Building Simulation 2009

Eleventh International IBPSA Conference
Glasgow, Scotland
July 27-30, 2009

INFORMATION-THEORETIC ENVIRONMENTAL FEATURES SELECTION FOR OCCUPANCY DETECTION IN OPEN OFFICES

Khee Poh Lam[1], Michael Höynck[2], Rui Zhang[1], Burton Andrews[2], Yun-Shang Chiou[1], Bing Dong[1], Diego Benitez[2]
[1]Center for Building Performance and Diagnostics, Carnegie Mellon University, Pittsburgh, PA, 15213
[2]Research and Technology Center, Robert Bosch LLC, Pittsburgh, PA, 15212

ABSTRACT

Knowing the presence or the actual number of occupants in a space at any given time is essential for the effective management of various building operation functions such as security and environmental control (e.g., lighting, HVAC). In the past, motion detection using Passive Infrared (PIR) sensors has been widely deployed in commercial buildings and can provide data on "presence" status. However, there are known limitations with PIR sensors, even for occupant presence detection, in that detection error can occur when the occupant is stationary or performing common tasks in the office space involving small movement such as typing or reading. Moreever, PIR can not detect the number of the occupants in the space.

As occupants "interact" with the indoor environment, they will affect environmental conditions through the emission of CO_2, heat and sound, and relatively little effort has been reported in the literature on utilizing this environmental sensing data for occupancy detection. This paper presents the findings of a study conducted at the Intelligent Workplace (IW) at Carnegie Mellon University (CMU) to address this question by exploring the most effective environmental features for occupancy level detection. A sensor network with robust, inexpensive, non-intrusive sensors such as CO_2, temperature, relative humidity, and acoustics is deployed in an open-plan office space in the IW. Using information theory, the physical correlation between the number of occupants and various combinations of features extracted from sensor data from a 10 week period is studied. The results show significant correlation between features extracted from humidity, acoustics, and CO_2, while little correlation with temperature data. Using features from multiple sensors increases correlation further, and over 90% information gain is acquired when at least six of the most informative features are combined. This work provides a foundation for future studies on using ambient environmental sensor data for occupancy detection.

KEYWORDS

Occupancy detection, environmental sensor network, information theory, feature selection.

INTRODUCTION

The actual number of occupants in a space at any given time is essential for the effective management of various building operation functions such as security and environmental control (e.g., lighting, HVAC). It is also important for first response decision making under emergencies such as an outburst of fire or poisonous gas leakage. For example, knowing the location of occupants and number of occupants before the emergencys' occurrence will help with deciding firefighter's rescue paths, and also facilitate the decision making of the operation modes of fans and fire-doors. The earlier the decision is made for fans and fire-doors, the higher the chance that the fire will be controlled or stopped.

Sensor network deployments have become more and more common practice in buildings. Numbers of research studies have been conducted on extracting higher level or abstract information from physical sets of data that are being collected. Sensor networks have been used for monitoring the structural health of buildings (Kinawi, Reda Tsha, and El-Sheimy (2002)), detecting a pollutant source in a building (Sohn et al. (2002)), detecting the strength of gas source (Federspiel (1997)), and detecting a pollutant source or leakage location of the water supply system (Laird, Biegler, and van Bloemen Waanders (2007)).

This study is based on the hypothesis that as occupants "interact" with the indoor environment, they will affect environmental conditions through the emission of CO_2, heat, moisture and sound, and thus indoor environmental variables such as CO_2 concentration, temperature, relative humidity, and acoustics. These environmental variables may be used for measuring these interactions and inferring the underlying occupancy level.

This paper presents the findings of a study conducted at the Intelligent Workplace (IW) at Carnegie Mellon University exploring the effectiveness of environmental sensing variables for occupancy level detection in an open plan office environment. A sensor network with robust, inexpensive, non-intrusive sensors such as CO_2, temperature, relative humidity, and acoustics is deployed in an open-plan office space in the IW. Using information theory, the physical correlation between the numbers of occupants and various combinations of features extracted from sensor data from a 10 weeks period is studied. After next discussing previous work on sensor-based occupancy detection, we present the experimental setup and methodology followed by results and conclusions.

