Khan A, Nicholson J, Mellor S, Jackson D, Ladha K, Ladha C, Hand J, Clarke J, Olivier P, Ploetz T. Occupancy Monitoring using Environmental & Context Sensors and a Hierarchical Analysis Framework. *In: BuildSys '14 Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*. 2014, Memphis, USA: ACM.

DOI link to article:

http://dx.doi.org/10.1145/2674061.2674080

Date deposited:

05/05/2016

Occupancy Monitoring using Environmental & Context Sensors and a Hierarchical Analysis Framework

Aftab Khan[1]*, James Nicholson[1], Sebastian Mellor[1], Daniel Jackson[1], Karim Ladha[1], Cassim Ladha[1], Jon Hand[2], Joseph Clarke[2], Patrick Olivier[1], Thomas Plötz[1]

[1]Culture Lab, School of Computing Science, Newcastle University, Newcastle upon Tyne, UK.
[2]Energy Systems Research Unit, University Of Strathclyde, Glasgow, Scotland, UK.

Email: aftab.khan@newcastle.ac.uk

Abstract

Saving energy in residential and commercial buildings is of great interest due to diminishing resources. Heating ventilation and air conditioning systems, and electric lighting are responsible for a significant share of energy usage, which makes it desirable to optimise their operations while maintaining user comfort. Such optimisation requires accurate occupancy estimations. In contrast to current, often invasive or unreliable methods we present an approach for accurate occupancy estimation using a wireless sensor network (WSN) that only collects non-sensitive data and a novel, hierarchical analysis method. We integrate potentially uncertain contextual information to produce occupancy estimates at different levels of granularity and provide confidence measures for effective building management. We evaluate our framework in real-world deployments and demonstrate its effectiveness and accuracy for occupancy monitoring in both low- and high-traffic area scenarios. Furthermore, we show how the system is used for analysing historical data and identify effective room misuse and thus a potential for energy saving.

Keywords

Occupancy estimation; Hierarchical modeling; Environmental sensing; Energy.

1 Introduction

Sustainability, and more specifically thoughtful use of energy, has become a major global concern due to diminishing natural resources and growing demands inflicted by changing lifestyles [12]. In the next 30 years alone it is predicted that the world-wide consumption of energy will increase by over 56% [9] and alternative means of generating electrical energy are not yet as efficient as the more established ways of power generation. As a consequence, shortages are predicted within the next few decades [24]. This leaves actively saving energy as the only realistic option for more sustainable consumption.

In particular, the operation of commercial and residential buildings requires substantial amounts of energy. This is mainly related to energy-hungry heating, ventilation and air conditioning (HVAC) operations [26], as well as to such mundane requirements as appropriate lighting. Strategies for effective energy management in such environments are related to understanding consumption and, based on this, developing procedures for saving energy. *Accurate* information about occupancy, both in temporal as well as spatial context, is key for smart control strategies that reduce energy consumption by adjusting HVAC appliances and lighting appropriately without disrespecting basic comfort levels. Automatically deriving occupancy information at a fine-grained level of both temporal and spatial detail remains a surprisingly challenging endeavour. Existing, straightforward approaches utilise behaviour prediction techniques [8] based on video surveillance or activity monitoring using wearable sensing [3]. As such these approaches depend on existing, often complex, infrastructures being in place. More critically, camera-based approaches, e.g., employing vision based attention systems [31], are almost always out of question for work-spaces due to ethical concerns [17].

Occupancy monitoring requires varying levels of detail depending on the application case. Whereas, for example, automatic lighting control is typically based on binary decisions on whether people — any number — are in a room or not, HVAC control requires more detailed information — how many occupants? — as such numbers indirectly drive the settings of HVAC systems when aiming for comfortable room climate. Apart from such situated control applications, long-term optimisation of energy consumption in, for example, office buildings is typically based on detailed and accurate room usage information with the aim of allocating resources appropriately to reduce the overall energy footprint.

In this paper we present a novel approach for accurate occupancy monitoring that is based on a wireless sensor network measuring temperature, light, humidity, Passive Infrared (PIR) readings, and audio levels. Our main **contributions** are as follow: *i)* A wireless sensing infrastructure that is straightforward to deploy – consisting solely of a single sensing device, which is no larger than a portable mouse – cost-effective, and only collects non-sensitive environmental information; *ii)* The sensing infrastructure is then combined with a novel hierarchical analysis framework that utilizes statistical classifiers in general and can integrate potentially uncertain contextual information (meeting schedules, computer usage). This framework is used to produce highly accurate occupancy estimates at different levels of granularity – from binary occupied/not-occupied decisions to accurate counts of the number of occupants in a room or an area of a building to cater for a variety of applications. All automatically derived estimates are accompanied with confidence measures based on level-specific posterior probabilities produced by the recognition framework. This confidence information is forwarded to subsequent classification levels, which is key for accurate and fine-grained occupancy estimation under uncertainty; *iii)* The occupancy and confidence information is then presented in a hierarchical fashion to support facilities managers in making decisions regarding energy consumption and/or resource allocation.

In real-world deployments we demonstrate the efficacy and accuracy of our occupancy monitoring framework. In two case studies we evaluate occupancy estimations for both a High Traffic Area (HTA) and Low Traffic Area (LTA) in an office building over a period of ten days each. Our framework is able to estimate occupancy with a very high level of precision, ranging from almost perfect binary classification of general occupancy to more than 75% accuracy on the most fine grained level of estimating the actual number of occupants. We also identify significant potential for optimising the booking schedule for a meeting room by linking meeting timetable information to predicted true occupancy with the purpose of focusing energy consumption properly and thus reducing the uneccessary energy consumption of seldom-used rooms.

2 Related Work

The problem of occupancy inference in the context of commercial buildings has received significant interest in recent years due to constant efforts to lower energy consumption [9]. Generally, occupancy inference can refer to detection – i.e., discovering whether one or more persons are in a room – or estimation – i.e., unveiling how many people are in a room. Traditionally, occupancy *detection* was carried out using camera-based methods [29], which are usually expensive regarding setup, equipment, and operation. Additionally, privacy concerns are associated with cameras recording individuals at all times [17]. As a consequence, networks of PIR sensors have been proposed as alternatives. In combination with probabilistic analysis methods such approaches have been trialled to improve building services [7]. Occupancy detection was then also used for energy management where similar setups led to energy savings for HVAC operation of at least 10% [1, 10].

Indicators for the number of people being in a room – occupancy *estimation* – can, in principle, be inferred from various types of information such as meeting schedules and network activity in addition to that collected by a sensor network. However, commonly available sources of data contain substantial levels of uncertainty that may negatively influence occupancy models. For example, a scheduled meeting may not necessarily take place or a person may forget to logout from their computer. Similarly, sensors deployed in opportunistic manners may also produce noisy measurements and could thus result in further inaccuracy – e.g. PIR sensors may fail to identify a person in a room if they remain still.

With this in mind recent work has focused on the analysis of multi-modal sensor data – including video footage. For example, the Sensor-Utility Network (SUN) method was developed to incorporate various, yet unreliable, sensor data and produced occupancy estimates on a building-wide level [23]. The CO_2 and PIR sensor data along with sound and people counts at entrances obtained from the recorded video footage was combined with historical usage data from the building. Whilst providing reasonable occupancy estimates, the approach becomes less accurate over time, which limits its practical applicability. Furthermore, the active use of video cameras severely limits its adoption due to privacy violations.

