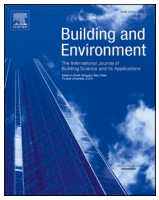
Building and Environment 243 (2023) 110714

Contents lists available at ScienceDirect

Building and Environment

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



Estimating the number of occupants and activity intensity in large spaces with environmental sensors

Xiaohao Zhang [a], Tongyu Zhou [a,*], Georgios Kokogiannakis [b], Liang Xia [a], Chaoju Wang [a]

[a] Department of Architecture and Built Environment, University of Nottingham Ningbo China, 199 Taikang East Road, Ningbo, 315100, China
[b] Sustainable Buildings Research Centre, Faculty of Engineering and Information Sciences, University of Wollongong, Wollongong, 2519, Australia

ARTICLE INFO

Keywords:
Occupant information
Number interval
Activity intensity
Large space
Occupant behavior

ABSTRACT

Recently, occupant-centered control models have been widely discussed, with intelligent control models looking for more refined and dynamic regulation based on occupant information. This article introduced a novel method for using environmental sensors and machine learning algorithms to identify the number of occupants and activity intensity in large spaces. A multi-functional space was monitored for approximately two months using PIR, CO_2, sound decibel, temperature and humidity sensors. To address the challenge of identifying the number of occupants in large spaces, the study proposed a non-uniformly distributed interval of occupant numbers that offsets the small fluctuations in the number of people. The study also demonstrated that PIR and CO_2 level measurements could be used to estimate the headcount interval with an accuracy rate of 84.5%. Furthermore, the study employed K-means clustering to identify low-, medium-, and high-level activities in the studied space, achieving an overall accuracy rate of 89.3%. A new metric of activity intensity was introduced to measure the activities carried out indoors, which incorporated CO_2 and sound decibel levels, PIR readings, and the number of occupants. This proposed metric was found to be appropriate for quantifying the activity intensity in the studied space. Overall, the method presented in this study provided a promising approach for enabling occupant-based control strategies that leverage advanced sensor data to optimize building service systems in large spaces.

1. Introduction

According to statistics, the building operation phase in China was responsible for 21.3% of the country's overall energy consumption in 2020 [1], demonstrating that long-term energy efficiency and sustainable building development are inseparable from an efficient building operation. In addition, increased awareness of environmental quality, health, and occupant productivity in the workspace has also become an inevitable consideration for building design and improvement. In order to promote energy efficiency and improve indoor comfort occupant-centric controls (OCC) have been recently developed, aiming to provide dynamic input to the building system control and drive building systems at a granular level according to the indoor occupant information changes [2,3]. Occupant-centric control strategies mainly include two categories; one is occupant-based control which is mainly based on the presence, counting of occupants, and location, the other is occupant behavior-based control, where the operation of the control system is adjusted based on the interaction of occupants with the indoor environment, such as indoor activity types and window-opening behavior [4].

In order to achieve dynamic management, detailed information about the indoor environment needs to be gathered first, typically via a wireless sensor network that aims to optimize occupancy-driven control strategies [5–7]. Passive infrared (PIR) sensors, for example, have been typically used to detect occupancy through signal interruptions [8,9]. Based on the PIR signal counting, a PIR sensor array could also be utilized to measure the activity intensity of residents [9]. However, the count is sometimes inaccurate when many people are passing through at the same time [10–12]. In addition, according to its operating principle, PIR sensors have a short masking effect after a motion is detected [13]. During the masking period, PIR sensors do not detect movement, which leads to errors in counting, especially when there is a continuous flow of people passing by. In addition, carbon dioxide (CO_2) concentration is also a suitable parameter to be used as a proxy for occupancy in a building [14], with several studies confirming a close relationship between CO_2 concentration and occupancy [15,16]. Additionally, according to Galván-Tejada et al. (2018), human activity sounds could instantly describe the dynamic changes in the indoor environment, such

* Corresponding author.
E-mail address: tongyu.zhou@nottingham.edu.cn (T. Zhou).

https://doi.org/10.1016/j.buildenv.2023.110714
Received 26 April 2023; Received in revised form 4 August 2023; Accepted 6 August 2023
Available online 7 August 2023

as event discrimination and location estimation [17]. Besides monitoring appliances, the energy meter and lighting switch sensors are also helpful in detecting occupant activities and identifying habits in relation to energy efficiency [18]. However, these methods highly depend on personnel habits and awareness [2]. In addition, cutting-edge image--based sensors such as accurate three dimensional (3D) stereo vision cameras have also been extensively studied. These high-precision sensors can achieve decimeter-level positioning accuracy [19]. However, one of the major problems inhibiting their wide promotion is the relatively high cost. For instance, image preprocessing in a symbol recognition system based on 3D stereo vision consumes a significant amount of logic and memory resources, thereby increasing operating expenses [20].

Based on the analysis of the performance and limitations of each type of sensor, multi-sensor fusion is necessary for detecting applications across a wide range of demands and scenarios to overcome the constraints of individual metrics [14,21,22]. Huang et al. demonstrated the efficacy of a combined CO_2 and light sensor (PIR-like) system in correcting the false estimation of occupant presence based on CO_2 concentration fluctuations, where a light sensor taped on a door frame produced a distinctive pulse response upon detecting entrances or exits when a person blocks the lighting [23]. Wang et al. selected CO_2 concentration to identify the number of occupants and then corrected the results based on video detection to remove cumulative errors [24]. Besides the environmental indicator combination methods, considering the accumulation effect of CO_2, the time factor can also be used as a variable. An experiment showed that the number of people detected using CO_2 concentrations could be highly accurate at night when there was no entry or exit of people and the air infiltration rates were relatively stable, so the whole day was divided into two periods from 00:00–07:00 and 7:00–24:00 [25]. From 00:00 to 07:00, CO_2 concentration was chosen to identify the number of occupants in the room of the study, and video was used from 07:00–24:00 [24], however privacy considerations were not of a concern in this study.

Machine learning approaches are often used to analyze measured environmental data and extract information about occupants, such as the occupant's presence, number and location. Support vector machine (SVM), decision tree (DT), and random forests (RF) are commonly used in identifying the presence and the number of occupants [24,25]. In addition, for detecting relatively uncomplicated event types, unsupervised data-driven methods are a choice [26], such as K-means clustering. Prior research has utilized K-means clustering to detect the occupancy presence [27]. Moreover, to investigate the energy-related behavior of occupants in residential buildings, K-means could cluster the time and places of energy-use patterns [28].

The literature review above highlights the prevalent use of environmental sensors for gathering occupant information. However, it also reveals some gaps in this research area. First, there is still a lack of definitive research on large open spaces with many occupants. Studies using CO_2-based data-driven models often focused on a small number of occupants, typically less than 10 [29–32]. For studies with a low number of occupants in small spaces, the distances between the occupants and the sensor positions are relatively fixed, and the types of personnel access are relatively uncomplicated, such as single entry or exit. In contrast, for areas with numerous people, the different positions of occupants relative to the sensors will significantly increase the complexity of the study due to the limitation of the required number of sensors. In addition, because of the considerable latency in the correlation between the number of occupants and the CO_2 concentration, it is difficult to detect small changes in a large space [16]. A few instructive concepts have been put forward on this issue. Considering the complexity of determining the exact number of occupants in a large space, the concept of number x-tolerance accuracy was proposed in an experiment to evaluate the occupancy estimator [33]. X-tolerance means that a small error in the number of occupants is acceptable within a given range [33,34]. For example, researchers suggested that minor differences would not significantly affect the indoor environment and the HVAC operating settings for an office with many occupants [33]. In another experimental study, Zuraimi et al. set a margin of error tolerance of 10 when determining the actual number of people in a room that could accommodate up to 120 people [30]. These studies indicate that it is reasonable to set an appropriate tolerance interval when determining the number of occupants in a large space [34].

