 doi:10.20944/preprints202008.0202.v1

Article

Occupancy prediction using low-cost and low-resolution heat sensors for smart offices

Beril Sirmacek and Maria Riveiro

Jönköping AI Lab (JAIL), Department of Computer Science and Informatics, School of Engineering, Jönköping University, Sweden
Email: beril.sirmacek@ju.se, maria.riveiro@ju.se

Abstract: In order to design efficient and sustainable office spaces and to automate lighting, heating and air circulation in these facilities, solving the challenge of occupancy prediction is crucial. In office spaces where large areas need to be observed, multiple sensors must be used for full coverage. In these cases, it is normally important to keep the costs low, but also to make sure that the privacy of the people who use such environments are preserved. Low-cost and low-resolution heat (thermal) sensors can be very useful to build solutions that address these concerns. However, they are extremely sensitive to noise artifacts which might be caused by heat prints of the people who left the space or by other objects which are either using electricity or exposed to sunlight. There are some earlier solutions for occupancy prediction that employ low-resolution heat sensors, however, they have not addressed nor compensate for such heat artifacts. Therefore, in this paper, we present a low-cost and low-energy consuming smart space implementation to predict the number of people in the environment based on whether their activity is static or dynamic in time. We use a low-resolution (8×8) and non-intrusive heat sensor to collect data from an actual meeting room. We propose two novel workflows to predict the occupancy; one based on computer vision and one based on machine learning. Besides comparing the advantages and disadvantages of these different workflows, we use several state-of-the-art explainability methods in order to provide a detailed analysis of the algorithm parameters and how the image properties influence the resulting performance. Furthermore, we analyze noise resources which affect the heat sensor data. We hope that our analysis brings light into understanding how to handle very low-resolution heat images in these environments. The presented workflows could be used in various domains and applications other than smart offices, where occupancy prediction is essential, e.g., for elderly care.

Keywords: heat sensors; smart offices; occupancy prediction; machine learning; computer vision; feature engineering; explainability

1. Introduction

Designing smart environments have gained significant attention over the last years due to advancements in sensor technology, decreased hardware cost (even edge devices) and ease of deployment (using sensor fusion, cloud and Internet of Things –IoT– technologies). Using sensor technologies combined with small and efficient hardware and intelligent algorithms, we are now able to provide smart solutions that support humans at home and work, e.g., in business offices, school buildings, hospitals, elderly facilities and wellness centers. Smart environments differ from traditional environments because of real-time interactions between the users and the facilities; one important research topic within this area is occupancy prediction and human activity detection/recognition, due to their broad application in automation, robotics and human computer interaction (HCI).

Detecting human activities accurately and in real-time is a challenging problem for several reasons [1]. There are several methods developed for the purpose of detecting human activity indoors, e.g., using multimedia-sources (such as audio/video), wearable devices (smartwatches, wristbands, etc.), and ambient sensing. Each method comes with its advantages and disadvantages based on its use cases. For instance, despite having high accuracy, audio/video -based solutions bring challenges in real implementations because of the privacy regulations. Similarly, wearable device-based methods can provide customized solutions but also might lead to discomfort issues, negatively affecting their feasibility. Ambient sensing (such as infrared sensors), on the other hand, relies on sensors that provide only ambient information about the environment. It does not have any privacy or discomfort issues, but requires careful thinking in terms of sensor placement and development of intelligent solutions that can extract information from sensor data which does not provide sharp visual features to understand the environment easily. Ambient sensing systems generally used for detecting human activities install a large number of sensors (same or of different modality) in the environment. Each sensor brings sensing noise and such noise needs to be handled before the output is fed to a sensor data fusion module, a decision-making system or a machine learning model.

Smart spaces are usually shared by multiple people. In order to automate the facilities, i.e., heating, ventilation, air conditioning (*HVAC*), it is important to know how many people is there, at which location and what do they do [2,3]. When smart offices are considered, this information is also important to address security issues and an efficient meeting room occupancy management. Real-time human activity detection and prediction can be used to effectively allocate resources to increase user comfort in conference rooms in offices [4] as well as study-rooms in households [5]. Success of such automation units can also be greatly improved with robotics systems [6].

On the one hand, existing solutions for occupancy prediction often require higher resolution sensors that are expensive and raise privacy issues of the people who use the environment (see background section for more details). On the other hand, solutions that make use of lower resolution sensors have not been widely tested and their performances are mostly illustrated within controlled environments where there is not sun exposure difference during the day, heat sources from other objects or heat prints left from people who used the chairs and desks for a long time. Therefore, a deeper analysis of the effects of this noise is necessary in order to deploy accurate solutions in real environments. We believe that low-cost and low-resolution heat sensors would be helpful to create easily-accessible solutions for many smart offices while avoiding privacy concerns of the people who use them (it is impossible to identify the users in the environment in these low resolution images). However, one of their major challenges is the high sensitivity to other heat sources; this must be considered in order to provide robust algorithms which can adapt to the changing conditions of the environment. Thus, in this paper, we explore how to use low-cost and low-resolution heat sensors looking for solutions that are robust to noise artifacts. We present two approaches, one computer vision-based and the other one, machine learning-based. We evaluate and compare their performance with data recorded in a real office environment, and we discuss their advantages and disadvantages, as well as their particularities regarding different heat artifacts.

In particular, this paper provides the following contributions:

1. We propose two different workflows (one based on computer vision and one based on machine learning) for predicting the number of people in the environment.
2. We evaluate our novel workflows and provide an exhaustive analysis of their performances on a real office environment where several meetings happen while the temperature and the sun illumination from the windows change during the day.
3. We explain the relation of the algorithm parameters to the installation position of the sensor in the office environment.
4. We use explainable AI methods in order to investigate which algorithm parameters and image properties provide useful information to the classifications.
5. Finally, we focus on understanding the heat artifacts in recordings which are acquired from an

 doi:10.20944/preprints202008.0202.v1

environment which is heavily exposed to the sunlight and used by many people in long meetings. We discuss the different effects of the environmental heat artifacts (like sunlight) and the heat prints of the people who sat on a specific location for a long time. We discuss the capabilities of our workflows to deal with such noise artifacts and we offer compensation methods to increase accuracy of the results.

2. Background

The human activity detection problem is solved in different ways in the literature. Activity detection has usually been performed using audio/video surveillance because multimedia data provides very accurate representations of human activities [7]. Another method to obtain activity detection is using wearable devices [8]. This method has gained a lot of attention due to that wearable devices have become widely-available and popular. Nevertheless, limitations of wearable device -based applications are obvious. First of all, people might forget or does not cooperate with the wearing sensors. Secondly, new people visiting the environment cannot be tracked when they do not have sensors.

Due to the high prices and the privacy issues raised by surveillance cameras, a lot of companies and researchers turned their focus on using low-cost and low-resolution sensors for keeping the prices low and keeping people's privacy in the scene. Passive infrared (PIR) and heat sensors became interesting for such applications and researchers have focused on developing algorithms which can extract information even when the data cannot obviously tell all important visual properties of the scene.

Wahl et al. [9] used PIR sensors to recognize activities, and to keep track of how many people where in a room. Murao et al. [10] used PIR sensors which are distributed in a house, in order to track the activation patterns of the sensors. They used these patterns to recognize activities, and to keep track of how many people where there. Even though PIR sensors are very low-cost, available to many researchers and they keep the privacy of the people involved, when they are used in a constrained area like a meeting room, the sensors need to be placed on all areas or objects of interest, as presented in [11]. Such an application requires many sensors even in one room, at all locations where they could interfere with the activity being performed. Multiple devices also means that installation, maintenance, sensor fusion etc. require more effort, even if the sensors are "tape and forget", as argued in [11]. Besides, even though one PIR sensor is much cheaper than one heat sensor, when so many of them are necessary in one room, the whole application might cost even more than a heat sensor alternative.

