Energy and Buildings 77 (2014) 270–280

Contents lists available at ScienceDirect

Energy and Buildings

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

A robust occupancy detection and tracking algorithm for the automatic monitoring and commissioning of a building☆

Huang-Chia Shih

Human-Computer Interaction Multimedia Laboratory, Department of Electrical Engineering, Yuan Ze University, Taoyuan, Taiwan, ROC

ARTICLE INFO

Article history:
Received 13 December 2013
Received in revised form 24 February 2014
Accepted 31 March 2014
Available online 12 April 2014

Keywords:
Building control
Commissioning
Building management system (BMS)
Building monitoring
Intelligent-controlled environment
Occupancy detection
Tracking algorithm

ABSTRACT

The focus of this study is the 24 h a day monitoring of buildings for commissioning purposes. Based on an image-based depth sensor and a programmable pan-tilt-zoom (PTZ) camera, the proposed monitoring system enables the continuous detection and tracking of the occupants, even under dim-lighting conditions. The proposed SVM-based observation measurement provides a more reliable tracking performance. This paper presents a robust day-and-night people tracking and counting algorithm. The function of large-scale field monitoring is realized using a PTZ camera network instead of a conventional fixed camera. Furthermore, based on the depth image sensor, the contour information of the occupant can be applied for more accurate activity recognition. In our experiments we demonstrated the positive result of the occupancy detection and tracking algorithms applied to count people and monitor a building.

1. Introduction

An intelligent-controlled environment is becoming standard for buildings. The task of monitoring and commissioning most of the common devices in a building such as air conditioning, ventilation, heating, and lighting has become a critical issue from the point of energy saving. In Taiwan for instance any self-produced energy is minimal, with over 99% of the energy being imported from abroad. Nevertheless, in 2010 the average CO_2 emission per capita/year was 11.66 tons, ranking it in 19th place in the world [1]. Based on the statistical analysis of the energy consumption by the Bureau of Energy, Ministry of Economic Affairs of Taiwan, in 2012, the buildings and business sectors were responsible for 30% of the total energy consumption, higher than any other non-industry production sector. Heating, ventilation and air conditioning (HVAC) accounts for 43% of the total energy consumption in commercial buildings, and lighting is responsible for 26%. Fig. 1 shows the typical power-demand of the building sector. The general intrusion of heat in the summer is 13.3%, which is generated indoor by the human body, light bulbs, and electrical/electronic equipment. Consequently, occupancy and the related activities in a building are highly related with the energy consumption of that building. Any offline strategy for pre-defined control parameters is unable to handle all variations of building configurations, not to mention the large numbers of humans and their various behaviors. This barrier to energy savings, the effective use, control and interaction of the facilities in a building is often overlooked. To construct a smart home, the energy-related information may have to be presented in a semantic form [2] that can bridge the gap between the subjective sensation of the people and the control parameters for the system. It is possible to construct a smart-home knowledge base that makes the knowledge required for commissioning readily accessible. For instance, a sustainable comfortable thermal standard can be developed indoor temperature, adjustable by means of an adaptive index [3]. The usual approach for studying the relationship between the subjective sensation of active people in transitional spaces and the commissioning of these spaces for energy savings has always been to collect the environment variables and compute the statistical correlations [4].

2. Background: the evolution, challenges, and motivations of the building monitoring system

The main goal of a building management system (BMS) is to maintain the environmental conditions such as lighting, temperature, and air quality of that building. As shown in Fig. 2, the evolution of BMS can be divided into five categories, namely manual controlling, timer scheduling, sensor controlling, visual

☆ The corresponding research project has received the award of the investigation grants from Pan Wen Yuan Foundation in May 2013.

E-mail address: hcshih@saturn.yzu.edu.tw

http://dx.doi.org/10.1016/j.enbuild.2014.03.069
0378-7788/

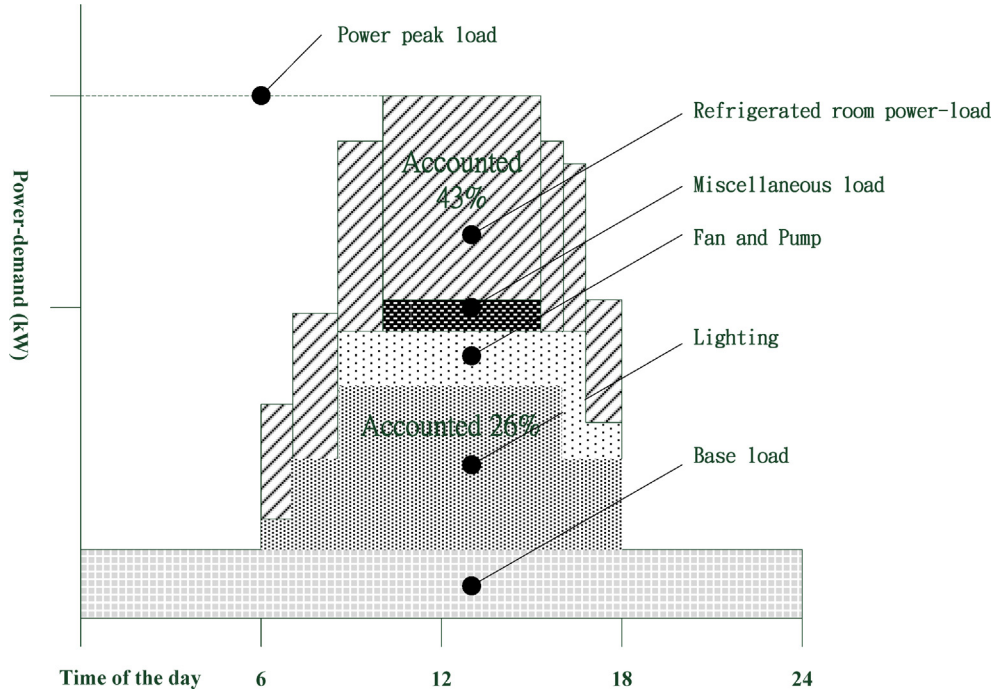


Fig. 1. Typical power-demand graph for commercial buildings in Taiwan.

recognition, and vision-based dynamic commissioning. Manual control is not only a waste of electricity, it is labor-intensive. Based on a clock timer, automatic control can be done following a pre-defined schedule, but it is both inefficient and inflexible. More advanced monitoring techniques use either a sensor-based, vision-based, or a dynamic vision approach. A camera network is used for large-scale field monitoring. It allows the BMS to be applied to a very wide field using online commissioning and dynamic hands-off devices, including cameras.

2.1. Sensor-based approaches

In recent years, the use of sensors for the reliable evaluation of the energy consumption of a building has attracted much attention [5–7]. Installing a sensing system throughout a building allows for a maintenance program with a higher degree of reliability and results in a faster commissioning process [8]. Dong et al. [9] presented a large-scale sensor network in a test-bed open office environment. They found that the characteristics of an open office plan, CO_2 and the acoustic parameters were all closely correlated with the number of occupants and the occupancy rate. Jang et al. [10] proposed a web-based system that would allow users to receive the deciphered data transmitted by the wireless sensor, including location of sensor, and the time of data acquisition. This information allowed engineers to easily monitor the conditions in and around buildings. In addition Dodier et al. [11] developed a statistical approach to the sensor network for the detection of building occupancy.

2.2. Static/dynamic model updating for vision-based presence detection

The sensor-based strategy is effective for monitoring a stable and closed environment. However, it cannot provide details and dynamic feedback on the appearance of the occupants and their activities. In addition, maintaining the sensing network of a large-scale environment is difficult. A great number of false alarms will occur due to the presence of animals or other moving objects.

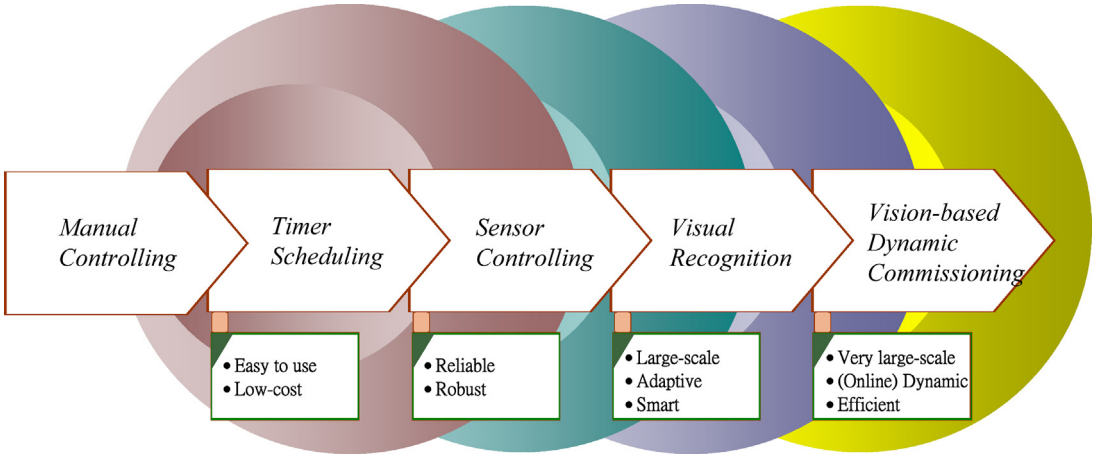


Fig. 2. Evolution of "building monitoring".

