Energy and Buildings 43 (2011) 305–314

Contents lists available at ScienceDirect

Energy and Buildings

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

Towards a sensor for detecting human presence and characterizing activity

Y. Benezeth [a], H. Laurent [b,*], B. Emile [c], C. Rosenberger [d]

[a] Orange Labs, 4 rue du Clos Courtel, 35510 Cesson-Sévigné, France
[b] ENSI de Bourges, Institut PRISME, 88 bd Lahitolle, 18020 Bourges Cedex, France
[c] Institut PRISME, Université d'Orléans, 2 av F. Mitterrand, 36000 Châteauroux, France
[d] Laboratoire GREYC, ENSICAEN - Université de Caen - CNRS, 6 bd du Maréchal Juin, 14000 Caen, France

ARTICLE INFO

Article history:
Received 20 May 2010
Received in revised form 6 September 2010
Accepted 18 September 2010

Keywords:
Human detection
Presence sensor
Occupancy number detection
Activity characterization
People localization

ABSTRACT

In this paper, we propose a vision-based system for human detection and activity analysis in indoor environment. The developed presence sensor is based on video analysis, using a static camera. Composed of three main steps, the first one consists in change detection using a background model updated at different levels to manage the most common variations of the environment. Second, a moving objects tracking, based on interest points, is performed. Third, in order to know the nature of the various objects that could be present in the scene, multiple cascades of boosted classifiers are used. The validation protocol, defined by the industrial partners involved in the *CAPTHOM* project focusing among other things on "Energy Management in Building", is then detailed. Three applications integrated into the *CAPTHOM* project illustrate how the developed system can help in collecting useful information for the building management system. Occupancy detection and people counting as well as activity characterization and 3D location extend to a wide variety of buildings technology research areas such as human-centered environmental control including heating adjustment and demand-controlled ventilation, but also security and energy efficient buildings.

1. Introduction

The building sector is now one of those that consumes most energy. For example, in France, the building sector is responsible for 21% of the CO_2 emission and for 43% of the total energy use. To economize energy, there are several solutions: first using renewable energies, second developing passive solutions such as insulation and third proposing solutions based on an active management of power consumption. This last approach requires to use reliable knowledge on buildings occupation. With this aim, we propose in this paper a new sensor to detect human presence and collect high level information on people activity.

Nowadays, the sensors available on the market are mostly based on passive infrared technology. These sensors detect the electromagnetic radiations emitted by humans of wavelengths between 6 and 14 μm. When a person moves in the detector field of view, the infrared radiation is focused by the Fresnel lens on the pyroelectric sensor. This technology is now well known and commonly used for home automation. However, it has several major flaws:

- static people cannot be detected,
- the sensor is sensitive to air flows or sunshine radiations, and
- the sensor is not able to distinguish between pets and humans.

The technological limits of these sensors, which are actually motion detectors, hinder the development of innovative solutions for energy consumption management. Conventional systems relying on a single occupancy sensor often suffer from a lack of data analysis of the measured sensor signals and cannot moreover differentiate between one or more occupants in the monitored space. In order to overcome these limits, several works have been conducted. They can mainly be gathered into two groups. The first one recommends the use of multiple low-cost, non-intrusive, environmental occupancy sensors, privileging the use of an independent distributed detectors network combined with a probabilistic data analysis. The second one applies for more advanced sensors such as video cameras.

The approaches belonging to the first group, that fuse together information from multiple sources, result in virtual sensors which are intended to be more powerful than single physical sensors [1,2]. One proposal consists in combining three traditional inexpensive PIR occupancy sensors with a sensor that determines when the telephone is off-hook [1]. If the use of probabilistic models offers improved capability of detecting occupant presence, the fundamental dependance on motion still remains. In order to address the

* Corresponding author.
E-mail addresses: yannick.benezeth@orange-ftgroup.com (Y. Benezeth), helene.laurent@ensi-bourges.fr (H. Laurent), Bruno.Emile@univ-orleans.fr (B. Emile), christophe.rosenberger@greyc.ensicaen.fr (C. Rosenberger).

