Energy and Buildings 130 (2016) 787–793

Contents lists available at ScienceDirect

Energy and Buildings

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

Establishing an image-based ground truth for validation of sensor data-based room occupancy detection

Steffen Petersen*, Theis Heidmann Pedersen, Kasper Ubbe Nielsen, Michael Dahl Knudsen

Department of Engineering, Finlandsgade 22, Aarhus University, DK-8200 Aarhus N, Denmark

ARTICLE INFO

Article history:
Received 14 April 2016
Received in revised form 31 August 2016
Accepted 5 September 2016
Available online 6 September 2016

Keywords:
Occupancy detection
Non-intrusive ground truth
HVAC control
Energy-efficiency

ABSTRACT

Occupancy controlled heating, ventilation and air condition in offices are a mean to avoid energy waste due to over-conditioning of rooms. Occupancy detection methods based on sensor data that is already available in a building for other reasons than occupancy detection are therefore of great interest. However, such methods need to be validated according to a ground truth. This paper presents an image-based method for automatic establishment of a reliable and anonymous ground truth at entrance areas where the room height only allow the camera to be installed at 2.3–3.0 m above the floor. The method deploys various image processing techniques to infrared depthframe images recorded with a ceiling-mounted Kinect camera to anonymously detect and track persons entering or leaving a room and, consequently, counting the number of occupants in the room. The method demonstrated an accuracy of more than 98% in a stress test, 100% in a range of relevant function test and 99% in a three week long room occupancy test.

1. Introduction

A simulation-based study by Dong and Lam [1] demonstrated an average energy saving of 18.5% using occupant behaviour-based control approach compared to conventional temperature set-point schedule. Another simulation-based study by Goyal et al. [2] identified energy savings of approx. 50% when applying HVAC controllers with integrated occupancy detection and prediction instead of rule-based temperature schedule. These findings indicate that the energy performance of HVAC systems can be improved by integrating real-time occupancy detection in the control of HVAC systems. The realization of these energy savings requires the development of reliable yet inexpensive occupancy detection technologies that can be deployed easily in practice.

The current development of methods for occupancy detection can be split into two groups: image-based methods and data-based methods. Image-based methods rely on camera technology and can be quite accurate compared to manually counted persons, e.g. 97.0% [3] and 94.8% [4] for methods where the camera is surveying a vertical scene and 99.7% [5] for a method where the camera is surveying a horizontal scene. However, occupants can perceive the use of cameras as a violation of their privacy. Furthermore, the establishment of cameras is often an additional investment and running cost for the building owner.

Data-based methods are a more cost-efficient and non-intrusive approach to occupancy detection which relies on data from sensors already installed in the building for other reasons than occupancy detection. One of the most commonly used sensors for the purpose of detecting occupancy is a passive infrared sensor (PIR) [1,6–10] which are often installed for activation of lighting. Other current data-based detection methods are based on GPS location information [11], cellular ID and WIFI AP location analysis [12], office appliance power consumption [13], radio-frequency identification systems [14], and data from indoor climate sensors used for HVAC control [1,7,15,16].

Some of the above mentioned data-based models assumes that their data is a 100% accurate expression of occupancy [8,9,15]. Other studies attempts to validate and calibrate their data-based occupancy detection model by comparing the output to a ground truth. The ground truth is usually generated by manual real-time counting of occupants [14,16] or by manual counting of persons in images from cameras [1,7]. One attempt to automate the recording of the ground truth is to equip test persons with pedometers enabling their personal computer to register their presence within a range of 6 m from the computer [13]. Two studies claim that their data for occupancy detection in fact is the ground truth [11,12]. Manual

* Corresponding author.
E-mail address: stp@eng.au.dk (S. Petersen).

http://dx.doi.org/10.1016/j.enbuild.2016.09.009
0378-7788/

counting of persons may lead to a reliable ground truth in many cases but it is an expensive method due to the man hours. Furthermore, it may not be an accurate method for long-term and/or large-scale experiments as it relies on the concentration of the persons counting.

An image-based detection method, as described earlier in this paper, may be ideal for automatic establishment of a reliable ground truth. The image-based method by Zhang et al. [5] seems appropriate as it has a detection accuracy of 99.7% when the camera is mounted above the entrance of a supermarket in a height of four meters above the floor. This paper presents an image-based method based on the method proposed by Zhang et al. [5]. The method is targeted for automatic establishment of a reliable ground truth in rooms with a room height of three meters or less which is often the case for offices and residential buildings. The outline of the paper is a description of the method followed by the description of different tests which were conducted to document the performance of the method, and a conclusion.