Pursuing a radically differing approach, Liao and Barooah [22] developed an agent-based framework to simulate the behaviour of a room's occupants. They extract reduced-order graphical models from Monte-Carlo simulations of their agent-based model, which was then validated with sensor data for one room and one occupant. Simulations were used to illustrate the effectiveness of the proposed method. However, due to the lack of formal occupancy measurements and especially by focusing on small office spaces where the nominal occupancy value was only one, the practicality of the approach beyond such rather simple application cases remains questionable.

In the context of energy saving, previous work has also focused on simplifying building infrastructures so that all sensors are able to communicate with each other using a common language. The Building Operating System Services (BOSS) [5] facilitates the development of building applications as data can be simply collated. Three applications were demonstrated to work with BOSS, including one that allowed the management of airflow and ventilation for HVAC control by using occupancy data. However, this application used data from a Google calendar feed for determining occupancy, which relies heavily on the users to keep the data updated and as such is a weak indicator of occupancy.

Of course, it is possible to estimate occupancy using methods other than networked sensors. For example, Crowd++ uses the microphone in smartphones to estimate the occupancy of any room [38]. While such a system is appealing due to its non-dependence on the building's infrastructure, it requires all residents to carry a specially-configured smartphone at all times which may not be realistic in a number of organisations. Smartphones can also be used for occupancy *prediction* to support HVAC systems [18], but while very accurate, such implementations are only feasible in a home environment where only a few people live. Wearable devices have also been used to estimate occupancy and have proven to be very accurate (e.g., [3, 33]), but the privacy of individuals is typically compromised using this approach [35]. Finally, Cyber-Physical Systems have been used to control room-specific HVAC systems with promising results and at a low cost [34], but such an approach has limited applications due to detecting rather than estimating occupancy.

3 System Overview

The purpose of an automatic occupancy estimation framework such as the one proposed in this paper is to support decision makers (e.g., building and facility managers) to optimise the use of rooms and meeting spaces in large, commercial office buildings. Their goal is to save resources while at the same time maintaining the comfort for occupants and users of a building. Due to the inherent uncertainty and the dynamics of office routines combined with the delays in operating, for example, HVAC systems this optimisation often results in a balancing act where administrators need to make decisions without having complete and reliable information. Different granularity levels of occupancy data are necessary for different applications – e.g., binary occupancy detection for lighting, categorical occupancy estimation for HVAC systems, or exact headcounts for room allocation optimisation. Our occupancy estimation framework takes these constraints into account and provides results of automatic occupancy analysis combined with confidence information, which serves as a basis for effective and well informed human decision making. Due to ethical concerns our framework is solely based on sensing facilities that only collect non-sensitive information thereby explicitly avoiding privacy-violating means such as cameras or microphones.

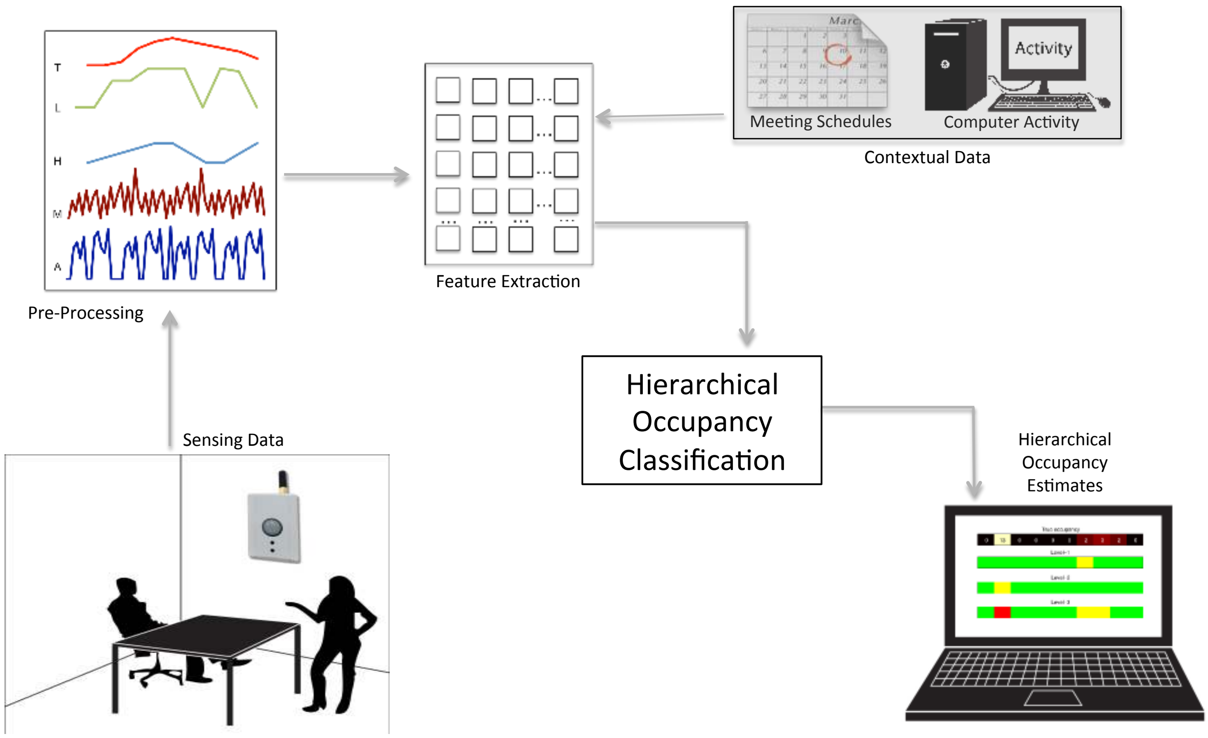

Figure 1: Hierarchical classification strategy for occupancy data, starting with the environmental sensors on the left and resulting in the hierarchical occupancy estimates on the right.

The core conceptual idea for our occupancy estimation framework is to break down the analysis in a hierarchical way and present the estimation results at different levels of granularity combined with confidence levels for the estimates at every level of the hierarchy. This concept is a response to the uncertainty and sometimes even unreliability of the analysed data, which may prevent accurate head counts for a room. In such cases a human operator could still make reasonable decisions if they had access to less detailed but more reliable information that would indicate whether a room is over- or under-crowded.

Our technical approach reflects this concept through introducing a hierarchical analysis framework. We consider occupancy estimation at three different levels of granularity: *i)* binary detection; *ii)* categorical occupancy estimation; and *iii)* counting the exact number of occupants. For this purpose we employ a hierarchy of statistical classifiers and use their associated posterior probabilities for representing confidence levels for decision making at subsequent levels. In an exemplary use case, the output of our framework may specify that a particular room is being used to capacity, but analysis level iii) states that ten people are in the room when the capacity is only 8. However, the confidence value for analysis level iii) may be low, which would tell an operator to not solely rely on that information. As a consequence, analysis level ii) may be a better indication for the room's actual usage.

The primary source of information for our analysis framework is environmental data, i.e. light, humidity, audio level, PIR etc., which are captured using a miniature sensing platform. In addition to this physical information, we integrate external data sources into our analysis, including computer activity and meeting schedules. Whilst the latter is by itself notoriously unreliable, its combination with the environmental sensor data provides additional cues making occupancy estimation reliable.