Heat sensors (infra-red sensors/cameras) provide heat images that can be conveniently used to tracking people, as they only can visualize heat reflections. Heat sensors can be used to detect and track people [12,13]. While higher resolution sensors provide more chances to algorithms to provide robust solutions, the prices of the heat sensors increase exponentially with respect to their increasing resolutions. Therefore, in order to keep the application low-cost, it is important to explore the possibilities of developing algorithms that can make some predictions relying on information extracted from images with very low resolution.

Some researchers looked for opportunities of using extremely low-resolution heat sensors for understanding human activity indoors. Jeong et al. [14] showed that an adaptive thresholding method could help to detect people using low-resolution heat sensors. Singh et al. [15] used adaptive thresholding and the binary results for understanding whether people are on the floor, in a standing or sitting position. Next, they extended their work towards a multiple sensor application [16]. Troost [17] compared different classifiers which are trained to predict the number of people in low-resolution heat sensor images. Gonzalez et al. [18] used a single sensor and trained a classifier to differentiate 21 different activities in the environment. The experiments gave successful results to show that it is possible to rely on such low-resolution image information to understand the usage of the environment. Johansson and Sandberg [19] looked for opportunities of training neural networks, specifically CNNs (Convolutional Neural Network) and LSTMs (Long Short-Term Memory), in order to have a mathematical model which can make predictions about the number of people visible in a

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 August 2020 doi:10.20944/preprints202008.0202.v1

low-resolution heat sensor. As we see in the literature related to the usage of very low-resolution heat sensors, the experiments are restricted to data sets which are acquired in controlled environments (such as places that are not exposed to changing sunshine conditions from large windows – which are common in office spaces). Robustness in changing environmental conditions and generalization of the different algorithms tested have not been discussed either and, moreover, the research on thermal noise and compensation methods is still quite limited. Thus, in this study, we focus on these aspects hoping to further understand how to solve them.

3. Aim and objectives

Given the challenges of the studies described in the previous section, the aim of this work is to explore low-cost and privacy preserving solutions for occupancy prediction that are robust and accurate and take into account the noise artifacts of the environment. After exploring possible solutions that fulfilled these criteria, we suggested two alternative possible workflows based on two different methods. One method heavily relies on the use of computer vision algorithms while the second method relies on feature engineering (finding good/representative features in data) and solving the occupancy prediction problem through machine learning. Considering these two different workflows that we present in this work, we herein tackle the following three research questions:

- *Can we rely on computer vision or machine learning based methods in order to provide robust occupancy prediction solutions when we use a low-cost and low-resolution sensor?*
- *What is the best occupancy prediction accuracy that we can achieve when computer vision based or machine learning based methods are used?*
- *What are the sources of the noise artifacts and what are the possible noise compensation methods to overcome errors caused by them?*

4. Scenario and data

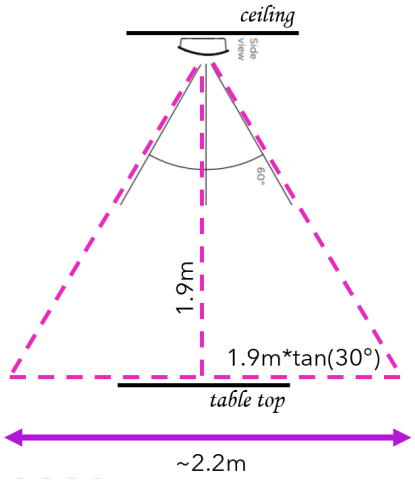
The low-cost low-resolution heat sensor used in this study[1] [20] as well as the environment used to collect the data are presented in Figs. 1 and 2. The sensor description is provided by the manufacturer at [21]. Table 1 shows the heat recording data set which we have used for testing our algorithms. The data set was collected in three different dates using the same sensor which was located at the office ceiling, as shown in Fig. 2. The figure also depicts the labels for the chair ID's which we have indicated whether there is a person or not in the data set description part of the Table 1. In the table, the blue highlighted rows correspond to the sensor recordings which are captured when the room is empty. Later, we will use these recordings for noise analysis and noise compensation. For the performance analysis of the workflow 1 and workflow 2, we have used the Day 1 and Day 2 data set. We have used the 24 hours recording of the Day 3 data for the assessment of the explainability methods since the recording includes significant environment temperature changes.

Figure 1. The low-cost and low-resolution heat sensor used in this study [20].

[1] https://www.rolergo.com/sensors

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 August 2020 doi:10.20944/preprints202008.0202.v1

(a)



(b)

Figure 2. (a) The meeting room where the test data streams were collected. The sensor was placed at the ceiling, 1.9 m above the table. The seat numbers correspond to the person presence numbers in Table 1. (b) Field of view calculation for the meeting room setup. Table top line corresponds to the black line at the table edge with 90 cm measurement in (a).

5. Methods

In this section, we introduce two different workflows in order to determine the number of people in each frame (image) of the heat sensor recordings. The first method uses traditional computer vision algorithms (aka computer vision based occupancy prediction), while the second one is a machine learning-based solution. The machine learning based approach creates a feature vector for each frame, and after that, a supervised classification algorithm learns how to fit an optimal model in order to be able to classify feature vectors for identifying the number of people in the scene (we refer to this solution as feature classification based occupancy prediction). In the reminder of this section, we focus on the theory of the modules which construct our workflows, and we provide detailed analysis with the different data sets in the *Experimental Results* section.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 August 2020 doi:10.20944/preprints202008.0202.v1

Data label	Description	Total people	Total frames
$h_1(x,y)$	Day1 before meeting	0	94
$h_2(x,y)$	Day1 after meeting	0	45
$h_3(x,y)$	Day1 present at 1	1	24
$h_4(x,y)$	Day1 present at 1 2	2	30
$h_5(x,y)$	Day1 present at 1 2 3	3	27
$h_6(x,y)$	Day1 present at 1 2 3 4	4	51
$h_7(x,y)$	Day1 present at 2 3 4	3	29
$h_8(x,y)$	Day1 present at 3 4	2	36
$h_9(x,y)$	Day1 present at 4	1	35
$h_{10}(x,y)$	Day2 before meeting	0	94
$h_{11}(x,y)$	Day2 present at 1 2	2	56
$h_{12}(x,y)$	Day2 present at 2 3	2	47
$h_{13}(x,y)$	Day2 present at 3	1	28
$h_{14}(x,y)$	Day2 present at 3 4	2	31
$h_{15}(x,y)$	Day3 24 hours	max 2	661

Table 1. Presence sensor recordings used for testing our algorithms.

5.1. Computer vision based occupancy prediction

Fig. 3 presents the workflow of the computer vision based occupancy prediction approach. This solution requires a baseline, i.e., a recording when the office is empty. Each module of the workflow is described in detail hereafter.