The second aspect to note is that while previous research primarily focuses on the number of occupants, the specific details related to indoor events have not been thoroughly researched, despite their critical role in determining the control settings for the indoor environment [25,35]. Occupant activity is considered a pivotal contributor to changes in energy consumption and indoor environments [36]. Naylor et al. suggested that the environmental preferences of occupants would change as they engaged in one task or another [37], so various scenarios required the control system to provide the matching control strategies and regulate the indoor environment in a more targeted manner [38]. In order to study the behavioral patterns, Khani et al. proposed the Occupant Activity Indicator (OAI) concept. OAI is defined as the activity level of the occupants in one study, which is evaluated by the motion of occupants, window opening behavior, and CO_2 concentration. A higher OAI value means more occupants in a space [28]. Nevertheless, the indicator is only used to determine the magnitude of the number of people in the room and an approximate estimation of the intensity of the occupant activity. In conclusion, the current research is limited to determining the occupant number in larger spaces and assessing indoor activities, which could be useful for providing richer information for optimizing control strategies. From the perspective of occupancy information and occupant behavior-based information, this study aims to overcome the limitations of previous research on the identification of occupants' numbers in large spaces. Subsequently, we focus on judging the types of indoor events and quantitatively evaluating these activities. This research is rooted in the concept of occupant-centered control and leverages a series of low-cost, non-intrusive wireless sensors to recognize indoor dynamic environment information. This approach paves the way for subsequent intelligent control and the development of refined control strategies. The study has been carried out in an existing open space, where the environmental parameters were monitored and used to identify the occupant information in the space.

2. Methodology

A design studio in the CSET (Centre for Sustainable Energy and Technologies) building at the University of Nottingham Ningbo China was chosen as the space of the study. The studio has an area of 374 m^2 and can accommodate up to 80 persons. The main function of the studio is teaching and self-study. The details of the CSET studio space can be found in Table 1.

Table 1
Detail information of CSET studio.

Indicator	Information
Location	CSET building, University of Nottingham Ningbo China Campus
Area	374 m^2
Number of occupants	Up to 80
Function	Classroom and Self-study room
Main user	Students
Main event	Self-study: This typically involves minimal, irregular movement of people entering and exiting. workshop: This has a regular influx at the beginning. As the workshop progresses, people may leave at varying times. Lecture: This typically has a set start and end time. Therefore, there is often a surge of people entering at the beginning and leaving at the end.
Noise source	Air conditioners and printers

2.1. Location of sensors

In the studio, a variety of sensors were placed inside and outside the space to collect occupant information and environmental parameters as shown in Fig. 1. As suggested by the literature review and statistics, the deployment of three to four CO_2 sensors would be optimal for a space with a floor area ranging from 200 m^2 to 399 m^2 floor [39]. Therefore, the main activity area was divided into four sub-areas, with each deployed one multi-functional environmental sensor (model: LT-CG-S/T-WIFI, brand: LTWY). To ensure accurate data collection, the placement of these sensors took into consideration occupant breathing areas, internal airflow, and potential disruptions such as ventilation systems or other openings [40]. Adhering to the International Standard Organization (ISO)'s indoor environmental sampling guidelines, sensors were situated at a height of 1.5 m and, wherever possible, at a distance of 0.7 m away from the major potential interference sources, including occupants, doors, windows, and the HVAC system to minimize their influence [41].

The multi-functional sensors were used to capture CO_2, sound decibels, temperature and relative humidity in the studio. One PIR sensor was installed at the studio entrance which can identify the direction of entry and exit. Additionally, a fisheye surveillance camera was installed on the lobby ceiling to monitor and collect ground truth information (number of occupants and indoor events). Table 2 shows the parameters and specifications of each sensor used in the study. The data collected in this study encompassed only environmental details in the space and information on human movement. It did not include any personal information about the individuals within the space. Prior approval for the implementation of this study was obtained from the Research Ethics Review Board of the University of Nottingham Ningbo China.

These sensors took continuous readings throughout the day, including periods when the space was unoccupied. To accurately identify these unoccupied periods, we manually cross-checked sensor readings against surveillance camera footage.

2.2. Data preprocessing

Data preprocessing is the process of converting raw data into clean, useful data sets and typically involves checking for missing values, noisy data, and other inconsistencies [42]. In this study, interpolation was used to supplement missing values [32], and Kalman filtering was applied to address state estimation and Gaussian errors in linear models for CO_2, temperature, and humidity measurements [43]. The Kalman Filter is an efficient optimal estimator, which can explicitly take account of the dynamic propagation of errors in the model [44].

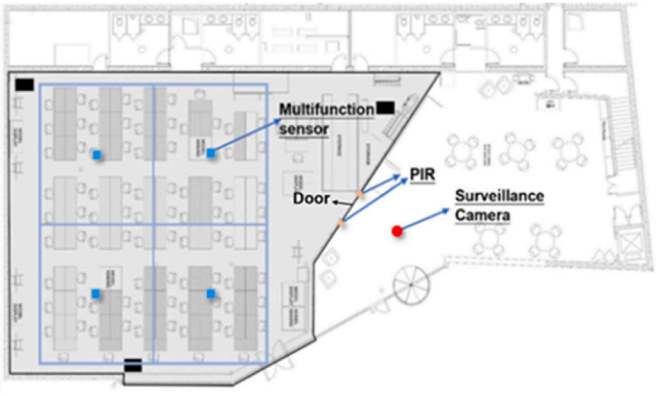


Fig. 1. Layout of the studio and the location of sensors.

2.3. Feature extraction

Based on the literature, a number of indicators were selected to describe the environmental parameters and their variations, as shown in Table 3. Changes in these indicators can reflect the variations in the indoor environment as a result of occupants' activities. However, the issue still remains on the error of such estimations.

To tackle this issue, this study employed information gain theory and measured each indicator with the value of information gain. Information gain quantifies the amount of information a particular variable or feature provides about the final outcome [46]. The following equations were used to calculate the relative information gain [45] based on the ground truth information recorded by the fisheye surveillance camera.

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

$$H(y/x) = \sum_{j=1}^{n} p(x_j) \sum_{i=1}^{n} -p(y_i|x_j)\log_2 p(y_i|x_j) \quad \text{Eq (2)}$$

$H(y)$ is a measure of the inherent uncertainty of the random variable. y and x represent the set of outcomes and features, respectively. y_i is the ith instance or possible outcome of random variable y. $p(x_j)$ is the prior probability for all ith values of y and $p(y_i|x_j)$ is the conditional probability of y_i given x_j.

$$IG(y,x) = H(y) - H(y|x) \quad \text{Eq (3)}$$

$IG(y,x)$ is the mutual information between y and x.

$$Rig(y,x) = \frac{IG(y,x)}{H(y)} \bullet 100\% \quad \text{Eq (4)}$$

$Rig(y,x)$ is the relative information gain between y and x.

2.4. Time granularity

The time granularity is the shortest time window in which the information about occupant behaviors is used for prediction. It depends on factors such as the granularity of the available data of sensors, the time frame in which the event has changed, and the expected predictive range. The time granularity of the recordings covers a wide range from about 30 s to several hours, and the most commonly used temporal resolution is between 10 and 19 min [47]. In this study, a time resolution of 10 min was selected for the occupancy monitoring. Shorter intervals may cause unnecessary fluctuation while longer intervals may result in a delay in reflecting the changes in the indoor environment and events [24].

2.5. Active period and inactive period

In this study, the time of a day was divided into an active period and an inactive period. Active periods are the periods of high occupancy rate, usually during the day when students are having class or discussion, while inactive periods are times for low- or non-occupancy, usually during the night. Based on the primary use of the studio, the active period of a day was defined as 9:00–23:59, and the inactive period of a day was defined as 0:00–8:59.

2.6. Occupant number interval

This study used the number interval instead of the exact number to classify the occupant number. Although, as has been pointed out in literature [4,34] that in a large space with many occupants, a small deviation from the exact number of occupants usually has little impact on the control strategies, an appropriate demarcation of number intervals is still a prerequisite for ensuring that deviations were within acceptable limits. As the number of occupants increases, the impact of

Table 2
Parameters of PIR and multi-functional sensors.

Device	Number	Measurement range	Measurement accuracy	Resolution	Detection interval	Recorded interval
PIR sensor	1	/	/	/	Real-time	10 min
CO_2 sensor	4	0 ~ 5000 PPM	± 4%	1 PPM	1 min	1 min
Decibel sensor	4	30 ~ 120 dB	± 1.5 dB	0.1 dB		
Temperature sensor	4	$-40 \sim 80°C$	$\pm 0.5°C$	0.1 $°C$		
Relative humidity sensor	4	0 ~ 100 RH%	± 3 RH%	0.1 RH%		

Table 3
Feature of different parameters.