Figure 1 gives an overview of the overall concept for our framework, consisting of *i)* capturing both environmental and contextual data; *ii)* data pre-processing and sensor fusion; *iii)* uni-level occupancy classification and confidence estimation; and *iv)* integration into our hierarchical analysis model; which we will discuss in detail in the subsequent sections.

3.1 Environmental Sensing

3.1.1 Sensing Platform

As the primary source of information for our occupancy estimation framework we analyse environmental parameters of the areas monitored: *i)* motion – through a PIR sensor; *ii)* acoustic noise – through a microphone that records *audio level* information only; *iii)* temperature; *iv)* light; and *v)* humidity. We employ bespoke, inexpensive (approx. \$8 – \$16) wireless sensing platforms that integrate aforementioned sensors in a compact device ($102 \times 51 \times 33$mm), and are durable, reliable and very easy to deploy without any prior calibration requirements (see Table 1 for an overview of the main technical features, and [16]). Depending on the size and the topology of the monitored area we deploy one or more of these platforms. Multiple devices communicate via an ad-hoc mesh network for communicating to a central hub. The reliable lifetime of each sensor is approximately four months before the (two AA) batteries need replacing.

These environmental sensors monitor the area they are deployed in by collecting only non-sensitive information which deliberately makes it impossible to infer identities of occupants through analysing the measurements. All integrated sensors record at very low sampling rates (6 – 24Hz). Furthermore, due to recording only acoustic noise *levels* (at 24Hz) it is ensured through the hardware design that no conversations can be overheard. As such, this approach holds an advantage over methods that implement a digital camera for data collection (e.g., [23,36]) with respect to user acceptance for implementation in commercial or residential buildings.

3.1.2 Processing sensor data

Environmental sensing data is typically very noisy and thus, cannot be directly used in its raw form but requires pre-processing and feature extraction.

Pre-processing Sensor data may contain missing data points, e.g., due to packet losses during transmission or errors. In order to address this issue, we perform linear interpolation to remove any gaps in our measurements. For multi-modal integration of all our data sources, which are sampled at varying rates (see Table 1), and for effective sensor fusion we re-sample all recordings in 1 second steps, which effectively leads to a uniformly sampled, multi-modal data stream. We use a target sampling rate of 1 Hz, a frequency that has been validated with a view on bandwidth preservation, and appropriate information content while at the same time preventing the identification of occupants or any inference regarding the acoustic details including what has been said by whom.

After interpolation and re-sampling, we then use a standard sliding window approach of 5 minutes length with a 50% overlap that produces equally sized and distributed analysis windows. Here the concrete choice of the parameters of the sliding window procedure is based on the assumption that the number of people in a monitored room remains static during this temporal window, which is reasonable for the addressed scenario occupancy estimation for meeting rooms and offices.

Feature extraction Once the analysis frames are established, we compute various statistical features implicitly quantifying data distribution for each sensor modality. Mean and standard deviation are the standard statistical features used in such a context, capturing the central tendency and variation from the expected value of the data distribution, respectively. We compute (μ, σ) pairs for every sliding window frame per sensing modality (see Table 1).

Prior to computing the (μ, σ) pairs for audio and PIR packets, we compute three additional features for quantifying change. This includes the sum of sample points ($S(a(r))$ – Equation 1), the sum of absolute sequential differences ($D_s(a(r))$ – Equation 2) – and the sum of squares of sequential differences ($S_d(a(r))$ – Equation 3). These values are computed for each of the sensor packet signals – $a(r)$ where $r \in 1, 2, ..., R$ – containing audio and PIR values ($R = 24$Hz in our experiments (see Table 1)). All of these features provide important information related to a change in signal values e.g., it is the change in PIR values that we are primarily interested in.

$$S(a(r)) = \sum_{r=1}^{R} a(r) \tag{1}$$

$$D_s(a(r)) = \sum_{r=1}^{R-1} abs(a(r+1) - a(r)) \tag{2}$$

$$S_d(a(r)) = \sum_{r=1}^{R-1} (a(r+1) - a(r))^2 \tag{3}$$

We also compute the sum of spectral coefficients $x(t)^2$ for each sliding window normalized by the length of the window (of length $T = 300$ (5 minutes) in our case) highlighting associated signal energy ($E(x(t))$ in Equation 4). Similarly, we also measure signal entropy, $I(x(t))$ (Equation 5) to quantify the expected value of the information contained in a given sliding window for each sensing

Sensor	Freq. (Hz)	Sensitivity
Motion (PIR)	24	3m/135deg
Sound	24	-45 to -39dB
Temperature	6	5 to 50C
Light	6	3 to 70 k.lux
Humidity	6	±4% RH

Table 1: Environmental sensors used in our setup (right) and their main technical specifications (left).

input. Both features provide a measure of unpredictability of input signals important for highlighting differences between no occupancy and occupancy.

$$E(x(t)) = \frac{1}{T} \sum_{t=1}^{T} x(t)^2 \tag{4}$$

$$I(x(t)) = \sum_{t=1}^{T} x(t) \cdot log(|x(t)|) \tag{5}$$

Finally, in addition to the above, we compute ECDF (Empirical Cumulative Distribution Function) coefficients and include them into our feature representation (Equation 6). ECDF features preserve crucial information about the distribution of data within a frame (introduced in [11, 27], and now widely used in wearable and ubiquitous computing applications of accelerometry, e.g., [4, 15, 19, 20, 28]). We use a $d = 10$ coefficient ECDF feature representation:

$$F_e(x(t)) = \{x \in \mathbb{R} : \exists i, p_i \leq P(x)\} \tag{6}$$

where,

$$\{p_i\} \in \mathbb{R}^d_{[0,1]}, p_i \leq p_{i+1} \tag{7}$$

Feature Combination After computing all of the aforementioned features for all of the sensing modalities, we concatenate them into a combined feature vector, which then represents a given input signal $x(t) \in \mathbb{R}^D$. To do this, we first compute 9 sub-feature sets SF_s per sensor modality where $s \in \{1, 2, ..., 9\}$ representing sets for temperature, light, humidity, and Equations 1 to 3 each computed for audio and PIR:

$$SF_s = \{\mu, \sigma, F_e, E, I\}_s \tag{8}$$

All of these sub-features are concatenated, such that:

$$SF_c = \{SF_1, SF_2, ..., SF_9\} \tag{9}$$

Finally, we can define the feature vector as:

$$\vec{f}(x(t)) = (SF_c, \mathcal{D}(SF_c))^T \tag{10}$$

Note that this feature vector also contains the mean first order derivative, $\mathcal{D}(SF_c)$, of the concatenated sub-features i.e., SF_c between consecutive frames, which captures the temporal aspects of occupancy. Overall we compute $D = 127$-dimensional feature vectors per analysis frame as extracted by the sliding window procedure (9 sensor modalities × 14 features) + 1 first order derivative).