Unfortunately, there are no sensors that can provide the central monitoring system with the accurate path and activities showing a skeletal outline. Moreover, sensors suffer from failing to monitor people that are stationary, and they also tend to be sensitive to bright sunlight, both of which can lead to significant commissioning errors. The main challenge is to model the environment more precisely by applying more advanced sensors such as a charge-coupled device (CCD) and infrared cameras to deal with occupancy detection and object tracking. The issue of vision-based intelligent monitoring has received a lot of attention in recent years, including research into human action recognition and activity identification, hand gesture recognition, gait identification, fall detection and others. For example, a fixed CCD camera is employed in [12] to detect presence of humans and characterize their activity based on video analysis techniques such as change detection, moving objects tracking, and classification algorithms. Nevertheless, with a CCD camera the monitoring ability depends on the focus of view (FoV) of the camera. In order to fill the gap between the subjective sensing of people and system control parameters, we proposed a hybrid method which combines vision-based and sensor-based cameras.

In a cluttered scene it is necessary to first build the object model before we can start with human presence detection and tracking. Because moving objects often overlap each other, McKenna et al. [13] proposed an adaptive Gaussian mixture model for describing the color distribution of a moving object. Mittal et al. [14] developed a multiple-camera human tracking system. They proposed the so-called "presence probability" to distinguish each person in the scene. They then combined the presence probability with the color model to classify each pixel to one of the objects in the scene, allowing for the object to be segmented. Kang et al. [15] presented a method for detecting and tracking more than one person by means of occlusion handling. They used time weighted color information called "temporal color", which is a set of color values and their associated weight regarding the size, duration and frequency of appearance in the color region. Krumm et al. [16] adopted two sets of color stereo cameras for tracking multiple persons during a live demonstration in a living room. The stereo images were used to detect people, and the color histogram models were used to identify them. Tsutsui et al. [17] proposed a method that used multiple cameras to track a single target in a cluttered indoor scene. The motion information among the cameras allows the velocity and the 3-D position of the target object to be estimated. To cope with the probability density propagation for motion parameters in a non-Gaussian distribution, Isard et al. [18] demonstrated a stochastic sampling method called particle filtering (PF), a method that has been widely used in computerized object tracking. The BarMBLe system [19] applied the "human shape" model with the boundary constraints from the camera calibration process. However, none of these methods provide an object-specific representation and consequently human objects are likely to be confused when overlapped or occluded.

2.3. Monitoring a large-scale field using a PTZ camera network

To try and understand the wide range of field applications, Pan-Tilt-Zoom (PTZ) cameras were used in this work. PTZ cameras enable a monitoring system to observe the dynamic factor changes in an environment. Based on the walking activities of the occupants, the system can dynamically track and commission the control parameters. The PTZ camera is a programmable camera that is capable of changing the FoV with directional and zooming adjustment, providing a highly dynamic active range for an administrator to inspect the environment. Advanced PTZ cameras have a built-in firmware program for automatically monitoring any change in pixels in the FoV. This allows the camera to track a moving object at the best possible observable scale and keeps it in the center of the FoV.

In the development of the adaptive tracking algorithm, particle filtering (PF) was extensively investigated because its target tracking is based on a nonlinear and non-Gaussian model, and tracks the trajectory of an object using the pre-defined model in vector form. In Ref. [20], the authors proposed an improved PF tracking system based on different posture instants (e.g. tilt and rotation). This improved PF complete with a re-sampling algorithm was then proposed to overcome when the object is moving randomly. Choi et al. [21] discussed the architecture using two PTZ cameras to track a human object and collect a face image for the identification process. For camera network applications, it is necessary to calibrate the real world coordination between cameras. Some of the literature [22–24] utilized a rotating coordination system to obtain the desired homographic matrix. Chen and Wang [23] proposed a vision-based commissioning method based on the tilt angle and the 3D to 2D coordinate transformation of the FOVs of the PTZ cameras to determine the relative position and orientation of the objects in real world coordinate. Based on the inclination angle, the horizontal position in the real world, the internal parameters of the PTZ camera and the location of the object being monitored can be estimated.

2.4. Handing off control

To track the object with a single camera usually results in ambiguity due to occlusions and the angle of the point of view. Handing off control between multiple PTZ cameras plays an important role in PTZ camera networks. Lee et al. [25] proposed a handing off control method for multiple PTZ cameras. Dinh et al. [26] adopted the observation angle and the change in light conditions to handle real-time tracking and the associated control of the camera network. Geng et al. [27] proposed an adaptive feature fusion method to cope with changes in the environment. Similar to Ref. [27], Shakhnarovich and Darrell [28] proposed a system to first establish the outline of the persons and their direction of travel. Chen et al. [24] demonstrated a mapping algorithm between the panoramic camera and the PTZ cameras. They started by localizing the target using the panoramic camera, then pass that data on to the PTZ camera for obtaining a close-up view. The relative positioning and orientation between two PTZ cameras is based on a unified polynomial approximation model. Lu and Payandeh [29] performed the event detection and human tracking system using a CCD camera and a PTZ camera. The fixed CCD camera observed the entire field, and when it detected a human motion event, it would send the approximate position to PTZ camera. Then a color-based particle filter algorithm was applied for tracking the object. Because the location of the object is obtained by a fixed CCD camera, the tracking region is restricted to the FoV of the camera. Chang et al. [30] proposed a real-time tracking system, applying a PTZ camera and a fixed camera. Their system used a modified mean-shift algorithm to track multiple colors of the target. The template matching method was then employed to identify the target, and they applied the motion history image (MHI) technique for dynamic object tracking. The MHI technique provides an efficient representation of moving objects, however, it does not work when the object remains stationary. In addition, when the color of the object is similar to the background, the same error occurs. Qureshi et al. [31] developed a planning framework for controlling the active PTZ cameras in order to acquire a seamless close-up video of pedestrians of interest when they move through a pre-defined region and enter and exit the observational range of different cameras. This approach enables the monitoring of over 1000 pedestrians and computes the optimal arrangement among objects and cameras. Krahnstoever et al. [32] comprised four fixed cameras and four PTZ cameras in a one-by-one mapping manner. To use the fixed CCD camera for detecting a target object in advance, the system determines an optimal planning

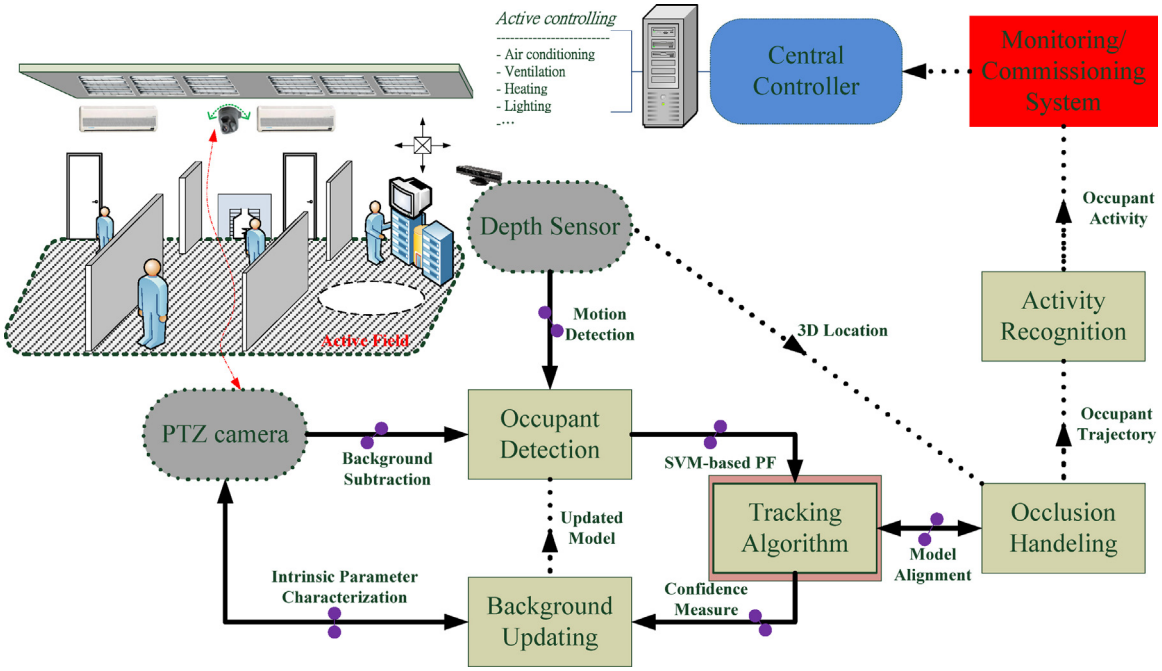


Fig. 3. Flow diagram and virtual scenario of the intelligent building monitoring system.

strategy, which is then send complete with control parameters to the corresponding PTZ for tracking.