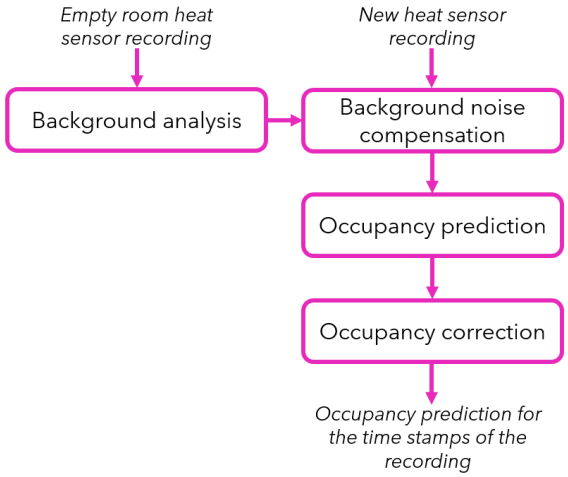


Figure 3. Workflow of the computer vision based occupancy prediction method.

5.1.1. Background analysis

We assume that we can improve the quality of each frame by removing the noise effects coming from other heated objects in the environment, such as computer like devices or other objects exposed to sunlight coming from windows. We assume that we can use a short recording of the empty room before any meeting for compensation of such heat artifacts. To do so, we calculate the mean and the standard deviation of each pixel in an empty room recording with our background analysis algorithm 1. Thus, the background analysis module provides the mean ($m(x,y)$) and standard deviation ($s(x,y)$) of each pixel value through the recording of the empty room. Later on, when a new heat sensor recording is captured, these two matrices are used for background noise compensation.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 August 2020 doi:10.20944/preprints202008.0202.v1

The blue highlighted rows in Table 1 indicate the recordings which are considered for doing the background analysis. Fig. 4 shows the empty room mean and standard deviation based on those three recordings ($h_1(x,y)$, $h_2(x,y)$ and $h_{10}(x,y)$).

Algorithm 1: Background analysis algorithm

$h_i(x,y)$**:** Input frames
$m(x,y)$ **and** $s(x,y)$**:** Mean and standard deviation
Data: Empty room recording where i is a constant and n represents each frame
Function `backgroundAnalysis`$(h_i^n(x,y))$**:**
    $m(x,y) = \text{mean}(h_i^n(x,y))$
    $s(x,y) = \text{std}(h_i^n(x,y))$
    return $m(x,y)$ and $s(x,y)$

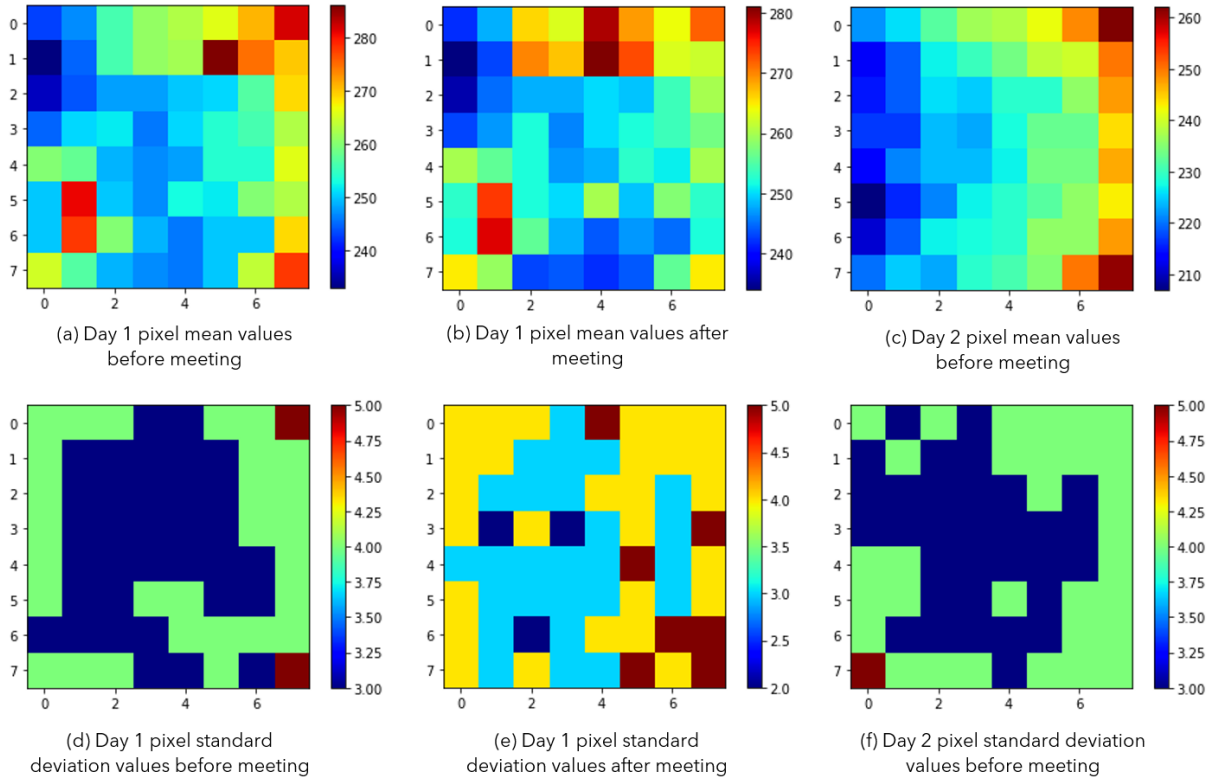


Figure 4. (a) and (b) show the pixel mean values ($m(x,y)$) for the empty room recording in Day 1 data ($h_1(x,y)$ and $h_2(x,y)$) before and after meeting, (c) shows the pixel mean values for the empty room recording in Day 2 data ($h_{10}(x,y)$) before meeting, (d) and (e) show the pixel standard deviation values ($s(x,y)$) for Day 1 data ($h_1(x,y)$ and $h_2(x,y)$) before and after meeting, (f) shows the pixel standard deviation values for ($h_{10}(x,y)$).

As seen in Fig. 4, there are two empty office recordings in Day 1 and one empty room recording in Day 2. The recordings of Day 1 are taken before and after a meeting that took place in the office. The Day 2 recording is taken only before the meeting. When we compare the *before meeting* recordings of two different days (*a* and *c*), we notice that the mean values are different. We understand that, even though there are no heat prints from people yet, the environment is heated differently because of the sunlight exposure from large windows. Day 2 looks sunnier and warmer. Standard deviations of the frames for those recordings (*d* and *f*) look similar. However, even though the sunlight exposure did not change during the recordings of Day 1, before and after meeting recordings show some differences. We see a slight increase in the mean values (*a* and *b*). Besides, the standard deviations look higher for

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 August 2020 doi:10.20944/preprints202008.0202.v1

the after meeting recording (d and e are compared). We believe that the difference is caused by the heat prints left on the seats and the table edges where people was located.

5.1.2. Background noise compensation

We use $s(x, y)$, calculated in the previous step, for removing the pixels that we cannot trust. For instance, if the heat fluctuation (standard deviation) of a pixel is larger than a pre-determined threshold value, we do not use that pixel in the further process (we simply change the corresponding pixel values to zero in the occupied room recordings). If all pixels of $s(x, y)$ have values lower than the threshold temperature, then $s(x, y)$ matrix has no impact to the further processing steps. In our studies, we have chosen the heat threshold as 50 (corresponding to 5 Celsius degrees), since we have noticed that a human heat print cause more than 50 standard deviations when we compared to the standard deviation of the empty office.

For removing the background noise, we have conducted experiments by employing the *Fine Grained Saliency approach* [22]. This method creates real-time saliency maps that indicate the image regions which look different than their surroundings. With this method, we expected to extract the heat resources which might indicate both people and also noisy pixels. Even though this saliency method gives very successful results on high resolution images, it did not perform well on our low-resolution data. Therefore, we believe that the noise compensation method that we proposed above is a reliable solution for this case, and, in general, when low resolution imanges are used.