PREVIOUS WORK

Methods of Occupancy Detection

The methods used for occupancy detection can be classified into three categories:

1. Direct sensing method: The direct sensing method of occupancy detection relates to the deployment of sensors or sensor networks for directly sensing or tracking the presence of occupants through, for example, the use of motion detectors or mobile unit tracking.

Some research efforts focused on occupant tracking include the Georgia Tech Aware Home (Lesser et al. (1999)), MIT Intelligent Room (Torrance (1995)), MavHome from University of Texas Arlington (Cook et al. (2003)) and University of Colorado, Boulder, Neural Network Adaptive Home (Mozer (1998)) which implemented an array of sensors to detect and determine occupancy locations. Harter and Hopper (1994) exploited the topology of a fixed network of wireless receivers as the basis for a location system. The position of a mobile wireless transmitter, which is being carried by an occupant, is determined by the identity of the receivers within the network. The locations boundaries are difficult to define, but if the wavelength in use is such that transmissions do not pass through the walls, then the walls become the location boundaries. The limitation of this system lies in the restriction that the sensors must be carried by occupants, and no location boundaries exist in an open plan office. Similar technologies have been used to locate objects in offices (Satish, Cengiz, and Deborah (2000)) and to track patients in hospitals (Bauer et al. (2000)).

Motion detection using Passive Infrared (PIR) sensors is commonly used in commercial buildings and can provide data on presence status. However, there are known limitations with PIR sensors, even for occupant presence detection, in that detection error can occur when the occupant is stationary or performing common tasks in the office space involving small movement such as typing or reading.

2. Modeling method: Some research has exploited the effectiveness of statistical data analysis methods in occupancy modeling. Page et al. (2008) have developed a stochastic model of occupant's presence states for building simulation tools. Based on an inhomogeneous Markov chain interrupted by occasional periods of long absence, the model will generate a time series of the states of presence of each occupant for each space in the building. Prior knowledge of the occupants' behavior is needed as the inputs of the model, which include a profile of the probability of presence state, a parameter of mobility, and a distribution of the period of long absence. The model's effectiveness is weakened by the findings from Wang, Federspiel, and Rubinstein (2005), who studied the statistical properties of occupancy in single occupied offices of a large office building in San Francisco. It is found that vacancy intervals follow an exponential distribution, while the distribution of occupancy intervals is time varying.

3. Combined method: The combined method is the combination of advanced analysis methods with physical measurements which includes direct or other types of occupancy measurement.

Our proposed approach falls into the combined method category. Wang, Burnett, and Chong (1999) developed both a dynamic and a steady-state model based on the air change rate of the office and the CO_2 concentration to predict the total number of occupants in an open floor office and a lecture theater. A related approach using the Kalman Filtering was used in Federspiel (1997). In Wang, Burnett, and Chong (1999), CFD simulation is utilized to both calibrate and further test the model. Both a dynamic and a steady-state model based on the mass balance of CO_2 are developed, which utilizes the CO_2 concentration of returning air, fresh air and the fresh air supply volume. Results demonstrated estimates of the number of occupants, but were not able to predict the on-off (present/absent) status due to the large scale of occupancy levels targeted in the study (occupant numbers in test cases were relatively large at around 100 people). Our proposed methods aim to detect the number of people in each cubical in an open plan office, with typical number of occupants around 2 to 3. Thus, the methods we propose will be more sensitive, and the on-off presence status will be easily derived. Furthermore, the model proposed by Wang, Burnett, and Chong (1999) is less effective when rooms are under natural ventilation since the volume of fresh air will not be available for calculation.

Dodier et al. (2006) also showed research efforts using a combined detection method. In their research, redundant PIR occupancy sensor networks are deployed in two private offices, and then a Bayesian belief network was utilized to extract the occupancy level from the information from the network. While this approach of using sensor data as an input to a stochastic model for occupancy detection was proved to be quite useful, it will be demonstrated in this paper that additional gains can be acquired through

Figure 1: The Robert L. Preger Intelligent Workplace

the use of a more diverse set of relatively inexpensive sensors targeted at different environmental domains.

