# AURES: A Wide-Band Ultrasonic Occupancy Sensing Platform

Oliver Shih Patrick Lazik Anthony Rowe
Electrical and Computer Engineering
Carnegie Mellon University, Pittsburgh PA, USA
{oshih, plazik, agr}@andrew.cmu.edu

## ABSTRACT

In this paper, we present a platform designed for low-power real-time sensing of the number of occupants in indoor spaces. The system transmits a wide-band ultrasonic signal into a room and then processes the superposition of the reflections recorded by a microphone. The system has two modes of operation, one for presence detection and one for estimating the number of occupants in a region. The presence detection uses the difference between multiple transmissions in succession with a set of general classifiers that make a binary decision about if the room contains occupants. We then use a semi-supervised learning approach based on Weighted Principal Component Analysis (WPCA) that requires minimal training data to estimate the number of occupants. We also present the design of an energy harvesting embedded platform and demonstrate that our algorithm can continuously execute using energy harvested from indoor solar panels. The platform has a dual Bluetooth Low-Energy and 802.15.4 interface to communicate with a gateway or nearby mobile phone that runs an interface that aids in collecting training data. We evaluate our algorithm on a wide-variety of indoor spaces as well as benchmark the hardware in terms of sampling rate given an energy budget. On more than three weeks of data, we see that we can detect motions with an average of 85% recall rate and perform occupancy counting with an average error of 10% in terms of maximum occupancy.

## CCS Concepts

•**Computer systems organization** → **Sensors and actuators;** *Sensor networks; Embedded hardware;* •**Hardware** → PCB design and layout; •**Computing methodologies** → Classification and regression trees;

## Keywords

Occupancy detection, ultrasonic sensing, machine learning



*BuildSys '16, November 16-17, 2016, Palo Alto, CA, USA*
© 2016 ACM. ISBN 978-1-4503-4264-3/16/11…$15.00
DOI: http://dx.doi.org/10.1145/2993422.2993580

## 1. INTRODUCTION

Heating, cooling, and ventilation of buildings represents about 17% of all energy used domestically, equivalent to about 16.7 QBtu ("quads") of energy annually. It has been shown that HVAC controls that are adaptive to fluctuations in occupancy density and distribution should allow optimization of air distribution and provide substantial energy savings of more then 1 QBtu annually [1]. In order to support applications like optimizing Variable Air Volume (VAV) control or perform real-time crowd detection for restaurant crowd and lines, we need to be able to sense not only motions in a space, but also the number of people. There are currently many approaches for measuring occupancy in spaces including: passive infra-red (PIR) sensors, ultrasonic ranging sensors, microwave sensors, smart cameras, break beam sensors and laser range-finders. These devices span a wide spectrum of cost and performance. Lower-cost solutions, like PIR and narrow-band ultrasonic ranging sensors, are typically error-prone and usually only detect binary occupancy values rather then estimating the number of people in a space. More expensive sensors like smart camera systems tend to require sophisticated site-specific installation and calibration. They also require wall power, pose privacy risks and are often hindered by obstructions.

We propose an active ultrasonic sensing technique that uses changes in a room's acoustic properties when occupied to estimate the number of people. Frequency dependent models of reverberation and room capacity are often used when designing auditoriums and concert halls. In our previous work [2], we leverage this property by using measured changes in the ultrasonic spectrum reflected back from a wide-band transmitter to estimate occupancy. A centrally located beacon transmits an ultrasonic chirp and then records how the signal dissipates over time. By analyzing the frequency response over a wide-band $1kHz$ chirp's bandwidth at a few known occupancy levels, we are able to extrapolate the response as the number of people in the room changes. We showed that it is possible to train a system to capture the nuances of a particular space with little training data (two points). One of the key techniques for maintaining performance even when features of the environment change, like when furniture moves, is to let the system periodically recalibrate on a known occupancy level. In this paper, we show an approach for accurately classifying when the room is empty and show how this can be used to periodically adjust the model of occupancy as room features change. We also both improve our previous occupancy estimation algorithm and design a supporting embedded platform called the Adaptive Ultrasonic Response Estimation Sensor(AURES), that is a stand-alone energy-harvesting transducer with wireless communication. AURES consists of an ARM Cortex M3 microcontroller with indoor solar energy harvesting, a wide-band ultrasonic transmitter, an ultrasonic MEMs microphone, 802.15.4 radio, and Bluetooth Low-Energy (BLE) wireless interface for communicating with nearby smartphones and gateway nodes. The installer can simply attach the sensor to the ceiling in a room and then label a handful of calibration points on a smartphone application to configure the system. We use a semi-supervised approach where we cluster data unsupervised over an extended period and then ask users to label a subset of the clusters.

The system has two main phases of operation: presence detection and occupancy counting. In the first phase we detect the presence of people using three different classifiers and in the second phase we estimate the number of occupants using a trained regression model. We use multiple transmissions of a single frequency tone in order to measure Doppler shift, changes in signal amplitude and changes in signal energy. To estimate the number of people, we utilize a wide-band chirp and a spectral approach that improves upon our previous approach to capture more sophisticated room features. If the room is classified as empty in the first phase, then the received signal in the second phase is used to re-calibrate the trained model for occupancy estimation in order to adapt to changes in the environment. The presence detector combines our three classifiers to identify both sudden movements and static changes in the presence of occupants. These presence features are general enough to be used in different indoor environments without training on known data or assuming prior knowledge. To estimate the number of occupants, we apply a semi-supervised machine learning approach that models the characteristics of the room under multiple loads as previously described in [2]. We improve the estimation accuracy of our original approach by using features that better capture room absorption over time by dividing the received signal into segments before transforming it into the frequency domain. This allows us to train multiple amplitudes across each segment rather than averaging them into a single window. Since the consistency of the background environment among sparsely labeled data points often determines the performance of the trained model, we also adopt a new technique to help the training process cope with noise in the training dataset.

One of the main challenges when installing occupancy sensors is the cost of running power and data cables. Many motion detectors can wirelessly transmit data to gateway nodes within a building. Some of these sensors can even operate for extended periods (years) off of batteries. Unfortunately these systems only detect motion and cannot count the number of people in space. More sophisticated occupancy estimation sensors like PIR arrays or smart cameras currently consume too much power to make prolonged battery operation feasible. Unlike PIR motion detectors, occupancy estimation sensors are significantly more difficult to aggressively duty-cycle since they often resort to tracking, frame differencing or have long warm-up and configuration times. The AURES platform is designed with an energy-harvesting sub-system that can power the system and charge on-board batteries using indoor lighting sources. A typical use-case is to place a solar panel inside a recessed lighting or florescent fixture and then run the low-voltage wire (which does not require a commercially certified electrician) to the main AURES module mounted nearby on the ceiling. In drop-down ceiling tile installations, the majority of the transducers can sit on the top of the tile with just the ultrasonic transducer protruding through the tile. In combination with our improved algorithm that can run on a microcontroller, this makes for an extremely effective, low-cost and easy to install sensing package.