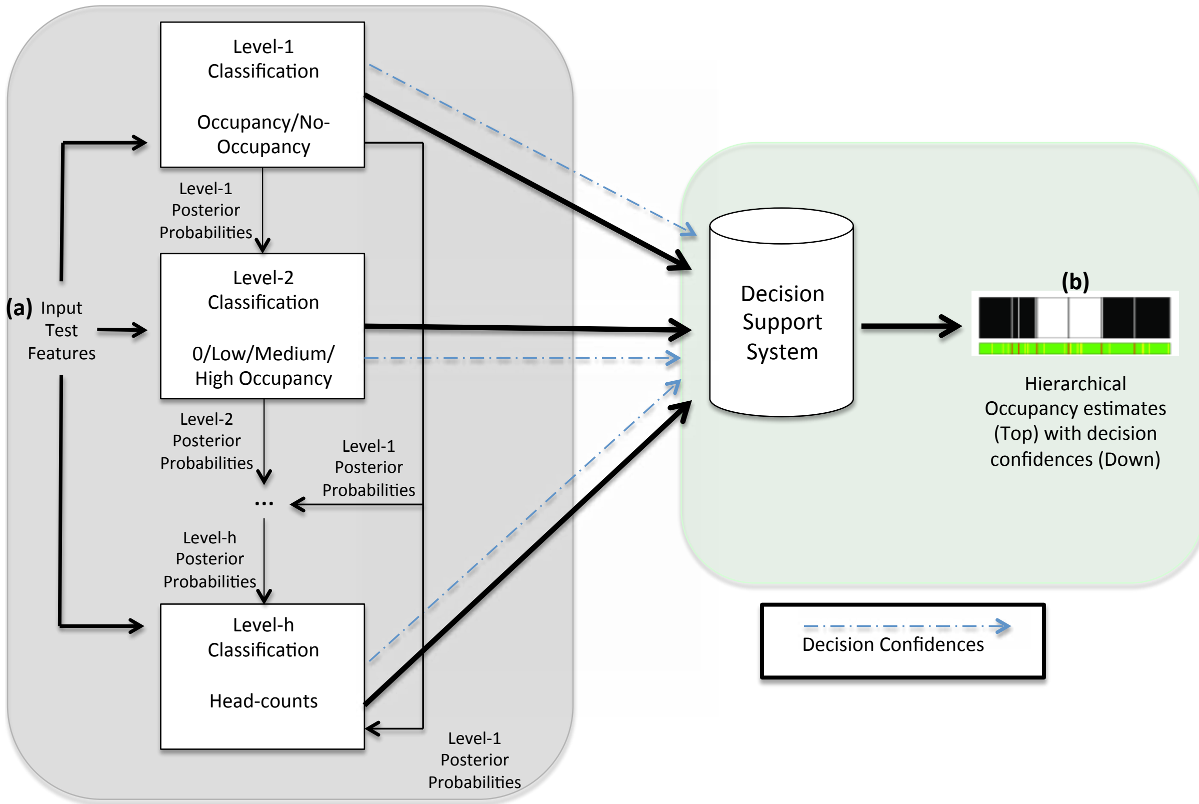

Figure 2: Overview of our hierarchical analysis framework (a) for occupancy estimation leading to a decision support system (b) for effective building management. See text for explanation.

3.2 Contextual Information

Whilst environmental sensors provide useful information for occupancy estimation, limiting the analysis to physical measurements can lead to ambiguous and potentially unreliable inferences. Examples of which are erroneous measurements due to environmental influences, e.g., skewed temperature readings on very sunny or overcast days, respectively, or simply incorrect with regard to actual room usage, e.g., short room visits of cleaning or security personnel would erroneously lead to occupancy detection although the room is not actually being used.

In order to disambiguate such inferences and to increase the general reliability of our occupancy estimation, we incorporate a second type of information into the analysis. Through opportunistically gathering contextual information from external data sources we enhance our observations, which shall lead to more accurate and more detailed occupancy analysis.

There are various forms of contextual data that could be used in combination with physical sensors such as location information, weather data, electricity consumption, meeting schedules, equipment usage, PC activity data, etc. In our experiments, we make use of meeting schedules, and PC activity information by appending this information to the feature vector (see Equation 10) as described in Section 4. Note that our framework is not restricted to specific contextual information as our sensor fusion technique ensures that all external information is effectively sampled at the same rate as the environmental sensors using the aforementioned pre-processing pipeline including re-sampling and interpolation.

3.3 Hierarchical Occupancy Estimation

We employ standard statistical classification procedures to classify occupancy at various levels of the hierarchical representation that is also capable of providing level-based confidences. These are used for providing decision confidence to the facilities manager as well as improving the occupancy estimates at subsequent levels.

After initial classification at every level, we extract the related posterior probabilities for each occupancy density prediction and use them as additional feature sets $a^h (h \in \{2,\ldots,H\}, H = \#\{levels\})$ for the subsequent levels (each with a decreasing level of abstraction – i.e., from level 2, down to the true number of occupants):

$$a^h = \{a^h_z; z \in O\}, \tag{11}$$

where $O = \{0,\ldots,C^{h-1}-1\}$, $C^h = \#\{\text{classes in layer } h\}$,

$$a^h_z = p^{h-1}(N = z|\vec{f}^{h-1}, \theta^{h-1}), \tag{12}$$

and $p^h(N = z)$ is the probability that N takes the class z in level h with feature vector $\vec{f}^h$ and parameters θ^h.

A block diagram for the hierarchical classification strategy is shown in Figure 2.

Such hierarchical classification strategy makes the hierarchy of occupancy representation more exact through enrichment of the feature space per level. Such that if, at the topmost level, the posterior probability of occupancy is high, then the proposed approach implicitly rejects the majority of anomalous outcomes such as zero or low occupancy values by using the posterior probabilities in the penultimate level.

3.4 Decision Support System

Decision making for optimising resource consumption is a non-trivial process in settings with slow technical reaction time while at the same time rather dynamic and hard to predict usage scenarios. For example, heating regulation systems have substantial power-up phases and even more substantial lagging times of the affected room temperature after powering down, which represents a substantial optimisation challenge with regards to occupants comfort and economic resource usage. Effective building management thus requires extensive expertise and insight into what is typically a very complex situation. A technical system outputting various levels of occupancy estimations can hardly replace human decision making due to complex scenarios and concessions that will have to be made from time to time, but such a system ought to support this decision making process, which is the main objective of our work.

We aim to provide effective support for facilities managers and our approach is to not only report general, i.e., binary occupancy states but also estimates of actual occupancy with as much detail as our system can confidently provide. Consequently, the outputs of our framework are not only occupancy estimations at the three contextual levels of our hierarchical analysis but also confidence values per level. The combination of these two result categories form the input to decision support systems that enables human operators to effectively manage the facilities for which energy consumption shall be optimised. An example of the decision support output is shown in Figure 2-(b). The confidence levels are achieved by using the posterior probabilities of the statistical classifier used, for a given class label by using the per class probability values. Consequently, the posterior probability associated with the predicted class label is divided between three levels of certainty (i.e., low, medium and high as shown in Figure 2) where green represents a highly trusted outcome and red represents a less trusted outcome. A human operator would then make a decision based on the level of confidence presented by the system.

4 Experiments

In this section, we report the experimental protocols followed by a comprehensive evaluation using both the accuracy of the occupancy estimation approach and the overall efficacy of the framework for practical applications.

4.1 Experimental Protocol

We deployed our sensing and analysis framework in a large commercial building that contained offices, meeting rooms, and the usual infrastructure such as bathrooms, stairs, hallways etc. With a view on the most common settings in such buildings we focus our evaluation on two scenarios: *i)* high-traffic area (HTA); and *ii)* low-traffic area(LTA).