We provide an example heat image frame before and after the background noise compensation in Fig. 5. There is one person in this data, however the original recording had very high heat levels on the right top and bottom pixels. The difference clearly shows that the background noise compensation algorithm can eliminate the noisy pixels and highlight the person in the scene.

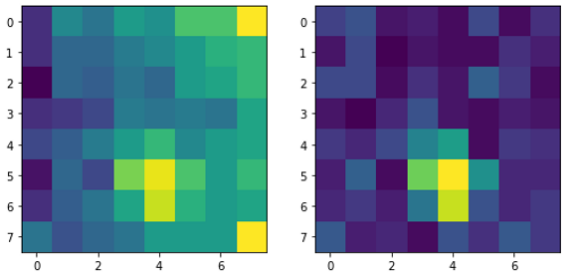


Figure 5. Left; the first frame of the $h_{13}(x, y)$ recording as an example of raw heat image having one person in the scene. Right; the same frame after applying the background noise removal technique.

5.1.3. Occupancy prediction

After the background noise compensation is applied, we apply adaptive filtering and create a binary image from each frame [23]. After obtaining the binary image, we apply a bounding box around each binary segment.

The field of view of the sensor is 60 degrees in both vertical and horizontal directions [20]. In Fig. 2 (b), we approximate the coverage area of each heat sensor frame as a side view. Considering the height of the sensor, we expect $2.2 \times 2.2m^2$ coverage area in the table top level. This means that, when we divide the area into 8×8, each pixel shows approx. an area of $0.27 \times 0.27m^2$. Assuming that one person covers between $0.5 \times 0.5m^2$ to $0.75 \times 0.75m^2$ (top view), we expect one person to cover from 2 to 3 pixels width in the image when a bounding box is located on the detected region.

Based on this assumption, when a bounding box width or height is larger than 3 pixels, we count the number of people by dividing the $(width \times height)$ to 3×3 and rounding the result to make it an integer value. In this way, we count the number of people in each frame. This predicted value can be presented real-time. However, if a correction is needed (because of any heat noise fluctuation during

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 August 2020 doi:10.20944/preprints202008.0202.v1

the recording), then the following correction module offers filtering and produces a post-processed result.

5.1.4. Occupancy correction

After the real-time computations of the predictions are done, we post-process the results for correcting possible miscalculated frames. After we extracted the number of people in each frame of the whole recording, we filter the predicted values using a 5 frame size median filter. This means that, if the predicted number of people in the ith frame is N_i, the filter considers $[N_{i-2}, N_{i-1}, N_i, N_{i+1}, N_{i+2}]$ for $i \geq 2$ for applying the median filter. Of course, this parameter should be chosen carefully, since it will affect the detection performance. If the frame size is small, e.g. 3, the filtering method would not be able to remove the noisy results which appear more than 3 times in sequential frames. If the frame size is large, then there are more chances to remove the noisy results which appear in sequential frames; however, the larger filter size can also remove true detection results when the number of people really changes in some frames for a very short time period. Therefore, this trade-off should be considered when tuning this parameter, depending on the expected noise and the expected human movement in the recordings.

5.2. Feature classification based occupancy prediction

The workflow of the second approach for occupancy prediction is presented in Fig. 6. In contrast to the previous approach, we do not need a baseline image that represents an empty room, but we need labelled heat sensor recordings in order to train our classifier. With a reasonable good amount of labelled sensor recordings as example, we could train the classifier only once to use it new sensor data later on. With "labelled heat sensor recordings", we mean that we need to know how many people are seen in the scene for each frame. It is not necessary to label the pixels of the frames in order to show where these people are located, since we are not concerned with the locations of the people in the scene. Nevertheless, we would like to extract representative features and train our classifier so it can identify the number of people in new frames. In the following sections, we describe each module of the workflow in detail.

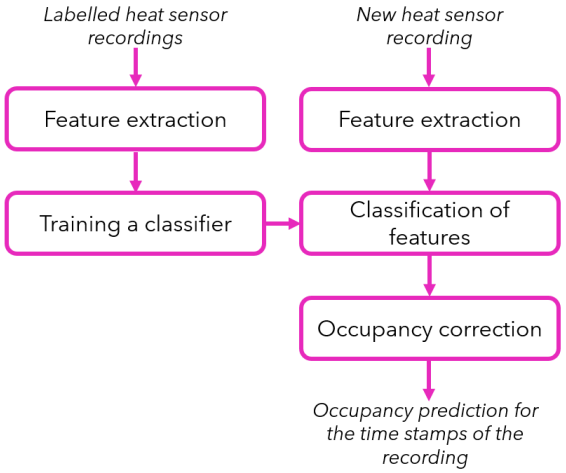


Figure 6. Workflow of the feature classification based occupancy prediction method.

5.2.1. Feature extraction

For each ith frame of a $h^i(x,y)$ heat recording, a feature vector is constructed as $f_i = [\sigma_i, \mu_i, min(h^i(x,y)), max(h^i(x,y))]$. Here, σ_i and μ_i correspond to the standard deviation and mean values of all pixel within the ith frame called $h^i(x,y)$. $min(h^i(x,y))$ and $max(h^i(x,y))$ correspond to the smallest and the largest pixel values within the ith frame. Naming the four components of the f_i

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 August 2020 doi:10.20944/preprints202008.0202.v1

feature vector as $[f_0, f_1, f_2, f_3]$, we have plotted the $[f_1, f_2, f_3]$ feature values for the Day 2 recordings in Fig. 7. Since we can only visualize 3 feature components at a time without applying feature reduction, we chose to visualize all but f_0; compared to the scales of the other feature components, f_0 varies in a very small scale which means that it is hard to visualize how the differences contribute to the differentiate classes.

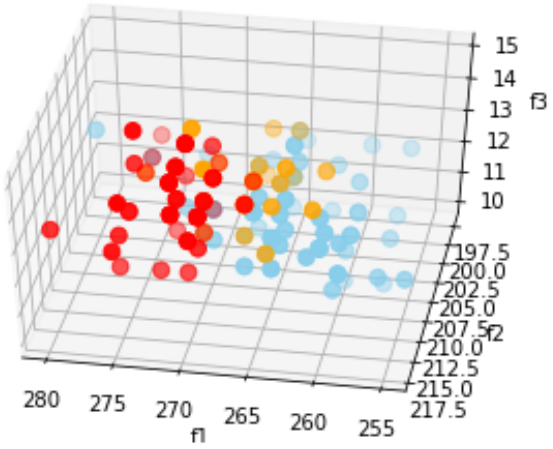


Figure 7. Training features extracted from heat sensor images of the office space when it is in three different states. Blue data points: features of the empty office recordings, Orange data points: features of the office recordings when it is occupied by one person, Red data points: features of the office recordings when it is occupied by two people.

Three components of the Day 2 recording features in Fig. 7, show that the feature space could be learned by a classifier in order to separate different classes showing different occupancy levels of the office.

5.2.2. Training a classifier

In order to classify f_i feature vectors, we use CatBoost classifier. CatBoost is a recently open-sourced machine learning algorithm [24]. It is able to work with diverse data types to help solving a wide range of problems that data scientists face today. "Boost" comes from gradient boosting machine learning algorithm as this library is based on a gradient boosting library. Gradient boosting is a powerful machine learning algorithm that is widely applied to multiple types of business challenges like fraud detection, recommendation systems, forecasting, etc. It can return very good results with relatively little data, unlike deep learning models that need to learn from massive amounts of data. Moreover, CatBoost works well with multiple categories of data, such as audio, text, images and historical data.