In summary, we have three main contributions in this paper:

1. We improved upon our previous occupancy counting algorithm and designed a version that can run on embedded targets.

2. We designed and evaluated a presence detection algorithm that can recalibrate our occupancy counting algorithm to account for changes in the background environment over time.

3. We design and evaluate a self-contained energy-harvesting platform with wireless communication that can execute our algorithm in real-time and leverage a smartphone for training.

# 2. RELATED WORK

In this section, we discuss the background related to acoustics followed by similar approaches that have been used to measure both presence and occupancy. Common commercially available occupancy sensors like PIR motion detectors; ultrasonic motion detectors and microwave sensors usually only detect the presence of one or more people in a room. Cameras and more advanced infrared systems attempt to estimate the actual number of people in a space, but are typically expensive, difficult to train and suffer from occlusion. Our proposed approach is comparatively low-cost, relatively easy to train, can perpetually operate from harvested energy and has the advantage of filling an entire space with sound, therefore making it more immune to obstacles.

## 2.1 Acoustics

A large body of work in acoustics has shown that human bodies in a space significantly impact reverberation and that reverberation is frequency [3] as well as room geometry dependent [4]. Over the last 120 years there have been countless efforts proposed to model these acoustic properties in order to improve concert hall performance. Recent work in this space relies on computer simulations [5–8]. Creating simple, generalized models of reverberation has remained quite challenging. For this reason, we propose using machine learning techniques to learn and classify the frequency response on a per-installation basis. In various recent profiles of reverberation [9], it is clear that given a particular room geometry, audience absorption follows relatively distinct curves that make it a powerful feature for occupancy detection.

Active acoustic approaches have shown great potential in multiple forms of sensing. In [10], the authors use a single speaker with multiple microphones to determine the shape of a room based on echoes. In [11], the authors show how reflected Doppler signals can be used to classify anything from speech, to walking motion and even gestures. To the best of our knowledge, this is one of the first end-to-end systems where ultrasound has been used to directly estimate occupancy.

## 2.2 Occupancy

Most of the related work on occupancy has used cameras or multiple sensors to measure the number of people in a space. All of these approaches generally fall into two categories based on slightly different goals. One group focuses on only detecting the presence of people [12–15], which often comes with analysis of more detailed user behavior and actions. The other categories focuses on people counting [16–19], usually involving more sophisticated algorithms for learning.

### *Presence Detection*

In the category of presence detection, many approaches fuse data readings from different sensor types. For example in [20], the authors combine multiple available sensors feeds of data to estimate occupancy. In [12], the authors propose a sonar system using four microphones and a single frequency sinusoid of $20kHz$ in order to detect the user's attention state and several pre-defined activities. The classifier is built by characterizing the delta in echoes, namely the variance in intensity, of the reflections from the user's body. Their experimental results show supportive evidence that a user's presence impacts the intensity of the echoes, which is a fundamental characteristic we assume in our approach. Nevertheless, this technique requires copious amount of training data to predict the pre-defined activity, and assumes the environment to be free from interference. In [21] the authors focus primarily on WiFi signals. In both cases, the approaches do not perform as well in large spaces like auditoriums, unless each occupant is carrying a mobile device that cooperates with the system. Two of the recent works use similar approaches by utilizing ultrasonic signals [12] [13].

[13] proposed an ultrasonic array sensor and tracking algorithm to detect presence and capture the movement of targets. This is achieved by taking the difference in the received echo signal to estimate direction-of-arrival (DoA) with the array of sensors, and utilizing the received signal to noise ratio (SNR) as an indicator of occupancy. A simple tracking algorithm is also proposed to increase performance of presence detection. While this method shows better performance than PIR sensors, the detection zone is limited to a certain area and confined by DoA angle. Other approaches proposed in [14] and [15] take advantage of using multiple co-located sensors. In [14], device nodes are deployed with pressure sensors, PIR sensors, and audio sensors. The system is able to predict pre-defined activities by correlating the binary readings from multiple sensors. The overall classification accuracy is more than 90%, but it requires careful deployment of multiple sensors at different locations in the room. Similar in the choice of sensors, the author in [15] adopts additional light and $CO_2$ sensors. Classification is done using a decision tree in order to determine which sensors are most important. The results indicate that the motion sensor is dominant, and accounts for 97% of accuracy, even when used alone.

Although most of the presence detection techniques have the advantage of low-cost and low-complexity, their applications are limited due to the coarse resolution. Based on the proposed methods, they also suffer from scalability and deployment difficulties due to the confined detection area of the sensors.

### *People Counting*

The most common solutions for people counting tend to use cameras [16–19]. Once configured well, camera systems tend to perform with a high degree of accuracy. These systems do however pose privacy concerns, consume large amounts of power making them difficult to run off of batteries and suffer from obstructions, shadows and a limited field-of-view. Early work for fine-grained indoor people counting is presented in [16], where the locations of the objects are first measured by their silhouettes from image sensors deployed around the room. The system shows accurate results up to 12 people moving in a room, but requires careful placement of multiple image sensors. Also, the computational complexity grows proportionally to the number of sensors. For counting larger groups of people, a crowd counting algorithm proposed in [17] shows accurate results for tens of pedestrians with an error of less than 2 people. The algorithm also claims to be privacy preserving by segmenting the crowd into groups using low-level features, and then using a regression model to count people within each segment. A pedestrian database is required for providing a large number of training images, which is often costly and thus makes it less feasible in more constrained use cases like on an embedded sensor. Recently in [18], the authors proposed a crowd estimation algorithm using IR-UWB radar sensors. The algorithm performs analysis tracking people going in and out of the sensing area by detecting impulse signals. The experiment result shows an accurate counting of up to 9 people in a classroom environment. In [19], the authors evaluated three different learning methods - Support Vector Machine (SVM), Neural Network (NN), and Hidden Markov Model (HMM) over dozens of different sensor inputs, and are able to estimate $0-3$ occupants in an open office area with 75% accuracy. Another different approach is proposed by Hnat et al. [22]. The author introduced the $Doorjamb$ tracking system that uses ultrasonic range finders mounted on door frames to monitor room access. By using probability inference and associating people's identity with their heights, the system performs well on people tracking in special environments, such as labs or residential homes with a 90% room tracking accuracy. However, the system is not designed for counting crowds of people and not suitable for environments with wide open entrances.