In both scenarios we deployed the wireless sensing system as described in Section 3.1. We recorded data for 14 days in the LTA, and for 10 days in the HTA. Both studies were run during normal office days where the rooms were typically used from 8am to 6pm. The LTA consisted of a small meeting room that was moderately used throughout the day. Building residents typically reserved the room in advance using an online booking system. The room hosted one table with 12 chairs and a networked PC. The HTA consisted of a large open space that was in constant use over the course of the day. 20 residents typically worked in the area, which included 22 personal desks and two large tables. Unlike the LTA, residents mainly used personal laptops that were not always connected to the mains power.

The two scenarios presented assessed the feasibility of employing the framework in different contexts. The deployment of the system at a building-wide level could be modelled after the HTA setup, but for the purposes of our evaluation such a large-scale implementation was impractical. We compared the accuracy of our proposed hierarchical occupancy estimation approach with a standard non-hierarchical approach resembling the current state of the art. We evaluated the hierarchical and non-hierarchical approaches using two established statistical classifiers – kNN and SVM – to demonstrate its broad applicability.

4.1.1 Data Collection

Environmental Sensors: Two sensors were placed at opposite ends of the HTA to cover the area of approximately 92 square meters. One sensor was placed in the LTA to cover the area of approximately 45 square meters. The sensors were fastened to the walls using double sided tape at 1.5 meters high.

Contextual Sensors: For the LTA scenario we integrated two additional contextual data sources: meeting schedule and PC usage information (both anonymised to preserve the privacy of the occupants). Binary indicators of whether meetings where scheduled or not were recorded at 15 minute intervals, which represents our assumption of a minimum meeting duration, and binary indicators of PC activity were recorded at hourly intervals. Both values are linearly interpolated to the frequency specified by the window length in the previous section and then appended to the feature vector of Equation 10, resulting in augmented representations $\vec{f}' \in \mathbb{R}^{D'}$ with $D' = 129$.

a) Meeting Schedules: We obtained historical timetabling records for the LTA in which the sensing system was deployed. The schedules detailed the times when the room had been officially booked. However, this schedule information is not necessarily reliable as cancelled meetings or early finishes typically do not lead to schedule updates and thus would represent faulty information. Furthermore, this source of information does not unveil anything about potential under-crowding or over-crowding of a room. While the former would result in waste of resources, the latter would lead to discomfort for the occupants – both cases an optimisation would try to avoid.

b) PC Usage: The LTA contained a single, networked PC, which is used for presentations. PC activity data was used as second type of contextual information, consisting of the number of minutes in each hour that the computer was active for. Active in this context, means that the machine was processing CPU cycles or had done so no more than five minutes before. Simply, if the machine entered sleep mode (after five minutes of inactivity) then it was classed as inactive. Note that it is possible for a computer to be classified as active when in fact it is not being used by a person – e.g. a program is left running or an active script in the browser prevents the computer from sleeping. Additionally, it is possible for a meeting to take place without the computer being used, e.g., a one-to-one supervision meeting.

This activity data was also collected in a privacy preserving way: only non-sensitive power-related activity was measured and no information about usernames or other identifiable information (such as application use or internet activity) was collected.

Ground Truth Annotation:

a) High Traffic Area data: We captured environmental sensing data in the HTA over 10 days and used it to validate the performance of these sensors for determining occupancy. The ground truth occupancy measurements were obtained using footage from a still camera in a corner of the room where all occupants could be captured. The camera was programmed to take one picture every five minutes. The researchers then annotated the images by counting the number of people in each photo. The data was further represented in a hierarchical fashion: *i)* at the top most level there is a binary representation of occupancy, i.e., $O^1 \in \{0,1\}$; *ii)* the second level categorises occupancy according to 'no-occupancy', 'low', 'medium', and 'high' occupancies, i.e., $O^2 \in \{0, low, med, high\}$ (determined using the maximum number of observed occupants); and *iii)* at the bottom level, occupancy values represent the actual number of people in the observed space, i.e., $O^3 \in \{0,1,2,...,n\}$.

b) Low Traffic Area occupancy: Ground truth occupancy measurements for LTA were obtained by observing for occupants every 15 minutes during the working hours (8am - 6pm) over the two weeks recording period. Due to confidentiality concerns raised by the building's management team, this process was done manually rather than using a camera. A member of the research team noted down the number of occupants after each visit to the room. Similar to the office space study, ground truth annotation was represented in a hierarchical, three-level manner (O^1, O^2, and O^3).

We note that the ground truth collection exercise can be simplified by temporarily installing a camera to monitor the workspace and the images produced can then be crowdsourced for headcounts after automatically blurring the faces, similar to [21].

4.1.2 Classification Mechanism

We ran two separate experiments using two standard classification procedures to infer occupancy at the three levels of hierarchical representation, which also serves as our baseline. Specifically we validated the effectiveness of k-Nearest Neighbour (k-NN; we used $k = 3$) and Support Vector Machine classifiers (SVM; we used an RBF kernel and selected the optimized slack and kernel parameter using a standard grid search procedure [32]). All of the experiments were conducted using a standard 10-fold cross-validation procedure.

We then ran two further experiments using the hierarchical approach that takes posterior probabilities for each level progressively into account by appending them to the feature vectors for classification. For k-NN (with $k = 3$), Equations 11 and 12 can be represented as:

$$\hat{Y^h} = \underset{y=1,..,C^h}{\operatorname{argmin}} \sum_{c=1}^{C^h} \hat{P}(c|f)C(y|c) \tag{13}$$

where $\hat{Y^h}$ is the predicted classification at level $h \in \{1,2,...,H\}$ and C^h represents the number of classes. $\hat{P}(c|f)$ represents the posterior probability of class c for observation $f \in F^v$ and $C(y|c)$ is the cost of classifying the given observation as y when its true class is c (Figure 2 shows a general representation of the classification mechanism).

For SVM experiments, posterior probabilities are estimated using a pair-wise classification method (detailed in [37]) and combined into a single distribution over all of the occupancy labels separately per level. Optimization of the slack and the kernel parameter is performed per level of the hierarchical representation using the grid search procedure of [32].

Comparisons are then made to the baseline results in different settings (highlighting various conditions) that are detailed in the next subsection.

4.2 Experimental Settings and Results

We conducted our experiments in three settings to highlight two main aspects of our framework: *i)* the impact of using the hierarchical approach and *ii)* the influence of using the contextual information. Our overall aim is to provide accurate occupancy estimates down to level 3 – i.e., head counts – with associated levels of confidence.

In the first two settings, we performed hierarchical occupancy estimation in the HTA and LTA respectively using the wireless environmental sensors that measured various modalities to highlight the impact of the hierarchical framework using solely the sensors. In Setting 3, we incorporated the contextual information in the LTA as described before. In addition to highlighting the impact of the hierarchical framework with the added contextual data, we also evaluated: *i)* how uncertain this information is; and *ii)* how much this additional information actually benefits the estimations.

4.2.1 Setting 1: Occupancy estimation in a High Traffic Area using only environmental sensors

Confidence values from the top-level were appended to the feature vector (Equation 10) for classification at the penultimate level. Mean accuracies per level with standard deviations are shown in Table 2-S1. Using our hierarchical classification method, we saw a significant improvement over the baseline classification strategy (independent samples t-test). This resulted in improvements only seen at level-2 onwards (and not level-1). With increased level of classification accuracy, we could more accurately estimate the actual count of people occupying the monitored area.