CatBoost is especially powerful in two ways:

1. It yields state of the art results without extensive data training typically required by other machine learning methods, and
2. It provides powerful out of the box support for the more descriptive data formats that are associated with many data science problems.

We use the CatBoost library [25] and trained a classifier which learns how to distinguish the feature vectors when they are labelled with the number of people in the room. As seen in the workflow from Fig. 6, the training is performed only once and the trained classifier is used to classify feature vectors when a new heat recording is processed.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 August 2020 doi:10.20944/preprints202008.0202.v1

5.2.3. Classification of features

When a new heat recording is captured, for each frame f_i a feature vector is extracted and the trained classifier is used to identify in which class the feature vector falls. The classification result indicates the number of people in the room for each frame. Since the CatBoost classifier performs at an excellent computational speed, even for classifications of large feature vectors, the classifier can be used in real-time applications.

5.2.4. Occupancy correction

As seen previously, the occupancy correction module can be used as a post-processing step in order to eliminate the noisy computations in frames. Occupancy correction is done with the same median filtering method introduced in the previous 5.1.4 section within the first workflow.

Hereafter, we introduce the results obtained after applying the two workflows previously described.

6. Experimental Results

In this section, we present the performances results of the two suggested workflows for occupancy prediction, discussing different use cases and their advantages and disadvantages. We use explainability methods in order to analyze the feature classification method of the second workflow in depth.

6.1. Analysis of the computer vision based method

We start with the performance analysis of the first workflow. Table 2 shows the performance of the computer vision based workflow for each heat recording used for testing. Here, TD stands for 'True Detection' which corresponds to the percentage of people who are detected correctly in an overall recording; FP stands for 'False Positives'. We notice that the TD rate is 100% and false detection rates are almost 0% in most of the Day 2 recordings. That is probably because of the low environmental noise artifacts (less outside heat and sun rays) during this day. In Day 2 recordings, only $h_{14}(x, y)$ is processed with a lower performance. When we visualize the detection results for this data, we see that a human heat print (because of a person who changed the seat) causes a false detection, see Fig. 8(f).

Data label	Total people	Total frames	TD (%)	FP (%)	FN (%)
$h_3(x, y)$	1	24	100	100	0
$h_4(x, y)$	2	30	100	86.60	0
$h_5(x, y)$	3	27	85.19	37.03	14.81
$h_6(x, y)$	4	51	94.11	0	5.89
$h_7(x, y)$	3	29	97.50	0	2.50
$h_8(x, y)$	2	36	80.55	2.77	19.45
$h_9(x, y)$	1	35	100	80	0
$h_{11}(x, y)$	2	56	100	0	0
$h_{12}(x, y)$	2	47	100	0	0
$h_{13}(x, y)$	1	28	100	0	0
$h_{14}(x, y)$	2	31	83.8	12	3.2

Table 2. Performance of the computer vision based method on each recording which includes at least one person in the scene.

6.2. Analysis of the feature classification based method

The CatBoost classifier has been trained on the 70% randomly chosen frames of the Day 2 data set, the remainder 30% of the frames were used for testing. We obtained 97.64% classification performance (TD) with those frames. Since Day 2 data was captured with less sun exposure (as it was reported

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 August 2020 doi:10.20944/preprints202008.0202.v1

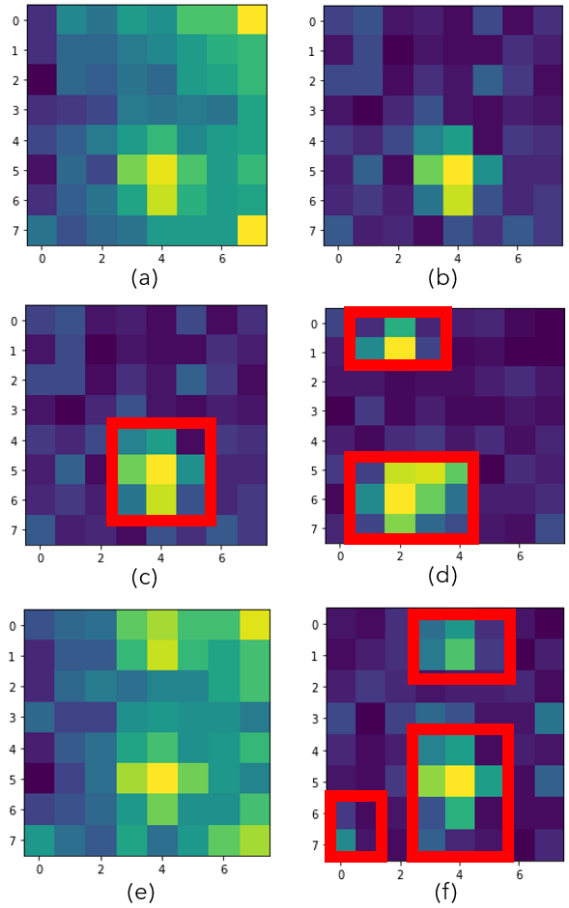


Figure 8. Analysis of the computer vision based method (workflow 1). (a) A raw image frame from $h_{13}(x, y)$ recording having one person in the scene. (b) Frame given in (a), after background noise compensation. (c) Occupancy detection result for the frame given in (a). (d) Occupancy prediction result of a frame from the $h_{12}(x, y)$ recording which has two people in the scene. (e) A raw frame from the $h_{14}(x, y)$ recording which again has two people in the scene but at different seats. (f) Occupancy prediction result of the frame given in (e) which, unfortunately, has a false positive due to the heated seat where a person was sitting earlier.

by the data collecting scientists), we have chosen Day 2 data to start and see the performance of our classifier on the data that we expect to find less noise. After seeing the classifier performed good, we have applied the same experiment on the Day 1 data. We re-trained the classifier (from scratch) on the the 70% randomly chosen frames of the Day 1 data set. We used the remainder 30% of the frames for testing. We achieved 95.48% performance. This result showed us that even in the presence of high noise artifacts, the classifier can be trained and give robust results.

Next, we have conducted experiments by using the classifier which was trained on the Day 1 data for making occupancy predictions on the Day 2 data. In this case, the we have achieved 94.70% performance on the considering the true detection percentage on the Day 1 data. When we mixed Day 1 and Day 2 data for training the classifier, we achieved 96.70% true detection performance on the Day 1 data.

Thus, experiments show a high performances of the feature based occupancy prediction method, even when it trained on a different day with different environmental conditions and used on another day again. However, if one classifier is going to be used in an environment which has significantly changing noise artifacts, the classifier will perform better when data from the different times of the day (or different days) are fused to train the classifier.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 August 2020 doi:10.20944/preprints202008.0202.v1

6.3. Analysis of the feature contributions

As mentioned earlier, the feature vector is constructed using four properties that we extract from each frame. However, it is possible that some of these properties make a very poor contribution to the overall classification. Therefore, we borrow methods from the recent eXplainable Artificial Intelligence (XAI) [26] field of research to investigate further the influence and effect of these proprieties. We use the well-known approach SHapley Additive exPlanations (SHAP) [27] in order to determine which properties of the frames are really effective in the classification module. The goal of SHAP is to explain the prediction of an instance x by computing the contribution of each feature to the prediction. The SHAP explanation method computes Shapley values from coalitional game theory. The feature values of a data instance act as players in a coalition; Shapley values tell us how to fairly distribute the "payout" (= the prediction) among the features.