To summarize, although most presence detection techniques have the advantage of low-cost and low-complexity, they only provide a coarse resolution of people within a space. In contrast, most people counting techniques are either more expensive in terms of cost and complexity, suffer from privacy issues, or require a large labeled databases. To the best of our knowledge there is no existing framework that can perform wide area people counting with a single cost-effective and versatile sensor.

# 3. OCCUPANCY ESTIMATION

In our previous work [2], we proposed an occupancy estimation algorithm based on the acoustic response of the environment over a range of ultrasonic frequencies. To quickly measure the acoustic response, we utilize chirps that capture reverberation and multipath characteristics across a large frequency bandwidth. When a room is occupied, sound impulses dissipate faster over time and results in a shorter reverberation time. By analyzing the frequency response over the chirp's bandwidth at a few known occupancy levels, we are able to extrapolate the response as the number of people in the room changes.

Our original occupancy estimation algorithm is composed of two parts. In the first part, a classical principal component analysis (PCA) is performed on the training dataset that contains data points collected from different occupancy levels. It allows us to reduce the dimensionality of the received data by projecting them into a lower-dimensional space and learning which of the principal components best characterize the absorption pattern of human bodies. In the second part, a regression model is built based on the projected data in order to interpolate/extrapolate the occupancy beyond the training data. This eliminates the need for copious amounts of labeled training data and improves scalability. At run time, the estimation of the occupancy level is determined based on the trained model with an auto-recalibrate to help the system adjusts itself over time to accommodate for background changes. The model is constantly re-zeroed using new empty room data points.

## 3.1 Training Features

The training features extracted from the data capture the frequency response of the chirp's bandwidth. This is computed by performing a Fast Fourier Transform (FFT) on the full received signal. In our previous work, we showed that the chirp's frequency and duration have a direct impact on the performance of the system, where increasing the chirp's frequency band and length improve the system performance. However, when building the platform, we saw that if the transmitter and receiver are physically close, then the system suffers from crosstalk. Recording after playback in turn defines an upper bound of the chirp length which must now be much shorter (originally $300ms$, now $30ms$). To compensate for the performance loss of using a shorter chirp, we segment the training features instead of computing the FFT over the whole received signal. This provides us with additional amplitude data across each segment. First, we separate the signal into segments with the same length as the chirp. Each segment is then transformed into the frequency domain individually, and later combined together to form the new training features. The difference can be essentially interpreted as reducing the FFT window size from the recording length to the chirp length. Since the chirp is much shorter than the recording, it not only generates features that better capture how the sound dissipates over time in amplitude, but also greatly reduce the memory required to perform the FFT. To prevent bias between features when performing PCA, all features are later normalized and subtracted by their means.

## 3.2 Weighted Principal Component Analysis

One constraint of our original algorithm is that it assumed the whole training dataset is collected in background environments with similar reverberation characteristics. If the acoustic response of the environment changes dramatically during data collection, which is likely to happen in practice, then PCA can perform poorly. The resulting PCA can erroneously produce principal components that explain the changes in the environment, rather than the desired ones that differentiate occupancy levels. To solve this problem, a weighted variation of PCA (WPCA) is adopted to target components that separate occupancy levels. Classical PCA is known to be sensitive to outliers and missing data, while WPCA increases the robustness of the system to outliers by assigning different weights to data points based on their estimated relevancy. For our application, we utilize the same idea to minimize the influence of the changing environments when performing PCA.

Assuming the training dataset is given by matrix $X$ where each of the $i$ rows represents a feature variable and each of the $j$ columns represents an observation. The goal of a classical PCA is to find a decomposition of the matrix

$$X = PC \tag{1}$$

where $P$ is the orthogonal matrix of principal component and $C$ is the principal coefficient matrix, such that the matrix $D$ given by

$$D = P^T X X^T P = P^T \sigma^2 P \tag{2}$$

is diagonal and has its variance maximized. The diagonals of $D$ are often re-arranged in order such that $D_{ii} >= D_{jj}$, $\forall i < j$ so that the first column of $P$ represents the first principle component that accounts for the most variance. Equivalently to maximize the variance, the principal components allows us to minimize the reconstruction error $\|X - PC\|_2^2$ when the data is projected into a lower-dimensional space. Similarly, the goal of WPCA is to minimize the weighted reconstruction error given by

$$\|W(X - PC)\|_2^2 = \sum_{ij} W_{ij}^2 (X_{ij} - PC_{ij})^2 \tag{3}$$

By assigning lower weights $w_j < 1$ column-wise to the empty room instances, which are identified by the presence detector, WPCA is biased toward finding principal components that best explain the variance between different occupancy levels. Other non-empty room instances are assigned with a fixed weight $w_j = 1$ to prevent bias. To center the dataset and calculate the covariance matrix, the weighted mean to be subtracted is given by

$$\bar{x} = \frac{\sum_j W_j X_j}{\sum_j W_j} \tag{4}$$

where $X_j$ denotes the $j$th column of the dataset $X$. More evaluation on how to select a proper weight is discussed in Section 5.3.

## 3.3 Presence Detector

The ability to detect whether a room is empty can improve the quality of WPCA and help to determine when the system should recalibrate. To automate the recurring recalibration process, we proposed using a single tone instead of a chirp to facilitate the detection of Doppler shift. In each sensing period, the system transmits 5 consecutive tones with a delay of $300ms$ in between to allow echoes to fully dissipate. The received signals are first transformed into frequency domain and then filtered to remove out-of-band noise. The presence detector is composed of three binary classifier (empty or non-empty), where each focuses on different features of the received signal. Note that since presence detection is now part of the people counting algorithm, the features and mechanisms used in presence detection are independent from those used in the determination of occupancy level. The first classifier is a Doppler motion detector