4.2.2 Setting 2: Occupancy estimation in a Low Traffic Area using only environmental sensors

Comparative results for mean occupancy estimation accuracies in a LTA setting using both of the non-hierarchical and hierarchical models for both of the classification mechanisms are shown in Table 2-S2. Again by using our hierarchical framework we observed a significant improvement over the baseline approach at both levels 2 and 3.

4.2.3 Setting 3: Occupancy estimation in a Low Traffic Area using environmental sensors and contextual information

We used the original feature vector of Equation 10 and appended the contextual features to it as described earlier (System Overview). For all further levels of our hierarchical model we similarly use confidences from the lower levels and augment the feature vectors accordingly. For the bottom level classification we then arrive at feature vectors $\vec{f}'' \in \mathbb{R}^{D''}$ with $D'' = 135$ (127 (original features) + 2 (contextual features) + 2 (level-1 confidences representing binary decisions) + 4 (level-2 confidences representing 0, low,medium, high occupancy categorisations). We used the same hierarchical ground truth annotation for evaluation.

Mean classification accuracies with associated standard deviations for hierarchical occupancy estimates using PC activity information as an additional feature are shown in Table 2-S3(a). We observed a significant improvement when using the hierarchical approach that implicitly dealt with associated uncertainty in noisy – sensor and context – data. Similarly, mean classification accuracies with associated standard deviations for hierarchical occupancy estimates using meeting scheduling information are shown in Table 2-S3(b). Similar improvement figures can be observed in favour of the hierarchical approach compared with the baseline. When

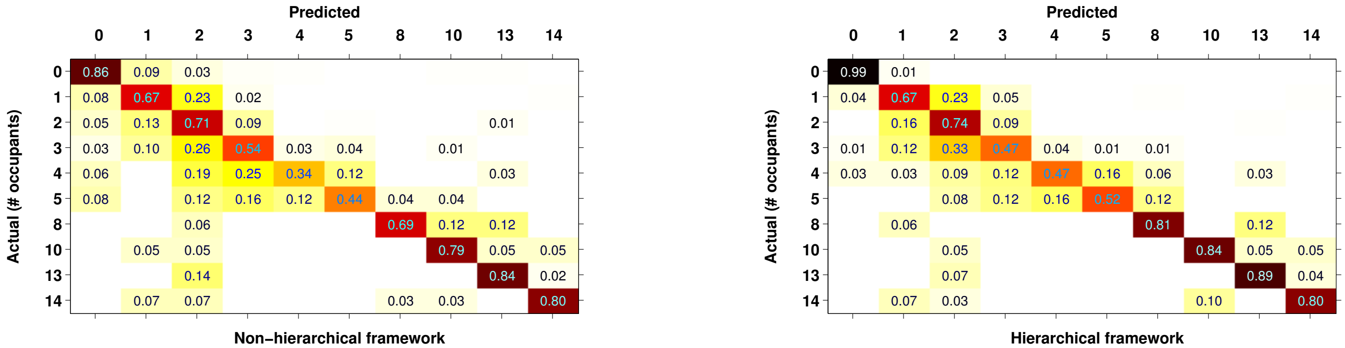

Figure 3: Confusion matrices for the baseline approach (left) and the hierarchical framework (right) for level-3 occupancy estimation.

		S1 - HTA No context		S2 - LTA No context		S3a - LTA PC activity		S3b - LTA Scheduling		S3c - LTA PC + Scheduling	
		Normal	Hierarchical	Normal	Hierarchical	Normal	Hierarchical	Normal	Hierarchical	Normal	Hierarchical
k-NN	L1	94.6 (1.1)	94.6 (1.1)	92.4 (2.1)	92.4 (2.1)	94.1 (1.8)	94.1 (1.8)	94.0 (1.9)	94.0 (1.9)	94.7 (1.9)	94.7 (1.9)
	L2	81.6 (2.0)	84.7 (1.9)	87.3 (3.0)	91.6 (2.4)	89.6 (2.5)	93.3 (2.3)	89.1 (2.7)	93.2 (2.1)	90.4 (2.4)	94.0 (2.1)
	L3	68.9 (2.4)	74.5 (2.1)	69.2 (4.4)	74.9 (4.4)	72.4 (4.3)	78.0 (4.2)	70.7 (4.0)	76.0 (3.9)	72.5 (4.2)	77.7 (4.2)
SVM	L1	93.3 (1.1)	93.3 (1.1)	91.7 (2.3)	91.7 (2.3)	94.9 (2.1)	94.9 (2.1)	94.0 (1.9)	94.0 (1.9)	95.5 (1.7)	95.5 (1.7)
	L2	78.9 (2.0)	82.0 (1.9)	84.0 (3.1)	89.1 (2.7)	90.5 (2.3)	91.6 (2.2)	89.0 (2.5)	90.1 (2.5)	91.3 (2.4)	92.3 (2.2)
	L3	66.3 (2.3)	71.3 (2.3)	64.6 (3.9)	71.3 (4.3)	70.4 (3.9)	75.2 (3.5)	71.4 (4.1)	74.0 (4.2)	73.8 (4.1)	76.8 (3.8)

Table 2: Mean hierarchical classification accuracies for all settings. Setting 1 (S1) represents a comparison made between the two methodologies in a High Traffic Area (HTA) using environmental sensors. Setting 2 (S2) shows the comparative results using only environmental sensors without context in a Low Traffic Area (LTA). Settings 3 (S3) represents comparative results using contextual data i.e., PC activity data (a), meeting schedules (b) and both (c). All of the results shown were significant ($p<.005$) per setting intra level.

all contextual information sources were incorporated into the feature vector we achieved further significant improvements over the standard approach, as well as over the configurations using a single contextual source (Table 2-S3(c)). Confusion matrices at the bottom level of occupancy representation (head-counts) – highlighting major improvement using the hierarchical approach as compared with the baseline – are shown in Figure 3.

In addition to highlighting performance improvement using our hierarchical approach, we also explored the influence of using contextual information in addition to the environmental sensors. The combination of all data sources again resulted in significant improvements of the occupancy estimation accuracy – for both the baseline and our hierarchical approach (Table 2-S2,S3(c)). The table shows consistent improvement of both the models in favour of using contextual information (in this case PC activity and meeting schedules). These results are interesting as with the addition of contextual information, occupancy estimation at every level of the hierarchical representation improves even though our contextual information does not explicitly represent head-counts in the case of level-3 occupancy estimates.

4.2.4 *Meeting schedule uncertainty*

Meeting schedules are potentially the biggest contextual indicators for general room occupancy. However, scheduled meetings may not take place (i.e., schedules are not altered for last minute cancellations) and similarly, people may be present in the room at a time when no meeting is scheduled. During the 14 days of data recording in our LTA we found the correctness of scheduled meetings to be 87.08% with respect to recorded occupancy. According to our experiments still this imperfect accuracy of the meeting schedule improves the overall accuracy of the occupancy estimation (Table 2-S3(b,c))

However, it remains to be explored up to which degree of erroneousness an uncertain meeting schedule can positively contribute to accurate occupancy estimation for a meeting room. In order to evaluate the influence of the uncertainty that is inherent to meeting schedules we systematically explored this effect by artificially and randomly falsifying meeting schedules e.g., randomly adding meetings in schedule when no actual meetings took place. This effectively translates into introducing noise into the meeting schedule. Figure 4 shows the accuracy of occupancy estimation at level 1 (binary occupancy) in dependency on the amount of noise added to the meeting schedule, starting with 0% up to 50% noise – i.e. by falsifying up to 5 out of the 10 days worth of meeting schedule data. It can be seen that the use of meeting schedules yields a benefit as long as this contextual data is correct for at least half of the contained meetings.