EXPERIMENT SETUP

Sensor Network Deployment

The Intelligent Workplace (IW) is a research and education facility at Carnegie Mellon University. It is an open plan office space, with sixteen rooms (bays) and one conference room accommodating faculty members, PhD students, conferences, classes and frequent visitors. The experimental system setup includes:

- An IP video camera network for recording real occupancy numbers;
- An iLON CO_2 sensor network with a sample rate of 2 minutes. An iLON 100 data server is used for data collection, and then Single Object Access Protocol(SOAP) and Java web service are used to export data into a mySQL database;
- A wireless ambient-sensing network, which measures illuminance, temperature, relative humidity (RH), motion and acoustics at a sampling rate of 1 minute. Units are, respectively, percentage (calibrated), degrees Celsius, percentage, binary, and percentage (calibrated);
- A centralized database for automatically retrieving data from different sources continuously.

Air Conditioning System in the IW

The correlation between number of occupants and environmental variables will be influenced by the air conditioning systems. In this study, the air conditioning systems in the Intelligent Workplace (IW) are comprised of: a radiant mullion cooling and heating system and floor based ventilation system. Heating and cooling is controlled by the indoor air temperature set points, and sensors for heating and cooling controls are located at two ends of the long axis of the IW. Ventilation control is based on CO_2 concentration set point in the space. The control sensors are located at various locations in the IW.

METHODOLOGY

An information theory-based feature analysis is carried out to select the most effective features from each of the sensing domains. Sensing domain here is defined as any measurement or feature derived from a specific physical variable in the environment such as CO_2 and temperature. The best combinations of features across different domains are then selected. Data mining and machine learning tool of Accelerated Statistical Learning (ASL) developed at the Auton Lab of the Robotics Institute at CMU is used to conduct the computation.

Information Theory

Information theory is generally considered to be founded in 1948 by Claude Shannon in his seminal work, "A Mathematical Theory of Communication" (Shannon (1948)). It was first applied in data compression and communication. Because of its rigorous mathematic proof, and effectiveness in quantifying the uncertainty of random variables, it has been used in a broad range of applications such as statistical inference, natural language processing, molecular codes and etc. (Anderson (2007)).

An underlying concept in information theory is that of entropy, which characterizes the amount of uncertainty associated with a random variable. High entropy corresponds to high uncertainty (e.g., in the case of a uniformly distributed random variable), and low entropy corresponds to low uncertainty (e.g., in the case of a variable that always takes the same value). Mathematically speaking, entropy is defined as the (negative) expected value of the log of the probability distribution:

$$H(y) = \sum_{i=0}^{n} -p(y_i)log_2 p(y_i). \quad (1)$$

where:

$H(y)$	entropy
y	a random variable
y_i	the i^{th} instance or possible outcome of random variable y
n	the size of the sample space associated with y
$p(y_i)$	the probability of $y = y_i$

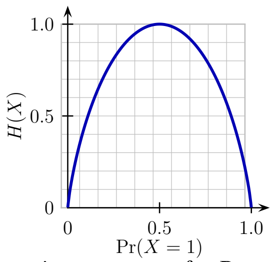

Figure 2: Information entropy of a Bernoulli trial X. If X can assume values 0 and 1, entropy of X is defined as H(X) = -Pr(X=0) log2 Pr(X=0) - Pr(X=1) log2 Pr(X=1). It has value 0 if Pr(X=0)=1 or Pr(X=1)=1. The entropy reaches maximum when Pr(X=0)=Pr(X=1)=1/2 (the value of entropy is then 1) (Brona and Damato (2007)).

As demonstrated in the example plot of Fig.2, entropy will be 0 if the probability of an outcome is zero or 1; there is no uncertainty in the random variable.