Fig. 9 presents the SHAP analysis results when Day 2 feature based CatBoost classifier training is considered. We can observe that f_2 feature components (minimum values) make the biggest contribution to the classification and that the f_3 feature components (maximum values) make the second largest contribution; the least contribution comes from the f_1 feature components (mean values). From this analysis, we conclude that we could slightly speed up the second workflow process by extracting only three feature components (f_2, f_3 and f_0) and training the classifier only with them.

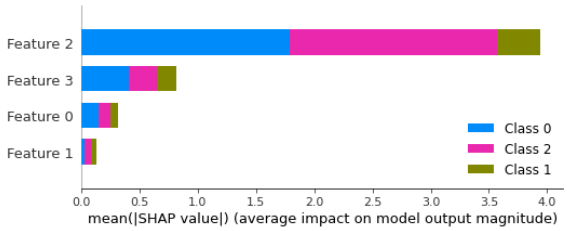


Figure 9. SHAP analysis results for the feature based classification method used in workflow 2.

6.4. Analysis of the noise artifacts

Day 1 recordings were acquired when it was very hot weather outside and the sun was shining straight through the meeting table from the large office windows (this information was provided by our industrial partner who collected the data in a real office environment). During the Day 2 data acquisition, it was rather cloudy and the meeting room was not exposed to the sun rays. We have observed the effects of the outside weather conditions in our Day 1 data set as noise artifacts. We have also achieved less performance on the Day 1 recordings than the Day 2 recordings with both two workflows. It has been obvious that the outside weather impact on the meeting room has caused noise artifacts in the data set.

Day 3 recordings were captured in a 24 hours time period and, thus, contain the largest outside weather condition variety. Fig. 10(a) illustrates the f_i feature components for each frame. Fig. 10(b) provides the plot of the outside temperature value in Malmö, Sweden –where the office is located– from the beginning to the end of the data acquisition process on the same date [28]. We notice a high correlation between the outside temperature and the f_1, f_2 and f_3 feature components (mean, minimum value and maximum value, seen with the blue, green and red colors respectively). However, f_0 feature component (standard deviation of each frame seen in the magenta color plot), looks just slightly affected by the outside temperature changes. Since all four features are affected in a similar direction (even though the f_0 feature component is affected less), we believe that the second workflow can do feature classification robustly regardless of the outside temperature changes. However, the first workflow requires background correction using the mean values collected from the empty office. If the outside temperature changes a lot, this means that those values need to be re-calculated for the empty office again. When we look closely at Fig. 10(a), we notice two spikes (plotted in red) at the final frames of the recording. Tiny spikes on the same frames are also visible on the standard

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 August 2020 doi:10.20944/preprints202008.0202.v1

deviation values (plotted in magenta). When we look at our RGB camera recording which we have left in the office for validation (not for computing any results, only for calculating the performances of our experiments), we notice that one person has visited the meeting room at the time when the first spike occurred. Another two people has sat on the meeting table at the time when the second spike occurred. When we used the classifier (trained on Day 2 data) on these frames, once again, we were able to predict the occupancy correctly. We have correct occupancy predictions on samples of the empty office recordings of the different time frames of the 24 hour recording as well. These results prove the robustness of the algorithm regarding the environmental noise artifacts.

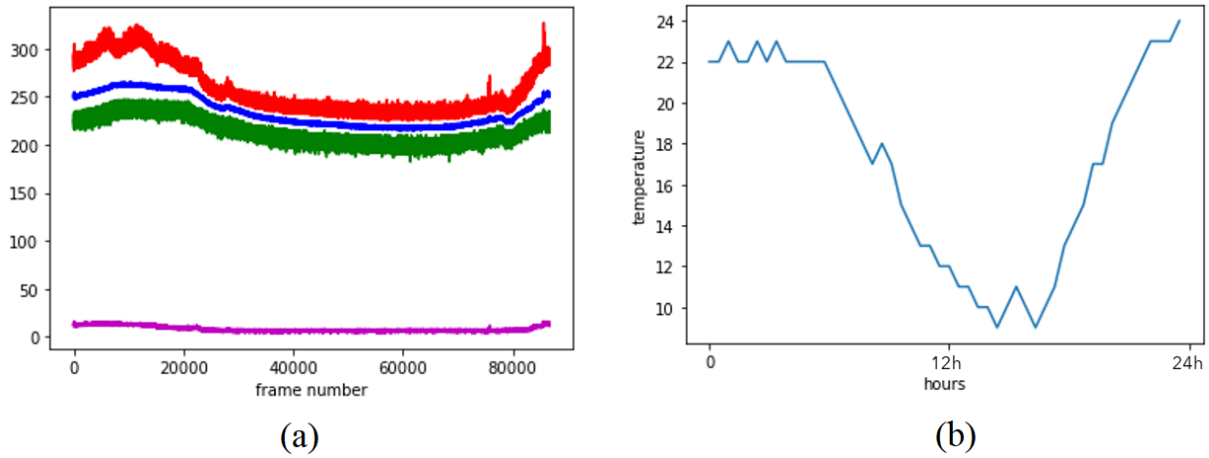


Figure 10. (a) Red; f_3, maximum pixel value of each frame (the maximum value is around 360, which corresponds to 36 degree Celsius). Blue; f_1 mean pixel values. Green; f_2 minimum pixel values. Magenta; f_0 standard deviation of the pixel values in each frame. (b) Outside temperature in Malmö, south Sweden, from the beginning to the end of the heat sensor recording used in (a).

We conclude that the first workflow (computer vision -based) is more sensitive to the environmental heat noise (changing outside temperature and the hitting sun rays) than the second workflow (feature classification -based). Looking at the empty office mean and standard deviations before and after the meeting in Day 1 (see Fig. 5), we can see that the heat prints which are left from people sitting for a long time highly affect the standard deviation calculations. Therefore, the second workflow (machine learning -based) is likely to be less sensitive to the environmental heat noise, but more sensitive to the human heat print related noise artifacts.

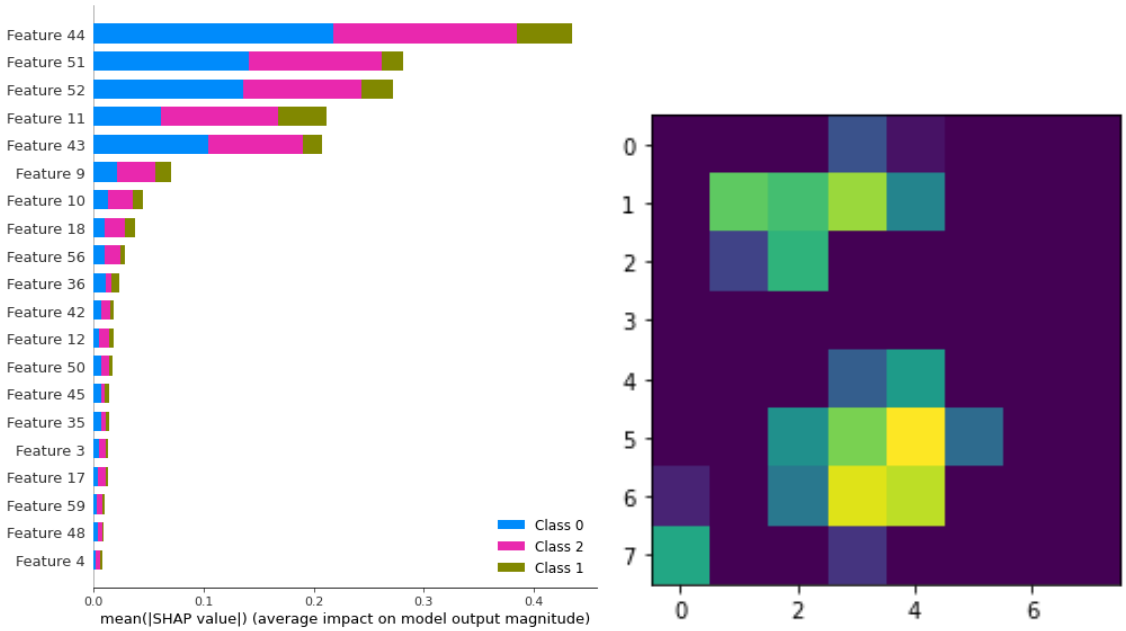
6.5. Analysis of the local pixel contributions