2.5. Motivations and contributions

The majority of past researches in building monitoring used presence detection and activity recognition. The above subsections comprehensively reported the relevant reviews for occupancy detection and tracking using a PTZ camera network and a handing off control mechanism to handle the problem of continuous tracking in a large-scale field. Unlike the aforementioned studies of the vision-based system, the present study aims to monitor a building for 24 h commissioning. Based on calibrated and synchronized depth sensors and PTZ cameras, the proposed monitoring system allows to continuously track the occupants even under dim-lighting conditions. By using the new SVM-based observation measurement, it provides a robust day-and-night occupant tracking and counting performance. To realize large-scale field monitoring, programmable PTZ cameras are used instead of conventional fixed CCD cameras. In addition, contour information of the occupant can be applied for activity recognition.

In summary, the contributions of this paper are as follows:

(1) A comprehensive survey, categorizing the evolution of the various techniques for building monitoring.
(2) A robust smart building management system is proposed, employing a PTZ camera network that is suitable for large-scale field commissioning during both daytime and nighttime.
(3) The proposed well-defined tracking algorithm with optimal observation model and refinement is capable of solving partial occlusion and other issues when the camera moves.
(4) The prototype of the building monitoring system is presented, characterizing the parameters of any activity, including trajectory, action behaviors, and human shape parameters.

3. System framework

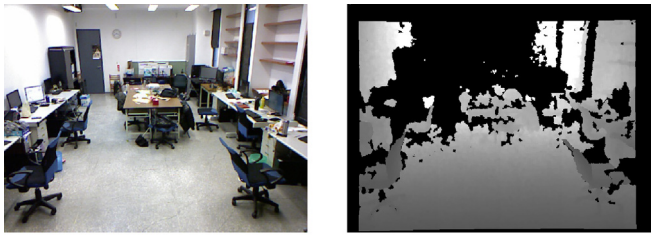
Detecting human presence and tracking techniques are the cornerstones of building monitoring and commissioning. As shown in Fig. 3, we applied two types of sensing device, namely the PTZ camera and the depth sensor. First, occupant detection can be performed using a refined background subtraction model. The depth sensor helps the system to rapidly detect any object motion and assists with dim lighting conditions. Second, a novel tracking algorithm provides a robust scenario when the camera moves. Third, the occlusion handling algorithm deals with the actions of the object. Trajectory and shape features are utilized to recognize the behavior of the object. Finally, we discuss the active control of the proposed smart building management system.

Particle filtering [18,19] is a powerful method for dealing with nonlinear and non-Gaussian visual tracking problems, and is sensitive to the propagation mechanism of the probability densities of the particles. Instead of a histogram-based similarity measure, we modified the support vector machine (SVM) classifier [33,34] to improve the efficiency of the probability density propagation. We also improved the localization for any object considered significant, and enhanced the tracking results for occlusion handling in a cluttered background.

4. Occupant presence detection and tracking

The conventional object presence detection algorithm is based on a pre-trained background model. Background subtraction is then used to segment the moving object. With the rapid development of the motion detection technique, we can expect that the vision capturing system will be enable to extract accurate human presence information in the very near future. However, the standard CCD cameras are unable to offer sufficient scene information in a dim-lighting environment, suffering from a number of difficulties in the background substation algorithm, including

(1) Shadows: shadows are usually classified as the foreground region due to their similar hue values.
(2) Frequent illumination changes: changes in illumination completely alter the color characteristics of the background, including the reflection of sun and flashlights. This results in



(a) (b)

Fig. 4. Background models of (a) the PTZ camera and (b) the depth camera.

a drastic increase in the number of false foreground regions being detected.

(3) Initialization with moving objects: in initialization mode, the presence of moving foreground objects is partially occluded by the background object.

To cope with these problems, an infrared depth camera can be a good solution. This is a costly solution, and therefore may impede its usability. Recently, the Kinect sensor was developed by Microsoft Corporation for the purpose of entertainment, which is an input device for XBOX 360 for capturing the parameters of human motion. In this study, we applied the Kinect sensor to obtain the depth information to improve the accuracy of the occupant presence detection and the efficiency of the tracking algorithm.

4.1. Presence motion detection with an updated dynamic model

In this study, the observation distribution is modeled by the multiple kernel of the probabilistic color model. The color model of the occupant is approximated by three Gaussian kernels. The Gaussian Mixture Model (GMM) is built by the mean and the variance, defining the observational likelihood on each of the three color components (i.e., HSI in this paper). Furthermore, each of the color components is defined as a single Gaussian distribution.

4.1.1. Background modeling

The initial process of human motion analysis is extracting the target object from the scene. Background subtraction is a well-developed algorithm to extract the object, with the main idea of subtracting the current video frame from the reference background model. The pixels of each frame in the video sequence are categorized into background pixels and foreground pixels, based on the statistical background model. Pixels with a large deviation from the background model are treated as foreground pixels. In this paper, the process of background modeling is not only employed in the depth camera sensor, but also in the PTZ camera. It should be noted that the PTZ camera moves based on the environmental conditions, and thus the background model needs to be updated periodically. Fig. 4 shows the results of the background images, which are used for presence detection under dim-lighting conditions.

4.1.2. Comprehensive motion detection

To initialize the proposed system, we start with a coarse searching mechanism, which simply is an exhaustively sampling and uniform picking-up of the particles in the frame. The state dynamic model in time instant t is defined as $x_t = Ax_{t-1} + v_t$, where v_t denotes a zero mean Gaussian random variable with variance σ_t, that is $v_t = G(0, \sigma_t)$, and the matrix A is defined manually and assumes that the object is moving at a constant velocity. Variance σ_t is adaptively adjusted, as well as being inversely proportional to the measurement score in the observation process with respect to the previous mean-state $\hat{\rho}_{t-1}$. Thus $\sigma_t = \sigma_{\max} \times (1 - \hat{\rho}_{t-1})$, where $\sigma_{\max}$ denotes the largest search range. When the occupant object is unable to be tracked successfully, the sampling range of the particle will be enlarged. Thus, when $\hat{\rho}_{t-1}$ is enhanced, we consider that the object was found and tracked well, and thus the search range will be shrunk. This operation can also be used to recover an object when it is lost. When the search range is continuously increasing, this indicates that the object may be lost or that it is occluded. In this case, the tracker will search for the object over the whole frame in the "initialization mode" which is switched to dynamically and supplies the necessary information for the tracking phase.

4.2. Robust tracking algorithm using the SVM-based observation model

The main idea of the particle filtering (PF) algorithm is to represent an incomplete probability density function (PDF) of the state for a set of weighted samples, called particles. If the number of particles is large, it indicates that the actual PDF can be observed. However, the size of the sample directly affects the computational cost, and should be kept as small as possible in order to preserve the efficiency of the PF algorithm.

4.2.1. Foundations of particle filtering

The PF algorithm requires two probabilistic models: the state model and the observation model. The state model describes the evolution of the system from its past state, while the observation model compares the current state observations with the current state of the system. Based on the observation model, the system measures the weight of each particle. Therefore, the measurement of the observation model is critical for updating the prior density. The PF-based visual tracking approach is mainly based on the color-based histogram similarity for achieving a measurement. Compared to using edge feature, using the color features is both more stable and distinguishable. To enhance the distinguishing ability of the observation measurements, the optimal classification algorithm in machine learning, support vector machine (SVM), was employed in this study. Here, we present a systematic method for occupant tracking, using the SVM-based discrimination method in the spatiotemporal domain. It is not only very capable of recovering a lost object, but it also substantially accelerates the convergence speed in a cluttered scene.