4.2.5 *Historical evaluation*

With our estimation system it is possible to analyse ongoing room occupancies as demonstrated above. Whilst this evaluation allows for a systematic analysis of the effectiveness of the proposed analysis framework, it misses out an important practical use case. For sustainable optimisation of the way a particular building is used, facility management would rather look into the analysis of long term monitoring data instead of focusing on short temporal contexts. A straightforward extension of our deployment study is, however, challenging as it is not acceptable – not even for system development and validation purposes – to obtain detailed ground truth using the intrusive procedure as described before over longer periods of time, which leaves us with the challenge of unreliable annotation for long-term monitoring data.

Given the results of our systematic evaluation we are able to effectively overcome this dilemma. Aiming for an estimation of

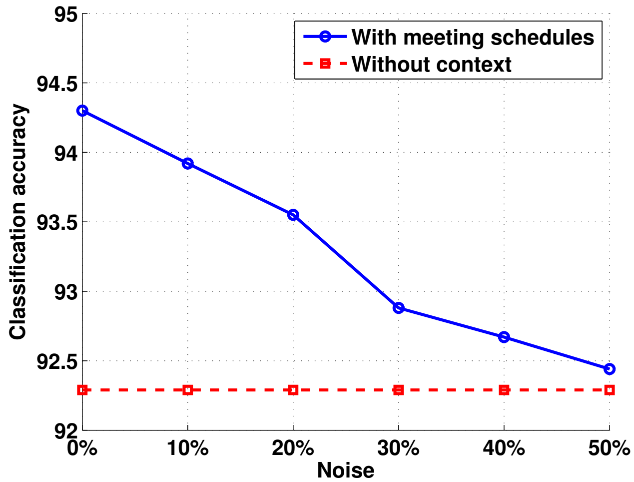

Figure 4: Binary occupancy estimation accuracy with regards to the amount of scheduling noise. It can be seen that, meeting schedules are beneficial to occupancy estimation as long as the scheduling information is correct at least half the time.

the degree of misuse of the meeting room we collected sensor and contextual data for a period of one month, i.e., we deployed the system "as is" and left it in the room untouched over the whole period of deployment. The resulting historical data is then analysed as described in the previous section and we use the room's meeting schedule as ground truth annotation. For the sake of clarity of presentation we only focus on high-level occupancy estimates here. We then use the results of the systematic validation study to effectively recalibrate the results, i.e., we explicitly weigh the prediction results using the results of the 10 days study:

i) w_1 = correlation between scheduled meetings and occupancy – $\hat{M} = 87.08\%$ (as established previously);

ii) w_2 = occupancy estimation accuracy – $\hat{O} = 92.30\%$ (Table 2-S1).

We also use the classifier accuracy for the one month deployment with regards to the room's meeting schedule $w_3 = 96\%$. Taking the correction factors into account, the effective correction factor results in:

$$\hat{C}_m = w_1 \times w_2 \times w_3 = 0.87 \times 0.92 \times 0.96 = 0.77 \qquad (14)$$

(assuming equal influence of errors w_1, w_2 and w_3).

Using the trained classifier, we predict binary occupancy for the one month deployment and find correlation with regards to the scheduling information. We found that approx. 30.22% of the time, a meeting was scheduled but people were not present and 10.56% of the time people were present when there was no meeting scheduled (both corrected using Equation 14).

4.2.6 *On Accurate Head-Counting*

In principle, the methodology discussed in this paper is trained using a certain number of people (in the case of the meeting room it was 14). However, in the context of a decision support system, it would be useful to consider an open-ended occupancy classification problem where the maximum number of occupants is not strictly pre-defined. Effectively this corresponds to an accurate head-counting problem with no limitations, which is very challenging when using non-intrusive sensing capabilities as it is dictated by the targeted application scenario. We approach this problem through the exploration of hierarchical posterior probabilities (as indicated in the previous section) again, this time however, looking at the inter-level inconsistencies.

Figure 5: True occupancy (top) and per-level decision confidences (top-centre, bottom-centre, bottom) showing inter-level inconsistency when number of people is 13 (highlighted).

In an exemplary evaluation scenario, we train our system for a target maximum occupancy of 10 people. This upper limit could, for example, come from the room specification. However, during evaluation it is presented with scenarios where up to 14 persons occupied the room, which corresponds to overcrowding and thus misuse of the room. Accordingly, Figure 5 shows the inter-level inconsistencies where the topmost bar shows the number of occupants in the meeting room, and per level confidence measures are shown in the three bars at the bottom where colours represent the confidence levels: red for low confidence; yellow for medium confidence; and green for high confidence. For 13 occupants in the room, level-1 provides high confidence in an occupant's presence and level-2 provides medium confidence about high occupancy presence. However, level-3 confidence measure is red indicating uncertainty about the number of people.

Having access to such a representation, a building manager can now reliably make sensible decisions related to resource allocation based on per-level confidences and inter-level anomalies.

5 Conclusions

We have presented a new approach for accurate occupancy monitoring based on a wireless sensing system, which measures the non-sensitive values of temperature, light, humidity, PIR and audio levels that is convenient to deploy and cost-effective. We developed a novel hierarchical analysis framework that combines environmental sensing with potentially uncertain contextual information (e.g., meeting schedules, computer activity) and demonstrated in real-world deployments that the proposed approach can monitor occupancy accurately up to head-count level, thereby performing significantly better than a baseline, non-hierarchical approach.

When monitoring room use over a longer period of time, we found that the studied meeting room was not always utilised appropriately – with people not being present at the time of a scheduled meeting nearly a third of the time resulting in potential wasted energy in other rooms. The system was designed to support human decision makers in optimising the use of the building's facilities with a view on sustainable use of resources whilst at the same time maintaining occupants' comfort. Our framework produces detailed yet easy to comprehend analysis results including confidences, which supports effective decision making.

The hierarchical occupancy estimation algorithm itself can also be incorporated into existing infrastructures to improve their accuracy for example, by replacing the Google calendar feed in BOSS [5] or the one used in Redwood's Room Tracker [30]. This would yield better results due to the improved occupancy estimation quantified in this paper. We can also use the proposed framework with other existing hardware such as the Nest thermostat [25] that measures similar modalities.

Alternative methods such as the hierarchical rule induction [14] (originally proposed for automated sports video annotation [6]) can also be deployed in this context that is able to specify the number of levels of hierarchical occupancy in a temporal fashion. Anomaly detection methodologies such as [2, 13] can also be explored to highlight anomalous number of occupants for a given location.

Acknowledgment

Parts of this work have been funded by a grant from the Engineering and Physical Sciences Research Council, UK (EPSRC, grant number EP/I000755/1 – TEDDI: Transforming Energy Demand through Digital innovation).

6 References

[1] Y. Agarwal, B. Balaji, R. Gupta, J. Lyles, M. Wei, and T. Weng. Occupancy-driven energy management for smart building automation. In *Proc. BuildSys*. ACM, 2010.

[2] I. Almajai, F. Yan, T. de Campos, A. Khan, W. Christmas, D. Windridge, and J. Kittler. Anomaly detection and knowledge transfer in automatic sports video annotation. *Detection and Identification of Rare Audiovisual Cues*, 2012.

