Virtual occupancy for real-time occupancy information in buildings

Citation for published version (APA):

DOI:
10.1016/j.buildenv.2015.06.019

Document status and date:
Published: 20/06/2015

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:
• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.
Virtual occupancy sensors for real-time occupancy information in buildings

Yang Zhao*, Wim Zeiler, Gert Boxem, Timi Labedan

Department of the Built Environment, Eindhoven University of Technology, Eindhoven, The Netherlands

Abstract

This study aims at developing a generic, feasible and low cost occupancy detection solution to provide reliable real-time occupancy information in buildings. Currently, various low cost or even free occupancy measurements are common in offices along with the popularization of information technologies. An information fusion method is proposed to integrate multiple occupancy measurements for reliable real-time occupancy information using the Bayesian belief network (BBN) algorithm. Based on this method, two types of virtual occupancy sensor are developed at room-level and working zone-level respectively. The room level virtual occupancy sensors are composed of physical occupancy sensors, chair sensor, keyboard and mouse amongst others. The working zone-level virtual occupancy sensors are developed based on real-time GPS location and Wi-Fi connection from smart device like smart phones and occupancy access information from building management systems. The developments of these two types of virtual occupancy sensors can be conducted automatically with functions of self-learning, self-performance assessment and fault detection. The performances of the developed virtual sensors are evaluated in two private office rooms. Results show that the developed virtual occupancy sensor are reliable and effective in providing real-time occupancy information. The paper also discusses application of the virtual occupancy sensors for demand driven HVAC operations.

1. Introduction

Demand driven operations of comfort systems like heating, ventilation and air conditioning (HVAC) systems in buildings can avoid significant amount of energy wastes compared with conventional operations. The term demand mainly refers to real-time occupancy requirements of indoor air quality and thermal comfort. Energy wastes occur when the supplies of HVAC systems are more than demands; an example could be in an instant when the HVAC system for a room is operational but the room is unoccupied for a long period. Past studies indicate that about 10%–20% of HVAC energy consumption and 30% of lighting energy consumption can be reduced using demand driven controls by considering real-time occupancy information in building management systems. This is because real-time occupancy population is usually less than the design population. For instance, the actual occupancy diversity factors in an office building could be as much as 46% lower than the values published in ASHRAE 90.1 2004 energy cost method guidelines. Further evidence of this trend is illustrated in the references.

Real-time occupancy information is crucial for demand driven HVAC operations. In recent years, various types of occupancy detection methods have been developed using tools such as passive infrared (PIR) sensors, cameras, wireless sensor networks, radio frequency identification (RFID) sensors and CO2 sensors amongst others. Each method has its own advantages and applicability; in addition each has specific limitation(s) in real practice. It is known that occupancy detection might be unreliable when only based on an individual physical sensor; this is due to uncertainties of occupant activities (for example steady occupants cannot be detected by PIR sensors) and sensitivities of occupancy sensors. Dodier et al. found that the six PIR sensors in their study detected 20%, 39%, 50%, 70%, 76% and 76% of occupancy events respectively. Duarte et al. found that 2.56% of data measured by 629 infrared/ultrasonic sensors in an office building should be removed since the measurements indicated continuous occupancy for 48 h or more. They estimated that the same amount of failure data should exist failing to detect occupants. Demand driven HVAC operations based on individual physical occupancy
sensor of such a quality cannot guarantee indoor air quality and thermal comfort. Occupants’ complaints would consequently result to change from demand driven operations back to conventional operations or manual operations.

Information fusion is a technique that combines use of independent sensor measurements and information from multiple sources to improve accuracy and reliability. Along with the popularization of information technologies, occupancy information/measurements of multiply sources are available in low costs. It is possible to consider introduction of information fusion in occupancy detection. Personal computers are presently quite common in offices. Subsequently, real-time status of keyboard and mouse could be used to provide meaningful occupancy status of the room in which the computer is located. The use of smart devices like smart phones are widely. Wearable smart devices such as smart watches and smart bands tend to be popular. With wireless technologies inside, smart devices have potentials to provide real-time individual occupancy-related information, for example. Wi-Fi connection from smart devices or from wireless routers (details can be found in Section 3.1), GPS location from smart devices and RFID information. The information teased from this can reveal individual occupancy status that may be useful for demand driven operations. For can be used in the HVAC systems for a room can be switched on prior to arrival of occupant if the GPS location shows that the occupant is approaching the building and vice versa. Some types of information cannot provide accurate occupancy status directly; however the information provided may still be useful in improving the reliabilities of occupancy detection through information fusion. For instance, the smart phone of an occupant always connects to a certain Wi-Fi hotspot during working hours. The belief of occupancy in the room would as a result be lower than in the case when there is no connection. Building management systems also provide useful information such as entrance information and car parking information. A straightforward idea of utilizing such type of information would be directly tap individual information source for demand driven HVAC operations. Problems would however arise when parts of these information sources fail to work; example would be in cases whereby the smart phone is powered off, or when faulty measurements are registered. Occupancy detection would thus be more reliable by fusing these measurements with multiple information sources. However, comprehensive literature survey reveal that information fusion for occupancy detection has not attracted much attention.

