Application of AI-based Image Processing for Occupancy Monitoring in Building Energy Management

Andreas Reichel, Jens Döge, Dirk Mayer and Jan Bräunig
Fraunhofer Institute of Integrated Circuits IIS, Division Engineering of Adaptive Systems EAS, Münchner Str. 16, 01187 Dresden, Germany

Keywords: Vision-based User Recognition, Machine Learning, Building Energy System.

Abstract: Smart Buildings enable significant savings in energy and CO2 emissions by model-predictive methods. The building users have a considerable influence on the energetic building management. On the one hand, they dictate the comfort parameters to be set. On the other hand, they generate internal thermal gains through their presence, affect humidity, consume oxygen and produce carbon dioxide. The more precisely the user behavior is known, the more precisely and resource-efficiently the room climate control can be adapted to this user behavior. In this paper, an intelligent vision-based sensor concept is proposed and tested that is capable to estimate occupancy and activity inside a building. The contribution initially concentrates on functional buildings, since here, compared to residential buildings, there is an even greater need for use-oriented room air conditioning, including savings potential.

1 INTRODUCTION

Buildings, and particularly their energy systems, are a significant contributor to the world-wide CO2 emissions. Besides conventional measures like enhanced insulation of walls or windows, digital technologies are an enabler towards greener buildings.

These smart buildings integrate data from external sources (e.g. weather forecasts) and sensors with powerful analytics and control systems to optimize energy consumption. One of the most important parameters is occupancy.

Traditionally, building automation systems use sensors to measure temperature, humidity and CO2 concentration in order to draw conclusions about the indoor climate. Such systems usually provide very inaccurate data on user behaviour, which makes user-adapted energy-optimised building control with advanced control strategies hardly feasible.

However, improvements in detection of user behaviour are still under development. Two questions need to be answered: technical possibilities and data-law constraints.

This paper aims to provide an overview of integrating a vision-based system into a smart building data infrastructure.

To this end, general system architectures for smart buildings are considered, with a focus on the distribution of data analysis to the edge of the network. Potential scenarios for the integration of a smart sensor for occupancy detection into building automation are discussed. Finally, the set-up and first results from a feasibility study at a real building are presented and directions for further research are given.

2 CURRENT STATE OF THE ART

2.1 Smart Building Technologies and System Architectures

Smart buildings integrate digital building control systems and networked building automation. Using multiple, distributed sensors for the building energy system enables advanced building control schemes. In

https://orcid.org/0000-0002-3971-1585
https://orcid.org/0000-0002-2891-984X
https://orcid.org/0000-0002-4972-6529
https://orcid.org/0000-0002-7282-723X
contrast to preset schedules, the energy system can be adapted to physical parameters, weather forecast or occupancy. This can lead to significant energy savings of 24% (King, 2017). The gathered sensor information can also be applied to fault detection and diagnosis.

In summary, the main hardware components of a smart building are:

- Sensors to acquire physical parameters
- An internal network to connect smart devices and control systems as well as an internet connection
- Centralized or decentralized computing capabilities

This infrastructure is complemented by analytics and control algorithms to extract relevant information from sensor data, classify the state of components, predict time series for relevant parameters and, last but not least, implement control actions. Obviously, these applications are very well-suited for machine learning algorithms.

Findings of a recent review study show that most control architectures are cloud-based (Yaïci, 2021). This enables using powerful machine learning algorithms, e.g. deep learning, that require high efforts for training. However, such an approach needs a reliable data connection from the building to the cloud; also, data privacy might be a concern.

An alternative to this is shifting parts of the data analytics and control to the edge of the network (Figure 1). Due to computing power and a network connection being available in nearly every building automation device, this enables a variety of system architectures. In most cases, data analytics and control functions are implemented on a unit that also serves as a gateway between the sensor-actuator-network and the internet (Curto Fuentes, 2021). In a building automation system, this would be the central building automation control platform (Rinaldi, 2020).