The conditional entropy is defined as :

$$H(y|x) = \sum_{j=1}^{A_X} P(x = x_j) H(y|x = x_j), \tag{2}$$

Where the entropy of y given $x = x_j$ is defined as:

$$H(y|x = x_j) = \sum_{i=0}^{n} P(y = y_i|x = x_j) log \frac{1}{P(y = y_i|x = x_j)}. \tag{3}$$

where:

1

$H(y\|x)$	conditional entropy of y given x
$H(y\|x = x_j)$	conditional entropy of y given $x = x_j$
x, y	random variable
x_j	the j^{th} possible outcome of x
A_x	the sample space of random variable x
$P(x = x_j)$	the probability of $x = x_j$
$P(y = y_i\|x = x_j)$	the conditional probability of y given x

The mutual information (referred to here as information gain, IG) between y and x is:

$$IG(x,y) = H(y) - H(y|x). \tag{4}$$

Relative information gain (RIG) is:

$$RIG(x,y) = \frac{IG(y,x)}{H(y)} \cdot 100\%. \tag{5}$$

Note that x and y are perfectly associated if $H(y|x) = H(x|y) = 0$, $IG(y,x) = H(y)$, $IG(x,y) = H(x)$. Knowing x reveals everything about y and vice versa.

On the other hand, x and y are perfectly disassociated if $H(y|x) = H(y)$, $H(x|y) = H(x)$, and $IG(y,x) = IG(x,y) = 0$. Knowing x, does not reveal any information about y and vice versa.

In this research work, information gain is used to study the correlation between occupancy levels and features extracted from various environmental measurements. The random variable y corresponds to the number of occupants in the space. The input variable x corresponds to a vector of environmental features both measured and derived.

ADtree Algorithm

The data mining and machine learning tool of Accelerated Statistical Learning (ASL) developed at the Auton Lab of the Robotics Institute at Carnegie Mellon University is used to run the feature selection. The algorithm behind the tool is based on the ADtree data structure (Moore and Lee (1998)). A more detailed description of the algorithm is available in Moore and Lee (1998) and B. and Moore (1998).

The search for the best features or combination of features used by this tool is an exhaustive search algorithm. The algorithm will evaluate the RIG for all possible combinations of a given number of input features. For example, for a data set with 10 features, a search for the best combination of 3 features will evaluate: $C(10,3) = 120$ possible combinations, where $C(10,3)$ is the 3-combinadic of 10.

ANALYSIS RESULTS

Raw Data and Data Processing

1. Raw data: The data sets from two IW bays (Bay 10 and 13) and two sensing periods are used, which are measured from Jan 29 to March 07, 2008, and from March 17 to April 04, 2008, mainly during the winter and spring seasons. The number of data points for each period is listed in Table 1. The first 5000 data points from each data set are used for analysis. Each of the features is divided into five ranges with nearly equal number of data instances in order to apply the information theory based feature selection.

2. Data Processing: Measurements of temperature, relative humidity and acoustic are reported from the sensor network both as an average value (from a 1 minute, 2 Hz-sampled signal) and an outlier value, which is the largest

Table 1: Data sets used for analysis.

Dataset	Bay	Starting Date	Ending Date	Number of Data Points
B13_P1	13	Jan-29-08	Mar-07-08	21528
B13_P2	13	Mar-17-08	Apr-04-08	7705
B10_P1	10	Jan-29-08	Mar-07-08	20702
B10_P2	10	Mar-17-08	Apr-04-08	7555

deviation from the average value during the 1 minute period. CO_2 data is sampled at a rate of 2 minutes. Outdoor CO_2 level is also recorded, and is denoted as CO_2_Out.

We also computed additional features derived from the raw sampled data, such as the first order difference, which captures temporal changes, and the difference between indoor and outdoor CO_2 concentrations. The features added and their notations are listed as follows:

- First order difference (notation _FD): raw(i) - raw(i-1);
- Second order difference (notation _SD): raw_FD(i) - raw_ FD(i-1);
- Variation of first order difference (notation _FD2): raw(i) - raw(i-2);
- Difference between indoor and outdoor CO_2, and difference between average values and outliers for temperature, acoustic, and relative humidity (notation _ Diff): raw_ average - raw_ outlier or raw_ indoor - raw_ outdoor;
- First order difference of the difference values (notation _Diff_FD): raw_ Diff(i) - raw_ Diff(i-1);
- Second order difference of the difference values (notation _Diff_SD): raw_ Diff_ FD(i) - raw_ Diff_ FD(i-1);
- 20 minutes moving average (notation _MA_20min): $H(y) = \left(\sum_{i-19}^{i} raw(i)\right)/20$;

Individual Sensing Domain Analysis

The purpose of the individual sensing domain analysis is to investigate the correlation between number of occupants and each individual sensing domain. A rank of the correlations of each of the sensing domains will be provided, and the three most effective features for each of the sensing domains will be selected for further combined domain analysis.