Bayesian belief network (BBN) is a powerful tool proposed in the early 1980s which has been successfully applied in the domain of information fusion, knowledge discovery and probabilistic inference [18,19]. In the domain of HVAC, applications can be found in fusing diagnostic information for fault detection and diagnosis [20,21], Hawarah et al. [22] developed a BBN to predict and diagnose user’s behavior in housing for home automation system. Compared with other information fusion algorithms in theory, BBN could be an outstanding in developing information fusion models for occupancy detection. Firstly, it is possible to develop a generic BBN for occupancy detection. A specific BBN for a certain situation can be easily developed based on the generic BBN considering various types and amounts of measurements/information. Secondly, parameters in BBN can be assigned by estimations of experts. In this way, the BBN can work without training at the beginning of usage. In practical applications, there are generally no historical occupancy data for training. Thirdly, BBN is a straightforward method which can be easily understood by operators. Also, the associated computation load is low. Dodier et al. [10] developed a BBN for occupancy detection using three PIR sensors and a detector of telephone handset status for occupancy detection in two private offices respectively. Results showed that the method improved occupancy detection accuracy significantly. In Dodier et al.’s work, occupancy patterns were not considered in the BBN. Only two types of sensors were used. Ideally, the BBN should be trained using ground truth values of occupancy status; this might be a major barrier in practical applications. The term ground truth value of occupancy status refers to the real occupancy status. Measurements from occupancy sensors cannot be regarded as ground truth values of occupancy statuses due to associated inaccuracies. Ground truth values of occupancy statuses in Dodier et al.’s work were obtained by human observers.

This study attempts to propose a generic, reliable and low cost occupancy detection solution to provide reliable real-time occupancy information for demand driven HVAC operations in buildings. Two types of virtual occupancy sensors are developed for individual occupants, i.e. room-level virtual occupancy sensors and working zone-level occupancy sensors. The term virtual sensor refers to an emerging form of multiply sensors and information, which has no differences compared with a physical sensor from the users’ perspective. The room-level virtual occupancy sensors provide occupancy status in a private office room, i.e. occupied and vacant. The working zone-level virtual occupancy sensors provide occupancy status of individual occupant in a working zone, such as leaving, coming, in the working zone, or not in the working zone. They are useful inputs for demand driven HVAC operations.

This study is based on Dodier et al.’s work [6] with following improvements: (i) various types of sensors are considered after a comprehensive survey; (ii) a new sensor and a new information source are proposed, i.e. chair sensor, and keyboard and mouse status; (iii) occupancy patterns are also considered in the BBNs; (iv) Expectation—maximization (EM) algorithm is introduced to train the BBNs. Its benefits include avoidance of requirements for ground truth values in the developments of BBNs. In addition a sensor performance assessment method is proposed to detect inefficient and faulty sensors/information sources. The proposed virtual occupancy sensors are evaluated in two office rooms.

2. Bayesian belief network (BBN) and expectation—maximization (EM)

This section presents an introduction of the two algorithms used in the development of virtual occupancy sensors.

2.1. Bayesian belief network theory

A BBN is defined by two components, i.e. structure and parameters. The structure of a BBN is a graphical and qualitative illustration of the relations among the modeled variables. It is a directed acyclic graph in which nodes represent variables and arcs represent direct probabilistic dependences among them. Each arc starts from a parent node and ends at a child node. An example is shown in Section 4.3. A node contains all possible states of the variable it represented. For instance, node InRoom? in Fig. 6 has two states, i.e. Occupied and Vacant. The parameters of a BBN represent quantitative direct probabilistic dependences among nodes. Each child node has a conditional probability table based on parental values. An example is shown in Section 3.4.

The inference in a BBN is to calculate posterior probability P(Xj/Xk), where Xj is the node of interest and Xk is a node or a set of nodes which is/are observed. In the case of the BBN in Fig. 6, Xi is the InRoom? node. Xj is a set of nodes whose states are observed like node TimeOfDay, PIR2 and Keyboard + mouse. The posterior probability P(Xj/Xk) is the belief (probability) of occupancy. More details about BBN are illustrated in Zhao et al. [21], Xiao et al. [20] and [18,19].
2.2. Expectation–maximization (EM) algorithm

In this study, the structures of BBNs are determined according to the available sensors/information sources, examples as shown in Figs. 2 and 3. The parameters in the BBNs can be assigned by experts or learned from historical data. The latter one is more effective. Historical data for the occupancy interaction info nodes, occupancy detection sensor nodes and time information nodes in Figs. 2 and 3 can be recorded in historical database. However, there are generally no records of the ground truth values of occupancy statuses which is represented by the node \textit{OnnRooms}? (Fig. 2) and \textit{InWorkingZone}? (Fig. 3). Such a problem can be solved by the EM algorithm which is powerful to estimate parameters of BBNs using incomplete training data [23,24].

In the EM algorithm, given the historical data and the structure of BBN, there are two steps to estimate parameters $\theta$, i.e. the expectation step (E-step) and the maximization step (M-step). The two steps are alternated iteratively until a stopping criterion is satisfied. Given the initial values of parameters $\theta$, the E-step is to compute the expected data frequencies. The M-step is to maximize the log-likelihood of the parameters under the expected data frequencies. More descriptions about EM algorithm can be found in Refs. [23,24].

3. Development of the virtual occupancy sensors

This section presents the methodology of the two virtual occupancy sensors. A comprehensive survey is made on the common occupancy sensors/information sources firstly, based on which two generic virtual occupancy sensors are proposed respectively. An offline builder is proposed to develop the virtual occupancy sensors automatically. Sensor performance assessment method and fault detection method are also proposed to guarantee the reliability of the sensors.

3.1. Common occupancy information/measurements

Table 1 shows a comprehensive survey of common occupancy sensors/information sources that are useful for real-time occupancy detection in office buildings. They are classified into five categories, i.e. time information (occupancy pattern), physical occupancy sensor, occupancy interaction information, occupancy access information and smart device information.

3.1.1. Time information

Field surveys demonstrated that an individual occupant always shows occupancy patterns statically. In this study, such patterns are...
described by the conditional probability tables of the time information nodes (e.g. node TimeOfDay). Working hours in a day are divided into a list of states with a constant interval, e.g. every 20 min. Duarte et al. found that, the occupancy in the day of the week has a significance difference with the middle of the week following a similar profile but statistically higher occupancy on Mondays and early departure on Fridays (about 30 min earlier) in an office building. Such kind of patterns can be presented by node DayOfWeek. However, much more training data are needed if DayOfWeek is introduced in the BBN. The conditional probabilities can be learned from historical data or assigned by experts.