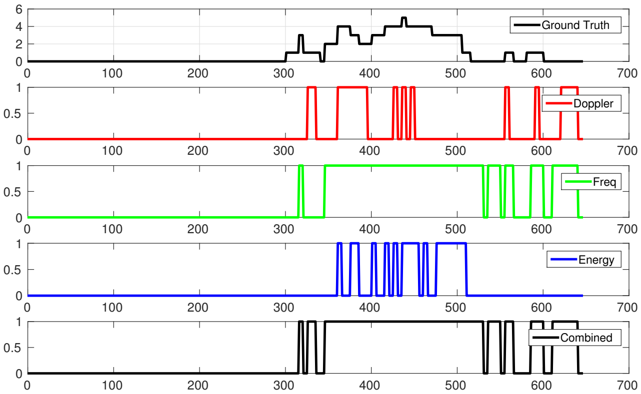



Figure 1: Comparison of presence detection result between three classifiers with one day of empirical data

that detects Doppler shift caused by the movement of bodies or gestures. Even though Doppler detectors work well at detecting sudden movements, it is often difficult to detect static changes such as different postures of the occupants or slow motions. To improve performance we apply two additional classifiers to calculate the variance of the spectral amplitude and the variance of the received signal energy respectively. Tuning the threshold of each classifier allows us to control the ratio between false-negative rate (FNR) and false-positive rate (FPR). To prevent the system from recalibrating on non-empty data points, lowering the rate of getting a negative feedback while the room is occupied is critical. Recalibrating on FN instances offsets the baseline of the model and introduces estimation error that would last until the next recalibration cycle. On the other hand, FP instances trigger the system to make estimation on the new environment, which does not introduce much estimation error in comparison since the model is trained to detect human bodies that absorb more power. Therefore, the thresholds in all three detectors are tuned to be conservative, and the final decision is obtained by taking an OR operation between the three binary results in order to achieve a low FNR. We discuss the performance of the presence detector in Section 5.2.

## 3.4 Volume Control

Reducing the power consumption is key for building a self-sustained energy harvesting platform. Based on the energy footprint of the device, signal playback and recording are the most significant power-consuming operations. The power consumption of recording is fixed, but the transmit power can be controlled by adjusting the speaker volume. We also generally want to decrease volume for scalability and to improve pet friendliness. Since the system relies on the amplitude of the received signal to estimate occupancy level, we observed a trade-off between the power consumption and the system performance. Figure 6 shows how volume impacts the clustering performance of WPCA in one of our test environments. Tighter clusters are easier to distinguish and hence perform better. The ideal output power is both environment and room geometry dependent. For this reason, we utilize signal-to-noise ratio (SNR) as a criteria to estimate the system performance in different environments. The duration of the received signal on which we calculate the SNR is an important factor since the received signal dissipates at different rates in different environments. Based on our experiment results, we find that the features from the first two segments of the received signal (i.e. the first reflection) are generally more significant in the generation of high-rank principal components, therefore we use them to define SNR. During installation, the volume of the transmitter is slowly increased until a particular SNR threshold of the reflected signal is achieved. This threshold is selected based on results described in Section 5.1.

# 4. PLATFORM IMPLEMENTATION

In this section, we discuss the hardware platform and the software processing workflow. This entails how data is captured and passed to a mobile device for installation and training.

## 4.1 Hardware Design

We developed an energy harvesting, embedded hardware platform for our ultrasound transceivers as shown in Figure 3. The platform was designed to have a low enough power consumption so that it can be powered using a 7x5.5$cm$ solar cell harvesting energy from artificial or natural light sources. This allows for a flexible installation at a low cost, since the transceivers do not need to be connected to AC wall power, which is often difficult to access at ceiling mounting locations.

The hardware platform features a single PCB design, which uses a TI CC2650 multi-standard BLE and 802.15.4 SoC connected to a 192$kHz$ audio codec, a MEMS microphone and a piezo ultrasound speaker connected to a Class D piezo speaker amplifier to transmit and receive ultrasound signals. An ultrasonic horn as described in [23] is attached to the speaker to disperse the emitted ultrasound in an omnidirectional fashion. 2Mbits of on-board SRAM is used to store recorded waveforms before they are processed and the results are sent to a gateway using 802.15.4 or BLE. Figure 2 shows a block diagram of the primary components of the hardware platform. The total cost of our current hardware design is around $30 at quantity 1000, including the energy harvesting module.

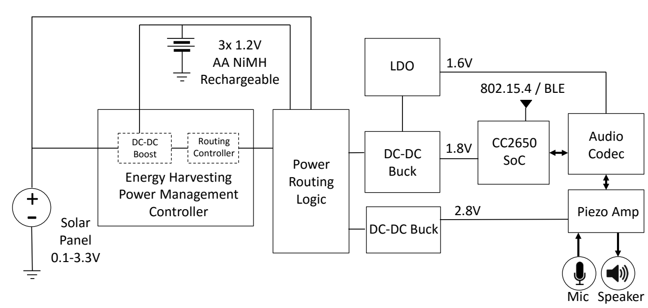



Figure 2: Block diagram of main hardware components

## 4.2 Processing Workflow

In this section, we discuss the processing workflow of our system starting with installation, training and then steady-state. All processing is performed on-board except the initial training, which is offloaded to a computer for processing due to memory constraint. An installer should first mount the AURES node to the ceiling in a central location with the

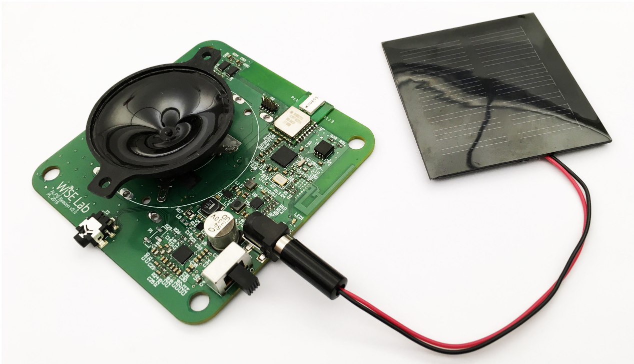


Figure 3: Hardware PCB design with external solar panel

Figure 4: Hardware mounted on hanging fluorescent light

solar panel near a lighting fixture. The installer can then configure the node using a BLE enabled device, like a smartphone, and bootstrap the volume configuration sequence on AURES where the transmitter profiles the room's SNR. After determining a sufficient volume threshold, the node periodically scans for presence followed by collecting an occupancy reading. Since initially there is no trained model, the node will store the output of the high-pass filtered spectrum response of the chirp in its flash memory as training data. This will be collected over an extended period and eventually all training data is transferred to a phone or computer to perform WPCA and regression. In cases where a gateway is available, this could also be done in a streaming fashion. During data collection, the installer should come back periodically to label a subset of the room occupancy levels. In our experiments we used only two labels, but at least one point should be above 10% of the room's capacity. When collecting data once every 10 minutes, the AURES node has enough storage to hold two weeks of data in its 4Mbits of flash storage which requires up to 30 seconds to transfer to a phone. The resulting model (<4KB) is then transferred back to the node over BLE at which point the system begins executing. The regression model is periodically updated afterward when the room is identified as empty.