Let n_Rule represent a combination that has n features in it. In order to investigate the RIG changes as more features are used in the combinations, RIGs for n_Rule with $n = 3, 4, ..., 9$ are calculated. The resulting trends are shown in the following subsections and corresponding figures for each domain. For each n_Rule, the n features resulting in the highest IG are selected from an exhaustive search over all of the possible combinations. The effectiveness rank of features for each individual sensing domain is then evaluated as the number of times that a given feature appears as one of the best features of the n_Rule computations throughout the four datasets. The three features with the highest effectiveness rank are then selected as the best features.

Table 2: The RIG from CO_2 for each n_Rule across four datasetsets.

♯ features	CO_2	CO_2_FD	CO_2_FD2	CO_2_SD	CO_2_Out	CO_2_Diff	$CO_2\_Diff\_FD$	$CO_2\_Diff\_SD$	$CO_2\_MA\_20min$	RIG
$B13\_P1$										
3			√		√	√				20.12%
4			√		√	√		√		28.37%
5			√		√		√	√	√	40.60%
6			√	√	√	√	√	√	√	52.84%
7			√	√	√	√		√	√	60.45%
8	√		√	√	√	√	√	√	√	67.14%
9	√	√	√	√	√	√	√	√	√	67.14%
sum	2	1	7	4	7	6	4	6	5	
$B13\_P2$										
3	√				√				√	13.68%
4	√		√		√				√	23.30%
5	√		√		√			√	√	35.09%
6	√		√	√	√			√	√	45.23%
7	√		√	√	√	√		√	√	53.58%
8	√		√	√	√	√	√	√	√	55.63%
9	√	√	√	√	√	√	√	√	√	56.60%
sum	7	1	6	4	7	3	2	5	7	
$B10\_P1$										
3					√	√			√	13.68%
4			√		√	√			√	23.30%
5			√		√	√		√	√	35.09%
6			√		√	√		√	√	45.23%
7	√		√	√	√	√		√	√	53.58%
8	√		√	√	√	√	√	√	√	55.63%
9	√	√	√	√	√	√	√	√	√	56.60%
sum	3	1	6	3	7	7	2	5	7	
$B10\_P2$										
3	√				√				√	21.57%
4	√		√		√				√	30.61%
5	√		√		√			√	√	42.44%
6	√		√	√	√			√	√	52.31%
7	√		√	√	√	√		√	√	59.43%
8	√		√	√	√	√	√	√	√	60.18%
9	√	√	√	√	√	√	√	√	√	60.64%
sum	7	1	6	4	7	3	2	5	7	
total	19	4	25	15	28	19	10	21	26	

1. CO_2 :The features selected for each n_Rule across the four datasets from CO_2 are listed in Table 2.

- The highest RIG from CO_2 is 67.14%.
- The trends of RIG are shown in Fig.3. As is shown, RIG increases as the number of features increases; the rate of increase of RIG reduces after the number of features is greater than 7.
- The best features from CO_2 are CO_2_Out, CO_2_FD2 and $CO_2\_MA\_20min$.

2. Temperature:

- The highest RIG from *Temperature* is 37.39%.
- The trends of RIG are shown in Fig.4. As is shown RIG increases as the number of features increases; The rate of increase of RIG reduces after the number of features is greater than 7.

Table 3: The total number of times (effectiveness rank) each feature appears in the n_Rule for Temperature, RelativeHumidity and Acoustic across four datasetsets.

Dataset	raw	raw_FD	raw_FD2	raw_SD	raw_Out	raw_Diff	raw_Diff_FD	raw_Diff_SD	raw_MA_20min
Temperature	18	6	20	22	18	25	13	25	22
Relative Humidity	14	13	26	23	16	27	18	21	10
Acoustic	21	8	17	21	6	28	17	24	26

- The number of times each feature appears in the n_Rule is listed in Table 3, and the best features from $Temperature$ are $Temp\_Diff\_SD$, $Temp\_Diff$ and $Temp\_MA\_20min$.