3.1.2. Physical occupancy sensor
Surplus sensors can improve the reliabilities of occupancy detection results significantly. It becomes acceptable to building owners to install surplus sensors in practical applications. It is worth noting that some wearable devices have been equipped with radio frequency identification sensors. This makes them have potential for application in occupancy detection.

3.1.3. Occupancy interaction information
Occupants always interact with items in office rooms such as chairs, windows, computers and keyboards and mouse. For instance, the touches/movements of keyboard or mouse of a desktop computer indicates that somebody is using the computer in the room. Such kind of information can be uploaded over internet to building management systems. The wireless technology also allows measuring occupancy interaction information with low costs like the proposed chair sensor.

3.1.4. Occupancy access information
Occupancy access information from building management systems is helpful in detecting arrival and exit events of individual occupant, e.g. parking access information and entrance guard information.
3.1.5. Smart device info

Information from smart device may assist in detecting an occupant in working zone, such as Wi-Fi connection, GPS location and radio frequency identification. Smart devices include smart phone, smart watches and smart bands, etc. Smart phones are very popular currently. Wi-Fi hotspots are also widely available in office environment. Every Wi-Fi hotspot has a unique and non-editable BSSID (Basic Service Set Identifier). From the phone side, phones can read the BSSID of the connected Wi-Fi hotspot and upload it to building management system. Similarly, from the Wi-Fi hotspot side, it can read the unique identifier of connected phones. For instance, the belief of an occupant in the working zone is high if his/her smart phone is connecting to a Wi-Fi hotspot in the working zone. On the contrary, the belief is low if the smart phone does not connect to the Wi-Fi hotspot. The belief is very low if the smart phone is connecting to another Wi-Fi hotspot outside of the working zone. In the Wi-Fi connections, it is available to get the global unique identifiers of phones and Wi-Fi hotspots.

3.2. Structure of the virtual sensors

The basic idea of the virtual sensors is to obtain accurate occupancy status by fusing various sensors (physical occupancy sensors), information (Occupancy interaction info, occupancy access info and smart device info) and occupancy patterns (time information) in Table 1 by information fusion using BBNs. Two virtual sensors are developed, i.e. room-level virtual sensor and working zone-level virtual sensor, as illustrated in Fig. 1.

3.2.1. Structure of the room-level sensor

A generic BBN for the room-level virtual occupancy sensor is proposed as illustrated in Fig. 2. The inputs of this BBN are observed...
states of the physical occupancy sensor nodes, occupancy reaction info nodes and time information nodes. *TimeOfDay* has a list of states. Current state of *TimeOfDay* is determined by the prevailing log time. A sensor node generally has two states, e.g. detected and undetected. In the BBN, the observed states are set to be observed in the corresponding nodes. For instance, the state undetected of the PIR sensor node is set to be observed (100%) if the PIR sensor detects occupant in Fig. 6. Thereafter, the BBN updates beliefs of states in other unobserved nodes. InRoom? has two states, i.e. occupied and vacant. Outputs are the beliefs of states in InRoom? For instance, the belief of state occupied is 0.9 in node InRoom? in Fig. 6.

### 3.2.2. Structure of the zone-level sensor

Similarly, a generic BBN for the working zone-level virtual occupancy sensor is proposed as illustrated in Fig. 3. Inputs of this BBN are the observed states of smart device info nodes, occupancy access nodes and time information nodes. More states can be considered in node *InWorkingZone?*, e.g. LeavingEvent (occupant is leaving), ComingEvent (occupant is coming) and occupied (occupant is in working zone), according to actual situations. The output is the belief of states in node *InWorkingZone*.

The two nodes, i.e. *InRoom?* in Fig. 2 and *InWorkingZone?* in Fig. 3, make the BBNs to be generic for practical applications. Available sensors or information sources can be integrated into the BBNs by adding an arc from these two nodes to sensor/information node. The selection of time information (occupancy pattern) nodes is determined by amount of training data. All of the three time information nodes can be retained if there are a large amount of training data available. If not, only *TimeOfDay* node is suggested for use.

### 3.3. Automatic development of the virtual occupancy sensors

An automatic offline development method is proposed as illustrated in Fig. 4. The structures of BBNs can be generated automatically according to available sensors or information sources. The parameters of BBNs can be obtained by two ways, i.e. automatic learning using historical data and using default values (without historical data).

If there are insufficient historical data (situation 1 in Fig. 4), parameters of the BBNs are set with default values which present general occupancy patterns and general probability dependences. The default values can be obtained by survey or estimations of experts. If there are enough historical data available (situation 2 in Fig. 4), parameters of the BBN can be obtained by learning from historical data using EM algorithm. Parameters of BBN can also be adaptively updated by learning from accumulated historical data. The BBN developed in situation 1 can be improved continuously using accumulated historical data by the method in Situation 2.

Compared with Dodier et al.’s work [10], the room-level virtual sensors have following improvements: (1) various types of sensors are introduced (as shown in Table 1); (2) occupancy patterns are considered in the BBN; (3) some new measurements are proposed such as chair sensor, and keyboard and mouse; (4) ground truth values are not required in the training data.

### 3.4. Performance assessment and fault detection indicators

The trained conditional probability table between a sensor/information source node and InRoom?/InWorkingZone? node actually indicates the performance of sensor/information source. The performance of a sensor can be presented by negative error ratio and positive error ratio. Taking the room-level occupancy detection for instance, a negative error occurs when a room is occupied but it is detected/regarded as unoccupied. On the other hand, a positive error occurs when a room is not occupied but it is detected/regarded as occupied.