Mathematically, PF approximates the posterior density using a discrete random measure which is composed of the particles and their weights, that is $S_t = \{(s_t^{(n)}, \pi_t^{(n)})\}_{n=1}^{N_s}$, where N_s denotes the number of particles used in the approximation, and $\sum_{n=1}^{N_s} \pi_t^{(n)} = 1$. Assume that the distribution of interest is $\tilde{p}(x_t)$ and its approximating random measure can be expressed as $\tilde{p}(x_t) \approx \sum_{n=1}^{N_s} \pi_t^{(n)} \delta(x_t - s_t^{(n)})$, where $\delta(\cdot)$ denotes the Dirac delta function, each particle is then weighted in terms of the observations. Then N_s samples are drawn and replaced by choosing a particular sample with corresponding weights,

$$\pi_t^{(n)} = \frac{p(z_t|x_t = s_t^{(n)})}{\sum_{n=1}^{N_s} p(z_t|x_t = s_t^{(n)})}, n = 1, 2, \ldots, N_s. \quad (1)$$

Finally, the estimation of the object-state at time instant t can be computed by the weighted mean over all sample-states as follows:

$$\tilde{p}(x_t) \approx E[S_t] = \sum_{n=1}^{N_s} \pi_t^{(n)} s_t^{(n)}. \quad (2)$$

4.2.2. Likelihood density propagation

To cope with the problem of tracking a nonlinear non-Gaussian object in a cluttered background, we first consider the evolution of the state sequence $\{x_t, t \in T\}$ in a dynamic system which is formulated as $x_t = f_t(x_{t-1}, u_{t-1})$ in the period of T. The state varies in time following the *system function* f_t which describes a Markov process driven by the i.i.d. process noise sequence $\{u_{t-1}, t \in T\}$. The

tracking process consists of recursively estimating the observation z_t from x_t with the *measurement function* h_t, that is $z_t = h_t(x_t, v_t)$, where $\{v_t, t \in T\}$ denotes the white *measurement noise* sequence. The goal of the PF algorithm is to approximate the posterior density $p(X_t|Z_t)$ based on the observation density $p(Z_t|X_t)$ and the predictive density $p(x_t|Z_{0:t-1})$. Consequently it can model the uncertainties of observation and incomplete priors. Let Z_t denote the state of the occupant for the history of the observations up to time instant t, i.e., $Z_t = \{z_1, \ldots, z_t\}$. Therefore, the desired distribution $p(x_t|Z_{0:t})$ contains all the information regarding the target and the effect of the current observations in time instant t. First *Bayes'* rule is applied to determine the posterior density from the predictive density $p(x_t|Z_{0:t-1})$ and current observation z_t, as follows.

$$\begin{aligned} p(x_t|Z_{0:t}) &= \frac{p(z_t|x_t, Z_{0:t-1})p(x_t|Z_{0:t-1})}{p(z_t|Z_{0:t-1})} \\ &= kp(x_t|Z_{0:t-1})p(z_t|x_t, Z_{0:t-1}) \end{aligned}, \tag{3}$$

where k is a normalization factor. Because the measurement at time instant t is independent from the previous measurements, we get $p(z_t|x_t, Z_{0:t-1}) = p(z_t|x_t)$. Therefore, the distribution $p(x_t|Z_{0:t})$ can be written as $kp(x_t|Z_{0:t-1})p(z_t|x_t)$. Based on the Bayesian sequential estimation, the prior density function $p(x_t|Z_{0:t-1})$ can be computed by the following two recursive stages:

$$\text{Prediction stage}: \quad p(x_t|Z_{0:t-1}) = \int p(x_t|x_{t-1})p(x_{t-1}|Z_{0:t-1})dx_{t-1}, \tag{4}$$

$$\begin{aligned} \text{Updating stage}: \quad p(x_t|Z_{0:t}) &= \frac{p(z_t|x_t)p(x_t|Z_{0:t-1})}{p(z_t|Z_{0:t-1})} \\ &= \frac{p(z_t|x_t)\int p(x_t|x_{t-1})p(x_{t-1}|Z_{0:t-1})dx_{t-1}}{\int p(z_t|x_t)p(x_t|Z_{0:t-1})dx_t}. \end{aligned} \tag{5}$$

In the prediction stage we use the previous posterior distribution $p(x_{t-1}|Z_{0:t-1})$, and transition probability $p(x_t|x_{t-1})$ to calculate the predictive distribution $p(x_t|Z_{0:t-1})$. In the updating stage, the new posterior distribution $p(x_t|Z_{0:t})$ is obtained from a direct consequence of *Bayes'* rule, which is computed by the predictive distribution $p(x_t|Z_{0:t-1})$ and the observation likelihood $p(z_t|x_t)$. Overall, the *prediction* and *updating* stages are applied alternately at each time step.

4.2.3. Observation model with online learning

Support vector machine (SVM), based on the statistical learning theory (SLT), was derived by Vapnik [33]. Unlike other machine learning algorithms, the SVM are trained by solving a constrained quadratic optimization problem using Lagrange multipliers to derive a unique optimal solution for each choice of the SVM parameters. In this study we applied SVM for measuring the observation density, which estimates the data likelihood (similarity) between target model and testing feature vector. The SVM was originally designed for binary classification and it finds a discriminating hyper-plane by maximizing the margin between two classes. The margin is defined as the distance between the closest points of two classes.

Let $S = \{(p_i, l_i), i = 1 \ldots N \text{ and } p_i \in \Re^d\}$ denote a set of samples. Each sample p_i with the label $l_i \in \{-1, +1\}$, where $l_i = +1$ denotes an object sample and $l_i = -1$ denotes a non-object sample. To solve a constrained optimization problem using quadratic programming, the separating hyper-plane can be represented as a linear combination of the training examples and testing feature vector $\bar{p}$. Therefore, the SVM score function can be defined as:

$$\varphi(\bar{p}) = \sum_{i=1}^{N} \alpha_i l_i k(\bar{p}, p_i) + b \tag{6}$$

where $k(\cdot)$ denotes a kernel function, v_i denotes the support vector, l_i denotes its sign, α_i denotes its Lagrange multipliers, with bias b. The sign $\varphi(\bar{p})$ determines the class membership of $\bar{p}$. The linear kernel $k(\bar{p}, p_i) = \langle \bar{p}^T \cdot p_i \rangle$ is applied in the measurement process.

Referring to (1), $\pi_t^{(n)}$ describes a possible object-state weight, where $\pi_t^{(n)} \propto p(z_t|x_t = s_t^{(n)})$. In this paper, the observation model is based on score function $\varphi(\cdot)$ in the measurement step. By using (6), the observation likelihood can be expressed as:

$$\pi_t^{(n)} \propto p(z_t|x_t = s_t^{(n)}) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(\varphi^2(s_t^{(n)}))/2\sigma^2} \tag{7}$$

The samples evolve over time due to the re-sampling mechanism which is proportional to the weight distribution and the dynamic state model. In each time step, we estimate the occupant object location (u_t, v_t) and add it to the trajectory. Therefore the estimated object-state $\tilde{p}(x_t)$ at time instant t can be calculated by the weighted mean over all sample-states as:

$$\tilde{p}(x_t) = \sum_{n=1}^{N_s} \tilde{\pi}_t^{(n)} x_t^{(n)}. \tag{8}$$

5. Simulations

This paper proposed a building monitoring and commissioning system which can be divided into three modules: (1) Occupancy detection; (2) People tracking and counting; (3) Action recognition. The proposed method is suitable for both daytime and nighttime applications. To demonstrate the usability and adaptability of the proposed system diverse environmental conditions, the experiments were conducted indoor as well as outdoor. In practice, the modules of occupant detection, tracking, and counting can be successfully applied to monitor a building.

5.1. Installation

As shown in Fig. 5, the Kinect sensor was deploy in the corner of the floor to capture the occupants in the FoV. The Kinect sensor utilized RGB video streaming with 8-bit VGA resolution and captured the video from the infrared sensor at a resolution of 640 × 480 pixels @30 Hz. The sensor had an angular FoV of 57° in the horizontal view and 43° in the vertical view. Its motorized pivot was capable of tilting the sensor at 27° either up or down. It should be noted that the working distance of a Kinect sensor is limited, and can only maintain the tracking algorithm from 0.7 to 4 m. Based on the received raw infrared signal, we extended the operating distance to 8 m. We utilized the PTZ camera to compensate for this limitation, so that the system had a more reliable occupant motion detection and tracking capability.