2. Method

An image-based method for unsupervised real-time counting of persons entering and/or leaving a room with a room height of 2.3–3.0 m is developed. The method relies on image data from the camera Kinect V2 for Windows [17] placed near the entrance of the room mounted at the ceiling facing the floor. The Kinect camera provides data from a range of different types of embedded sensors but only the *depthframe* image is used in this method. The depthframe image is generated by infrared sensors that calculate the distance in millimetres from the camera plane to any object in the camera field of view in a grid corresponding to an image resolution of 512×424 pixels. The main argument for placing the camera at the ceiling with the lens facing the floor and the use of the depthframe image is privacy: It is difficult to recognize a specific person in the infrared depthframe image especially when recorded from above. This is best illustrated by comparing the RGB image in Fig. 1(a) with the depthframe image in Fig. 1(b). Furthermore, the use of the depthframe image is a reliable source for person detection. Persons usually emit significantly more infrared radiation than other objects in a room and are therefore easily captured by the infrared sensors in the Kinect camera. However, a range of image processing techniques are needed for unsupervised real-time detection of people in the depthframe image. The description of the techniques used in the method proposed in this paper can be divided into a detection part which has the purpose of detecting persons in the camera field of view, and a tracking part which determines whether detected persons enter or leave the room.

2.1. Detection

The primary technique for detection of persons in the depthframe image is the so-called *waterfilling* method [5]. However, we recommend applying the image processing techniques described in this section to increase the efficiency of the waterfilling algorithm in relation to real-time people detection. The processes must be applied in same chronological order as they are described in the case of implementation.

2.1.1. Pixel filter

Limitations of the Kinect camera sensor technology may result in zero-value pixels in the depthframe image which means that the camera was unable to determine a distance from the camera to this pixel. This is illustrated by the black pixels in Fig. 1(b). These undetected pixel depths will become a disturbance to the people detection algorithm (the waterfilling algorithm described in Section 2.1.4). A pixel filter as described by Shen et al. [18] is therefore applied to remove these pixels. The pixel filter scans the depthframe image pixel by pixel: Let $u_{i,j}$ be the element value in the matrix of the depthframe image. Whenever $u_{i,j} = 0$ the filter counts the number of non-zero pixels, n, of the 24 neighbouring pixels to $u_{i,j}$, see Eq. (1). These pixels are denoted p_i. If n is above a user-defined threshold, r, then a new value is assigned to $u_{i,j}$ calculated as the average value of p_i, see Eq. (2).

$$\mathrm{n} = x_i \overset{\sum}{\underset{24}{i = 1}} \quad \text{where} \quad x_i = \begin{cases} 0 \ \ if \quad \mathrm{p}_i = 0 \\ 1 \ \ if \quad \mathrm{p}_i > 0 \end{cases} \tag{1}$$

$$u_{i,j} = \frac{\sum_{i=1}^{24} \mathrm{p}_i}{n} \tag{2}$$

The pixel filter effectively removes nearly all the disturbance from undetected pixel depths as illustrated in Fig. 1(c) where the black edges and areas of the raw depthframe image in Fig. 1(b) are removed.

2.1.2. Background subtraction

The waterfilling algorithm for people detection described in Section 2.1.4 could be applied directly after the pixel filter described in Section 2.1.1. However, in this case the waterfilling algorithm has to process the whole image which in terms of computation time could pose a practical problem. A static background subtraction technique is therefore introduced (see Eq. (3)) to separate the people moving in the foreground which are of interest.

$$M_t = D_t - B \tag{3}$$

where D_t is the pixels of the depthframe image at the time t, B is the averaged pixel values of the background depthframe image, and M_t is the moving foreground of the image at the time t. The waterfilling algorithm is then only applied for the non-zero pixels of M_t which is significantly less than the whole image thus reducing computation time. The establishment of the background image B requires an algorithm since the depthframe image of a static scene differs from frame to frame due to instability in measurements of the infrared sensors. Eq. (4) calculates B as the average value of the individual image pixel of m consecutive frames, D_i, plus K which is a matrix with the same dimension as D_i in which all elements has the value of n millimetres. The matrix K is added to stabilise the background image but it can also be used to leave out e.g. pets walking on the floor in the room if n is set to be higher than the pet in question.

$$B = K + \frac{\sum_{i=1}^{m} D_i}{m} \tag{4}$$

The result of background subtraction is illustrated in Fig. 1(d). Notice that B must be updated if the static background changes, e.g. if the furnishing in the camera field of view changes. In the suggested method, B is updated when no movement is registered in the camera field of view according to the procedure described in Section 2.1.2 for 5 min and every following half hour.

2.1.3. Frame weighted filter

The detection of persons might be disturbed if persons are moving very fast across the camera field of view. A frame weighted filter is therefore applied to each pixel u of the depthframe image, see Eq. (5). The filter is an exponential weighted average of the pixel in present frame, u_t, and the previous frame, u_{t-1}. The weighing factor α is user-defined.

$$u = \alpha \cdot u_{t-1} + (1 - \alpha) \cdot u_t \tag{5}$$

The frame weighted filter results in a lack in the image as illustrated by the pixels close to the moving arm in Fig. 1(e). This lack improves the detection of fast moving persons.