Parameter	Indicators	Description
CO_2	$AVE\_CO_2\_10min$	Average CO2 per 10 min [29,30,45]
	$FD\_AVE(t) = AVE\_CO2(t) - AVE\_CO2(t-1)$	First-order difference [29,30,45]
	$SD\_AVE(t) = FD\_AVE\_CO2(t) - FD\_AVE\_CO2(t-1)$	Second-order difference [29,30]
	$FD2\_AVE(t) = AVE\_CO2(t) - AVE\_CO2(t-2)$	First-order shifted difference [30,45]
PIR	*PIR_IN*	Signal for personnel entry [29,30]
	PIR_OUT	Signal for personnel leaving [29]
Sound decibel	*AVE_dB_10 min*	Average decibel per 10 min [29]
Temperature	*AVE_Temperature_10 min*	Average temperature per 10 min [29]
Relative humidity	*AVE_Relative humidity _10min*	Average relative humidity per 10 min [29]

the same change in numbers on the indoor environment tends to be diluted. For example, when there is a high density of occupants, the access of a few more people will not make much difference. On the other hand, when only a small number of occupants is present in the space, minor changes could be of high significance. Therefore, the number intervals should not be evenly distributed for a large space.

We examined many studies about occupancy identification, and summarized their space types, space capacities and evaluation indicators in Table 4. Based on the error tolerances of different space capacities, the distribution of the number intervals (range of occupants' number) in this study was set in a stepwise increasing manner, as follows: [0], [1–2], [3–4], [5–6], [7–9], [10–12], [13–15], [16–18], [19–22], [23–26], [27–30], [31–34], [35–39], [40–44], [45–50], [51–56], [57–62], [63–69], [70–77], [78–86], [87–96]. If an estimate falls within the neighboring interval, it is also considered an acceptable tolerance. For example, if a headcount is [13–15] while in reality the number of occupants is [10–12], we consider it an acceptable estimation for the purposes of the study.

3. Model development

Three models were developed in order to identify the number of occupants and the activity intensity in the space, respectively, as shown in Fig. 2.

3.1. Identification of the number of occupants

Despite the widespread use of PIR sensors in public spaces for passenger flow detection, their accuracy can be affected when multiple people pass through simultaneously [10–12]. This may lead to a rapid accumulation of errors when using only PIR to count people in large spaces. To overcome this issue, this study adopted a data-driven method to establish a number recognition model based on the combination of PIR and environmental parameters. The data of PIR and environmental sensors were collected for the training set, and the following three methods were tested for processing them: Support vector machine (SVM), Decision Tree, and Random Forest. SVM is a popular tool in classification, forecasting, and regression of random data sets. It can model non-linear time series by mapping non-linear functions to a higher dimensional space using a kernel function [30]. Decision Tree classification is a method of selecting a class by descending a tree of

Table 4
Number interval setting.

Range	Function	Maximum Capacity	Error tolerance	Accuracy	RMSE	Reference	Interval ranges
1–5	Cell office	4	≤1	≥70%	/	[45]	1–2
	Cell office	2	≤1	≥95%	/	[48]	
	Open office	5	≤2	67%; 69%	1.01; 0.91	[29]	
	Open office	4	≤1	≥80%	0.49	[34]	
	Laboratory	3	≤1	≥80%	0.15	[49]	
	Office	3	≤1	/	0.59	[50]	
6–10	Lab	9	≤2	≥85%	1.2/1.5	[51]	2–3
	Lab	9	≤3	≥65%	2.3/2.74		
	Laboratory	6	≤2	≥75%	1.08	[52]	
	Open office	10	≤3	≥70%	1.58/1.63	[53]	
	Laboratory	6	≤3	≥80%	/	[54]	
	Lab	8	≤2	/	1.1	[50]	
11–20	Lab	20	3–4	≥90%	0.31	[55]	3–4
	Office	15	3–5	/	/	[56]	
	Hospital Patient room	20	≤3	/	/	[13]	
	Office	20	≤3	≥85%	2.3	[28]	
20–30	Office	25	≤4	≥80%	1.89	[13]	4
30–40	Open office	35	3–4	≥84%	1.77–3.23	[33]	4–5
40–60	Office	45	4–5	/	3.46	[56]	5–6
60–80	Lecture theatre	85	6–7	≥73%	4.68	[57]	7–8
80–100	Lecture theatre	100	8–9	>70%	4.3/5.5/12.6	[57]	9–10
100–150	Lecture theatre	120	9–10	/	12.1–27.4	[30]	10–13
	Lecture theatre	150	9	50%–70%	12.7	[58]	

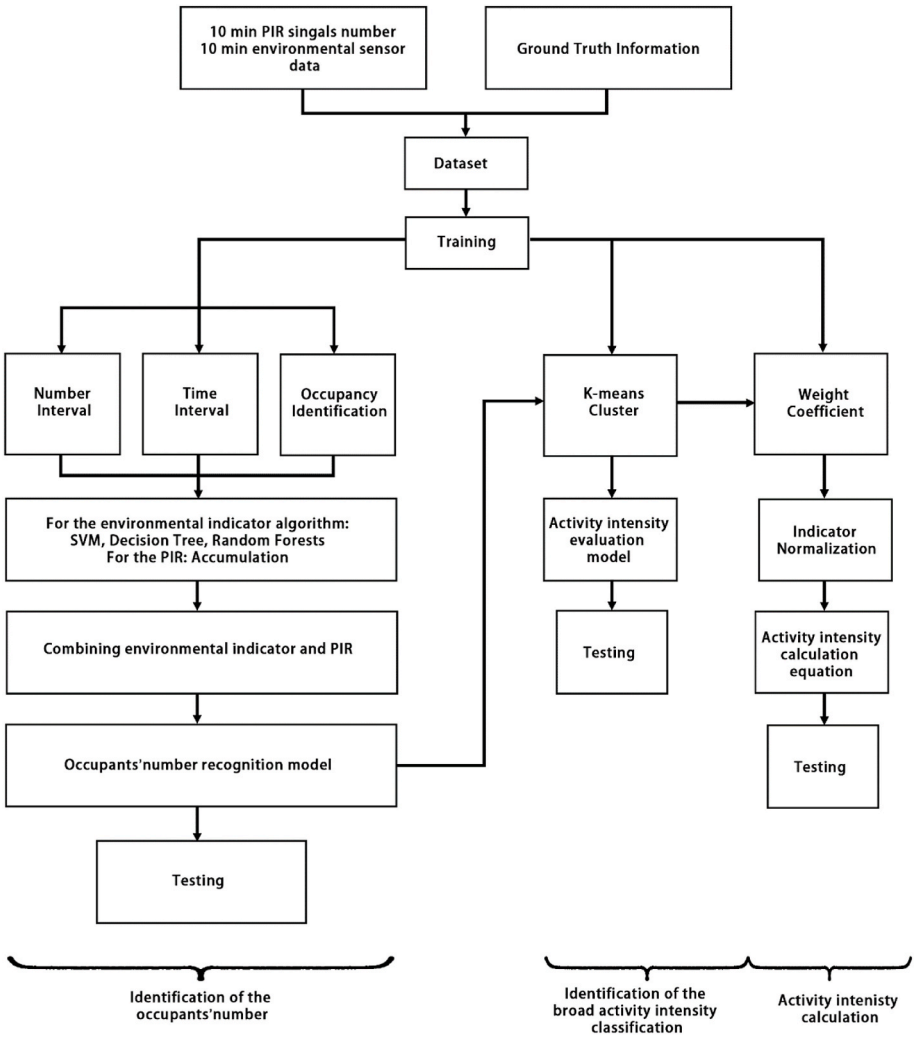


Fig. 2. Model development flow chart.

possible decisions, where at each internal node, a single feature value is compared with a threshold value until a leaf node is reached [59]. The Random Forest is an ensemble prediction model that improves the accuracy and robustness of regression models by utilizing a collection of different regression trees, trained through bagging and random variable selection [60].

3.2. Activity classification and activity intensity

3.2.1. Activity classification

According to the ASHRAE Handbook [61], humans produce different amounts of heat and have different metabolic rates when performing various activities. Therefore, even with the same number of people, the indoor environment control strategies should be different, highlighting the importance of identifying the type of activities in addition to the quantification of the number of occupants. Internal heat gains, metabolic rate and work intensity should be considered in evaluating activity classification.

Given the main function of the CSET studio, three activity classifications were predetermined as low, middle and high levels, in correspondence with the three representative activities of self-study, lecture and workshop, respectively, as shown in Table 5 and Fig. 3.

K-means clustering algorithm is chosen to cluster the indicators' data separately and then identify the three levels of activities, allowing for identifying regular events in a specific environment.

Table 5
Event classification.

