Distributed Sensor Network for Indirect Occupancy Measurement in Smart Buildings

Colin Brennan*†, Graham W. Taylor*†, and Petros Spachos*
*School of Engineering, University of Guelph, Guelph, ON, N1G 2W1, Canada
E-mail: cbrennan@mail.uoguelph.ca, {gwtaylor, petros}@uoguelph.ca
†Vector Institute for Artificial Intelligence

Abstract—Many areas like smart buildings, crowd flow, action recognition, and assisted living rely on occupancy information. Although the use of smart cameras can alleviate the problem and provide accurate occupancy information, at the same time it can be cost prohibitive, invasive and not easy to scale or generalize to different environments. An alternative solution should bring similar accuracy while minimizing the previous problems. This work presents a candidate wireless sensor network for indirect occupancy measurements in a smart building. A prototype was built, that consists of $CO_2$, temperature and humidity sensors. The prototype was placed in a complex indoor environment to collect data for a week. Then, the data was analyzed to examine potential correlation between sensor data and occupancy information. According to experimental results, $CO_2$ can be used for indirect occupancy measurements.

Index Terms—Room Occupancy, WSN, $CO_2$.

I. INTRODUCTION

As the trend in technology continues to make compute power and integration among devices more accessible, there is an increase in demand for intelligent, connected systems. Consumers seek systems that can anticipate their needs and act accordingly to in some way make their lives easier. In order for any smart system to comply with this request it needs to have enough information about what specific action is required of them and the state of the environment in which they may act [1]. This critical context must be observed and measured in some way to be integrated into the system for consideration.

Specific applications for these smart systems are already being addressed, including systems for smart homes and smart offices, crowd behaviour analysis, and human action recognition. Common to each of these applications is that these systems require context on where the user or users are within the observable space. While some occupancy capable systems do perform well for some of these applications, it is a challenge to produce a system that can generalize to work for any of these settings. The infrastructure available for a large office building seeking smart office control is quite different from that available in a smart home. Smart Camera systems may be well suited for such an office setting, but may be impractical for large crowds and may not be welcome within a user’s home. In addition to generalizing across applications, a suitable solution should thus include consideration for anonymous and non-invasive sensing as required.

Sensors which observe indoor environment state present a viable solution to this problem. These sensors track occupancy by identifying which changes in environment state are caused by changes in the number of persons present. When deployed as a Wireless Sensor Network (WSN), this capability can be extended to cover multiple environments such as the various rooms in a smart home or office. Distributed occupancy capable sensing nodes may be deployed as a network to a target area to serve as the context source for each individual space, as well as for information between those spaces. The scalability of the modular networks allows for generalization to different size application domains.

In this work, a WSN prototype is presented which is capable of interacting with its local environment to provide the necessary information to work with occupancy context information. The prototype has three sensor units for $CO_2$, temperature and humidity in order to provide the necessary environment monitoring. All the data is forwarded to a server in real time for further processing. The prototype was tested in a complex indoor environment and collected data for a week. Data analysis and correlation provide some useful insights regarding the levels of $CO_2$ and the number of the people in the monitoring area.

The major contributions of this work are listed below:

- A wireless prototype consisting of three sensors was built for experimentation.
- Experiments were conducted for a week, with six prototypes in two laboratories with different layouts. Each prototype contributed an average of 54, 140 samples. The samples were labelled for further processing.
- Data correlation between the sensor measurements and the number of the occupants in the room was examined. According to experimental results, $CO_2$ can be used for indirect occupancy measurements.

The rest of this paper is organized as follows: Section II discusses the related work in this area. Section III provides a brief discussion on the proposed system implementation. The experimental results are presented in Section IV while Section V concludes this work.

II. RELATED WORK

For applications that focus on the actions and behaviours of few or individual occupants, the state of occupancy can be obtained as a consequence of tracking the users’ locations.
Wearables such as a user’s smart-phone or other communication sensor beacons can be used to this end. The accelerometer and coarse location sensing abilities of a standard smart phone have been used for both the applications of human action recognition [2] and area departure prediction [3]. The work in [4] also showed that it is possible to obtain location accuracies of five meters from machine learning models based on received signal strength from statically placed beacons throughout a target space. WSN systems based around the inclusion of a wearable component can provide advantages in terms of specific user location accuracy. However, the per person hardware requirements may limit the scale of the installation, and may not abide by desire to provide an anonymous and non-invasive solution.

