12; 2012.
[21] Srinivasan V, Stankovic J, Whitehouse K. Using height sensors for biometric identification in multi-resident homes. Pervasive computing. Springer; 2010.
[22] Jenkins J, Ellis C. Using ground reaction forces from gait analysis: body mass as a weak biometric. International conference on pervasive computing. Springer; 2007. p. 251–67.
[23] Ranjan J, Yao Y, Whitehouse K. An RF doormat for tracking people's room locations. In: Ubicomp'13; 2013.
[24] Subbu KP, Thomas N. A location aware personalized smart control system. In: IPSN'14; 2014.
[25] Zhao W, Chellappa R, Phillips PJ, Rosenfeld A. Face recognition: a literature survey. ACM Comput Surv (CSUR) 2003;35(4):399–458.
[26] Kale A, Sundaresan A, Rajagopalan A, Cuntoor NP, Roy-Chowdhury AK, Kruger V, et al. Identification of humans using gait. IEEE Trans Image Process 2004;13(9):1163–73.
[27] Pan S, Wang N, Qian Y, Velibeyoglu I, Noh HY, Zhang P. Indoor person identification through footstep induced structural vibration. In: HotMobile'15; 2015.
[28] Jenkins J, Ellis C. Using ground reaction forces from gait analysis: body mass as a weak biometric. International conference on pervasive computing. Springer; 2007. p. 251–267,.
[29] Liau W-H, Wu C-L, Fu L-C. Inhabitants tracking system in a cluttered home environment via floor load sensors. IEEE Trans Autom Sci Eng 2008;5(1):10–20.
[30] Elrod S, Shrader E. Smart floor tiles/carpet for tracking movement in retail, industrial and other environments. US Patent App. 11/236,681; 2005.
[31] Zhang J, Wei B, Hu W, Kanhere SS, Tan A. Human identification using WiFi signal. 2016 IEEE International conference on pervasive computing and communication workshops (PerCom Workshops). IEEE; 2016. p. 1–2.
[32] Ogden CL. N.C. for Health Statistics (US), et al. Mean body weight, height, and body mass index: United States 1960–2002, Department of Health and Human Services; 2004.
[33] Liu H, Yu L. Toward integrating feature selection algorithms for classification and clustering. IEEE Trans Knowl Data Eng 2005;17(4):491–502.
[34] Doak J. An evaluation of feature selection methods and their application to computer security, University of California, Computer Science; 1992.
[35] Ghodsi A. Dimensionality reduction a short tutorial, Department of Statistics and Actuarial Science, Univ. of Waterloo, Ontario, Canada.
[36] Zelnik-Manor L, Perona P. Self-tuning spectral clustering. In: Advances in neural information processing systems; 2005. p. 1601–8.
[37] Schroff F, Kalenichenko D, Philbin J. Facenet: a unified embedding for face recognition and clustering. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2015. p. 815–23.
[38] Huang GB, Ramesh M, Berg T, Learned-Miller E. Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Tech rep 07-49, University of Massachusetts, Amherst; 2007.
[39] Amos B, Ludwiczuk B, Satyanarayanan M. OpenFace: a general-purpose face recognition library with mobile applications. Tech Rep, CMU-CS-16-118, CMU School of Computer Science; 2016.
[40] Manning CD, Raghavan P, Schütze H. Introduction to information retrieval. New York, NY, USA: Cambridge University Press; 2008. ISBN 0521865719, 9780521865715.
[41] Kim J, Deuble M, Candido C, de Dear R. Thermal comfort in a large property portfolio: what can we do before surrendering to AC?, Available from: researchgate.net [accessed June 2015].
[42] A. ASHRAE, Standard 55-2004, thermal environmental conditions for human occupancy, atlanta: american society of heating, refrigerating, and air-conditioning engineers, Inc., USA.
[43] Hoyt T, Arens E, Zhang H. Extending air temperature setpoints: simulated energy savings and design considerations for new and retrofit buildings. Build Environ 2015;88:89–96.
[44] Ghahramani A, Zhang K, Dutta K, Yang Z, Becerik-Gerber B. Energy savings from temperature setpoints and deadband: quantifying the influence of building and system properties on savings. Appl Energy 2016;165:930–42.
[45] Zeng Y, Pathak PH, Mohapatra P. WiWho: WiFi-based person identification in smart spaces. 2016 15th ACM/IEEE international conference on information processing in sensor networks (IPSN). IEEE; 2016. p. 1–12.