For room-level occupancy detection, the value of probability between the state Detected in a sensor node and the state Vacant in InRoom? is given by the positive error ratio. Similarly, the value of the probability between the state Undetected in a sensor node and the state Occupied in InRoom? is the negative error ratio. Two thresholds are set for them respectively, these are adopted in the study as 0.05 and 0.4 according to experiments from Dodier et al. [10]. An example of a conditional probability table is shown in Table 2. The probabilities are learned by EM using the experimental data in Section 4.

The value 0.62 between states Occupied and Detected is the estimated probability of correct occupancy detection when the room is occupied. The estimated negative error ratio is 0.38. Similarly, the value between states Vacant and Undetected is the estimated probability of correct occupancy detection when there is nobody in the room. The estimated positive error ratio is 0.01.

There are two methods to detect faulty sensors/information sources. The first method is to train BBNs using the data in recent n days and then check the performance of sensors/information sources. The second method is to collect online fault reports from the BBNs. An example is illustrated in Fig. 5. The occupancy detection result is determined according the beliefs of states InRoom? in every time slot. A fault is labeled if the measurement of a sensor is different from the result. A faulty sensor is alarmed if the fault ratio in a moving window exceeds a threshold. Moving windows for fault detection start from an interval before current time and end in current time.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Common occupancy sensors/information sources useful for real-time occupancy detection.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category</td>
<td>Occupancy information</td>
</tr>
<tr>
<td>Time information (occupancy pattern)</td>
<td>Time of the day Day of the week Holiday? Agenda</td>
</tr>
<tr>
<td>Occupancy interaction info</td>
<td>Computer Keypad &amp; mouse Chair sensor Light switch Window HVAC controller</td>
</tr>
<tr>
<td>Occupancy access info</td>
<td>Parking access Entrance guard</td>
</tr>
<tr>
<td>Smart device info</td>
<td>Wi-Fi connection GPS location Radio frequency identification</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2</th>
<th>An example of a conditional probability table.</th>
</tr>
</thead>
<tbody>
<tr>
<td>PIR2</td>
<td>InRoom?/InWorkingZone?</td>
</tr>
<tr>
<td>Occupied</td>
<td>Vacant</td>
</tr>
<tr>
<td>Detected</td>
<td>0.62 0.01</td>
</tr>
<tr>
<td>Undetected</td>
<td>0.38 0.99</td>
</tr>
</tbody>
</table>
4. Experimental validations and performance evaluations

This section presents the evaluations of the virtual occupancy sensors. Various virtual occupancy sensors were developed automatically by a software and conducted in two office rooms. The experimental data collected were then used to evaluate the EM-based parameter learning method and performance of virtual occupancy sensors at room level and working zone level respectively.

4.1. Implementation of the virtual occupancy sensors

A software was developed in Java to implement the proposed virtual occupancy sensors according to the flow charts in Fig. 4. The software applied BBN reasoning based on SMILE reasoning engine for graphical probabilistic model [29]. The database used is based on a SQL database engine named SQLite [30].

4.2. Description of the experiments

Experiments were made in two private office rooms in the 6th floor of Vertigo building in Eindhoven University of Technology, Netherlands. The rooms, named Room A and Room B, were used by seven volunteers for one or two weeks respectively. All volunteers used were employees of the University. The measurements and configurations of the two rooms used are as listed in Table 3. The volunteers were asked to record the time of arrival and exit events in the room and working zone. The records in the rooms were validated manually using measurements from the three extra PIR sensors installed around working desk. The records acted as ground truth values. It is noted that the ground truth values were only used to evaluate the performance of the virtual occupancy sensors.

In the experiments, every physical sensor was sampled independently. There were two different sampling intervals in the experiments, e.g. 20 s and 3 min. The sampling interval for virtual occupancy sensors was 20 s. Measurements were determined at the end of each time slot. For instance, the measurement of a PIR was occupied if it detected occupant within the time slot concerned. In each time slot of a virtual occupancy sensor, there should be at least one measurement from each physical sensor if the sampling interval of the physical sensor was not longer than that of the virtual occupancy sensor. For the physical sensors that had longer sampling intervals (e.g. WiFi sensor), measurement of the physical sensor determined the values of this sensor in current and coming n – 1 time slots of the virtual occupancy sensor. The sampling interval of physical sensor divided by that of virtual occupancy sensor equals n.

4.3. Evaluation of the EM-based parameter learning

The EM-based parameter learning method is evaluated by comparing the values which are estimated using EM algorithm with the recorded ground truth values.

Table 4 shows the estimated parameters of BBNs in four room-level virtual occupancy sensors in which only one sensor was introduced. The node TimeOfDay was used. The data were continuously measured in Room A for two weeks.

The estimated parameters are accurate enough compared with the ground truth values. For instance, PIR1 was estimated to detect 81.2% of the occupied events, while the ground truth values was 83.7%. The TimeOfDay node makes it possible to estimate parameters if there was only one sensor introduced. The PIR sensors were not sensitive to tiny movements of occupants. PIR1 and PIR2 detected 81.2% and 62.2% of occupancy events respectively. The chair sensor was very effective to detect occupants. It is because volunteers in the experiments were seated on the chair at all times. If the volunteer was in the room but not on the chair, the movements would have been significant enough to be detected by PIR sensors. Similarly, the volunteers did not have significant movements when using computers. It indicates that the combination of keyboard and mouse and PIR sensor can achieve high accurate occupancy detection results, as well as the combination of a chair sensor and a PIR sensor.

Table 5 shows the estimated parameters of a BBN in a room-level virtual occupancy sensor. There are two sensor introduced including a PIR sensor PIR1, and keyboard and mouse. The data are from the same volunteer in Table 4. The estimated values are also near to the ground truth values. There are no significant differences when the node TimeOfDay was introduced (Case I) or not (Case II).