Occupancy and location information have also previously been obtained through sequences of activations of unique sensors with known, fixed locations. In [5], a heterogeneous sensing network is used to identify and predict the activities of smart home residents. To identify the Activities of Daily Living (ADLs) of interest, some of the sensors considered in this type of network include: motion, light, door, contact, and temperature. Many of the ADLs, such as sleeping, cooking, and personal hygiene, can be uniquely associated with a single room of the smart home and thus provide room occupancy information for the resident. Once the ADLs of a user are obtained, they can be used to provide services customized for that user. Long-term trends of these ADLs have been shown to correlate with clinical mobility health assessments for senior residents [6], allowing for remote assessment. In a similar manner, the work in [7] has been used more generally for identifying deviations from nominal resident activity in a target smart home. Sensor activation based location has also been applied to producing energy saving recommendations through the incorporation of appliance energy sensors [8].

For the action activation approach, location information is obtained from known relationships between the recognized actions and the locations of the network sensors. It is often the case that confidence in action recognition comes from a sequence of multiple sensor events and from the proceeding and following actions. Depending on the complexity of the installation, this may impose restrictions on the delay before occupancy information is known. This approach is suitable to maintain anonymity but the invasiveness depends on the sensors required to observe the desired ADLs.

In contrast to the user tracking approaches, area centric methods obtain occupancy information through changes in environmental readings in a local proximity. Environmental assessment WSNs, such as those for indoor air quality applications [9]–[11], make use of sensors that can observe volatile organic compounds, temperature and humidity. These environmental readings can then be parsed for changes based on known or learned relations to occupancy behaviours [12], [13]. As with the action activation method, knowledge of the sensor locations and their effective regions is required. Unlike the action activation method, nodes in this approach are not restricted in deployment location by a competing sensing objective and thus can be adapted differently during installation of the system. Environment sensing methods have an additional requirement of knowing or learning the relationship between occupancy and the indirect measurements they conduct [14]. By not focusing on specific users, the environmental methods satisfy anonymity and the indirect nature avoids requiring user attention and interaction.

This work follows the area centric methods through a WSN with CO₂, temperature and humidity sensors for the desired indirect occupancy measurements.

III. SYSTEM IMPLEMENTATION

This section provides a brief description on the prototype that was built followed by the data collection procedure. This occupancy modelling network consists of a number of hardware components for each node. These components were selected to facilitate the required sensing and distributed processing requirements of the system. The deployment and coordination of nodes throughout the target space establishes the sensing network.

A. Prototype

The developed prototype is shown in Fig. 1 and has the following two main units:

i. Sensing Unit: To achieve the desired environmental observation, two sensors are utilized, the IAQ-2000 indoor air quality sensor and the DHT22 Temperature and humidity sensor. The IAQ-2000 Volatile Organic Compound (VOC) sensor is used to measure the local concentration of CO₂. While capable of reading other compounds such as methane, carbon monoxide and alcohols, in the intended environment CO₂ from occupants will be the most prevalent.
This sensor outputs continuous voltage values relating to the parts-per-million concentration of the applicable VOCs. These reports are in the range of 450ppm to 2000ppm [15] which encompasses the expected range for an office setting [16]. The second sensor, the DHT22, provides readings of temperature and humidity within the ranges of -40°C to 80°C and 0-100% humidity [17] at a sensing period of 2 seconds. These operation conditions are more than sufficient for the indoor setting. These sensors together permit the observation of the complex relation between environmental readings and occupancy [12].

ii. Processing Unit: For each sensing node it is necessary to have interfacing capabilities with the appropriate sensors, the capability to both store and communicate readings to a centralized location, and the capacity to serve as a filtering point for the data. For these reasons, it was decided that the nodes shall include an Arduino UNO and a Raspberry Pi. Both of the selected sensors are well supported on Arduino but the Arduino alone is insufficient for the storage and communication capabilities required in the designed experiment. As such, the Arduino code used in the prototype focuses on collecting and formatting the sensor data. Then the data are forwarded to the Raspberry Pi. The Raspberry Pi board satisfies the remaining requirements by enabling network connectivity and on-device data backups for the sensor nodes. All nodes transmit data to a shared Sparkfun server.

The selected sensors are wired to the Arduino for both power and communication. The IAQ-2000 sensor runs from the available 5V source, drawing between 22mA and 30mA during operation [15], and reads using an available analog I/O pin. The DHT22 sensor, attached to the 3.3V source, uses 2.5mA during a data request [17]. This request is communicated through the assigned digital I/O pin. The Arduino Uno board receives power from a USB connection to the Raspberry Pi which also permits serial communication between the two devices. As the main processing unit, the Raspberry Pi is powered through a wall adapter, and for this experiment, is provided network connectivity through an ethernet connection to minimize any collisions or interference with wireless communication during data collection process.