5.2. Robust occupancy detection for 24 h a day scenario

A robust occupancy detection system is needed with 24 h a day functionality. Most research has focused on daytime applications using an active camera network, since nighttime applications are more difficult in a vision-based monitoring system. A conventional system uses the background subtraction method to find the moving occupant. By using a pre-constructed background model, we can efficiently and precisely determine the region the person of interest is in. In our occupancy detection module the depth background model used background subtraction. Fig. 6 shows an indoor scenario, however, the flickering of the fluorescent lighting condition of the field is troublesome for a CCD camera, and because this light source does provide for homogeneous light emission, the background subtraction approach would have been easy to fail. The depth-based background subtraction however resulted in a high

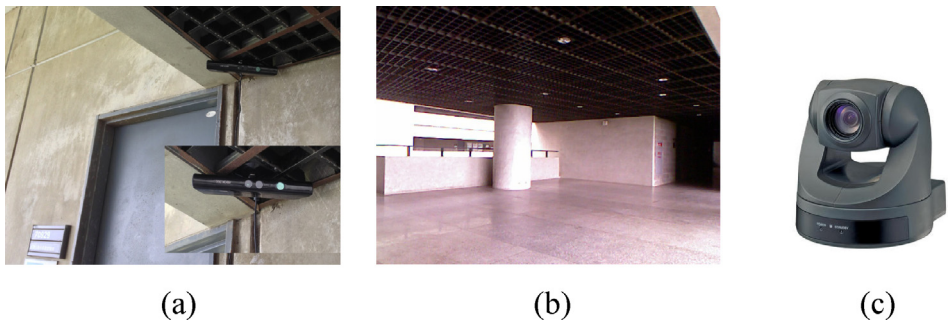


Fig. 5. Deployment of the Kinect sensor, (a) Kinect setup, (b) focus of view, (c) PTZ camera.

level of reliability and adaptation. As shown in Fig. 6(f), the images underwent a filtering operation to reduce false alarms. Occupants could be precisely extracted and tagged with a unique minimum-bounding box (MBB). Normally, the MBB tends to overlap due to the angle-of-view of the cameras. In the present study, the central pole of the occupant (i.e., the vertical line in Figs. 6(d)–(f)) was used to handle the occlusion problem. According to the depth information, we can determine which occupant should be labeled in the occluded region. Another advantage of the depth camera is that, when an occupant is stationary, its presence is still detected.

As shown in Fig. 7, in the outdoor scene, the background model is invisible due to the limits of the depth sensor. However, the regions of the occupants are obtained by subtracting the test images from the depth background image. Because the outdoor field is large, this method helps the system to easily distinguish identities among the occupants. At the same time, because the occupants are usually in an upright position, the aspect ratio can also be used as a significant feature for recognizing their activities.

Fig. 8 shows an example of dim-lighting condition. Because depth background subtraction is applied, occupant detection works well, contrary to the conventional CCD camera that only works

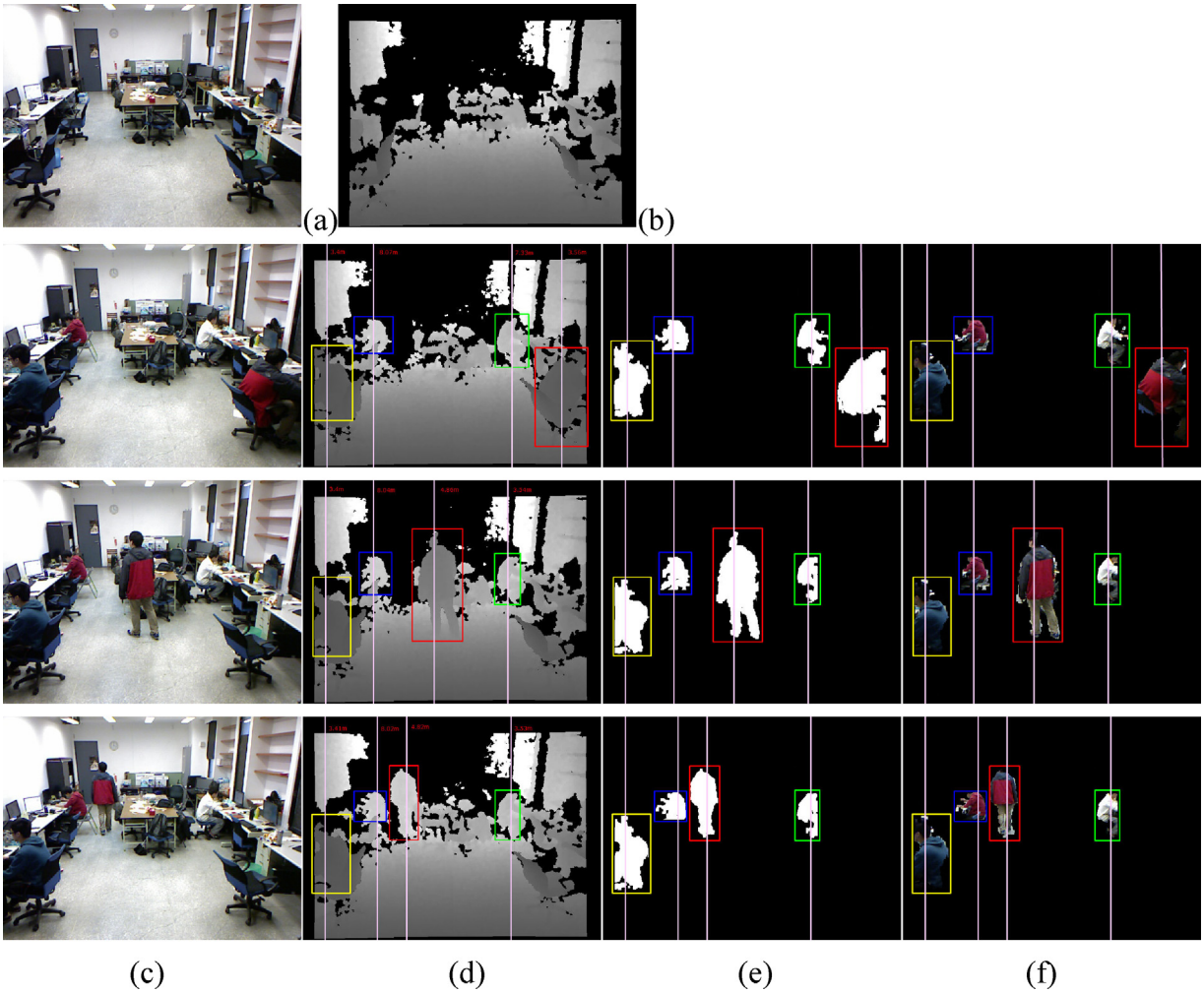


Fig. 6. The indoor scenario for occupant detection. (a) The RGB background model, (b) the depth background model, (c) the test images, (d) the depth images, (c) the subtracted images, (d) the resulting images.