# 5. EVALUATION

In this section, we discuss experimental results using data captured by our system. In order to collect raw waveform with ground truth, we connected the AURES transceiver to a BeagleBone Black Linux platform with a fish-eye camera. During the sensing period, our system starts the recording of $300ms$ right after each signal transmission and samples at a rate of 192KHz[1]. The recording length is selected to be significantly longer than the time required for the chirp to dissipate fully in the room [3]. The ideal chirp length should be shorter than the acoustic round-trip time of the room. Assuming the smallest room of operation is $3m^2$, the maximum chirps length thereby corresponds to $20ms$. To prevent audible artifacts in low-cost speakers that could be detected by humans [24], we added an additional $5ms$ of fade-in and fade-out time to the chirp and ended up with a chirp length of $30ms$.

We conducted experiments in ten environments of different room sizes over the campus[2]. In each room we mounted the system on the ceiling close to the center of the room to allow a better coverage[3]. A camera with a fish-eye lens (shown in Figure 5b) was installed next to the system and configured to take a low-resolution snapshot (shown in Figure 5c) right after each signal transmission to capture ground truth. The system was configured to collect 5 samples for both the presence detection and occupancy estimation every 10 minutes through out the day, which correspond to $\sim$ 1300 samples per day. We collected data between 3-14 consecutive days in each room and periodically offloaded the collected data to a remote server. Once the data collection was completed, we trained a model using the data collected from the first day with two occupancy levels manually labeled. To generalize the evaluation results, we classified these rooms into 3 categories based on their sizes. Rooms occupy less than $10m^2$ are classified as small rooms, the rest that occupy between $10m^2 - 100m^2$ are classified as medium rooms, and the rest that occupy more than $100m^2$ are classified as large rooms.

## 5.1 Volume Control

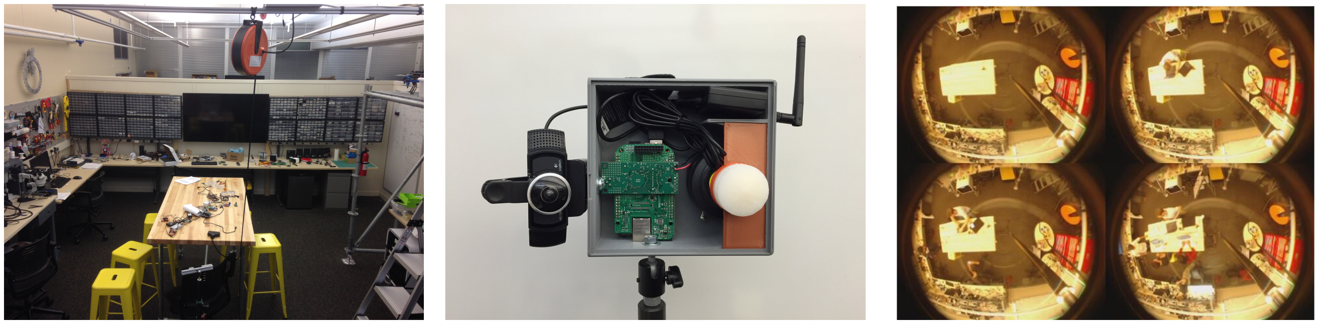
Figure 6 shows how speaker volume directly impacts the difference in received signals collected from three different occupancy levels. For the purpose of visualization, the data are presented in 2-D space using WPCA. Each data point represents an observation and its color reflects the occupancy level. We see that data collected at low volumes are more difficult to be separated by their occupancy levels while data collected with higher signal strength can be easily categorized into clusters. To better understand how the volume affects the system performance in different environments, we also calculate their corresponding SNR as discussed in Section 3.4. Figure 7 shows the average received SNR at different output volumes in different sizes of rooms. One could imagine using this property to estimate room size. We see received SNR increases exponentially with higher output volume, and the increasing rate is higher in smaller rooms. Figure 8 shows the system performance and the SNR of the received signal in different environments. We see a positive correlation between the SNR and estimation accuracy, and we find that the mean error is greatly reduced once the received SNR pass the $10dB$ threshold. At installation, the system slowly increased the volume until this $10dB$ threshold is reached.

---

[1]The sampling rate can be reduced to 96KHz without much performance loss [2].

[2]Our IRB declared this data collection to be non-human subject research.

[3]The location of the transceiver has little impact on the system performance [2].



(a) (b) (c)

Figure 5: Experimental setup (a) Lab with highly variable furniture and equipment positions (b) AURES node connected to a BeagleBone Black with a fish-eye camera (c) Ground truth camera snapshots of the lab at different occupancy levels

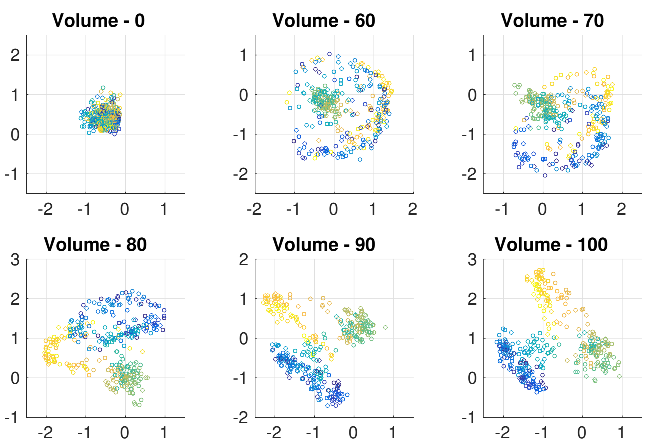



Figure 6: Effect of different speaker volumes on data clustering in 2-D space derived by WPCA

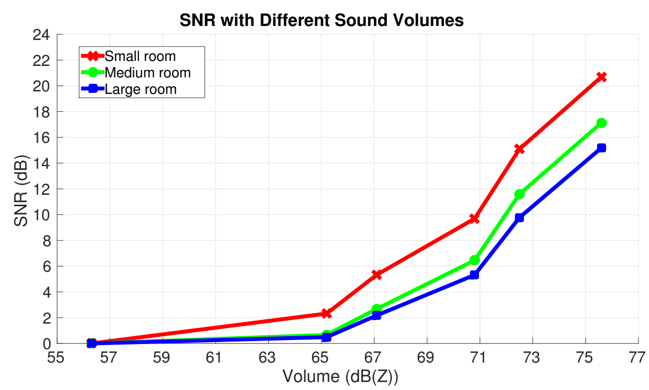



Figure 7: Received SNR with different output volumes

## 5.2 Presence Detector

In Figure 1, we show the classifiers' sensitivity to the room occupancy level. We can see that the Doppler-based classifier is sensitive to movements regardless of the number of the occupant in the room, while the variance-based classifiers are more accurate when there are more occupants. The overall performance of the presence detector is summarized in Table 1, which includes the accuracy, false positive rate (FPR), false negative rate (FNR), precision and recall. We see that the overall accuracy decreases as the size of room increases, which is not surprising since multipath reflections are much weaker and noisier in large spaces. In our ten different room environments, each classifier has an accuracy of $65 - 75\%$ on average, but when combined the overall accuracy increases to 80%. Since the detector is designed to reduce false positive instances, we are able to achieve a recall of 85%. For the remaining 15% false positive instances we analyzed the distribution over the number of occupants in different room environments to see the negative impact on occupancy estimation. Figure 9 shows the FNR as the number of occupants increases. We see that the detector suffers the most from single person instances, especially in cases where the only person is still like when typing or using laptop. However, the false positive rate decreases exponentially as the number of occupants increases. This indicates the introduced error on the successive occupancy estimation is minimal even if the system erroneously recalibrates on false negative instances. Also, it should be noted that in practice we can further improve the detection accuracy by extending the sensing period and/or increasing the number of tones used for detection.