The effectiveness of each individual sensing domain is ranked according to their maximum average RIG, which is the average maximum RIG of the four datasets with 9_Rule. Results are shown in Fig.7. The rank of the four sensing domains individually from high to low is $Relative$ $Humidity$, $Acoustic$, CO_2 and $Temperature$.

Combined Sensing Domain Analysis

The three most effective features for each of the four sensing domains from the previous individual sensing domain analysis are combined to form a new feature space for occupancy detection, and a similar analysis as used for the individual sensing domains is used on the combined domain.

As shown in Fig.8, the highest RIG is greater than 99% in all of the four datasets. Such high information gain can in part be attributed to"over-fitting" to a particular bay given the large number of feature combinations possible. However, these results strongly suggest the ability to capture a high correlation between appropriate features extracted from ambient sensor data. Furthermore, sufficient information can be gained from only a relatively small number of features. For instance, in Fig. 8, when the number of features is greater than 6, the RIG reaches 90%, and the rate of RIG decreases when the number of features is greater than 6. Thus, only marginal benefit will be achieved with features more than 6. The six most effective features are:

- $CO_2\_MA\_20min$
- CO_2_Out
- CO_2_FD2
- RH_Diff
- $Acou\_Diff$
- $Acou\_MA\_20min$

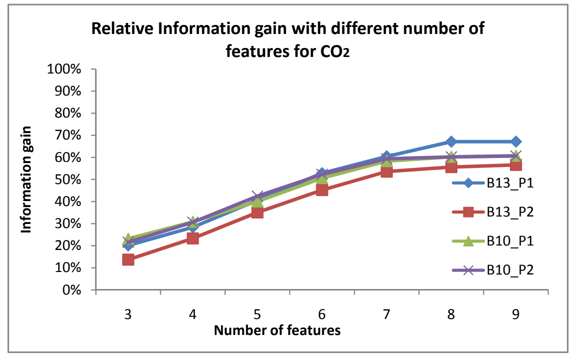

Figure 3: RIG under different number of features for CO_2.

3. Relative Humidity:

- The highest RIG from $Relative$ $Humidity$ is 77.65%.
- The trends of RIG are shown in Fig.5. As is shown RIG increases as the number of features increases; The rate of increase of RIG reduces after the number of features is greater than 7.
- The number of times each feature appears in the n_Rule is listed in Table 3, and the best features from $Relative$ $Humidity$ are RH_Diff, RH_FD2 and RH_SD.

4. Acoustic:

- The highest RIG from $Acoustic$ is 73.42%.
- The trends of RIG are shown in Fig.6. As is shown RIG increases as the number of features increases; The rate of increase of RIG reduces after the number of features is greater than 7.
- The number of times each feature appears in the n_Rule is listed in Table 3, and the best features from $Acoustic$ are $Acou\_Diff$, $Acou\_Diff\_SD$ and $Acou\_MA\_20min$.

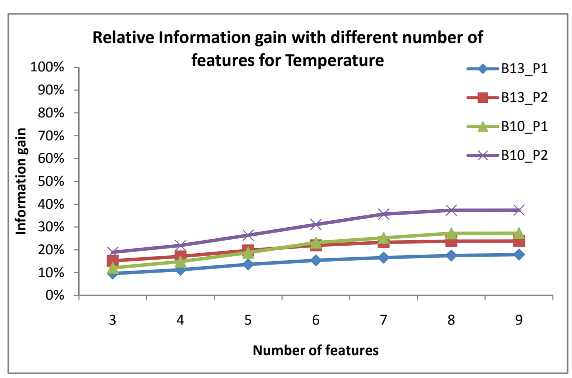

Figure 4: RIG under different number of features for $Temperature$.