Such an approach has several advantages. The distribution of computing power enhances the scalability of the system. The control system is more robust, which is particularly relevant in case of extreme weather events: If the connection to the cloud is lost, external information (e.g. weather forecast) is missing, but the control system is still operating. Furthermore, the building owner is not forced to share raw data with a remote cloud server and can implement higher privacy standards.

Still, smart functions can be shifted further towards the edge of the network by using smart sensors (Figure 1). The concept of a smart sensor is well known for three decades (Najafi, 1991). Their main characteristics are:

- Analog preprocessing of the input signals (amplification, filtering)
- Analog to digital conversion
- Bidirectional communication possibility
- Protocol based communication interface and a unique identification (network address)
- Possibility to implement data analysis algorithms, e.g. compressed sensing, auto-calibration and self-diagnosis

For building automation applications, the digital networking capabilities alleviate the installation of a scalable sensor network that can be spread also on larger facilities. Compression of the acquired signals can be useful, if a low-bandwidth wireless network is used and a large number of sensors have to be connected.

In certain applications, the sensor can also be shut down into sleep mode, enabling long-term operation on batteries or even energy harvesting. This is especially relevant when the sensors are wireless, which is an attractive solution for retrofitting of existing buildings.

There are several potential energy sources in buildings to power energy harvesting systems:

- Floor vibrations can transmit power to piezoelectric oscillators, that also can be used as a sensor measure to detect occupancy behavior (Jung, 2018).
- Also, indoor solar panels have been investigated for powering wireless sensors, e.g. temperature and light sensors (Fraternali, 2018). Last but not least, radio frequency (RF) energy harvesting should be mentioned. (Bjorkqvist, 2018) developed a harvester that is able to generate power in the micro-Watt range from surrounding Wifi and cellular networks in a building.

![Figure 1: System architecture for deployment of AI to the edge and the sensor.](image-url)
Usually, sensor platforms detect a variety of physical parameters. For instance, the BiB (Building in Briefcase) platform presented by Weekly (Weekly, 2018) integrates temperature, humidity, ambient light, CO2, infrared motion detection, and inertial measurement units in a wireless platform. Another example for a wireless multi-sensor platform was developed and tested by De Donno et al. (De Donno 2018), acquiring humidity, temperature, CO2, air quality, and occupancy by passive infrared sensors. Yet, power measurements showed that the device would operate just about 1 month until a battery replacement is necessary, i.e. a significant maintenance effort would be necessary.

Especially for the case of occupancy detection or other perhaps sensitive personal data, preprocessing directly at the sensor avoids transmission of raw data to remote devices, thus increasing the privacy of the acquired data.

### 2.2 Inclusion of User Behavior in Building Automation

For some years there has been an effort to use model predictive control strategies (MPC) in the field of building automation, which can save up to 40% of the energy used compared to conventional control (Serale, 2018). These include, for example, model-predictive zone and individual room controllers (Paschke, 2018; Kelman, 2011; Gwender, 2010), the design of which, however, requires knowledge of a prediction model for the air comfort parameters. This can be generated automatically using recorded measurement data (Paschke, 2019), but the number of people present in a room remains unknown. Since air comfort parameters are significantly influenced by the number of people, measuring room occupancy is needed for model identification and thus for the design of an MPC control strategy.

The recording of high-resolution measurement data from building occupancy as well as control and operation of technical systems with the help of decentralized sensor technology has become an important part of building control technology and the planning process in the last 10 years due to the increasing spread of bus systems and IoT applications in buildings. In this way, engineers are given a tool to cope with the increasing complexity of efficient energy supply systems. In this context, the implementation of multi-year monitoring campaigns for system optimization after commissioning has become established in selective research and innovation projects.

### 2.3 Determination of User Behavior

Several options exist to determine the occupancy and user behavior in-situ as an input to model predictive building control systems.

A well-established technology are CO2 sensors, that directly acquire the quality of the air inside a room as an important control input to the ventilation.