The estimated parameters in Tables 4 and 5 also indicate performance of physical sensors. They can be used for performance assessment of physical sensors and fault detection. PIR2 had the worst occupancy detection performance. The negative error ratio was found to be 0.378 whereas positive error was revealed as 0.010. This sensor could still be regarded as not regarded as it did not exceed the thresholds (0.4 and 0.05 respectively).

---

Table 3

Measurements and configurations in the experiments.

<table>
<thead>
<tr>
<th>Sensor/measurement</th>
<th>Description</th>
<th>Sampling interval</th>
<th>Sampling method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Passive infrared sensor (PIR)</td>
<td>Two PIR sensors were installed in ceiling (PIR1) and front wall (PIR2) respectively. Three PIR sensors were installed around the occupant within 0.5 m to check the recorded ground truth values</td>
<td>20 s</td>
<td>Recorded by hardwares named Arduino Yun [26]</td>
</tr>
<tr>
<td>Keyboard &amp; mouse sensor</td>
<td>A java application was developed to listen global keyboard and mouse events. It started up automatically. It is based on a library named Jnativehook [27]</td>
<td>20 s</td>
<td></td>
</tr>
<tr>
<td>Chair sensor</td>
<td>Wireless magnet switch were installed in the chair to monitor the deformation of chair [28]</td>
<td>20 s</td>
<td></td>
</tr>
<tr>
<td>Door sensor</td>
<td>Wireless magnet switch were installed on the door to monitor the open and close of door [28]</td>
<td>20 s</td>
<td></td>
</tr>
<tr>
<td>Light switch sensor</td>
<td>A button was put near the light switch. Volunteers were asked to press the button when turning on/off the light</td>
<td>20 s</td>
<td></td>
</tr>
<tr>
<td>Wi-Fi connection</td>
<td>An android application was developed to read BSSID of the connected Wi-Fi hotspot and upload it to a server in real time. The application was installed in volunteers’ smart phones in the same android application, functions were developed to obtain GPS location and upload it to a server in real time. It is based on a library named Google maps Android API v2 [25]</td>
<td>20 s</td>
<td>Recorded by an android application</td>
</tr>
<tr>
<td>GPS location</td>
<td>In the same android application, functions were developed to obtain GPS location and upload it to a server in real time. It is based on a library named Google maps Android API v2 [25]</td>
<td>3 min</td>
<td></td>
</tr>
</tbody>
</table>
4.4. Performance evaluation of developed sensors

4.4.1. Evaluation of the room-level virtual occupancy sensors

The room-level virtual occupancy sensors were evaluated using experimental data. Virtual occupancy sensors of various sensor combinations were generated and trained by the software automatically, and then evaluated using test data. Results indicate virtual occupancy sensors are better in occupancy detection performance than individual use of sensors/information sources for all cases. A comprehensive analysis was made for a virtual occupancy sensor with a combination of keyboard and mouse, and a PIR sensor. The structure of the BBN is as illustrated in Fig. 6.

The evaluation used two weeks of experiment data for Room B. The historical data were divided into two parts. One part was the test data which were recorded in a typical working day. It contained 1620 series of data. Another part was the training data which included data for the remaining nine working days. It contained 14580 series of data. The trained BBN is analyzed firstly. Then, evaluations are made for the typical working day using the test data.

As shown in Fig. 6, the belief of Occupied is 0.90 when the observations are Touched in node keyboard and mouse, and Undetected in node PIR2 during 8:00 am to 8:20 am (t_0820). For different time periods, the beliefs of Occupied are 0.34 (t_before), 0.99 (t_1100), 0.80 (t_1220) and 0.96 (t_1440) respectively with the same observations. Although the observations from sensors are the same, the beliefs of Occupied are various in different periods for the reason of different occupancy patterns. For instance, the belief of Occupied is low (0.34) if only keyboard and mouse detect occupant. It is because the prior probability of occupancy is very low (0.02) before 7:20 am (t_before). In such a situation, the sensor, which detects occupant, might be faulty with a high probability. Given the same observations, the belief of Occupied is high during 10:40 am to 11:00 am (t_1100) although only one sensor detects occupant. The sensor, which does not detect occupant, might be faulty with a high probability. On the contrary, if the observations are Untouched in node keyboard & mouse, and Detected in node PIR2, the beliefs of Occupied are 0.17 (t_before), 0.78 (t_0820), 0.97 (t_1100), 0.62 (t_1220) and 0.90 (t_1440) respectively in different time. The beliefs of Occupied are evidently lower. If the observations are Touched in node keyboard & mouse and Detected in node PIR2, the belief of Occupied is 0.99 (t_before) and 1.0 (t_0820). The belief of Occupied is still high in t_before because both sensors detect the occupant, although it is a rare event (the conditional probability is only 0.02). If the observations are Untouched in node keyboard & mouse and Undetected in node PIR2, the belief is 0.00 (t_before) and 0.03 (t_0820). Conclusions can be made that the occupancy patterns, which are learned from historical data by EM algorithm in the BBN, are helpful to improve the performance of information fusion.

Fig. 7 shows the occupancy detection results in the typical day by the BBN illustrated in Fig. 6 using the test data. The threshold was 0.75. The threshold was determined to balance the trade-off of positive error and negative error in the historical data and test data. The minimum negative error ratio was guaranteed. The correct occupancy detection ratio was increased from 73.4% (PIR2) and 92.8% (keyboard and mouse) to 96.7% with the help of the BBN.