For explaining the local contributions of each pixel to the classification results, we decided to use LIME (*Local Interpretable Model-agnostic Explanations*) by Ribeiro et al. [29]. The explanations given by LIME can help us to understand the trained model, and thus, transform an untrustworthy prediction into a trustworthy one. In particular, the pixels which do not make any contribution to the classification can be removed and more descriptive features can be extracted. Hence, LIME supports and complements the previous analysis carried out using SHAP, but, instead of looking at the contribution of each feature, we look at the contribution of each pixel to the classification.

7. Discussion

Herein, we will address our research questions and discuss the answers we have found with our experiments. We also highlight the lessons learned by our empirical evaluations and the application of the explainable methods.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 August 2020 doi:10.20944/preprints202008.0202.v1



(a) **(b)**

Figure 11. (a) Local contributions of pixels for performing good classifications considering heat recordings as time series for input. (b) Local contributions of each pixel are highlighted based on the explanations given in (a). Brighter pixels contribute more to the classification outcomes.

7.1. Can we rely on computer vision or machine learning based methods in order to provide robust occupancy prediction solutions when we use a low-cost and low-resolution sensor?

We answer this question by offering two different workflows in our study. We saw that, using a traditional computer vision method we can offer a robust real-time solution. However, the computer vision based method that we designed works when an empty room recording is available. When it is not possible to find such recording, our feature classification based second workflow can provide reliable results to identify the number of people within the scene.

7.2. What is the best occupancy prediction accuracy that we can achieve when computer vision based or machine learning based methods are used?

We have reached to over 80% occupancy prediction performance with the computer vision based workflow and over 90% performance with the feature classification based workflow. Choosing the right workflow for the specific use case will lead to the best performance. For instance, if it is possible to record an empty office (without human heat prints), before each meeting, we could provide reliable performances with the computer vision based method. If empty office recordings are not available, feature based algorithm are recommended.

7.3. What are the sources of the noise artifacts and what are the possible noise compensation methods to overcome errors caused by them?

We have noticed two significant noise factors which cause artifacts in our data. One is caused by the sun rays hitting directly the furniture and the objects in the office. Highly heated surfaces appear almost like a person and cause confusion to our algorithms. The second noise factor is the heat prints of people who used the room for a while (sitting at a meeting chair for half an hour, for instance). We saw that our first (computer vision based) workflow is more robust to this second kind of noise factor. However, we believe that using activity recognition methods in time series of heat recordings could increase the occupancy prediction accuracy. A time series analysis approach might also allow

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 August 2020 doi:10.20944/preprints202008.0202.v1

to identify heat prints which become cooler in time (like a chair where the heat print might be lost slowly).

7.4. Comparison of the two different workflows

Table 3 lists a comparison between the two proposed workflows according to the results of our experiments. The table highlights the advantages and disadvantages of each workflow.

Condition	Workflow 1	Workflow 2
Dependency on an empty room recording	Good performances obtained only by having an empty room recording before meetings	It is not necessary to have an empty room recording
Dependency on a training data set	The algorithm only needs an empty room recording as a condition. Other, training data is not necessary.	The algorithm needs a training data set which includes recordings with different numbers of people in the scene.
Dependency on the environmental noise (i.e. sun exposure)	The algorithm behaves robust when similar environment is recorded when the office is empty. Otherwise, it is prone to provide false positives.	If the algorithm has such examples in the training data set, it provides reliable results.
Dependency on the human heat prints	The algorithm behaves robust, since such prints can be eliminated with the background compensation module.	The algorithm cannot easily distinguish human heat print. The algorithm is prone to provide false positives.
Dependency on other heated objects (i.e. computer, hot drinks)	The algorithm behaves robust, because of the position of our sensor (shown in Fig. 2), such objects appear too small in the images and they do not contribute as a noise factor.	Same as the Workflow 1.

Table 3. Comparison of the different occupancy prediction workflows (1 computer vision -based, 2 machine learning -based.

7.5. Further recommendations

Looking at the results provided in Fig. 4, we can make the following two suggestions in order to achieve good accuracy values when using the first workflow (computer vision -based):

- For the background analysis, $m(x, y)$ and $s(x, y)$ values should be calculated when the office has been empty for a while (to make sure that there are no heat prints from people)
- We recommend to re-calculate $m(x, y)$ and $s(x, y)$ when the outside heat is significantly different or the sun rays are illuminating more/less powerfully than the earlier empty room recordings.

When the second workflow is used, we recommend to train the classifier when the noise artifacts are low (no heat prints from people who left a seat and no heavy sun light exposure on the objects). If the temperatures are varying significantly during the day, we recommend to train the classifier with data from different times of the day.

8. Future work

Algorithm-wise, we would like to focus on trying our solutions with multiple sensor data. In this case, our current efforts include: optimization of the sensor distribution in the environment, synchronization of the input data, overcoming challenges coming from the overlapping views and providing results based on the fused information. Secondly, we will focus on developing new algorithms for identifying human activity in our low-resolution sensor recordings. We are currently working on a solution based on the use of a deep neural network, specifically, Mask R-CNN [30]). We

 doi:10.20944/preprints202008.0202.v1

have conducted some early experiments training a Mask R-CNN network which can learn identifying person positions from low-resolution heat sensor images.

Application-wise, we are interested in bringing these smart office solutions to other fields, e.g., elderly care environments. The rapidly growing old age population in many countries has increased the research that addresses challenges in ambient assisted living (*AAL*). Robust smart office solutions, like those presented here, can be adopted in this area [31].

9. Conclusions

Herein we have proposed two new algorithm workflows for real-time occupancy prediction in smart office spaces using very low-resolution heat sensor data that take into account the background noise generated by the surroundings. The first workflow is based on computer vision algorithms, while the second one, on feature extraction and machine learning. We have conducted experiments in order to answer three major research questions related to finding reliable prediction methods, which accuracy can be achieved and the identification of noise artifacts in the environment. Additionally, we have used state-of-the-art explainability methods to understand how the performance of the algorithms are affected by various parameters. We believe that the proposed workflows can be used in many application areas related to the automation and efficient use of offices, spaces and buildings, in order to improve human well-being and the efficient and sustainable use of the environment. The analysis and methods employed and developed to understand the parameters and the environmental noise effects on the occupancy prediction problem might help other researchers to develop robust algorithms to process very low-resolution sensor data.