A major difficulty here is that measurements are usually taken at only one or very few individual points. This is particularly problematic in large rooms (lecture halls, gymnasiums...), as very strong local differences can occur, which can negatively influence the functionality of the control system and thus the energy consumption of the building if the sensor is placed in an unfavorable position. Further disadvantages of conventional CO2 sensor technology include:

- No direct statement about the number of people present.
- Delayed/indirect measurement of human activity
- Relatively high susceptibility to interference
- High calibration effort for CO2 sensor technology, comparably strong sensor drift

However, intelligent sensing can enhance the performance of such systems: The number of occupants can be estimated using an observer that processes the current CO2 concentration by a physical model (Jin, 2015). Another approach is the fusion of a CO2 and a light sensor (Huang, 2017).

Besides CO2, also other modalities are suitable to measure occupancy. (Ahmad, 2020) gives a review on several technologies, including passive infrared sensors, ultrasound or wireless network connections.

As demonstrated in (Naylor, 2017), integration of CO2 with other sensor modalities like Wi-Fi tracking by machine learning based data fusion can drastically increase the precision of the occupancy estimation.

The classic optical approach is based on simple presence or activity sensing using passive IR sensors (PIR, 2020). Here, one to three individual point sensors per sensor head with upstream Fresnel optics detect the movement of people through the room based on the change in their thermal signatures. This approach can be used very well to activate light in small rooms, but even transferring it to larger rooms is not possible due to the limited resolution. Likewise, people who do not move “disappear”, as they are no longer detected by the system without a change in their whereabouts in the room.

Thus, this sensor technology is very inaccurate, especially in large rooms, and is neither able to provide information about the actual number of...
people in a certain area of the building, nor information about their level of activity. It is also not possible to reliably detect the absence of people, so a predefined period of inactivity in the room is usually used as an equivalent.

Alternatively, there are surveillance cameras in security-critical building sections that transmit or record real images and thus technically provide the possibility to also derive information about the actual number of people and their activity. However, the computer-aided storage and automatic analysis of image data must usually be carried out in compliance with strict data protection guidelines.

Doppler radar-based solutions (Radar, 2020) try to circumvent this disadvantage, as they react much more sensitively to movements. Since these sensors also only provide integral information about larger areas, they are also not suitable for detecting specific numbers of people.

According to the state of the art, the use of systems based on area scan cameras is currently the only possibility to solve the resolution problem, however, especially in connection with the automatic evaluation of the image data, it rightly places very high demands on data protection. Thus, the output and storage of real image data and their external processing is a fundamental problem that is difficult to solve technically and, moreover, also has to contend with considerable acceptance problems among users.

In certain areas with increased security requirements, such as the gate environment of airports, autonomous 3D camera systems based on a stereo arrangement (Stereo 2020) are sometimes used to count and track people. These systems usually have a very high accuracy, but can only be used in quite small areas - usually up to 5m x 5m at typical room heights - and have to be installed and set up in a complex way. Due to the two camera heads required for 3D reconstruction and the high computing effort, these systems are also quite expensive. The data protection aspect is the same as with classic camera solutions.

To prevent using raw image data, that can include private information, encryption methods have been proposed that shuffle the image data in the regions containing persons (Ahmad 2020). Yet, such algorithms should be implemented directly at the pixel sensor. This way, an intelligent sensor is realized that does not transmit privacy-relevant information.

3 SOLUTION APPROACH AND FIRST FEASIBILITY STUDY

For the real-world demonstrator the lecture hall center “Bergstraße” of the TU Dresden was chosen and equipped. First, in a standard lecture hall, user behaviour can be recorded with sufficient accuracy over a certain period of time and processed with regard to the activity level.

As a key component for the vision-based occupancy sensor, the vision system-on-chip (VSoC) depicted in Figure 2 has been used (Döge, 2015).