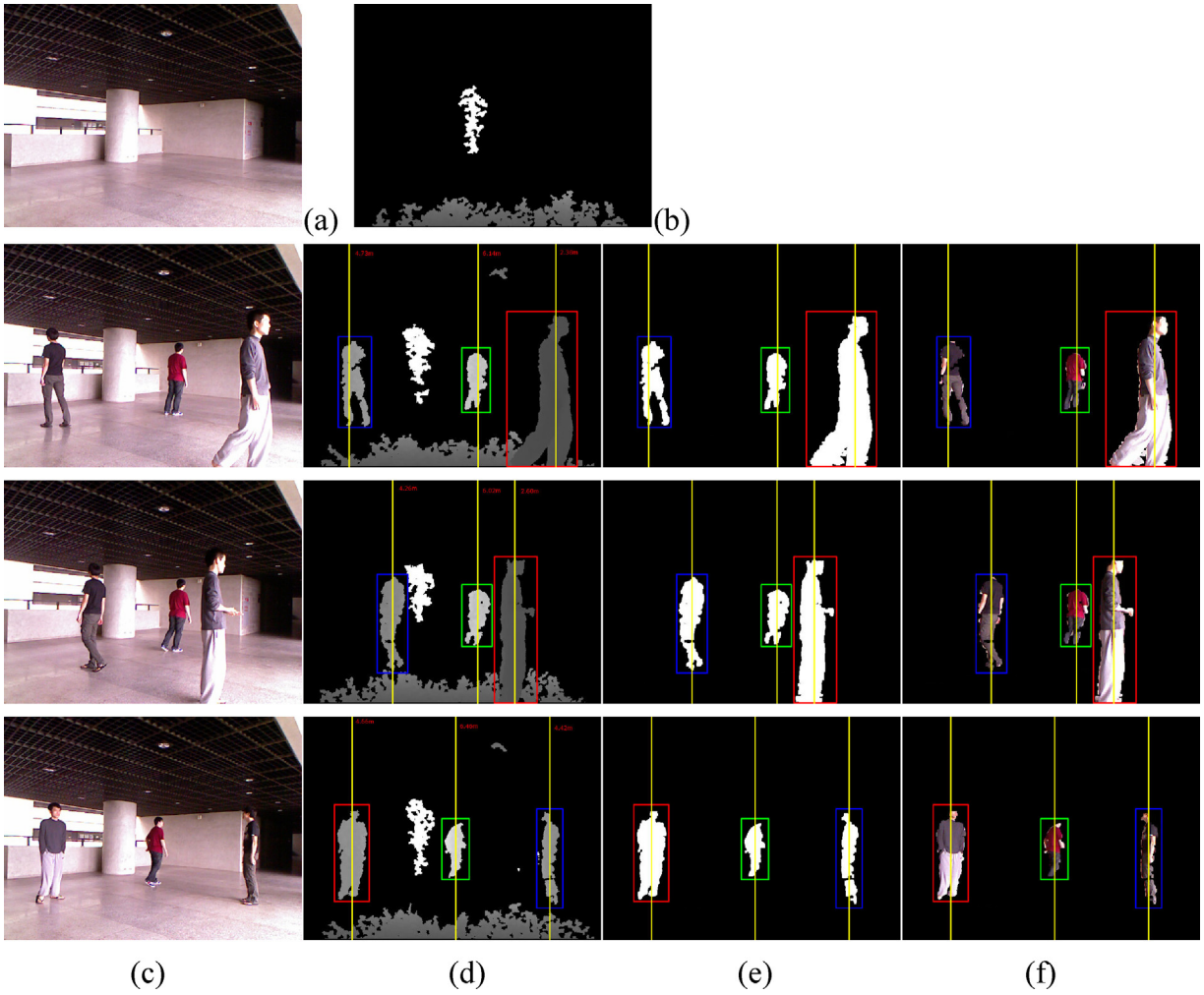


Fig. 7. The outdoor scenario for occupant detection. (a) the RGB background model, (b) the depth background model, (c) the test images, (d) the depth images, (c) the subtracted images, (d) the resulting images.

well in a high illumination environment. As shown in Fig. 8(f), when lighting conditions are unstable due to the flickering of the fluorescent lighting, objects are in a shadow, or a shiny floor throws a reflection, the MBBs of the occupants still available. This demonstrates that even when the view of the occupants is partially obstructed or not visible and the CCD camera has difficulty to restore the original appearances, occupancy detection never becomes a problem in our system.

5.3. Application for tracking and counting people

5.3.1. Occlusion handling

Monitoring objects with a single camera usually results in ambiguity due to occlusion and the angle of view-point. Handing off control between multiple cameras plays an important role in handling occlusions. However, the vision-based tracking algorithm is unable to cope with stationary objects when the background model gets updated, since the MHI method is only suitable for moving objects. In practice, people that live or work in a building do not move all the time. As a result, any vision-based process will yield frequent errors in object detection. In order to handle the occlusion problem, we employed a depth sensor in our system to acquire the presence of the object and any information regarding its motion.

Fig. 9 shows the different scenarios of the weight adjustment between Kinect sensor and PTZ camera. The Kinect sensor and the PTZ camera are deployed in a right angle to each other. The weighting factor provides the confidence in the captured data to the occlusion handling module. By applying the weighting factors the observation model is modified, and the object tracking obtained is robust. Generally speaking, weight adjustment is classified into four scenarios:

(1) Monotonous approach: As shown in Fig. 9(a) and (b), when the occupant walks toward the Kinect sensor, the weight of the PTZ camera will reduce, while if the occupant walks toward the PTZ camera, the weight of the Kinect sensor is reduced.
(2) Occluded in the beginning or at the end: When two occupants are overlapped in the view of the PTZ camera, it results in a detection fault because an unobservable object appears. In this situation, the Kinect sensor can supply information to the PTZ camera regarding appearance, shape, motion, and the actual distance between the two occupants.
(3) Occluded somewhere in between: In the tracking process, occupants may cross each other in an intermediate location, and in that case the depth information from the Kinect sensor becomes very critical. The occlusion handling module enables the system to recognize the occluded occupant whether walking toward or away from the PTZ camera. The same holds true for the FoV of the PTZ camera, whether the preceding occupant is walking toward or away from the PTZ camera.

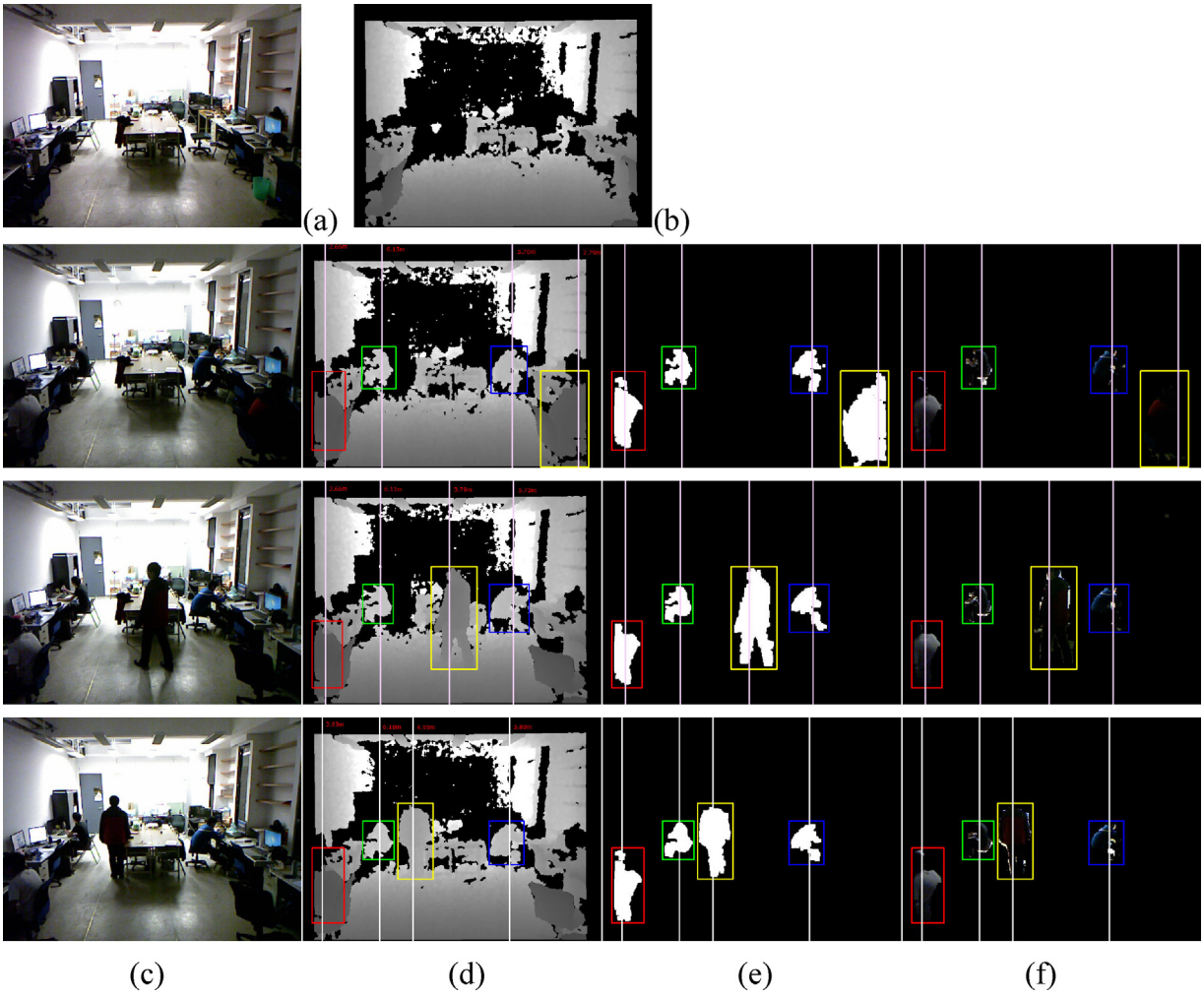


Fig. 8. Indoor scenario in a dimly lit scene. (a) the RGB background model, (b) the depth background model, (c) the test images, (d) the depth images, (c) the subtracted images, (d) the resulting images.