Author Contributions: Beril Sirmacek and Maria Riveiro have both conducted research and did authorship to write this article. Beril Sirmacek has designed given workflows, prepared software solutions and conducted the experiments in order to analyze the proposed workflows on the data set. Maria Riveiro has done previous work for the administration of the initial project proposal and for acquisition of the project funding.

Funding: We gratefully acknowledge the grant from The Knowledge Foundation under the project MAPPE (Mining Actionable Patterns from complex Physical Environments using interpretable machine learning).

Acknowledgments: Prof. Niklas Lavesson for supporting this research and our industrial partner, ROL Ergo[2], for their support and for collecting the data set used in this study.

Conflicts of Interest: The authors declare no conflict of interest.The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
HCI	Human Computer Interaction
HVAC	Heating Ventilation Air Conditioning
IoT	Internet of Things
LSTM	Long Short Term Memory
ML	Machine Learning
PIR	Passive Infrared
SHAP	SHapley Additive exPlanations
TD	True Detection

[2] https://www.rolergo.com/

References

1. Teixeira, T.; Dublon, G.; Savvides, A. A Survey of Human-Sensing: Methods for Detecting Presence, Count, Location, Track, and Identity. *ACM Computing Surveys* **2010**, *5*, 59.
2. Agarwal, Y.; Balaji, B.; Gupta, R.; Lyles, J.; Wei, M.; Weng, T. Occupancy-driven energy management for smart building automation. *BuildSys'10 - Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings* **2010**, pp. 1–6. doi:10.1145/1878431.1878433.
3. Liang, Y.; Zhang, R.; Wang, W.; Xiao, C. Design of Energy Saving Lighting System in University Classroom Based on Wireless Sensor Network. *Communications and Network* **2013**, *05*, 55–60. doi:10.4236/cn.2013.51B013.
4. Offermans, S.; Kota Gopalakrishna, A.; van Essen, H.; Ozcelebi, T. Breakout 404: A smart space implementation for lighting services in the office domain. *9th International Conference on Networked Sensing Systems, INSS 2012 - Conference Proceedings* **2012**, pp. 1 –4. doi:10.1109/INSS.2012.6240532.
5. Hsu, Y.L.; Chou, P.H.; Chang, H.C.; Lin, S.; Yang, S.C.; Su, H.Y.; Chang, C.C.; Cheng, Y.S.; Kuo, Y.C. Design and Implementation of a Smart Home System Using Multisensor Data Fusion Technology. *Sensors* **2017**, *17*, 1631. doi:10.3390/s17071631.
6. Piyathilaka, L.; Kodagoda, S. Human activity recognition for domestic robots. *Springer: Berlin, Germany* **2015**, *26*, 395—-408.
7. Sung, J.; Ponce, C.; Selman, B.; Saxena, A. Unstructured human activity detection from rgbd images. *In Proceedings of the 2012 IEEE International Conference on Robotics and Automation (ICRA)* **2012**, *15*, 842—-849.
8. Mukhopadhyay, S. Wearable sensors for human activity monitoring: A review. *IEEE Sens. J.* **2015**, *15*, 1321—-1330.
9. Wahl, F.; Milenkovic, M.; Amft, O. A Distributed PIR-based Approach for Estimating People Count in Office Environments. *2012 IEEE 15th International Conference on Computational Science and Engineering* **2012**, pp. 640–647.
10. Murao, K.; Terada, T.; Yano, A.; Matsukura, R. Detecting Room-to-Room Movement by Passive Infrared Sensors in Home Environments. *AwareCast 2012: Workshop on Recent Advances in Behavior Prediction and Pervasive Computing* **2012**.
11. Tapia, E.; Intille, S.; Larson, K. Activity Recognition in the Home Using Simple and Ubiquitous Sensors. *Pervasive Comput.* **2004**, *3001*, 158–175. doi:10.1007/978-3-540-24646-6_10.
12. Olmeda, D.; de la Escalera, A.; Armingol, J.M. Detection and tracking of pedestrians in infrared images. *2009 3rd International Conference on Signals, Circuits and Systems (SCS)* **2009**, pp. 1–6.
13. Doulamis, N.; Agrafiotis, P.; Athanasiou, G.; Amditis, A. Human Object Detection using Very Low Resolution Thermal Cameras for Urban Search and Rescue. *Proceedings of the 10th International Conference on Pervasive Technologies Related to Assistive Environments* **2017**, pp. 311–318. doi:10.1145/3056540.3076201.
14. Jeong, Y.; Yoon, K.; Joung, K. Probabilistic method to determine human subjects for low-resolution thermal imaging sensor. *2014 IEEE Sensors Applications Symposium (SAS)* **2014**, pp. 97–102. doi:10.1109/SAS.2014.6798925.
15. Singh, S.; Aksanli, B. Non-Intrusive Presence Detection and Position Tracking for Multiple People Using Low-Resolution Thermal Sensors. *Journal of Sensor and Actuator Networks* **2019**, *8*, 40. doi:10.3390/jsan8030040.
16. Shelke, S.; Aksanli, B. Static and Dynamic Activity Detection with Ambient Sensors in Smart Spaces. *Sensors* **2019**, *19*, 804. doi:10.3390/s19040804.
17. Troost, M.A. Presence detection and activity recognition using low-resolution passive IR sensors. *Student thesis: Master, Eindhoven University of Eindhoven, The Netherlands* **2013**.
18. Gonzalez, L.I.L.; Troost, M.; Amft, O. Using a Thermopile Matrix Sensor to Recognize Energy-related Activities in Offices. *Procedia Computer Science* **2013**, *19*, 678–685. doi:10.1016/j.procs.2013.06.090.
19. Johansson, A.; Sandberg, O. A comparative study of deep-learning approaches for activity recognition using sensor data in smart office environments. *Bachelor Thesis, Malmö universitet/Teknik och samhälle, Sweden* **2018**.
20. Ergo, R. Presence sensor technical documentation. https://www.rolergo.com/wp-content/uploads/2019/09/PS_RIO_PRESENCE_68011023-B.pdf, 2020.

21. PanasonicChip. Panasonic Infrared Array Sensor Grid-EYE (AMG88) Data sheet. https://cdn.sparkfun.com/assets/4/1/c/0/1/Grid-EYE_Datasheet.pdf, 2020.
22. Montabone, S.; Soto, A. Human detection using a mobile platform and novel features derived from a visual saliency mechanism. *Image and Vision Computing* **2010**, *28*, 391–402. doi:10.1016/j.imavis.2009.06.006.
23. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics* **1979**, *9*, 62–66.
24. Dorogush, A.V.; Ershov, V.; Gulin, A. CatBoost: gradient boosting with categorical features support, 2018, [arXiv:cs.LG/1810.11363].
25. Repository. CatBoost Library. https://github.com/catboost/catboost, 2020.
26. Gunning, D.; Aha, D.W. DARPA's explainable artificial intelligence program. *AI Magazine* **2019**, *40*, 44–58.
27. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems* **2017**, pp. 4765—-4774.
28. Time.; Website, D. Outside temperature in Malmö. https://www.timeanddate.com/weather/sweden/malmo/historic, 2020.
29. Ribeiro, M.; Singh, S.; Guestrin, C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA* **2016**, pp. 1135–1144.
30. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. *arXiv* **2017**, [arXiv:cs.CV/1703.06870].
31. Bourke, A.; O'brien, J.and Lyons, G. Evaluation of a threshold-based tri-axial accelerometer fall detection algorithm. *Gait Posture* **2007**, *26*, 194—-199.