Activity classification	Representative activities	Total heat gain from occupants (W) [61]	Typical metabolic heat generation (W/m^2) [62]
Low level	Self-study	130 (Seated, very light work)	Between 55 and 70
Middle level	Lecture; Discussion (small-scale)	140 (Moderately active work)	Between 70 and 100
High level	Workshop	160 (Walking, standing)	More than 100

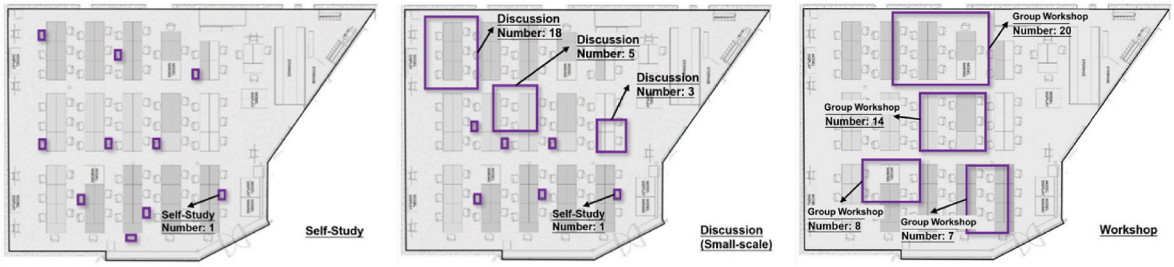
3.2.2. Activity intensity

A new variable of activity intensity is proposed, which is utilized to evaluate the activity level quantitatively. The equation to calculate the activity intensity is shown as Eq. (5) [28], and the range of activity intensity is between 0 and 100%. These variable complements and refines the K-means classification model mentioned in the previous section.

$$Activity\ intensity = \sqrt{f_1 I_1^{'2} + f_2 I_2^{'2} + f_3 I_3^{'2} + f_4 I_4^{'2}} \times 100\% \quad \text{Eq (5)}$$

where f is the weight of the indicator on the activity intensity, and the sum of weights is 1, and $I^{'}$ is the normalized value of each indicator (i.e., CO_2, Decibel, PIR and occupant number).

The weights f were determined by the relevance of each indicator to



(a) Low level activity example (b) Middle level activity example (c) High level activity example

Fig. 3. Activity classification examples.

identifying activity intensity. The Pearson correlation coefficient was utilized to examine the correlation between two variables, as in Eq. (6).

$$Cor(y_{true}, y_{iden}) = \frac{\sum(y_{true} - \overline{y_{true}})(\sum y_{iden} - \overline{y_{iden}})}{\sqrt{\sum(y_{true} - \overline{y_{true}})(\sum y_{iden} - \overline{y_{iden}})}} \quad \text{Eq (6)}$$

where y_{true} is the ground truth information and y_{iden} is the identified activities by each indicator.

The data normalization method uses the min-max normalization method, and the normalized value of the indicator was obtained by Eq. (7).

$$I_i^{'} = \frac{I_i - I_{min}}{I_{max} - I_{min}} \quad \text{Eq (7)}$$

Where I is the data and i is the category of each indicator (i.e., CO_2, Decibel, PIR and occupant number).

3.3. Performance evaluation indices

In order to evaluate the model performance, the precision and accuracy are calculated as per Eqs. (8) and (9).

$$Precision_i = N_{(T,i)}/N_i \quad \text{Eq (8)}$$

$N_{T,i}$ is the number of samples correctly recognized for each category i (such as [1,2] number interval), and N_i is the amount of the category i.

$$Accuracy = N_T/N \quad \text{Eq (9)}$$

N_T is the number of correctly identified samples, and N is the number of samples from all categories.

In order to further measure the model performance, Root Mean Square Error (RMSE) is introduced. RMSE aims to assess the difference between estimated and actual results and provides a more comprehensive view than the accuracy Eq. (9). RMSE is calculated as in Eq. (10).

$$RMSE(y_{true}, y_{iden}) = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(y_{true} - y_{iden})^2} \quad \text{Eq (10)}$$

y_{true} is the ground truth information and y_{iden} is the identified predicted value.

4. Occupant number recognition

Data collection was conducted from 25 February to 13 April 2022, in which the data from 25 February to 1 April was taken as the training set, and data from 2 April to 14 April was taken as the testing set. The data was recorded every 10 min, and each data consisted of PIR count, CO_2 level, sound decibel, room temperature and humidity at the time. In total, the training set included about 5000 groups of data and the testing set included about 1800 groups of data. The ground truth information in this paper includes the occupied situation, number of occupants and indoor events that were identified with the surveillance camera.

4.1. Occupant number recognition with PIR

As mentioned in the previous sections, errors in the PIR method accumulate quickly for a large space with continuous personnel flow, and the counting needs to be reset frequently in practice. Fig. 4 clearly shows that the PIR method achieved a satisfactory recognition rate after calibrating when the indoor occupancy was low (marked as "correction" in Fig. 4) on the first day. However, the recognition rate significantly deviated from the actual value after consecutive measurements due to error accumulation. Therefore, a method is needed to automatically calibrate the PIR count when there are fewer people present each day.

4.2. Feature extraction

The information gain method was used to establish the correlation between the features of environmental parameters and the actual number of occupants. Table 6 revealed that Ave_CO_2_10min obtained the highest information gain when compared to all the metrics, and thus this feature was selected for subsequent occupant number estimation. The information gains based on temperature and humidity were significantly lower than that of the CO_2 series features, which may be due to their low impact of thermal disturbances on the overall environment.

4.3. Occupant number recognition with CO_2

Upon determination of the most relevant input feature, we implemented SVM, Decision Tree and Random Forest algorithms to establish the occupancy identification models. Initially, these models were trained using 'Ave_CO2_10min' and corresponding occupant number intervals from the training dataset. Following this training phase, we utilized the 'Ave_CO2_10min' values from the testing dataset. This input enabled each model to generate output data in the form of identified occupant number intervals based on the testing group. The recognition results with CO_2 concentration are shown in Table 7. It suggested that relying solely on CO_2 concentration as a single parameter is insufficient to accurately identify the number of occupants in a large space during active periods, due to the complex and dynamic nature of both the environment and occupancy patterns. On the other hand, during inactive periods, where occupancy changes are relatively minimal, estimates derived from CO_2 concentration demonstrate a reasonable degree of accuracy. Additionally, the SVM method exhibited better recognition rates when compared to the other two algorithms, thus it was chosen to be combined with PIR to obtain a more accurate occupancy recognition rate.

4.4. Occupant number recognition with PIR and CO_2

In the PIR + CO_2 method, the CO_2 level was used for occupant

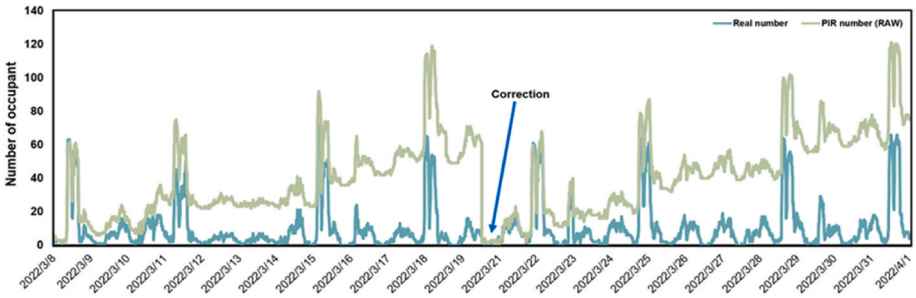


Fig. 4. Comparison of the truth and the occupant number detected by PIR only.

Table 6
Information gain ratio of each feature.