5.3.2. People tracking

In the monitoring system, the trajectory of people walking is one of the most important pieces of information for computing the energy scheduling for smart home applications. The test environment of our experiment was located on a corner site on the 9th floor of a building. Fig. 10(c)–(e) shows three trajectory maps constructed from three full days. It is obvious that there are four major trajectories. In the right-hand side of the test environment is a balcony. A trajectory was created by people walking to the balcony for a panoramic view of the city below. People entering and exiting the laboratory were responsible for two trajectories. The remaining trajectory was from people walking through the corridor. In Fig. 10(a)–(b), the MBBs of the occupants can be detected, and the color of the MBB represents their identity. Trajectories are generated by connecting the centroids of the MBBs.

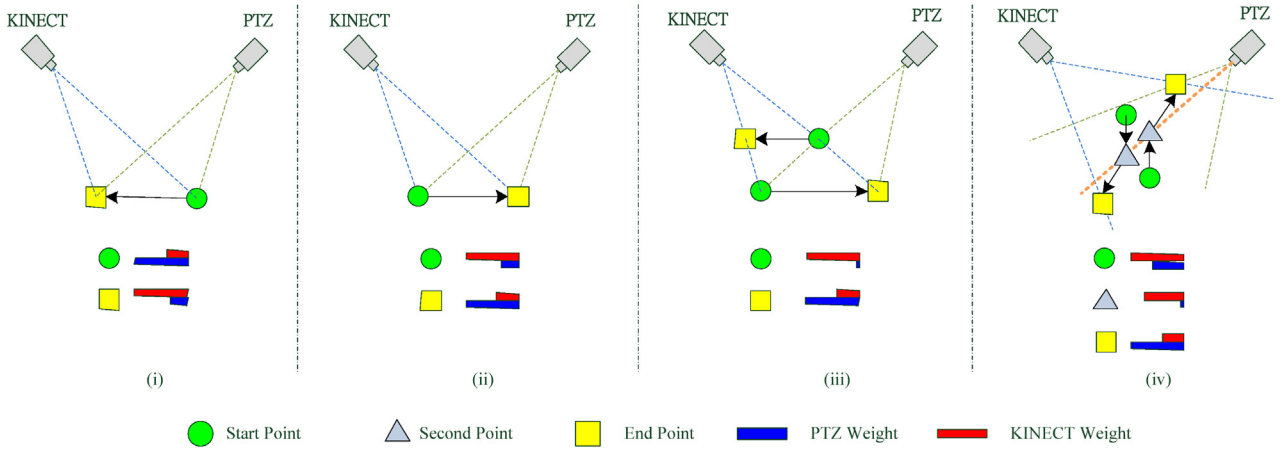


Fig. 9. Different scenarios for weight adjustment.

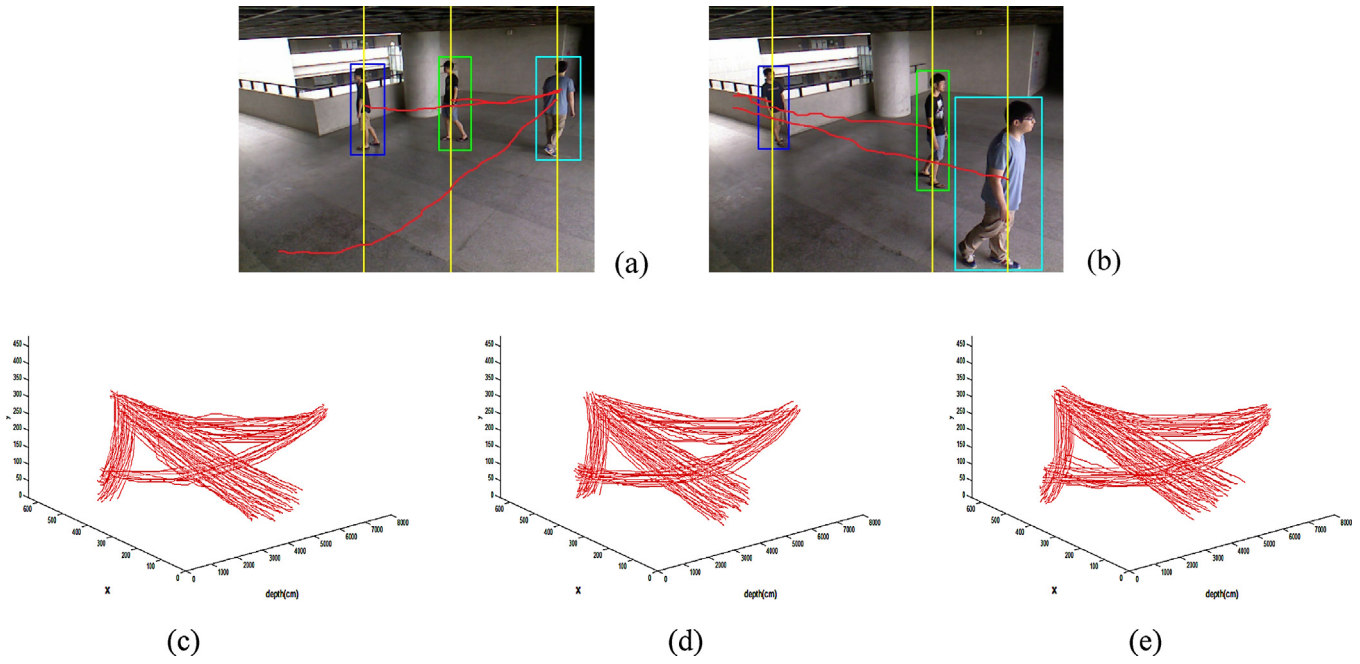


Fig. 10. People trajectory construction, (a, b) original video frames, the trajectory captured in (c) 0:00am–8:59am, (d) 9:00am–4:59pm, (e) 5:00pm–11:59am.

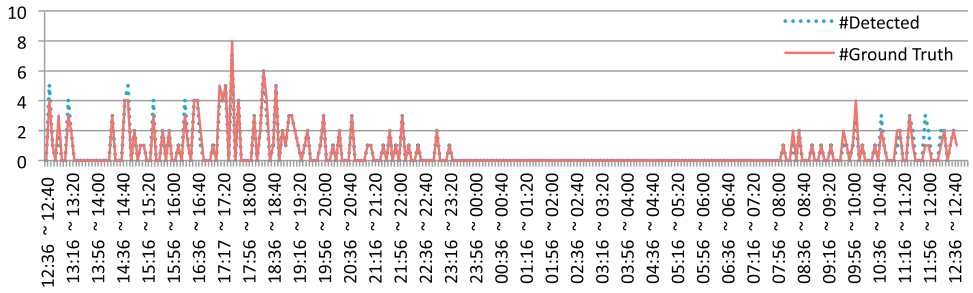


Fig. 11. Results of people-counting for testing and for collecting ground-truth data over a 24 h period (on May 28, 2013).

5.3.3. *People counting*

In this experiment, we captured three video sequences over 24 h. Because the minimum requirement of the receiving energy for activating the sensor cell is constant, a CCD camera is unable to capture a clear video at night. If there is no sufficient illumination, the camera sensor is unable to create a digital signal. As a result, it is difficult to obtain a good video sequence in a dim environment, and a conventional CCD camera device usually fails. Fig. 11 shows the results of the people-counting. We manually labeled the ground truth data of the number of people present. The depth background model was constructed in advance. Using the background subtraction algorithm enables the system to identify the moving people in real-time. Based on our observation, it appears that there were no occupants from 00:30am to 07:30am. The largest number of people walked through the testing field between 04:30pm and 06:00pm. However, there were some false alarms due to sunlight confusing the receiver of the Kinect device. The background model was updated every minute. Based on the results, we can conclude

Table 1
Comparisons of the system functionalities and features among related work.

	Robust to nonhuman objects	Support tracking	Dimly lit scene	Provide real-world coordination	Features	Sensor deployed	Scale of monitoring
[8]	N	N	Y	N	Temperature monitoring; Sensor overlay system	Temperature sensor, BMS and overlay sensors.	Large (with more sensors)
[11]	N	N	Y	N	Probabilistic inference; Low cost	Passive Infrared, handset sensor	Large (with more sensors)
[12]	Y	Y	N	N	Vision-based system; Activity characterization	Static camera	Median
[30]	N	Y	N	N	Template matching; Motion history image	Single PTZ camera	Large
[32]	Y	Y	N	Y	Multi-camera multi-target tracking; Best-first estimation	PTZ cameras	Large
Ours	Y	Y	Y	Y	Vision-based system; Robust tracking algorithm; 24 h a day monitoring	Kinect sensor, PTZ cameras	Large

that our system is robust for counting people in all-day applications, since the amount of people that were counted by the system was very close to the ground truth.