![Figure 2: Chip microphotograph of the Vision System-on-Chip (VSoC).](image)

This VSoC is capable to analyze image data directly on the sensor chip, extract information that can no longer be personalized and output only the features required for further processing. Based on this, the embedded image processing system performs the next sensor-external processing steps only on pre-selected image descriptors - such as specific gradient and texture features or local activity levels, and calculate the approximate number of people in regions of increased activity. The implemented activity and occupancy measurement system thus takes special account of the data protection required for public buildings.

The auditorium observed is shown in Figure 3 top. Due to its size and geometry, it has to be observed from two overlapping perspectives (see Figure 3 bottom) in order to achieve a complete coverage of the room with the used optics on the one hand and on the other hand to avoid that the effective subject size in pixels varies too much out of the camera perspective (see Figure 4 right). By choosing a height of the vision system's viewpoint of about 5m and splitting it into two views for two installed vision
systems, the minimum number of pixels per monitored seat could be slightly improved. However, the resolution in the rear areas of the room is still too low for detection and classification of the head-shoulder region, which is common according to the state of the art. Dynamic and static methods were investigated as a basis for occupancy analysis, but all of them basically work without object detection, which is potentially problematic in terms of data protection.

In dynamic scene analysis, it is assumed that people are generally not completely motionless. Regardless of the lighting situation, the corresponding regions are only learned as background on the assumption of the absence of movements over a certain period of time. Regions of a defined dimension, in which deviations from the current background model are detected, are by definition foreground and thus people.

In static methods, a large number of images of the unoccupied space are used to train the classifier. The training data differ mainly in the possible illumination situations that have been automatically recorded in advance. An extension of the training data set by data augmentation is also possible but was not necessary in the presented example. The trained model of the room is then used in operation by detecting deviations from it as people.

As a basis for both approaches, various descriptors feasible on the VSoC were evaluated, such as:

- gradients and angles,
- histograms based thereon (Histogram-of-Oriented-Gradients - HOG), and,
- Local Binary Patterns (LBP).

The static approach using principal component analysis (PCA) based on gradient data proved to be the most robust to the confounding effects of illumination. For this purpose, a static background model is initially trained by dividing a set of training gradient images into rectangles and, for each rectangle, calculating the first five principal components. Using more than five components did not improve performance any further. At runtime, each rectangle’s gray-values are projected onto its respective principal components and back into image space. An activity map is then obtained from the absolute difference between this reconstructed image and the pre-trained background model. Binary thresholding and morphological operations are used to close holes and suppress false positives. Including angular information provided very little added value, but introduced a significant noise contribution due to uncertainty in low-contrast areas.
The determined active regions were then converted into an occupancy level using the spatial resolution illustrated by the size of the seats in Figure 4 (right) as an example. Although this approach is more of an estimate than a precise count, especially when the room is very crowded with clusters of people connected or overlapping, it does provide a reliable indication that is considerably more accurate than required in terms of the air turnover and thermal load of the people present. In particular, considering random samples at low occupancy levels (less than 10 people), the estimated occupancy differs from the true occupancy by up to 20–25%. Empty rooms were consistently detected as empty, though. At higher occupancy levels (50–100 people), the relative estimation error was consistently below 5%. For reference, an example sequence of detected occupancy and roughly estimated error is depicted in Figure 5. It can be clearly observed that students start trickling into the lecture hall at around 04:27 pm, with an even larger group entering at around 04:35 pm. As expected, the estimated number of people fluctuates only mildly until 06:10 pm, when lecture time ends and students start leaving the lecture hall rather hastily. A more in-depth evaluation would have required either acquiring high-resolution gray-value image data instead of gradient data or attending events in the lecture hall in-person. The former was discarded due to data protection concerns raised by the proprietor of the building. The latter was not possible because most events taking place at the time were examinations with strict access control.

The vision-based occupancy sensor system and the building management system operate without feedback and independently of each other. The occupancy percentage determined in the building area being monitored is assigned to the available building operating data and statements can then be derived on comfort and energy optimization of the operational management. The aerosol load in the rooms is also to be reduced by adjusting the implemented air volume. Machine learning algorithms are used for this optimization task.