Feature	Relative information gain (Inactive period)	Relative information gain (Active period)
Ave_CO_2_10min	0.53	0.43
FD_Ave_CO_2	0.48	0.40
SD_Ave_CO_2	0.47	0.38
FD2_CO_2	0.44	0.37
Ave_Temperature_10min	0.23	0.17
Ave_Relative humidity_10min	0.20	0.16
Ave_dB_10min	0.14	0.08

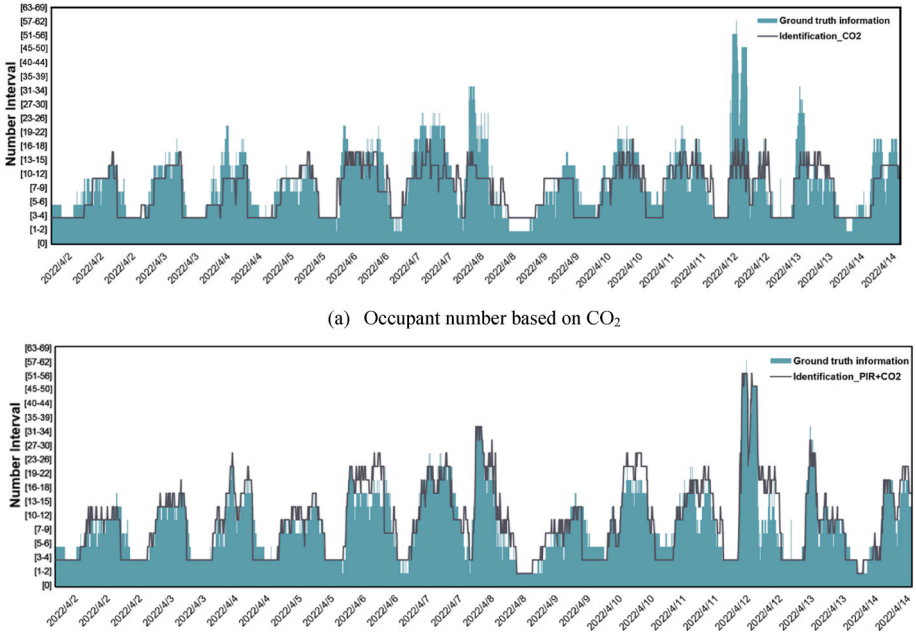
number recognition during the inactive period (0:00–08:59), while PIR was used during the active period (9:00–23:59). By calibrating based on the number of people identified during non-active periods, the accuracy of the number of people identified by PIR during the day can be improved. The testing set between 2022/4/2 and 2022/4/14 was used for occupant number identification and the comparisons with the ground truth are shown in Fig. 5. It is clear that, PIR + CO_2 allows better identification than using CO_2 alone, particularly for large numbers of occupants.

4.5. Comparison of CO_2 and PIR + CO_2 methods

Based on the results above, the integration of PIR and CO_2 achieved a higher accuracy than using CO_2 alone. In addition, Fig. 6 reveals that the integrated method performed better in most intervals, with a minimum

Table 7
Comparison of the performance between different data-driven models.

Model	Inactive period		(Tolerance with ±1 interval)		Active period		(Tolerance with ±1 interval)	
	accuracy	RMSE	accuracy	RMSE	accuracy	RMSE	accuracy	RMSE
SVM	78.9%	0.70	94.8%	0.56	30.2%	2.78	59.4%	2.29
Decision Tree	73.3%	0.75	88.1%	0.62	21.1%	3.02	42.3%	2.51
Random Forest	71.2%	0.79	83.2%	0.65	20.3%	3.25	40.7%	2.57



(a) Occupant number based on CO_2

(b) Occupant number based on PIR+ CO_2

Fig. 5. Identification of the occupant number based on the PIR and CO_2 (testing set).

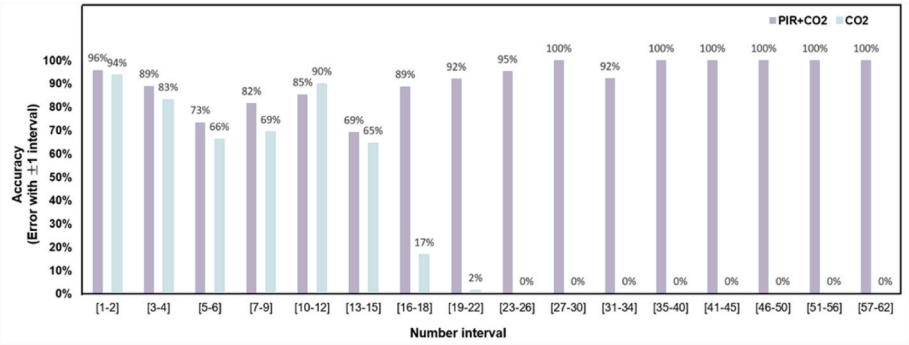


Fig. 6. Evaluation of recognition of each number interval.

accuracy of 69% in Refs. [13–15] interval. When the number of occupants was greater than 18, the accuracy (tolerance with ±1 interval) of interval recognition was 0 for the CO_2-based model, which demonstrated the weak applicability of using only CO_2 level for headcount in a large space with a crowd.

4.6. Comparison of PIR + CO_2 method and midnight reset PIR method

Some studies have suggested that automatically resetting PIR readings at midnight, a time typically characterized by fewer or no occupants, can be used as a strategy to minimize the cumulative error of PIR [63–65]. This approach, referred to as the midnight reset PIR method, was adopted as a baseline for comparison with the proposed PIR + CO_2 method. Considering the actual usage of the CSET studio, we set the PIR readings to automatically reset at 0:00 every day. The accuracy of the number recognition achieved by this method on the testing dataset was then evaluated in comparison to the results yielded by the PIR + CO_2 method.

Table 8 provides a comparative analysis of the performance of the PIR + CO_2 method and the midnight reset PIR method. The overall accuracy of the test set using PIR + CO_2 method is approximately 42.9%, which significantly increases to 85% when the recognition with an adjacent interval is acceptable. The RMSE value is 1.21 number intervals, suggesting that the recognitions are reliable. In contrast, the midnight reset PIR method, while simpler, underperforms in all three metrics, achieving 36.4%, 77.3% and 1.39, respectively.

5. Activity classification and activity intensity

5.1. K-means clustering

CO_2, Sound decibel, PIR, and the occupant number identified by the PIR + CO_2 model were selected for clustering. Three clusters were predetermined, representing low, middle and high activity classifications. In this study, the K-means clustering algorithm was implemented to establish the low, medium, and high clustering centers. The classification principle was based on the distance from these cluster centers. The specific thresholds for each parameter and activity intensity are outlined in Table 9.

The confusion matrix, Kappa coefficient and F1-Score were used to evaluate the performance. Kappa coefficient is a commonly used measure for consistency tests, and it varies from 0 to 1, indicating poor consistency to good consistency [66]. F1-measure considers both accuracy and recall, and the value of F1-Score is from 0 to 1, with 1 being the best and 0 being the worst [32].

Table 8
Comparison between PIR + CO_2 method and midnight reset PIR method.

Methods	Accuracy	Accuracy (Tolerance with ±1 interval)	RMSE
PIR and CO_2	42.9%	84.5%	1.21
Midnight reset PIR	36.4%	77.3%	1.39

Table 9
Identification threshold of each indicator.

Parameter	Activity Intensity		
	Low	Middle	High
CO_2 (ppm)	$\leq$ 488.94	(488.94–587.43)	$\geq$ 587.43
Decibel (dB)	$\leq$ 43.58	(43.58–47.64)	$\geq$ 47.64
Number of occupants	$\leq$ 11	(11–31)	$\geq$ 31

Table 10 shows that the overall accuracy reached about 89%, and the Kappa coefficient and F1-Score also proved a high degree of consistency between the recognized results and the ground truth information. The confusion matrix reveals variability in the model's performance across different intensity levels. For low-intensity events, the model correctly identified them 89% of the time. However, it misclassified 11% of low-intensity events as middle-intensity. For middle-intensity events, the model had a correct rate of 90%, with 10% of the events misclassified as low-intensity. The model was highly accurate for high-intensity events, correctly identifying them 96% of the time. Only 1% of high-intensity events were misclassified as low-intensity, and 3% as middle-intensity.

Indeed, the evaluation of low-intensity and medium-intensity events exhibits approximately a 10% error rate. The potential for bias in the clustering results may be due to the small number of data sets, the low frequency of activities and the lack of typical event characteristics. The small number of datasets might lead to overfitting, resulting in the model performing poorly on unseen data. The low frequency of activities can result in the underrepresentation of certain activity patterns, thereby leading to biased model predictions. Moreover, the lack of typical event characteristics could result in ineffective feature extraction, impeding the model's ability to accurately classify activity levels.

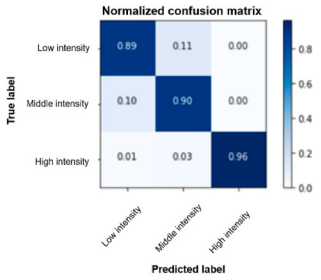
5.2. Validation of clustering

The ground truth information of the activities in the CSET studio was identified manually by the footage from the surveillance cameras. Validation was carried out for the testing group (2022/4/2–2022/4/14). The incorrect identification was colored in white, as shown in Fig. 7.

It is demonstrated that the period between 0:00 and 9:00 a.m. was dominated by low-level activities, while the middle-level activities occurred mainly in the afternoon and evening, and the high-level activities were concentrated between 9:00 to 16:00 in two days. The incorrect identification mainly occurred between 17:00 and midnight, during which time the number of occupants was usually at the critical threshold between low and middle intensity, causing a bias in judgment.