## 5.3 Occupancy Estimation

To evaluate our automatic retraining technique, we collected three weeks of data in a noisy semi-opened laboratory environment (shown in Figure 5a) which frequently changed due to everyday use. We show the estimation traces of the first five days of the collected data in Figure 10, where the estimation model is trained using the first 500 samples with two labeled occupancy levels. Without periodic self-retraining, we see an offset of estimation error right after the lab is being used on the first day. Moreover, the error offset begins to accumulate over time and prevents the system from accurately estimating the occupancy levels for the following days. However, when the system retrains itself with presence sensing, it is able to re-zero the baseline according to the new environment sporadically and thus greatly reduce the estimation error. Using our presence detector, the system is able to reduce the mean error from 2 to 0.5 people. In comparison, with a perfect presence detector, the estimation error can be further reduced to 0.3 people. As previously discussed in Section 3.3, the amount of improvement the presence detector provides depends mainly on its accuracy and the error distribution of the false negative cases.

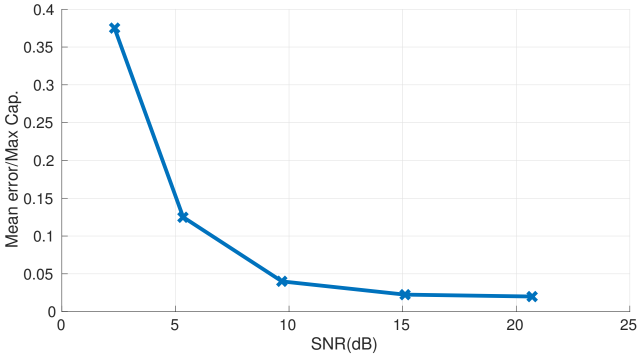



Figure 8: System performance with different SNR in small rooms

| Sizes \ Param. | Acc. | FP | FN | Prec. | Rec. |
|---|---|---|---|---|---|
| Small room(s) | 0.85 | 0.11 | 0.09 | 0.90 | 0.91 |
| Medium room(s) | 0.82 | 0.27 | 0.12 | 0.76 | 0.88 |
| Large room(s) | 0.75 | 0.29 | 0.21 | 0.72 | 0.79 |

Table 1: Presence detection performance with different room sizes

In Figure 11, we show the system performance with varying SNR of the received signal and weights assigned to the empty room instances. The assigned weights help the system cope with noisy environmental data in the training dataset. With a fixed SNR, we see that assigning overly high or low weights both negatively impact the system's performance. Assigning too much weight causes the WPCA to take into account the variance between different environments, and thus biases the estimator away from counting people. In contrast, an overly low weight would produce dominating principal components poorly extrapolate the occupancy levels, and the estimator would overfit and often predict the room to be full or empty. This negative impact is more noticeable when the SNR decreases, which is not surprising since with a low SNR the amplitudes along are not correctly estimating the occupancy level. At this point, increasing weights exacerbates the problem. Based on the experiment results, one should never use an overly low weight to prevent overfitting and for our evaluation we choose weights equal to 0.5 since it works well in most configurations. The overall system performance in different environments is summarized in Table 2. The error is calculated by taking the absolute difference between our estimation and the actual number of people in the room. The overall error slightly increases with the room size since large rooms result in lower received signal strength and higher variance in multipath delay. On average, the absolute error is no more than 3 people across different room sizes, and the error in percentage to the maximum number of the participated occupants is around 10%.

## 5.4 Open Space Performance

Unlike in enclosed rooms, in open air environments a large portion of the transmitted signal will be scattered away after the first reflection and only a small amount of signal can be captured by the receiver. The amplitude of the reflected signal is highly dependent on its distance from the transceiver and the surface material of the ground. To test the system's performance and sensing range in an open air environment,

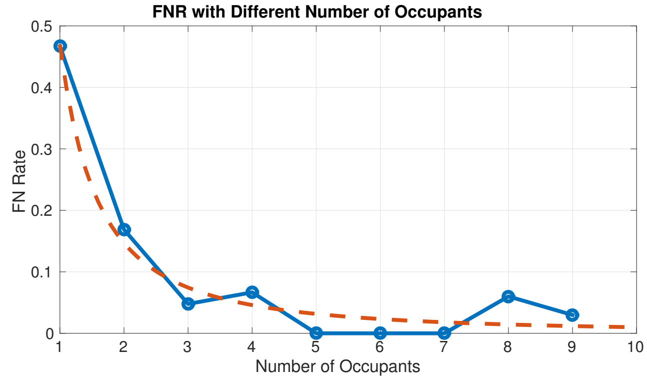



Figure 9: False negative rate with different number of occupants

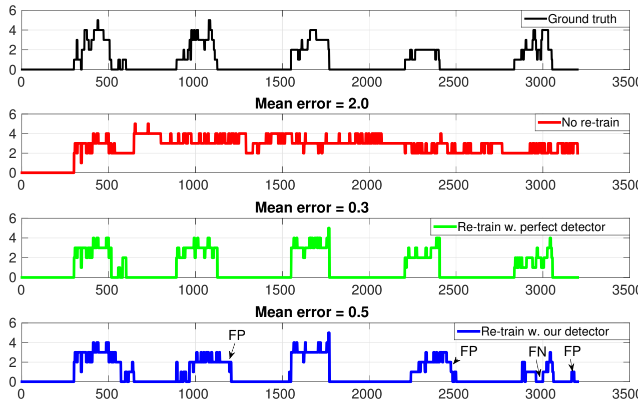



Figure 10: Comparison of occupancy estimation over 5 days of empirical data with (1) No-retrain (2) Retrain with perfect detector (3) Retrain with our detector

we collected a dataset of people standing in lines and in clusters at different distances away from the transceiver. The transceiver was placed 3.5 meters above the ground on a tripod in an open parking lot. Table 3 shows that performance is good for occupants standing closer than $6m$ in diameter from the transceiver with an 8% estimation error. However, as occupants move further away, the estimation error increases to 27% with a large performance drop-off beyond a $10m$ diameter. In our experiments, we also noticed several blind spots at certain transmission angles that have a shorter detection range, which is likely caused by the imperfect beam pattern of our horn speaker design. In comparison to enclosed environments, the system's performance in open air is noticeably worse except at close range. This supports the notion that our training feature is based on the reverberation and the decay of many multipath reflections. This experiment does show that our sensor could be used for estimating occupants in smaller regions, even in open environments which might be a powerful tool for estimating line length in a food court or detecting people in cubical areas.