4 PERSPECTIVES

The operator of a single functional building or city district with very heterogenous building usage (e.g. lecture hall, classrooms and seminar rooms, sports hall, shopping center, exhibition hall) receives information about when and where exactly how many people are in a quarter/building area at what activity level (walking, sitting, physical work, playing sports, etc.). For each usage scenario, the system provides statements on optimal building operating parameters and target specifications that can be applied manually by the system operator. As a result, the energy demand (and thus the CO₂ emissions) for the building/district area under consideration is reduced while the required comfort parameters are ensured through use-specific system operation and minimisation of aerosol dispersion.

In the future, the system can be expanded in such a way that, after a pilot phase, manual tracking of the operating parameters can be rolled out to other buildings/districts and finally integrated fully automatically into the building management system. An OPC-UA interface to commercial building management systems (BMS) and energy management systems (EMS) systems, which is operated together with these systems as a value-added service, is ideal for this purpose. The operator of a building gains potentially comprehensive benefits through such a scenario (Focus on energy and cost savings):

- Long-term forecast
- Model- and user-specific optimisation
- Communication on necessary maintenance
- Increase in system availability through wear prediction adapted to the usage behaviour

The algorithms implemented at the VSoC level have been tested in a lecture hall with fixed furniture positions and limited activity of the occupants. Future research should include improvements towards rooms with movable furniture (resulting in varying background images and greater variations in positions and activities of the occupants).

In addition to the determination of the occupancy levels of buildings, the presented vision technology can also be used in a variety of other applications, e.g. for the analysis of walkways in stores for the evaluation of interest in presented advertising measures or for the demand-driven regulation of ventilation and air-conditioning of booths at trade fairs. In particular, the large dynamic range of the sensor with a linear-logarithmic characteristic makes it predestined, e.g. in combination with other
wavelength ranges, for outdoor use for movement sensing in public areas (Reichel, 2019).

Due to the high compression rate of the intelligent sensor from raw images to the estimated number of occupants, a wireless connection is a possible improvement. This would facilitate the retrofit installation of the sensors at hard-to-reach positions in existing buildings. However, in turn an energy management on the sensor platform will have to be implemented to enable a long-term operation on batteries. Even more ambitious would be the development of a low power system that could be supplied by energy harvesting technologies.

5 CONCLUSIONS

In this paper, a concept for using machine learning methods on sensor level to provide a privacy-oriented vision-based occupancy detection has been proposed and tested in a first feasibility study. The article makes it clear that there is great potential in the application of vision technology in the building sector with regard to energy saving.

Furthermore, the approach provides an important building block for intelligent neighbourhood monitoring, also with the aim of recording the utilisation concepts in the neighbourhood and deriving further savings potential from them.

In the perspective, in addition to CO2-compatible modes of operation, it is also necessary to find ways of reducing the aerosol load individually when there is a high density of people in order to reduce the potential for viral contagion.

ACKNOWLEDGEMENTS

On the basis of the TGM.plus project (SAB Sächsische Aufbaubank), among others in cooperation with the partner of the SIB (Staatsbetrieb Sächsisches Immobilien- und Baumanagement), initial results on image processing-based presence detection could be elaborated: The aim of the project was to develop a value-added module for existing building automation systems. Its application enables a reduction of resource consumption in technical building operation through model-based forecasts of energy consumption and the derivation of appropriate measures. The basic knowledge created there makes a significant contribution to this paper. The authors gratefully acknowledge the ENLIGHT project (http://www.enlight-project.eu/), funded by ENIAC and the German Federal Ministry of Education and Research, for the financial support of the development of the Vision System-on-Chip.

Last not least the authors wish to thank Dr. Jürgen Haufe, Chief Scientist at EAS, for initiating the work and acting as a mentor for the research team.

REFERENCES