In Fig. 7, PIR2 did not detect the occupant severally for as long as 22 min between 9:00 am and 10:00 am. It would be unreliable if the demand driven HVAC operation was based on PIR2 only. There were delays in the keyboard and mouse measurements when the occupant came into the room. But, PIR2 detected the occupant immediately. Sometimes, keyboard and mouse detected the occupant but PIR2 failed because the movements of occupant were too tiny to be detected by PIR2, for instance in most of the time between 9:00 am and 10:00 am. Continuous negative errors occurred for 5 min around 11:00 am. During that time, both PIR2 and keyboard and mouse did not detect the occupant. This is still acceptable because there is always a delay period (e.g. 5 min or 10 min) in the demand driven HVAC operations.

For the same day, use of PIR1 sensor and, keyboard and mouse yielded an occupancy detection ratio increased from 85.1% (PIR1) and 92.8% (keyboard and mouse) to 97.6%. If PIR1 and PIR2 were used, the correct occupancy detection ratio could be increased from 85.1% (PIR1) and 73.4% (keyboard and mouse) to 91.3%. Similar results were observed using data from other volunteers.

There were no positive errors observed from all physical sensors during the experiments. Positive errors of faulty PIR sensors are generally caused by performance degradations or thermal disturbances. They can be assumed to be evenly distributed. Evidences can be found from Dodier et al.’s experiments [10]. In this study, simulated positive errors were added to a PIR in Room A. In this case, the virtual occupancy sensor contains two PIR sensors, i.e. PIR3 and PIR4. Similarly, the virtual occupancy sensor was trained using data of nine working days. Test data were collected in the remaining day. Positive errors were generated randomly with a probability of 0.3. The positive errors were added to the measurements of PIR4. Evaluation results are as shown in Fig. 8.

The correct occupancy detection ratios were 86.2%, 75.0% and 91.2% for PIR3, PIR4 and the virtual occupancy sensor respectively. The virtual occupancy sensor was robust to the positive errors in PIR4. The detected positive error ratios of PIR4 and the virtual occupancy sensor was 15.3% and 0.0% respectively. The positive errors of PIR4 did not lead to positive errors of the virtual occupancy sensor, since the beliefs of Occupied were still too low to report

<table>
<thead>
<tr>
<th>Sensor</th>
<th>PIR1</th>
<th>PIR2</th>
<th>Keyboard &amp; mouse</th>
<th>Chair sensor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TrueValue</td>
<td>Estimated (case I)</td>
<td>Estimated (case II)</td>
<td>TrueValue</td>
</tr>
<tr>
<td>Vacant</td>
<td>99.9%</td>
<td>98.7%</td>
<td>99.9%</td>
<td>99.0%</td>
</tr>
<tr>
<td>Occupied</td>
<td>83.7%</td>
<td>81.2%</td>
<td>62.4%</td>
<td>62.2%</td>
</tr>
</tbody>
</table>

Table 4

EM-based parameter learning: only a sensor in each case with the node TimeOfDay.

<table>
<thead>
<tr>
<th>Sensor</th>
<th>PIR1</th>
<th>Keyboard &amp; mouse</th>
<th>Chair sensor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TrueValue</td>
<td>Estimated (case I)</td>
<td>Estimated (case II)</td>
</tr>
<tr>
<td>Vacant</td>
<td>99.9%</td>
<td>99%</td>
<td>100%</td>
</tr>
<tr>
<td>Occupied</td>
<td>83.7%</td>
<td>82%</td>
<td>80%</td>
</tr>
</tbody>
</table>

Table 5

EM-based parameter learning: two sensor nodes with node TimeOfDay (Case I) and without node TimeOfDay (Case II).
Occupied events. The measurements of PIR3 and occupancy patterns (presented by \textit{TimeOfDay} node) did not support the beliefs of \textit{Occupied} very much when positive errors occurred in PIR4. The beliefs of \textit{Occupied} were almost 0 during 8:00 am to the arrival of occupant (8:22 am) and after 16:55 pm. Most of beliefs were also low during the unoccupied period in the lunchtime. It is because the prior probabilities of \textit{Occupied} were low during these periods. In the afternoon, when PIR4 had positive errors, the beliefs of \textit{Occupied} increased to around 0.6 since the prior probabilities of \textit{Occupied} were higher. The beliefs were still lower than the threshold (0.75). It is very interesting to find that positive errors in PIR3 would lead to positive errors in the virtual occupancy sensor. The beliefs were around 0.9 in the afternoon, and around 0.95 in the morning during the unoccupied periods, if the positive errors were added to measurements of PIR3 instead of PIR4. It is because PIR3 were more reliable than PIR4 in the training data set. The probability of positive errors in PIR3 was lower than the probability of positive errors in PIR4. Therefore, the \textit{Detected} events from PIR3 supported the beliefs of \textit{Occupied} strongly. BBNs in virtual occupancy sensors can be updated continuously using recent historical data. Parameters in BBNs would be adjusted if sensor performance degradations were presented in historical data. By this way, performance degradations in PIR3 would not lead to positive errors in the virtual occupancy sensor.

Fig. 9 presents the results when three physical sensors were introduced in the virtual occupancy sensor. Measurements of keyboard and mouse were considered. Evaluations were made using the same test data as the case in Fig. 8 for comparisons. The correct occupancy detection ratio increased from 91.2\% to 97.1\%. The beliefs of \textit{Occupied} decreased slightly when positive errors occurred. It seems that the two PIR sensors were robust enough to positive errors in one sensor. Further researches are necessary to evaluate the robustness of virtual sensors in the conditions that performance degradations slowly.

4.4.2. Evaluation of the working zone-level virtual occupancy sensor

The working zone-level virtual occupancy sensors are evaluated using experimental data in this section. Firstly, a specific BBN was developed for the experiments based on the generic BBN in Fig. 3. Evaluations are made later.