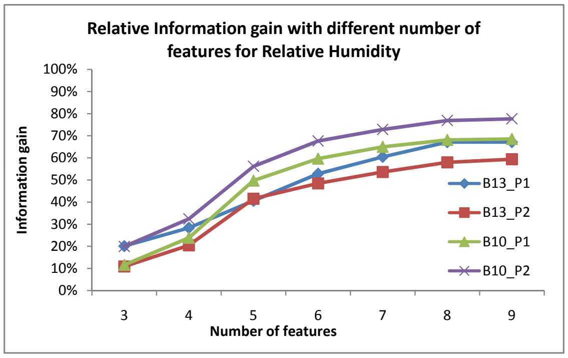

Figure 5: RIG under different number of features for Relative Humidity.

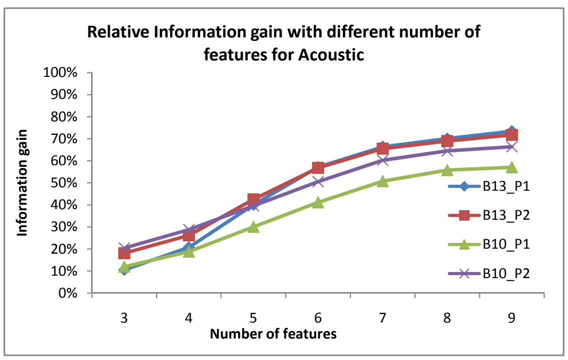

Figure 6: RIG under different number of features for Acoustic.

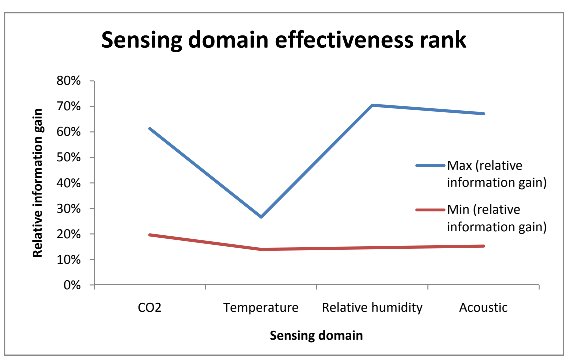

Figure 7: Rank of the effectiveness of individual sensing domain.

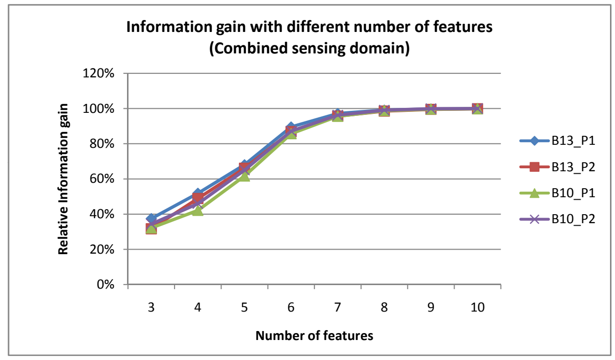

Figure 8: RIG with combined features from four sensing domains.

The results indicate that combinations of environmental variables can indicate the number of occupants in the space to a resolution of 1 to 4 people.

CONCLUSION

Knowing the presence or the actual number of occupants in a space at any given time is essential for the effective management of various building operation functions such as security, emergency response and environmental control. This paper has proposed a hypotheses that indoor environmental variables such as CO_2 concentration, temperature, relative humidity, and acoustics can be utilized to detect occupancy levels. A sensor network with robust, inexpensive, non-intrusive sensors such as CO_2, temperature, relative humidity, and acoustics is deployed in an open-plan office space in the IW. Using information theory, the physical correlation between the numbers of occupants and various combinations of features extracted from sensor data from a 10 week period is studied.

The following conclusions are drawn:

- The results show significant correlations between the environmental variables and the number of occupants in the space. The correlation between the number of occupants and each individual environmental variable ranks as 77.65% for humidity, 73.42% for acoustics, 67.14% for CO_2, and 37.39% for temperature.
- Sufficient correlation has been established with the current combination of features, with marginal benefit from adding more features. Using features from multiple sensors increases correlation significantly, and over 90% RIG is acquired when six features are combined across different domains. The best 6 features across the different sensing domains are $CO_2\_MA\_20min$, CO_2_Out, CO_2_FD2, RH_Diff,

Acou_Diff and *Acou_MA_20min*.

ACKNOWLEDGMENT

This work is supported in part by the Bosch Research and Technology Center, Pittsburgh, PA.