Table 11 compares the performance of different machine learning models in identifying the activity classifications. Although the accuracy of low-level activity was lower than that of the other algorithms, the K-means method outperformed in the identification of middle and high-

Table 10
Evaluation of the activity classification (testing group).

Activity classification	Duration (min) [proportion]	Accuracy	Kappa coefficient	F1-Score	Confusion matrix
Low Middle High	13210 [74%] 3870 [22%] 790 [4%]	89.3%	0.92	0.97	

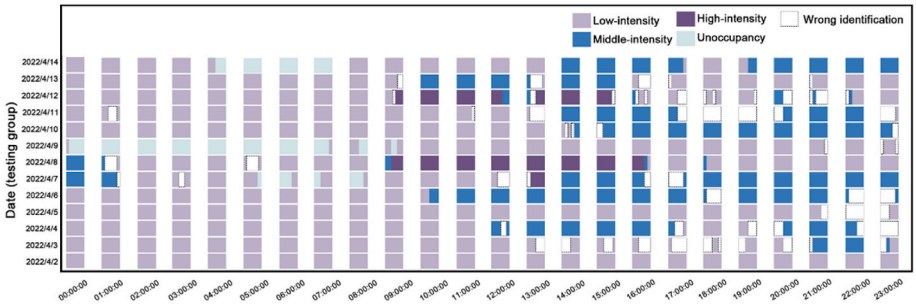


Fig. 7. Event classification based on clusters (testing group).

Table 11
Comparison of the activity classification identification model.

Model	Precision			Accuracy
	Low	Middle	High	
K-means	88.7%	89.7%	96.2%	89.3%
SVM	100.0%	50.1%	50.6%	85.3%
Decision Tree	95.6%	48.6%	50.6%	83.4%
Random Forest	96.5%	49.6%	52.7%	84.4%

level activities and achieved the highest overall accuracy.

5.3. Activity intensity recognition

As shown in Table 12, the Pearson correlation coefficients were first calculated between CO_2 levels, sound decibels, PIR readings, identified occupant numbers and activity classifications. Then, these correlation coefficients were normalized such that their sum equals 1. These values were then employed as weighting coefficients in Eq. (5) to determine the activity intensity.

The activity classification and the activity intensity from 2022/3/8–2022/4/11 are shown in Fig. 8. The daily activity intensity exhibits a generally oscillating pattern. The mean activity intensity for low, middle and high activity classifications were 18.8%, 34.7% and 75.2%, respectively, showing the distinctions between different activities. The classification corresponds well to the activity intensity. It can therefore be concluded that activity intensity as defined by Eq. (5) is a valid, quantitative measure of indoor activities.

Table 12
Weight coefficient of each indicator.

Indicator	Pearson correlation coefficient	Normalized Weight coefficient
CO_2	0.30	0.11
Decibel	0.62	0.23
Occupant number	0.89	0.33
PIR	0.87	0.32

6. Discussion

6.1. Occupant number analysis

The first part of this paper aims to identify reasonable intervals of the number of occupants in a large space. Even in the inactive period, when the number of people was relatively small, the accuracy rate was still less than 80%, implying that there were limitations in discriminating the number of people in spaces with an extensive range of activities based on environmental parameters. Moreover, although PIR sensors are widely used for detecting passenger flow in public spaces, it has been found in this paper that it was difficult to determine the number of people in pursuit of accuracy, due to the cumulative nature of PIR errors. An effective way to minimize the errors is to calibrate the data timely. Therefore, in this study, CO_2 level measurements were applied to calibrate the PIR method in the idle time, achieving a high accuracy (tolerance with ± 1 interval) of 84.5%.

In addition, the number interval setting may appear to compromise the accurate number identification. However, the overall benefits from the application of the proposed method with the low-cost, non-intrusive equipment could lead to improvements on the robustness of the HVAC control system and its ability to cope with unexpected environmental changes and occupant requirements.

6.2. Activity intensity analysis

It was found that directly acquired environmental information, such as CO_2 and sound decibels, is relatively weak for direct identification of activities, hence the proposal of a new indicator of activity intensity. For large spaces, room temperature changes are sometimes not obvious and timely in relation to the number of people. If an HVAC control strategy is based only on the number of occupants detected and the room temperature monitored, this may result in a delay in meeting the actual

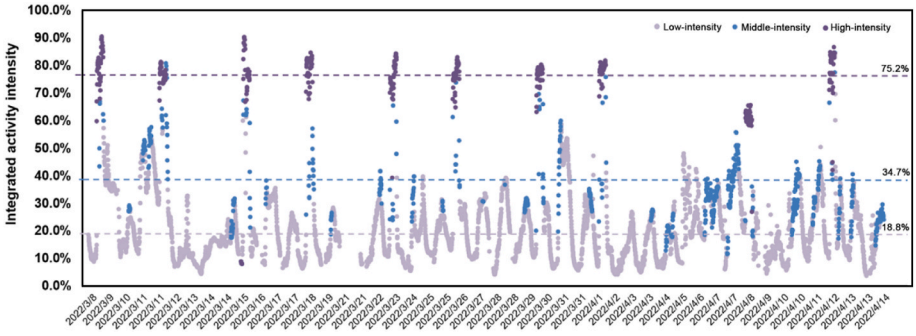


Fig. 8. Corresponding relationship between activity classification and activity intensity.

demand. The introduction of the concept of activity intensity can be used to measure the activities being carried out indoors. It can contribute to occupant-based control strategies that provide HVAC with advanced information on indoor activities.

6.3. Novel contributions of the study

This study introduced two contributions that advance the field of occupancy estimation. The first novelty of this study lies in the use of occupant number intervals in a stepwise increasing manner in large spaces. This methodology provides a more realistic model for occupancy estimation in large spaces where the number of occupants can vary significantly. It recognizes that the indoor environmental impact of occupants does not scale linearly with the number of people. In a large space, the environmental impact of each additional occupant is not the same and tends to diminish as the total number of occupants increases. Furthermore, a small deviation from the exact number of occupants usually has little impact on the control strategies for heating and cooling systems. This understanding is an advancement in improving the efficacy of occupancy detection in large spaces.

The second contribution is the introduction of a novel approach for the quantitative measurement of indoor human activity intensity, which is based on multiple environmental variables. This approach is distinct from existing studies that focus mainly on the count of occupants. Our method accounts for the variability in human metabolic rates and heat production when performing different activities, hence affecting the indoor environment differently.

7. Conclusion

This study developed an approach to calibrate PIR sensed headcount and to identify occupant activity information in large spaces based on measured environmental parameters such as CO_2 and sound decibel levels. The following conclusions are drawn.

- For a large space, the initial presence of people will have a large impact on the indoor environment, whereas once the people reach a certain amount, a small change in the population will not have a significant impact on the environment. Thus, this study proposed a non-uniformly distributed interval of occupant number for large spaces to replace the conventional expectation of predicting the exact number of people.
- For a large space with highly variable occupancy, the headcount using PIR or CO_2 alone can often be inaccurate. To tackle this problem, this study used PIR + CO_2 level to estimate the headcount interval and was able to increase the identification accuracy to 84.5% with RMSE of 1.21, even when the maximum population was up to about 60.
- To categorize the main activities in CSET, K-means clustering was used. The results demonstrated that the model achieved more than 85% accuracy in identifying low-, medium- and high-level activity.
- This study proposed a new concept of activity intensity to measure the activities being carried out indoors. The proposed metric takes a value between 0 and 1 and it considers CO_2 levels, sound decibels, PIR readings and the number of occupants. The metric was found to be appropriate for quantifying the activity in the room of the study. Using this metric, this study identified very clearly the three representative activities that occur in the CSET studio.

This study analyzed the intensity of differences in a large space; however, it did not explore identifying the precise location of the occupants or the environmental conditions locally around them. Such information could help improve the control system to supply location-based adjustment, especially in public spaces with multiple thermal zones. Furthermore, the potential use of machine learning approaches for combining and interpreting multiple environmental data could lead to more accurate occupancy detection and prediction. Exploring these avenues will form an important part of our future work.

Statement

During the final editing stages of the manuscript the authors used ChatGPT in order to improve readability and language. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

CRediT authorship contribution statement

Xiaohao Zhang: Writing – original draft, Investigation, Formal analysis. **Tongyu Zhou:** Writing – review & editing, Supervision, Conceptualization. **Georgios Kokogiannakis:** Writing – review & editing. **Liang Xia:** Writing – review & editing. **Chaoju Wang:** Software.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

This research was funded by Ningbo Commonwealth Funding Scheme, grant number: 2021S081 and internal seed funding of the University of Nottingham Ningbo China, grant number: RESI202203005.