[3] A. Almudevar, A. Leibovici, and A. Tentler. Home monitoring using wearable radio frequency transmitters. *Artificial Intelligence in Medicine*, 42(2):109–120, 2008.

[4] S. Bhattacharya, P. Nurmi, N. Hammerla, and T. Plötz. Using unlabeled data in a sparse-coding framework for human activity recognition. *Pervasive and Mobile Computing*, May 2014.

[5] S. Dawson-Haggerty, A. Krioukov, J. Taneja, S. Karandikar, G. Fierro, N. Kitaev, and D. Culler. Boss: Building operating system services. In *Proc. USENIX Symp. on Networked Systems Design and Implementation (NSDI)*, 2013.

[6] T. E. deCampos, A. Khan, F. Yan, N. FarajiDavar, D. Windridge, J. Kittler, and W. Christmas. A framework for automatic sports video annotation with anomaly detection and transfer learning. In *Machine Learning and Cognitive Science, collocated with EUCOGIII*, 2013.

[7] R. H. Dodier, G. P. Henze, D. K. Tiller, and X. Guo. Building occupancy detection through sensor belief networks. *Energy and buildings*, 38(9):1033–1043, 2006.

[8] B. Dong and B. Andrews. Sensor-based occupancy behavioral pattern recognition for energy and comfort management in intelligent buildings. In *Proc. IBPSA*, 2009.

[9] Energy Information Administration. International energy outlook 2013, July 2013.

[10] V. L. Erickson, Y. Lin, A. Kamthe, R. Brahme, A. E. Cerpa, M. D. Sohn, and S. Narayanan. Energy efficient building environment control strategies using real-time occupancy measurements. In *Proc. ACM BuildSys*, 2009.

[11] N. Y. Hammerla, R. Kirkham, P. Andras, and T. Plötz. On preserving statistical characteristics of accelerometry data using their empirical cumulative distribution. In *Proc. ISWC*, 2013.

[12] R. Howarth and S. Bringezu. Biofuels & environmental imp- acts: scientific analysis and implications for sustainability. **http://unesdoc.unesco.org/images/0018/001831/183113e.pdf**, 2009.

[13] A. Khan, D. Windridge, T. de Campos, J. Kittler, and W. Christmas. Lattice-based anomaly rectification for sport video annotation. In *proc. of IEEE ICPR*, 2010.

[14] A. Khan, D. Windridge, and J. Kittler. Multilevel chinese takeaway process and label-based processes for rule induction in the context of automated sports video annotation. *IEEE Trans. Cybernetics*, 2014.

[15] R. Kirkham, A. Khan, S. Bhattacharya, N. Hammerla, S. Mellor, D. Roggen, and T. Plötz. Automatic Correction of Annotation Boundaries in Activity Datasets by Class Separation Maximization. In *Adjunct Proc. UbiComp – HASCA workshop*, 2013.

[16] R. Kirkham, S. Mellor, D. Green, J.-S. Lin, K. Ladha, C. Ladha, D. Jackson, P. Olivier, P. Wright, and T. Plötz. The break-time barometer: an exploratory system for workplace break-time social awareness. In *Proc. ACM UbiComp*, 2013.

[17] P. Klasnja, S. Consolvo, T. Choudhury, R. Beckwith, and J. Hightower. Exploring privacy concerns about personal sensing. In *Proc. Pervasive*, 2009.

[18] C. Koehler, B. D. Ziebart, J. Mankoff, and A. K. Dey. Therml: Occupancy prediction for thermostat control. In *Proc. ACM UbiComp*, 2013.

[19] C. Ladha, N. Hammerla, E. Hughs, P. Olivier, and T. Plötz. Dog's Life: Wearable Activity Recognition for Dogs. In *Proc. UbiComp*, 2013.

[20] C. Ladha, N. Hammerla, P. Olivier, and T. Plötz. ClimbAX: Skill Assessment for Climbing Enthusiasts. In *Proc. UbiComp*, 2013.

[21] W. S. Lasecki, Y. C. Song, H. Kautz, and J. P. Bigham. Real-time crowd labeling for deployable activity recognition. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 1203–1212. ACM, 2013.

[22] C. Liao and P. Barooah. An integrated approach to occupancy modeling and estimation in commercial buildings. In *Proc. ACC*, 2010.

[23] S. Meyn, A. Surana, Y. Lin, S. Oggianu, S. Narayanan, and T. A. Frewen. A sensor-utility-network method for estimation of occupancy in buildings. In *Proc. Int. Conf. Decision and Control*, 2009.

[24] C. Nelder. The end of fossil fuel. **forbes.com/2009/07/24/peak-oil-production-business-energy-nelder.html**. Accessed: 26/09/2013.

[25] Nest: Thermostat. **https://nest.com/thermostat/life-with-nest-thermostat/**. Accessed: 14/02/2014.

[26] L. Pérez-Lombard, J. Ortiz, and C. Pout. A review on buildings energy consumption information. *Energy and buildings*, 40(3):394–398, 2008.

[27] T. Plötz, N. Hammerla, and P. Olivier. Feature Learning for Activity Recognition in Ubiquitous Computing. In *Proc. IJCAI*, 2011.

[28] T. Plötz, N. Hammerla, A. Rozga, A. Reavis, N. Call, and G. D. Abowd. Automatic Assessment of Problem Behavior in Individuals with Developmental Disabilities. In *Proc. UbiComp*, 2012.

[29] H. Ramoser, C. Beleznai, T. Schlögl, M. Winter, and H. Bischof. Shape-based detection of humans for video surveillance applications. In *Proc. ICIP*, 2003.

[30] Room tracker. **www.redwoodsystems.com/produces/applications/room-tracker**. Accessed: 05/02/2014.

[31] B. Schauerte, J. Richarz, T. Plötz, C. Thurau, and G. A. Fink. Multi-Modal and Multi-Camera Attention in Smart Environments. In *Proc. ICMI*, 2009.

[32] B. Schölkopf and A. J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA, USA, 2001.

[33] J. R. Smith, K. P. Fishkin, B. Jiang, A. Mamishev, M. Philipose, A. D. Rea, S. Roy, and K. Sundara-Rajan. Rfid-based techniques for human-activity detection. *Commun. ACM*, 48(9):39–44, Sept. 2005.

[34] T. Sookoor and K. Whitehouse. Roomzoner: Occupancy-based room-level zoning of a centralized hvac system. In *Pro. ICCPS*, 2013.

[35] L. Thomas, L. Little, P. Briggs, L. McInnes, E. Jones, and J. Nicholson. Location tracking: views from the older adult population. *Age and ageing*, 42(6):758–763, 2013.

[36] R. Tomastik, S. Narayanan, A. Banaszuk, and S. Meyn. Model-based real-time estimation building occupancy during emergency egress. In *Proc. Int. Conf. Pedestrian and Evacuation Dynamics*, 2010.

[37] T.-F. Wu, C.-J. Lin, and R. C. Weng. Probability estimates for multi-class classification by pairwise coupling. *Journal of Machine Learning Research*, 5(975-1005):4, 2004.

[38] C. Xu, S. Li, G. Liu, Y. Zhang, E. Miluzzo, Y.-F. Chen, J. Li, and B. Firner. Crowd++: Unsupervised speaker count with smartphones. In *Proc. ACM UbiComp*, 2013.