The working zone in the experiments is the entire 6th floor. There are three Wi-Fi hotspots in the floor, named WiFi_1, WiFi_2 and WiFi_3 in this study. The areas they covered are named WorkingZoneA, WorkingZoneB and WorkingZoneC respectively. There are some places covered by two Wi-Fi hotspots. The two experimental rooms are mainly covered by WiFi_1. Two meeting rooms are covered by WiFi_2. WiFi_3 covers the other side of the floor and another meeting room. If a volunteer went to other floors, the smart phone could connect to Wi-Fi hotspots in other floors automatically. There are two major entrances in the building. The first entrance is on the 1st floor. Smart phones always connected to Wi-Fi hotspots in the 1st floor automatically during the period between the entrance event and the event taking elevator to the 6th floor. The second entrance is on the 0th floor. Smart phones were always disconnected to any Wi-Fi hotspot since the Wi-Fi signal was weak and the residence periods were always short. But there is an entrance guard which requires checking in using staff card. Due to the difficulty to get data of entrance guard, it was not considered in the BBN in this study. The specific BBN for application in this working zone is illustrated in Fig. 10. Two common measurements were used, i.e. Wi-Fi connection information and GPS location, which were available from smart phones. Smart phones switched to 3G network automatically if there were no Wi-Fi connections.

In the node \textit{InWorkingZone?}, the states \textit{InWorkingZoneA}, \textit{InWorkingZoneB} and \textit{InWorkingZoneC} presents the locations covered by WiFi_1, WiFi_2 and WiFi_3 respectively on the 6th floor. If the observed state is \textit{InWorkingZoneA}, it means the occupant...
might be nearby or inside the experimental room.

In the node Wi-Fi connection, the state ApproachingWorkingZone means that the occupant is coming to the building. It is observed when the smart phone is connected to Wi-Fi hotspots on the 1st floor and the occupant was not in the building in last 10 min. On the contrary, the state LeavingWorkingZone is observed when the sequence of Wi-Fi connections is in the building (e.g. on 6th floor), 1st floor (optional) and disconnected. ConnectedUnknownWiFi means that the smart phone is connected to Wi-Fi hotspot outside of the building. It is strongly against the belief that the smart phone is in the building.

In the experiments, the GPS measurements from smart phones could obtain location accurately outside of the building concerned. In the building, the GPS measurements were a fixed location that is within the area of the building. It might be caused by the specific algorithm in Google maps API (the algorithm is not published but evidences showed that Google maps consider nearby Wi-Fi hotspots in GPS service). In node GPS location, the state ApproachingWorkingZone means that the GPS location indicates that the occupant is approaching and near to the building. The state LeavingWorkingZone has a contrast meaning. The state OutsideOfWorkingZone means that the occupant is outside of the working zone for more than 10 min (it is the boundary between OutsideOfWorkingZone and ApproachingWorkingZone or LeavingWorkingZone). The state Unknown means the GPS does not work, e.g. power off or no internation connection. The states in node TimeOfDay have a larger interval than the nodes in BBN of room-level. It is because there are few coming events and leaving events compared with detection events by physical occupancy sensors like PIR sensors. A larger interval means there are more events in every interval, and therefore more training data for every state in node TimeOfDay.

A working zone-level virtual occupancy sensor was trained using experimental data of nine working days based on the BBN in Fig. 10. Data of the left day were used to evaluate the virtual sensor. There were sufficient measurements during working period but few leaving events and coming events in the experiments. To solve this problem, the initial values of the parameters in the BBN illustrated in Fig. 10 were estimated by the authors according to expert experiences. In the EM algorithm, the sample size of the initial parameters was one third of training data. Then, the parameters were updated in the training process. Parameters related to the Occupied events were trained sufficiently since there were sufficient training data. Parameters related to arrival events and exit events were not trained sufficiently since there were few training data. Further researches are necessary to evaluate the improvement
of performance with increasing amount of training data. The performance of this virtual occupancy sensor is evaluated by analysis firstly. The virtual sensor is shown in Fig. 10. The belief of InWorkingZoneA is 0.87 if the GPS location is unknown and Wi-Fi connection is ConnectedWiFi_1. If the state of TimeOfDay changed, the belief would be different, i.e. 0.53 (t_before), 0.87 (t_7_9am), 0.95 (t_9_11am), 0.93 (t_13_15pm), 0.89 (t_15_17pm), 0.77 (t_17_19pm) and 0.53 (t_after). It is because of the prior possibilities of occupied in working zone are various in different time. If both Wi-Fi node and GPS location node are observed ApproachingWorkingZone, the belief of ComingEvent is 1.0 at any time. On the contrary, the belief of LeavingEvent is 1.0 at any time if both nodes are observed LeavingWorkingZone. Similarly, the belief of StillOutside is 1.0 if the smart phone connects Wi-Fi hotspot in other floors. If the smart phone is power off, i.e. state Unknown in GPS location node and Disconnected in Wi-Fi connection node, the belief of InWorkingZoneA is 0.01 (t_before), 0.21 (t_7_9am), 0.59 (t_9_11am), 0.21 (t_11_13pm), 0.56 (t_13_15pm), 0.34 (t_15_17pm), 0.12 (t_17_19pm) and 0.01 (t_after).