References

[1] China association of building energy efficiency, China Building Energy Consumption Research Report, 2022.

[2] T. Yang, A. Bandyopadhyay, Z. O'neill, J. Wen, B. Dong, From occupants to occupants: a review of the occupant information understanding for building HVAC occupant-centric control, Build. Simulat. 15 (2022) 913–932, https://doi.org/10.1007/s12273-021-0861-0.

[3] H. Chen, P. Chou, S. Duri, H. Lei, The design and implementation of a smart building control system, in: 2009 IEEE International Conference on E-Business Engineering, 2009.

[4] M. Kong, B. Dong, R. Zhang, Z. O'neill, HVAC energy savings, thermal comfort and air quality for occupant-centric control through a side-by-side experimental study, Appl. Energy 306 (2021), 117987, https://doi.org/10.1016/j.apenergy.2021.117987.

[5] T. Labeodan, C. De Bakker, A. Rosemann, W. Zeiler, On the application of wireless sensors and actuators network in existing buildings for occupancy detection and occupancy-driven lighting control, Energy Build. 127 (2016) 75–83, https://doi.org/10.1016/j.enbuild.2016.05.077.

[6] S.H. Ryu, H.J. Moon, Development of an occupancy prediction model using indoor environmental data based on machine learning techniques, Build. Environ. 107 (2016) 1–9, https://doi.org/10.1016/j.buildenv.2016.06.039.

[7] Z. Tu, C. Hong, H. Feng, EMACS: design and implementation of indoor environment monitoring and control system, in: 2017 IEEE/ACIS 16th International Conference on Computer and Information Science, ICIS, 2017.

[8] T.H. Pedersen, K.U. Nielsen, S. Petersen, Method for room occupancy detection based on trajectory of indoor climate sensor data, Build. Environ. 115 (2017) 147–156, https://doi.org/10.1016/j.buildenv.2017.01.023.

[9] J. Zhang, T. Zhao, X. Zhou, J. Wang, X. Zhang, C. Qin, M. Luo, Room zonal location and activity intensity recognition model for residential occupant using passive-infrared sensors and machine learning, Build. Simulat. 15 (2022) 1133–1144, https://doi.org/10.1007/s12273-0210870-z.

[10] P. Liu, S.-K. Nguang, A. Partridge, Occupancy inference using pyroelectric infrared sensors through hidden Markov models, IEEE Sensor. J. 16 (2015) 1062–1068, https://doi.org/10.1109/JSEN.2015.2496154.

[11] Y. Agarwal, B. Balaji, R. Gupta, J. Lyles, M. Wei, T. Weng, Occupancy-driven energy management for smart building automation, in: Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building, 2010, pp. 1–6.

[12] T. Labeodan, W. Zeiler, G. Boxem, Y. Zhao, Occupancy measurement in commercial office buildings for demand-driven control applications—a survey and detection system evaluation, Energy Build. 93 (2015) 303–314, https://doi.org/10.1016/j.enbuild.2015.02.028.

[13] F. Wahl, M. Milenkovic, O. Amft, A distributed PIR-based approach for estimating people count in office environments, in: 2012 IEEE 15th International Conference on Computational Science and Engineering, 2012, pp. 640–647.

[14] L.M. Candanedo, V. Feldheim, D. Deramaix, A methodology based on Hidden Markov Models for occupancy detection and a case study in a low energy residential building, Energy Build. 148 (2017) 327–341, https://doi.org/10.1016/j.enbuild.2017.05.031.

[15] Y. Agarwal, B. Balaji, R. Gupta, J. Lyles, M. Wei, T. Weng, Occupancy-driven energy management for smart building automation, in: Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building, 2010, pp. 1–6.

[16] S. Zikos, A. Tsolakis, D. Meskos, A. Tryferidis, D. Tzovaras, Conditional Random Fields - based approach for real-time building occupancy estimation with multi-sensory networks, Autom. ConStruct. 68 (2016) 128–145, https://doi.org/10.1016/j.autcon.2016.05.005.

[17] C. Galván-Tejada, F. López-Monteagudo, O. Alonso-González, J. Galván-Tejada, J. Celaya-Padilla, H. Gamboa-Rosales, R. Magallanes-Quintanar, L. Zanella-Calzada, A Generalized Model for Indoor Location Estimation Using Environmental Sound from Human Activity Recognition, ISPRS International Journal of Geo-Information, 2018, p. 7.

[18] W.K. Chang, T. Hong, Statistical analysis and modeling of occupancy patterns in open-plan offices using measured lighting-switch data, Build. Simulat. 6 (2013) 23–32, https://doi.org/10.1007/s12273-013-0106-y, 2003.

[19] S. Yang, J. Liu, X. Gong, G. Huang, F. Yin, An adaptive smartphone hybrid indoor positioning solution incorporating heterogeneous sensors, in: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2022.

[20] R. Rabiee, J. Karlsson, Multi-Bernoulli tracking approach for occupancy monitoring of smart buildings using low-resolution infrared sensor array, Rem. Sens. 13 (2021) 3127, https://doi.org/10.3390/rs13163127.

[21] T. Labeodan, W. Zeiler, G. Boxem, Y. Zhao, Occupancy measurement in commercial office buildings for demand-driven control applications—a survey and detection system evaluation, Energy Build. 93 (2015) 303–314, https://doi.org/10.1016/j.enbuild.2015.02.028.

[22] R.H. Dodier, G.P. Henze, D.K. Tiller, X. Guo, Building occupancy detection through sensor belief networks, Energy Build. 38 (9) (2006) 1033–1043, https://doi.org/10.1016/j.enbuild.2005.12.001.

[23] Q. Huang, C. Mao, Occupancy estimation in smart building using hybrid CO_2/light wireless sensor network, Journal of Applied Sciences and Arts 1 (2) (2017) 5. http://opensiuc.lib.siu.edu/jasa/vol1/iss2/5.

[24] F. Wang, Q. Feng, Z. Chen, Q. Zhao, Z. Cheng, J. Zou, Y. Zhang, J. Mai, Y. Li, H. Reeve, Predictive control of indoor environment using occupant number detected by video data and CO_2 concentration, Energy Build. 145 (2017) 155–162, https://doi.org/10.1016/j.enbuild.2017.04.014.

[25] P.F. Pereira, N.M.M. Ramos, Detection of occupant actions in buildings through change point analysis of in-situ measurements, Energy Build. 173 (2018) 365–377, https://doi.org/10.1016/j.enbuild.2018.05.050.

[26] G. You, Y. Li, Environmental sounds recognition using tespar, in: 5th International Congress on Image and Signal Processing, 2012, pp. 1796–1800.

[27] M. Dorokhova, C. Ballif, N. Wyrsch, Rule-based scheduling of air conditioning using occupancy forecasting, Energy and AI 2 (2020), https://doi.org/10.1016/j.egyai.2020.100022.

[28] S.M.R. Khani, F. Haghighat, K. Panchabikesan, M. Ashouri, Extracting energy-related knowledge from mining occupants' behavioral data in residential buildings, J. Build. Eng. 39 (2021), https://doi.org/10.1016/j.jobe.2021.102319.

[29] T. Ekwevugbe, N. Brown, V. Pakka, D. Fan, Real-time building occupancy sensing using neural-network based sensor network, in: 2013 7th IEEE International Conference on Digital Ecosystems and Technologies, DEST), 2013, pp. 114–119.

[30] M.S. Zuraimi, A. Pantazaras, K.A. Chaturvedi, J.J. Yang, K.W. Tham, S.E. Lee, Predicting occupancy counts using physical and statistical CO2-based modeling methodologies, Build. Environ. 123 (2017) 517–528, https://doi.org/10.1016/j.buildenv.2017.07.027.

[31] A. Ebadat, G. Bottegal, D. Varagnolo, B. Wahlberg, K.H. Johansson, Estimation of building occupancy levels through environmental signals deconvolution, in: Proceedings of the 5th ACM Workshop on Embedded Systems for Energy-Efficient Buildings, 2013, pp. 1–8.

[32] A. Arora, M. Amayri, V. Badarla, S. Ploix, S. Bandyopadhyay, Occupancy estimation using non intrusive sensors in energy efficient buildings, in: 14th Conference of International Building Performance Simulation Association, Hyderabad, India, Dec. 2015, pp. 7–9.

[33] C. Jiang, M.K. Masood, Y.C. Soh, H. Li, Indoor occupancy estimation from carbon dioxide concentration, Energy Build. 131 (2016) 132–141, https://doi.org/10.1016/j.enbuild.2016.09.002.