B. Data Collection

The experiment took place at the third floor of School of Engineering, at the University of Guelph, Canada, shown in Fig. 2. The duration of the experiment was one week.

1) Monitoring area: In total, six prototypes were placed in two laboratories with different layouts and number of occupants. Four prototypes were placed in Laboratory 1, shown in Fig. 3(a), each with effective coverage of 4 desks. Two prototypes were positioned to cover the doors to the lab. The remaining two prototypes were placed in Laboratory 2, shown in Fig. 3(b).

2) Filtering: Each prototype takes environmental readings at a rate of 0.1Hz. Messages were omitted from transmission to the central server if the sensor readings had not changed by at least 5 ppm of CO₂, and if there was no observed change in temperature and humidity readings. The data was reduced to an effective rate of once per minute by taking the median...
of any readings occurring in the same minute, on a per node basis.

3) Data Labelling: Occupancy was labelled manually throughout each day at asynchronous intervals. This was done to ensure a sufficient breadth of normal activity would be captured for over the experiment period. Due to expected rate of change of occupancy, each label was propagated to a window of 10 minutes centred around the time of observation. This propagation was not applied if another label occurred within that time frame, in which case the label propagated half way to the neighbouring sample.

IV. EXPERIMENTAL RESULTS

The experiment conducted serves to assess the ability of the proposed prototype to support sufficient environmental observation for occupancy based models.

A. Acquired Data

The experiment was conducted between the dates of April 8, and April 15, 2017. In this time, each of the sensor nodes contributed an average of 54,140 samples to the centralized server. Variation in specific contribution counts is attributed to sensor threshold policy for transmission omission under differing levels of activity. The combined samples amount to a total of 15.37MB.

As volatile organic compounds have been shown previously to be a strong environmental measure for occupancy [13], the acquired data is considered with respect to a common basis, being the observed \( \text{CO}_2 \) concentration. In consideration of the correlations presented in works of indoor air quality, the concentration measurements are shown along with the temperature readings in Fig. 4 as well as the humidity readings in Fig. 5 for the complete test duration. The manually obtained occupancy labels are shown in Fig. 8. Together these figures allow for inspection of independent and combined contributions of each respective signal. These figures present the readings of a subset of the sensor nodes for clarity. In each figure, \( \text{CO}_2 \) concentration is shown as the lower data series.

Throughout the span of the experiment, the sensor nodes common to each laboratory setting show minimal deviation in the measurements of both temperature and humidity. For this indoor setting, nodes in the same space would be subject to the same ventilation conditions and thus a well mixed thermal volume is a reasonable model. Due to the scale and rate of changes in \( \text{CO}_2 \) as a result of local area occupants, it is still possible to observe unique events in the shared air volume. This is seen in the events in the evening of April 9, and in the afternoon of April 13 where each sensor samples unique, rapid changes in concentration.

Figure 6 shows the unique variations in concentration observed by the four sensors deployed in the first laboratory. With fewer deployed sensors, the contrast is immediately apparent in the sensors of the second laboratory setting, shown in Fig. 7. The second laboratory is a more compact space and the distance from a sensor node to each occupant is reduced compared to the distances in the larger laboratory.

A consequence of this proximity is seen in the sustained elevation in concentrations observed by sensor five. This compact placement is also more likely to create a larger overlap in the observation areas of the sensor nodes. Optimizing the union
Following a strong precipitation system on the morning of April 11, the outside windows in both laboratory settings were opened at 13:00, remaining open until 9:00 on April 12. This period aligns with the sensor readings of above $22^\circ C$ and greater than 30% humidity seen in Fig. 4 and Fig. 5. The mix of outside air following the storm is the cause of the peaks in both temperature and humidity at $24.4^\circ C$ and 45.9% respectively, in the evening of April 11. During these elevated environmental measurements, an increase in variance of $CO_2$ concentration was observed.

The occupancy labels shown in Fig. 8 are those acquired from the windowed neighbour propagation of the original, manual occupancy labels. Of particular interest is the narrow crest in $CO_2$ concentration occurring in the afternoon of April 10 in the period of high label density. This region presents a key example of the complex relationship between environmental measures and occupancy as this period is also subject to significant increases in both temperature and humidity. This limits the effectiveness of single variable separation methods.