REFERENCES

Anderson, D. R. 2007. Some Background on Why People in the Empirical Sciences Mmay Want to Better Understand the Information-Theoretic Methods.

B., Anderson, and A Moore. 1998. "ADtrees for Fast Counting and for Fast Learning of Association Rules." *In Proceedings Fourth International Conference on Knowledge Discovery and Data Mining*. ACM Press, 134–138.

Bauer, P., M. Sichitiu, R. Istepanian, and K. Premaratne. 2000. "The Mobile Patient: Wireless Distributed Sensor Networks for Patient Monitoring and Care." *Proceeding of the IEEE Conference on Information Technology Applications in Biomedicine (IEEE ITAB 2000)*, November.

Brona, and A Damato. 2007. Information Entropy of a En:Bernoulli Trial X.

Cook, D. J., M. Youngblood, Heierman E. O. III, Gopalratnam K., S. Rao, A. Litvin, and F. Khawaja. 2003. "MavHome: An Agent-Based Smart Home." *Pervasive Computing and Communications, IEEE International Conference on* 0:521.

Dodier, R. H., G. P. Henze, D. K. Tiller, and X. Guo. 2006. "Building Occupancy Detection Through Sensor Belief Networks." *Energy and Buildings* 38 (9): 1033 – 1043.

Federspiel, C. C. 1997. "Estimating the Inputs of Gas Transport Processes in Buildings." *IEEE Transactions on Control Systems Technology* 5:480–489.

Harter, A., and A. Hopper. 1994. "A Distributed Location System for the Active Office." *IEEE Network, Special Issue on Distributed Systems for Telecommunications*, January.

Kinawi, H., M. M. Reda Tsha, and N. El-Sheimy. 2002. "Gpsr: Greedy Perimeter Stateless Wouting for Wireless Networks." *27th Annual IEEE Conference on Local Computer Networks*.

Laird, C. D., L. T. Biegler, and B. G. van Bloemen Waanders. 2007. "Real-time, Large Scale Optimization of Water Network Systems Using a Subdomain Approach." *Real-Time PDE-Constrained Optimization, SIAM Series in Computational Science and Engineering*, pp. 291–308.

Lesser, V. Atighetchi, M., A. Raja, V. Regis, Wagner T., P. Xuan, and S. X. Zhang. 1999. "The Intelligent Home Testbed." *Proceedings of the Autonomy Control Software Workshop*.

Moore, A, and M. S. Lee. 1998. "Cached Sufficient Statistics for Efficient Machine Learning with Large Datasets." *Journal of Artificial Intelligence Research* 8 (March): 67–91.

Mozer, M. 1998. "The Neural Network House: An Environment that Adapts to its Inhabitants." *Proceedings of the American Association for Artificial Intelligence*.

Page, J., D. Robinson, N. Morel, and J.-L. Scartezzini. 2008. "A Generalised Stochastic Model for the Simulation of Occupant Presence." *Energy and Buildings* 40 (2): 83 – 98.

Satish, K., A. Cengiz, and E. Deborah. 2000. "Scalable Object-Tracking Through Unattended Techniques (SCOUT)." *Proceedings of the 8th International Conference on Network Protocols(ICNP)*, November.

Shannon, C. E. 1948. "A Mathematical Theory of Communication." *Bell System Tech. J.* 27:379–423, 623–656. http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html.

Sohn, M. D., P. Reynolds, A. J. Gadgil, and R. G. Sextro. 2002. "Rapidly Locating Pources and Predicting Contaminant Dispersion in Buildings." *Proceeding of Indoor Air*, November, 17–21.

Torrance, M. C. 1995. "Advances in Human-Computer Interaction: The Intelligent Room." *In Working Notes of the CHI 95 Research Symposium*.

Wang, D., C. C. Federspiel, and F. Rubinstein. 2005. "Modeling Occupancy in Single Person Offices." *Energy and Buildings* 37 (2): 121 – 126.

Wang, S., J. Burnett, and H. Chong. 1999. "Experimental Validation of CO2-Based Occupancy Detection for Demand-Controlled Ventilation." *Indoor and Built Environment* 8 (6): 377–391.