Fig. 11 shows occupancy detection results in the typical day using the BBN illustrated in Fig. 10. The ground truth values in zone-level were not recorded during the experiments. The ground truth values in room-level were introduced here for reference. The beliefs of all states in InWorkingZone? node are shown in Fig. 11. The sum of beliefs at every moment was 1.0. Given the threshold 0.8, the correct detection ratio was 84% when the occupant was in room. In the experiments, measurements were discarded if smart phones failed to upload them on time. This ratio could be higher if the sampling interval was smaller and/or the delay of uploading data was allowed. The results indicated that the Wi-Fi connections were stable in the working zone. During 14:03 pm and 15:24 pm, the volunteer was in a meeting room that was covered by WiFi_2. The beliefs were 0.69, 0.30 and 0.11 for InWorkingZoneA, InWorkingZoneB and InWorkingZoneC respectively. Beliefs of these three states could be combined together to indicate whether the occupant was in working zone or not. For the zone-level virtual occupancy sensors, arrival events and exit events are more valuable to demand controls. In the morning, the arrival event was detected in the GPS and Wi-Fi measurements. The beliefs of ComingEvent were around 0.98. In the end of that day, the exit event was detected correctly by GPS only. The training data contained limited amount of arrival events and exit events. The performance would get better by self-learning from accumulated historical data.

5. Discussions on application issues

5.1. Discussion about the virtual detection sensors

The room-level virtual occupancy sensors and working zone-level virtual occupancy sensors can also work together in an information fusion way. For instance, a room-level virtual occupancy sensor can assess performance of a working zone-level virtual occupancy sensor. For instance, if the occupant is detected in the room, the working zone-level virtual occupancy sensor should report occupied in working zone.

This study reveals that Wi-Fi connection information is effective in estimation of occupants location. Currently, Wi-Fi hotspots are widely in use. Also, use of smart devices inform of smart phones and smart watches is a popular trend. This makes it convenient to obtain Wi-Fi connection information at individual level. Consequently, Wi-Fi connection information would seem ideal for working zone-level virtual occupancy sensor in the near future. Results from this experiment shows that parameter learning using the EM algorithm is effective. Subsequently, it becomes unnecessary to record ground truth values as training data. Further research is needed to demonstrate whether parameter learning using the EM algorithm would still yield credible results for cases whereby occupancy patterns are much more random than the ones in this study. It is recommended to introduce two measurements at least in practical applications; for example one PIR sensor and a keyboard and mouse or two PIR sensors.

The proposed method is only valid for private office rooms with one occupant. Adjustments are necessary for applications in rooms with more than one occupant. In such a situation, sensors of private usage would still valid for individual occupancy detection such as chair sensor, keyboard and mouse. Sensors like PIR sensor and ultrasonic sensor would only valid for Boolean results of occupancy status in the whole room. This makes it necessary to adjust the BBN in room-level virtual occupancy sensor for applications in rooms with more than one occupant.

5.2. Discussion about building demand driven operations

In demand driven HVAC operations, the HVAC system for a room concerned could be fully open if the room level-virtual occupancy sensor detects occupant, partially open if the room level-virtual occupancy sensor does not detect occupant but the working zone-level virtual occupancy sensor finds the occupant is still in working zone, and pre-open/shut off if the working zone-level virtual occupancy sensor detects an approaching/leaving event respectively. The term partially open means that the set-points in HVAC system are changed to optimal values to reduce energy costs.

Moving window concept is widely used in demand driven HVAC operation at room level. Moving windows for controls start from current time and end after an interval. If sensor detects occupant in current time, all the time within the moving window is regarded to be occupancy detected. It can reduce negative error ratios significantly but also increase positive error ratios. In the experiment, the
correct detection ratio of PIR2 could be increased from 62.2% to 95.0% if a moving window with a 10 min interval is introduced. The positive errors produced by moving window actually do not cause significant energy waste. It is because the errors always occur in the delay time of demand driven control. But, the positive error produced by sensors might lead to significant energy waste when a moving window is adopted. It is because the resultant positive errors might be evenly distributed. In such cases, the HVAC system for a room would open and close frequently when the room is unoccupied. One of the major contributions of the proposed virtual occupancy sensors is the reduction of positive error ratios for demand driven HVAC operations.

5.3. Discussion about feasibility and privacy

It is possible to easily collect occupancy information using wireless networks or internet/Ethernet. For instance, it is more efficient to get Wi-Fi connection information of the smart phone of an occupant from the Wi-Fi hotspots than from the smart phone. In such a case, the sampling intervals could be optimized to save battery power in smart devices. For instance, GPS tracking leads to high energy consumption. It therefore makes sense to passively activate it only when the GPS location is needed.

Privacy is an important issue to consider in the usage of personal information from smart devices and computers. The measurements could be discretized to limited states to avoid revealing detailed personal information. For example, the GPS location could be categorized into three states, i.e. inside, outside and unknown. On the same note, keyboard and mouse status could be classified as touched, untouched and unknown.

6. Conclusion

In this study, multi-source occupancy measurements and information are fused to obtain reliable occupancy information of individual occupant. The two virtual occupancy sensors provide real-time occupancy information at room level and working zone level respectively. The sensors can be automatically developed in the building management systems with self-learning, self-performance assessment, and fault detection functions. Only low cost measurements are considered in this study. Evaluations for the study were made in two private office rooms. The EM algorithm proved effective in the estimation of parameters for BBNs without use of ground truth values. The estimated parameters could be used to assess performance of sensors and information sources. The node TimeOfDay was useful in providing occupancy patterns. Performance assessment of sensors for instances with only one sensor introduced in BBNs also proved useful. The resultant errors of estimated correct occupancy detection ratios of four sensors are within 5.6% when the room was occupied. The chair sensors were highly effective in detecting occupant during experiments. The correct occupancy detection ratio was revealed as 93.3% in the experiment. Combinations of different sensors types in BBNs improved detection performance as was the case in using a PIR sensor together with keyboard and mouse sensor. Accuracy in occupancy detection ratio was increased from 73.4% (PIR2) and 92.8 (keyboard and mouse) to 96.7% by the BBN for a typical day.

The virtual occupancy sensors developed are ideal for use with demand driven control of HVAC systems and lighting systems at private office room levels. Based on this methodology, it would be possible to develop virtual occupancy sensors for applications at public office level and building level in further researches.

References