### B. Discussion

A number of outlier samples were observed in $CO_2$ concentration throughout the duration of the experiment. These rapid changes in concentration were most noticeable in the readings from sensors nodes 1 and 2 which were located near the doors of Laboratory 1. For this reason it is expected that these deviations are a consequence of transitions of occupants not local to the sensor’s area, resulting values beyond those expected for static occupancy. For this experiment, the median filtering for minute scale resolution is the primary means of noise reduction for the model. Incorporating a moving average filter and further tuning the threshold policy for transmission is expected to diminish the significance of these events.

Table I shows the correlation of $CO_2$ and Humidity measurements at each sensor location in the varying states of occupancy. This relation of environmental measures is considered in air quality applications [12] but not considered for previous occupancy applications. Further investigation of the environmental state is shown in Table II which contains a similar correlation study targeting instead, the role of temperature.

For the majority of cases, temperature was observed to have a less significant correlation to $CO_2$ than that of humidity in the same cases. Infrequently observed situations, such as three occupants in the area of sensor two which was observed at only three instances over the week, were omitted from this study.

The experimental results have presented examples of the non-linear and multivariate relationships between the environmental measurements and local area occupancy. In contrast to the previous work, it is thus expected that correlation, a linear evaluation, will have limitations in effectively representing the complex relationship observed in standard user settings. In consumer applications, lab quality control of the temperature and humidity in the environment cannot be guaranteed, thus this complexity must be measured and considered to best model the underlying $CO_2$ occupancy model. Whereas air quality applications have investigated the relation between

<table>
<thead>
<tr>
<th>Sensor ID</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Any</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.5178</td>
<td>0.4350</td>
<td>0.4070</td>
<td>0.6130</td>
<td>0.4734</td>
</tr>
<tr>
<td>2</td>
<td>0.5038</td>
<td>0.4022</td>
<td>0.2294</td>
<td>-</td>
<td>0.3370</td>
</tr>
<tr>
<td>3</td>
<td>0.6339</td>
<td>0.5325</td>
<td>0.5560</td>
<td>0.2675</td>
<td>0.3761</td>
</tr>
<tr>
<td>4</td>
<td>0.0160</td>
<td>0.4160</td>
<td>0.4065</td>
<td>0.6793</td>
<td>0.4286</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sensor ID</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Any</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0836</td>
<td>0.2916</td>
<td>0.1356</td>
<td>0.4977</td>
<td>0.2671</td>
</tr>
<tr>
<td>2</td>
<td>0.4071</td>
<td>0.2915</td>
<td>-0.2807</td>
<td>-0.1425</td>
<td>0.1425</td>
</tr>
<tr>
<td>3</td>
<td>0.6557</td>
<td>0.2903</td>
<td>-0.1629</td>
<td>-0.3784</td>
<td>0.1380</td>
</tr>
<tr>
<td>4</td>
<td>0.4120</td>
<td>0.0833</td>
<td>0.1282</td>
<td>0.4369</td>
<td>0.2644</td>
</tr>
</tbody>
</table>
humidity and CO₂ for static occupancy, and occupancy applications consider primarily changes in CO₂ with occupancy, this prototype permits the incorporation of both into a single network deployment.

Among sensor nodes in a shared space there is variation in the correlation for equivalent occupancy counts. This is attributed to the limitations in the sensor placements in that not all sensors could be placed exactly equidistant to the four desks in their local areas. As a result, the influence of a given occupant location is not guaranteed to be equal, leading to variation in readings at for each occupancy count depending on the observed permutation of seating.

Not all possible influences and correlations between the available signals where considered in this experiment. While not considered as the basis for this work, temperature and humidity were observed to have a minimum correlation of 0.84 across all sensor nodes. For future applications of this prototype, a combined heat index measurement may serve as a suitable replacement in order to reduce the dimensionality of the relational problem. For this prototype, it is better to have the individual metrics available and to perform this operation on the centralized data so that the prototype network may be feasibly transferred to other applications.

V. CONCLUSION

In this work a prototype WSN for environmental monitoring in a smart building was presented. Through localized measurement of temperature, humidity and CO₂ concentration at each sensing node, the prototype provides the necessary information to enable occupancy context for a smart system. In addition, the proposed prototype possesses the capability to attribute known environmental impacts on CO₂ concentration, not regularly considered previous occupancy models. As an anonymous, non-invasive, and simple solution, refinement of the prototype would enable it to serve as the basis for a smart management system.

To better investigate the specific complex relationships existing between environmental measurements and occupancy, further tests of extended duration are required. Once sufficient data is acquired, the prototype may be equipped with the necessary functionality to learn and model these relationships, allowing for performance validation against any competing occupancy methods. In this way, the proposed prototype may be upgraded to serve as a test bed for learning algorithms with the intent of improving the generalized deployment of the system.

REFERENCES