## 5.5 Energy Harvesting and Consumption

The AURES hardware platform uses a power management IC to charge three low-self-discharge $2100mAh$ NiMH cells to provide sufficient power for transmitting ultrasound whether or not solar power is currently available. The cells

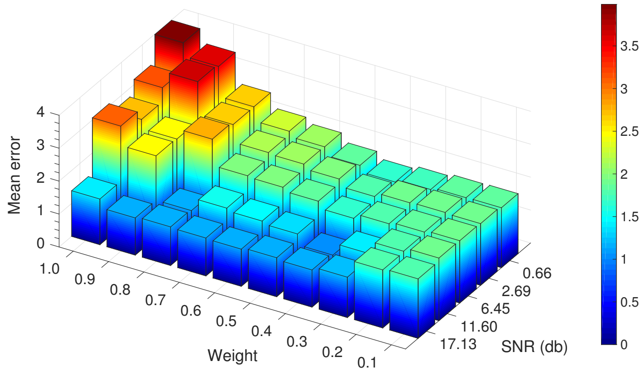



Figure 11: Mean estimation error with different received SNR and weights in WPCA

| Room Sizes | Tested/Max Capacity | Avg. Error | Error/ Tested Cap. |
|---|---|---|---|
| Small(s) | 4/5 | 0.26 | 6.5% |
| Medium(s) | 10/20 | 0.94 | 9.4% |
| Large(s) | 21/100 | 2.36 | 11.2% |

Table 2: System performance with different room sizes

| Sensing Diameter | Error/# of Occ. | Acc. Error |
|---|---|---|
| <6m | 0.08 | 0.08 |
| 6m-10m | 0.27 | 0.21 |
| >10m | 0.48 | 0.35 |

Table 3: System performance in open air environment

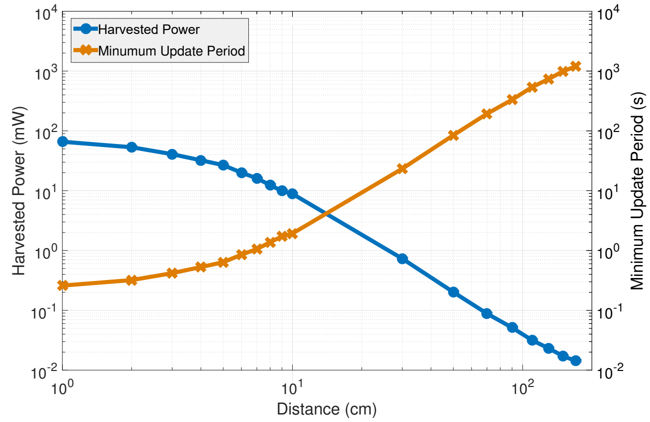



Figure 12: Power output from solar cell vs. distance to 100W equivalent CFL bulb vs. minimum update period

are able to retain 70% of their charge over 10 years, hence making their self-discharge rate negligible. The IC starts harvesting at an input voltage level as low as $100mV$ and features a Maximum Power Point Tracking (MPTT) algorithm, which modulates the load on the solar cell to maximize its power output. Voltage levels for all subsystems are regulated by highly efficient buck converters.

Figure 13 shows the typical power consumption of a transceiver waking up from sleep and activating its audio codec and piezo amplifier (1-2), transmitting a $40ms$ long ultrasound transmission ($30ms$ chirp with $5ms$ fade-in and fade-out time to prevent audible artifacts) at maximum volume ($86.5dB(Z)$ at $1m$) (2-3), recording for $300ms$ at a sampling rate of $96kHz$ (3-4), processing the recording and sending the result over the radio (4-5) and then going back to sleep (5). This sequence of operations consumes a total of $18.56mWs$. We designed our transceivers to be able to be installed close to light fixtures on the ceiling. Solar energy can be harvested directly at the bulb for an improved update rate, or simply from ambient light. Figure 12 shows the power output at the maximum power point of our $7$x$5.5cm$ solar cell at various distances from a single $100W$ equivalent CFL bulb. Based on these numbers and a negligible sleep power consumption on the order of micro-watts, we estimate the minimum update period of the system. When the solar cell is placed in close proximity to a lighting source, an update rate on the order of seconds is possible, while ambient light energy harvesting allows for an update rate on the order of tens of minutes.

## 5.6 Processing Microbenchmarks

The most CPU demanding part of our system's operations is performing 10 2048 point FFTs on 10 $30ms$ long chunks of the $300ms$ recording. Each segment is fetched from external SRAM and then processed using ARM's CMSIS-DSP library. Benchmarking the time duration of this process using the microcontroller's clock, we see that this typically requires $144.45ms$, of which $44.88ms$ are spent fetching the data and $99.56ms$ are spent calculating the FFTs. From each FFT result, 205 $16bit$ samples from the frequency band of interest are sent back to a base station via RF. It takes approximately $595ms$ to transmit, record, process and radio the result of an occupancy sample.

# 6. LIMITATIONS

Our system has a few practical limitations. We chose our frequency range because it is supported by low-cost commercial audio codecs and it is the lowest inaudible frequency that attenuates significantly less than higher frequency narrowband transducers. For this reason, the signal is likely perceptible to service animals. Though more testing is required, our target duty-cycles and volume levels are designed to aggressively optimize energy and should be almost undetectable to most animals. Many commercial ultrasonic motion detectors already produce louder harmonics in our target frequency band than what we require for sensing. Our system also requires labeling of training data. While mobile phone interfaces can simplify this process, an installer still needs to capture a snapshot when the room has a reasonable (>10%) occupancy level which might be difficult in some cases.