[34] W. Wang, J. Chen, T. Hong, N. Zhu, Occupancy prediction through Markov based feedback recurrent neural network (M-FRNN) algorithm with WIFI probe technology, Build. Environ. 138 (2018) 160–170, https://doi.org/10.1016/j.buildenv.2018.04.034.

[35] S. Dedesko, B. Stephens, J.A. Gilbert, J.A. Siegel, Methods to assess human occupancy and occupant activity in hospital patient rooms, Build. Environ. 90 (2015) 136–145, https://doi.org/10.1016/j.buildenv.2015.03.029.

[36] S. Zhan, A. Chong, Building occupancy and energy consumption: case studies across building types, Energy Built Environ 2 (2) (2021) 167–174, https://doi.org/10.1016/j.enbenv.2020.08.001.

[37] S. Naylor, M. Gillott, T. Lau, A review of occupant-centric building control strategies to reduce building energy use, Renew. Sustain. Energy Rev. 96 (2018) 1–10, https://doi.org/10.1016/j.rser.2018.07.019.

[38] T.A. Nguyen, M. Aiello, Beyond Indoor Presence Monitoring with Simple Sensors, PECCS, 2012, pp. 5–14.

[39] N. Mahyuddin, H. Awbi, A review of CO2Measurement Procedures in ventilation research, Int. J. Vent. 10 (4) (2012) 353–370.

[40] E.A.B. Maldonado, J.E. Woods, A method to select locations for indoor air quality sampling, Build. Environ. 18 (4) (1983) 171–180, https://doi.org/10.1016/0360-1323(83)90025-2.

[41] G. Pei, D. Rim, S. Schiavon, M. Vannucci, Effect of sensor position on the performance of CO2-based demand controlled ventilation, Energy Build. 202 (2019), 109358, https://doi.org/10.1016/j.enbuild.2019.109358.

[42] P. Singh, N. Singh, K.K. Singh, A. Singh, Diagnosing of disease using machine learning, in: K.K. Singh, M. Elhoseny, A. Singh, A.A. Elngar (Eds.), Machine Learning and the Internet of Medical Things in Healthcare, E-Publishing Inc., 2021, pp. 89–111.

[43] Z. Yang, B. Becerik-gerber, Modeling personalised occupancy profiles for representing long term patterns by using ambient context, Build. Environ. 78 (2014) 23–35, https://doi.org/10.1016/j.buildenv.2014.04.003.

[44] A. Kaneko, X. Zhu, J. Lin, Data assimilation, in: A. Kaneko, X. Zhu, J. Lin (Eds.), Coastal Acoustic Tomographye, E-Publishing Inc., 2020, pp. 95–106.

[45] B. Dong, B. Andrews, K.P. Lam, M. Höynck, R. Zhang, Y.S. Chiou, D. Benitez, An information technology enabled sustainability test-bed (ITEST) for occupancy detection through an environmental sensing network, Energy Build. 42 (2010) 1038–1046, https://doi.org/10.1016/j.enbuild.2010.01.016.

[46] K. Qu, J. Xu, Q. Hou, K. Qu, Y. Sun, Feature selection using Information Gain and decision information in neighborhood decision system, Appl. Soft Comput. 136 (2023), 110100, https://doi.org/10.1016/j.asoc.2023.110100.

[47] B. Dong, R. Markovic, S. Carlucci, Y. Liu, A. Wagner, A. Liguori, C. van Treeck, D. Oleynikov, E. Azar, G. Fajilla, J. Drgoňa, J. Kim, M. Vellei, M. De Simone, M. Shamsaiee, M. Bavaresco, M. Favero, M. Kjaergaard, M. Osman, M. Frahm, S. Dabirian, D. Yan, X. Kang, A guideline to document occupant behavior models for advanced building controls, Build. Environ. 219 (2022), 109195, https://doi.org/10.1016/j.buildenv.2022.109195.

[48] L.M. Candanedo, V. Feldheim, Accurate occupancy detection of an office room from light, temperature, humidity and CO_2 measurements using statistical learning models, Energy Build. 112 (2016) 28–39, https://doi.org/10.1016/j.enbuild.2015.11.071.

[49] Y. Yuan, X. Li, Z. Liu, X. Guan, Occupancy estimation in buildings based on infrared array sensors detection, IEEE Sensor. J. 20 (2) (2019) 1043–1053, https://doi.org/10.1109/JSEN.2019.2943157.

[50] C. Liao, P. Barooah, An integrated approach to occupancy modeling and estimation in commercial buildings, in: Proceedings of the 2010 American Control Conference, 2010, pp. 3130–3135.

[51] Y.Z. Yang, N. Li, B. Becerik-Gerber, M. Orosz, A multi-sensor-based occupancy estimation model for supporting demand driven HVAC operations, in: Proceedings of the 2012 Symposium on Simulation for Architecture and Urban Design 2, 2012, pp. 1–2.
[52] Z. Han, R.X. Gap, Z. Fan, Occupancy and Indoor Environment Quality Sensing for Smart Buildings, IEEE international instrumentation and measurement technology conference proceedings, 2012, pp. 882–887, 2012.
[53] T. Ekwevugbe, N. Brown, V. Pakka, Real-time Building Occupancy Sensing for Supporting Demand Driven HVAC Operations, 2013.
[54] B. Ai, Z. Fan, R.X. Gao, Occupancy estimation for smart buildings by an auto-regressive hidden Markov model, in: 2014 American Control Conference, 2014, pp. 2234–2239.
[55] M.K. Masood, C. Jiang, Y.C. Soh, A novel feature selection framework with Hybrid Feature-Scaled Extreme Learning Machine (HFS-ELM) for indoor occupancy estimation, Energy Build. 158 (2018) 1139–1151, https://doi.org/10.1016/j.enbuild.2017.08.087.
[56] S. Meyn, A. Surana, Y. Lin, S.M. Oggianu, S. Narayanan, T.A. Frewen, A sensor-utility-network method for estimation of occupancy in buildings, in: Proceedings of the 48h IEEE Conference on Decision and Control (CDC) Held Jointly with 2009 28th Chinese Control Conference, 2009, pp. 1494–1500.
[57] I.G. Dino, E. Kalfaoglu, O.K. Iseri, B. Erdogan, S. Kalkan, A.A. Alatan, Vision-based estimation of the number of occupants using video cameras, Adv. Eng. Inf. 53 (2022), 101662, https://doi.org/10.1016/j.aei.2022.101662.
[58] J. Yang, A. Pantazaras, K.A. Chaturvedi, A.K. Chandran, M. Santamouris, S.E. Lee, K.W. Tham, Comparison of different occupancy counting methods for single system-single zone applications, Energy Build. 172 (2018) 221–234, https://doi.org/10.1016/j.enbuild.2018.04.051.
[59] H.E. Hailemariam, R. Goldstein, R. Attar, A. Khan, Real-time occupancy detection using decision trees with multiple sensor types, in: Proceedings of the 2011 Symposium on Simulation for Architecture and Urban Design, 2021, pp. 141–148.
[60] Z. Wang, Y. Wang, R. Zeng, R.S. Srinivasan, S. Ahrentzen, Random Forest based hourly building energy prediction, Energy Build. 171 (2018) 11–25, https://doi.org/10.1016/j.enbuild.2018.04.008.
[61] ASHRAE, ASHRAE Handbook: Fundamentals, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, Ga, USA, 2017.
[62] ASHRAE, ASHRAE 55-2010: Thermal Environmental Conditions for Human Occupancy, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, Ga, USA, 2010.
[63] S. Zikos, A. Tsolakis, D. Meskos, A. Tryferidis, D. Tzovaras, Conditional Random Fields - based approach for real-time building occupancy estimation with multi-sensory networks, Auto, Constr. Met. (CTICM) 68 (2016) 128–145, https://doi.org/10.1016/j.autcon.2016.05.005.
[64] N. Li, G. Calis, B. Becerik-Gerber, Measuring and monitoring occupancy with an RFID based system for demand-driven HVAC operations, Autom. ConStruct. 24 (2012) 89–99, https://doi.org/10.1016/j.autcon.2012.02.013.
[65] T. Kitzberger, J. Kotik, T. Pröll, Energy savings potential of occupancy-based HVAC control in laboratory buildings, Energy Build. 263 (2022), 112031, https://doi.org/10.1016/j.enbuild.2022.112031.
[66] C.H. Yu, Test–retest reliability, Encyclopedia of Socia. Measur. (2005) 777–784, https://doi.org/10.1016/B0-12-369398-5/00094-3.