5.3.4. System comparisons

In this section we compare the functionality and feature to published state-of-the-art systems. Table 1 shows the critical properties of the monitoring and commissioning system: (1) does the system robust to nonhuman objects to avoid the false alarm? (2) Does the system support the tracking function? (3) Does the system remain valid in dimly lit scene? (4) Does the system provide the real-world coordination? (5) What is the featured functionality of the system? (6) What types of the sensors are deployed? (7) How large of the field can be monitored. Compared to recently published results we achieved the best system ability. Note that when the image-based sensor used, the nonhuman objects such as the cats and dogs can be recognized. When the infrared-based sensor deployed, we can reach the object tracking with more reliabilities and enable to monitor the dimly lit scene. Because the Kinect sensor deployed in our system, we can obtain the depth information of the scene. As a result, the real-world coordination can be extracted and supported for further occupant positioning and activity recognition applications.

6. Conclusions and future work

This study aimed to develop a system to monitor buildings for 24 h a day commissioning. Using image-based depth sensors and programmable PTZ cameras, the proposed system enables continuous detection and tracking of the occupants even under low-light conditions. The proposed SVM-based observation measurement provides a more dependable tracking performance. A robust day-and-night people tracking and counting algorithm has been presented in this paper. The function of monitoring a large-scale field can be realized using a PTZ camera network instead of conventional fixed cameras. In future work, the shape and skeletal features of the occupant could be used for more accurate action recognition. It would be useful to utilize body configuration as a control factor in the analysis of the energy consumption of air-conditioning. Gait recognition technique will enable the system to calculate the heat loss into the atmosphere by a person. The activities of the occupant, such as standing, sitting, walking, and running should be taken into consideration as per the fluid dynamics theory. In addition, the walking speed, frequency of arm swinging, and the size of a group of people could affect the commissioning parameters of a building monitoring system.

Acknowledgements

The author would thank Mr. Meng-Hsuang Chang for setting up the camera and collecting occupancy data. This work is supported in part by the Ministry of Science and Technology of Taiwan, under grant NSC102-2221-E-155-037-MY2.

References

[1] IEA/OECD Key World Energy Statistics, 2012 edition. [Online] http://www.iea.org/pub- lications/freepublications/publication/kwes.pdf (accessed 15.10.13).
[2] M.J. Kofler, C. Renisch, W. Kastner, A semantic representation of energy-related information in future smart homes, Energy and Buildings 47 (2012) 169–179.
[3] J.F. Nicol, M.A. Humphreys, Adaptive thermal comfort and sustainable thermal standards for buildings, Energy and Buildings 34 (2002) 563–572.
[4] N. Ghaddar, L. Ghali, S. Chehaitly, Assessing thermal comfort of active people in transitional spaces in presence of air movement, Energy and Buildings 43 (2011) 2832–2842.
[5] Q. Dong, L. Yu, H. Lu, Z. Hong, Y. Chen, Design of building monitoring systems based on wireless sensor networks, Wireless Sensor Networks 2 (2010) 703–709.
[6] B. Krausse, M.J. Cook, K.J. Lomas, Environmental performance of a naturally ventilated city centre library, Energy and Buildings 9 (7) (2007) 792–801.
[7] R.J. Meyers, E.D. Williams, H.S. Matthews, Scoping the potential of monitoring and control technologies to reduce energy use in homes, Energy and Buildings 42 (2010) 563–569.
[8] B. Painter, N. Brown, M.J. Cook, Practical application of a sensor overlay system for building monitoring and commissioning, Energy and Buildings 48 (2012) 29–39.
[9] B. Dong, B. Andrews, K.P. Lam, M. Hoynck, R. Zhang, Y.-S. Chiou, D. Benitez, An information technology enabled sustainability test-bed (ITEST) for occupancy detection through an environmental sensing network, Energy and Buildings 42 (2010) 1038–1046.
[10] W.-S. Jang, W.M. Healy, M.J. Skibniewski, Wireless sensor networks as part of a web-based building environmental monitoring system, Automation in Construction 17 (2008) 729–736.
[11] R.H. Dodier, G.P. Henze, D.K. Tiller, X. Guo, Building occupancy detection through sensor belief networks, Energy and Buildings 38 (2006) 1003–1043.
[12] Y. Benezeth, H. Laurent, B. Emile, C. Rosenberger, Towards a sensor for detecting human presence and characterizing activity, Energy and Buildings 43 (2011) 305–314.
[13] S.J. McKenna, Y. Raja, S. Gong, Tracking color objects using adaptive mixture models, Image and Vision Computing 19 (1999) 225–231.
[14] A. Mittal, L.S. Davis, M2tracker. A multi-view approach to segmenting and tracking people in a cluttered scene, International Journal of Computer Vision 51 (3) (2003) 189–203.
[15] S. Kang, B. Hwang, S.W. Lee, Multiple people tracking based on temporal color feature, International Journal of Pattern Recognition and Artificial Intelligence 17 (6) (2003) 931–949.
[16] J. Krumm, S. Harris, B. Meyers, B. Brumitt, M. Hale, S. Shafer, Multi-camera multi-person tracking for EasyLiving, in: Proceedings of the IEEE International Workshop on Visual Surveillance, 2000, pp. 3–10.
[17] H. Tsutsui, J. Miura, Y. Shirai, Optical flow-based person tracking by multiple cameras, in: Proceedings of the International Conference on Multisensor Fusion and Integration for Intelligent Systems, 2001, pp. 91–96.
[18] M. Isard, A. Blake, CONDENSATION-conditional density propagation for visual tracking, International Journal Computer Vision 29 (1) (1998) 5–28.
[19] M. Isard, J. MacCormick, BraMBLe: Bayesian multiple-blob tracker, Proceedings of the IEEE International Conference on Computer Vision 2 (2001) 34–41.
[20] T. Zhang, S.M. Fei, Improved particle filter for object tracking, in: Proceedings of the IEEE Chinese Control and Decision Conference (CCDC), 2011, pp. 3586–3590.
[21] H.-C. Choi, U. Park, A. Jain, PTZ Camera assisted face acquisition, tracking and recognition, in: IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS10), 2010.
[22] M.A. Haj, A.D. Bagdanov, J. Gonzàlez, F.X. Roca, Reactive object tracking with a single PTZ camera, in: Proceedings of the IEEE Pattern Recognition (ICPR), 2010, pp. 1690–1693.
[23] I.-H. Chen, S.-J. Wang, Efficient vision-based calibration for visual surveillance systems with multiple PTZ cameras, in: Proceedings of the IEEE Computer Vision Systems (ICVS), 2006.
[24] C.-C. Chen, Y. Yao, A. Drira, A. Koschan, M. Abidi, Cooperative mapping of multiple PTZ cameras in automated surveillance systems, in: Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR), 2009, pp. 1078–1084.
[25] C.-Y. Lee, S.-J. Lin, C.-W. Lee, C.-S. Yang, Adaptive camera assignment and hand-off algorithm in multiple active camera surveillance system, in: Proceedings of the IEEE Machine Learning and Cybernetics (ICMLC), 2010, pp. 3015–3020.
[26] T. Dinh, Q. Yu, G. Medioni, Real time tracking using an active pan-tilt-zoom network camera, in: Proceedings of the IEEE Intelligent Robots and Systems (IROS), 2009, pp. 3786–3793.
[27] X. Geng, L. Wang, M. Li, Q. Wu, K. Smith-Miles, Adaptive fusion of gait and face for human identification in video, in: Proceedings of the IEEE Applications of Computer Vision (WACV), 2008, pp. 1–6.
[28] G. Shakhnarovich, T. Darrell, On probabilistic combination of face and gait cues for identification, in: Proceedings of the IEEE Automatic Face and Gesture Recognition, 2002, pp. 169–174.
[29] Y. Lu, S. Payandeh, Intelligent cooperative tracking in multi-camera systems, in: Proceedings of the IEEE International Conference on Intelligent Systems Design and Applications (ISDA), 2009.
[30] F. Chang, G. Zhang, X. Wang, Z. Chen, PTZ camera target tracking in large complex scenes, in: Proceedings of the World Congress on Intelligent Control and Automation, 2010.
[31] F.Z. Qureshi, D. Terzopoulos, Planning ahead for PTZ camera assignment and handoff, in: Proceedings of the ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC), 2009.
[32] N. Krahnstoever, T. Yu, S. Lim, K. Patwardhan, P. Tu, Collaborative real-time control of active cameras in large-scale surveillance systems, in: Proceedings of the ECCV Workshop on Multi-Camera and Multi-Modal Sensor Fusion, 2008.
[33] V.N. Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag, Berlin, 1995.
[34] N. Cristianini, J. Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods, Cambridge University Press, Cambridge, 2000.