# 7. CONCLUSION

In this paper, we presented an indoor occupancy sensing platform that is lower-power, accurate, privacy preserving, and easy to train. The system operates by transmitting wide-band ultrasonic signals into a room and measure the superposition of the reflections over time to determine occupancy level. To help the system adapt to versatile background environments and improve system performance, we use a combination of Doppler shift, variance of spectral amplitudes and variance of signal energy for presence detection. We reduce the training effort while improving the system performance by using WPCA to cope with noisy training

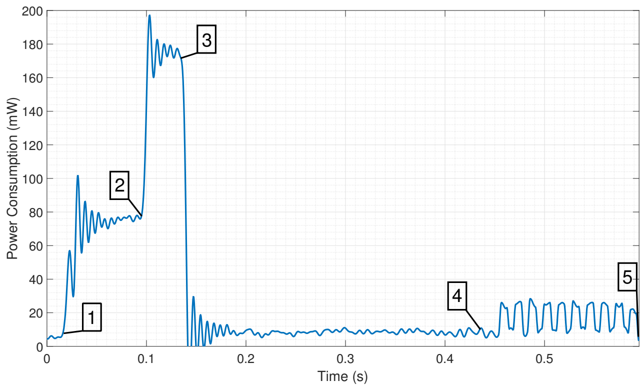



Figure 13: Power consumption of AURES at full volume

data. To improve energy efficiency and scalability, we propose a volume control mechanism and an energy-harvesting subsystem with benchmark test. Finally, we evaluated our system in 10 rooms with different sizes and collect data of daily use totaling over $60,000$ data samples. Our result shows an average of 85% recall rate for presence detection and less than 20% of estimation error on people counting.

# 8. ACKNOWLEDGEMENTS

This research was funded in part by TerraSwarm, one of six centers of STARnet, a Semiconductor Research Corporation program sponsored by MARCO and DARPA as well as the NSF as part of grant #1534114 and Bosch RTC Pittsburgh.

# 9. REFERENCES

[1] Z. et al., Energy Savings for Occupancy-Based Control (OBC) of Variable-Air-Volume (VAV) systems. PNNL-22072, 2013.

[2] O. Shih and A. Rowe, "Occupancy estimation using ultrasonic chirps," in *Proceedings of the ACM/IEEE Sixth International Conference on Cyber-Physical Systems*, ser. ICCPS '15. New York, NY, USA: ACM, 2015, pp. 149–158.

[3] W. C. Sabine, "Collected papers on acoustics," *Harvard University Press*, 1923.

[4] C. A. Andree, "The effect of position on the absorption of materials for the case of a cubical room," *Journal on the Acoustics Society of America*, 1932.

[5] T. Hidaka, N. Nishihara, and L. L. Beranek, "Relation of acoustical parameters with and without audiences in concert halls and a simple method for simulating the occupied state," *The Journal of the Acoustical Society of America*, vol. 109, 2001.

[6] T. Hidaka and N. Nishihara, "Reverberation time, mean-free-path, and sound absorption in concert halls-numerical examination by computer simulation," *The Journal of the Acoustical Society of America*, vol. 119, no. 5, pp. 3430–3430, 2006.

[7] M. R. Schroeder, "Computer models for concert hall acoustics," *American Journal of Physics*, vol. 41, no. 4, pp. 461–471, 1973.

[8] W. J. Davies, Y. W. Lam, and R. J. Orlowski, "Predicting theater chair absorption from reverberation chamber measurements," *Journal of the Acoustical Society of America*, vol. 93, no. 4, pp. 2238–2240, April 1993.

[9] L. L. Beranek, "Analysis of sabine and eyring equations and their application to concert hall audience and chair absorption," *The Journal of the Acoustical Society of America*, 2006.

[10] I. Dokmanic, R. Parhizkara, A. Walthera, Y. M. Lub, and M. Vetterli., "Acoustic echoes reveal room shape," *Proceedings of the National Academy of Sciences of the United States of America*, 2013.

[11] B. Raj, K. Kalgaonkar, C. Harrison, and P. Dietz, "Ultrasonic doppler sensing in hci," *Pervasive Computing, IEEE*, vol. 11, no. 2, pp. 24–29, Feb 2012.

[12] S. P. Tarzia, R. P. Dick, P. A. Dinda, and G. Memik, "Sonar-based measurement of user presence and attention," *Ubicomp*, 2009.

[13] D. Caicedo and A. Pandharipande, "Ultrasonic array sensor for indoor presence detection," in *Signal Processing Conference (EUSIPCO)*, 2012.

[14] T. A. Nguyen and M. Aiello, "Beyond indoor presence monitering with simple sensors," in *2nd International Conference on Pervasive and Embedded Computing and Communication Systems*, 2012.

[15] E. Hailemariam, R. Goldstein, R. Attar, and A. Khan, "Real-time occupancy detection using decision trees with multiple sensor types," in *Symposium on Simulation for Architecture and Urban Design*, 2011.

[16] D. B. Yang, H. H. Gonzalez-Banos, and L. J. Guibas, "Counting people in crowds with a real-time network of simple image sensors," in *International Conference on Computer Vision*, 2003.

[17] A. B. Chan, C. La Jolla, Z.-S. J. Liang, and N. Vasconcelos, "Privacy preserving crowd monitoring: Counting people without people models or tracking," in *Computer Vision and Pattern Recognition*, 2008.

[18] J.W.Choi and S. Cho, "A crowdedness measurement algorithm using an ir-uwb radar sensor," in *International Conference on Future Communication, Information and Computer Science*, 2014.

[19] K. P. Lam, M. Hoynck, B. Dong, B. Andrews, Y. shang Chiou, D. Benitez, and J. Choi, "Occupancy detection through an extensive environmental sensor network in an open-plan office building," in *Proc. of Building Simulation 09, an IBPSA Conference*, 2009.

[20] L. Yang, K. Ting, and M. Srivastava, "Inferring occupancy from opportunistically available sensor data," in *Pervasive Computing and Communications (PerCom), 2014 IEEE International Conference on*, March 2014, pp. 60–68.

[21] D. Li, B. Balaji, Y. Jiang, and K. Singh, "A wi-fi based occupancy sensing approach to smart energy in commercial office buildings," in *Proceedings of the Fourth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, ser. BuildSys '12. New York, NY, USA: ACM, 2012, pp. 197–198.

[22] T. W. Hnat, E. Griths, R. Dawson, and K. Whitehouse, "Doorjamb: Unobtrusive room-level tracking of people in homes using doorway sensors," in *ACM Conference on Embedded Network Sensor Systems*, 2012.

[23] P. Lazik, N. Rajagopal, B. Sinopoli, and A. Rowe, "Ultrasonic time synchronization and ranging on smartphones," in *Proceedings of the 21st IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS 2015)*, ser. RTAS '15, IEEE. IEEE, 2015.

[24] P. Lazik and A. Rowe, "Indoor pseudo-ranging of mobile devices using ultrasonic chirps," in *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*, ser. SenSys '12. New York, NY, USA: ACM, 2012, pp. 99–112.