0378-7788/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.enbuild.2010.09.014

problem of discerning the actual number of people in a room, complex sensor networks have been proposed. A second alternative, composed of a wireless ambient-sensing system complemented with a wired CO_2 sensing system and a wired indoor air quality sensing system, is considered to determine which parameters have the greatest correlation with the occupancy level [2]. The conducted tests show that, for the considered open office plan, CO_2 and acoustic parameters have the largest correlation with the number of occupants, complications arising however with acoustics because of the affect of sound by activities in nearby bays. Even if the proposed system achieves reasonable tracking of an actual occupancy profile, the achieved accuracy does not exceed 75% and, for certain days, it remains relatively low (e.g. 60%). Moreover, further exploration of sufficient training set sizes is needed.

Belonging to the second group, approaches relying on video cameras are also common especially when access to people activity or recognition are pursued [3,4]. The main drawbacks of this solution are the need of large amounts of data storage and above all to interfere with privacy concerns. That is why several works, using low-resolution cameras [5] or even developing “reduced” sensor from camera with a very different appearance from conventional video camera, have been proposed to obtain enough information to detect person’s position and movement status but reducing the psychological resistance of having a picture taken [6].

Other approaches focus on occupancy location detection using ubiquitous and pervasive computing environments, often requiring non-ambient sensors such as wearable devices for tracking inhabitant activity [7].

If the major drawback of users psychological resistance to image capture can be overcome, the use of video camera remains the single sensor allowing to extract at the same time a wide range of information from low-level to high level interpretation. We can finally notice that most of the approaches that use multiple low-cost detectors exploit a camera network to establish the true occupancy information. The true occupancy numbers are then manually counted from the camera pictures [1,2]. In order to facilitate the validation of new designed occupancy sensors, whatever can be the chosen technology, the design of an automated occupancy detector exploiting movies is needed.

The work presented in this paper takes place within the *CAPTHOM* project which fits with the theme “Energy Management in Building” and aims at developing a reliable and efficient sensor to detect the presence of humans in indoor environments. The foreseen concerned sectors could be residential or tertiary areas. The main objectives of such a system are the power consumption management and the increase of comfort for residents, including for example heating adjustment considering their activity. Adaptability is also pursued. The *CAPTHOM* component must be easily integrated into other application areas related to security or supervision and assistance for elderly or disabled people, at home or in institutions.

The sensor must be able to detect the presence of a person in its environment without being disturbed by the presence of animals, other moving objects or by the morphology or activity of people. The sensor must be robust to light changes, heat sources and be able to detect people up to 15 m. The complexity of the proposed algorithms, the used memory must be consistent with material constraints so that the treatment could be carried out in an embedded architecture. The developed system must therefore be an optimal compromise between false detection rate and algorithmic complexity. The proposed solution must be cheap, have a short time response and respect the European directive *2005/32/EC* of 6 July 2005 establishing a framework for the setting of eco-design requirements for energy-using products. Energy consumed by the sensor must be very low. Finally, the use of cameras imposes the respect of privacy and also implies the acceptance problem of this sensor by users. These issues were considered in the *CAPTHOM* project and it has been decided that the camera and its processing unit will only return useful information for the building management system. For the application, none image will be transmitted by the sensor to external devices.

We propose in this paper, algorithmic solutions to detect people from image sequences. Several challenges must be settled:

1. an image is a 2D representation of a 3D scene: the same object, observed from different views, may look very different,
2. image acquisition conditions may change from one environment to another; they can also vary all time long,
3. backgrounds can be very complex; possibilities of false detections are numerous and the contrast between people and background may possibly be very low,
4. many occlusions may appear between the person and the environment or among several individuals, and
5. the main difficulty encountered for people detection is the very high intra-class disparity: through their clothes, size, weight, outline, haircut etc., two individuals may appear very different. Moreover, the human body being highly articulated, the number of possible postures is great and the characteristics will vary in time.

We developed a system using video analysis to interpret the content of a scene without doing strong assumptions about the nature of objects that could be present. Objects nature is determined by statistical tools derived from object detection. The next section describes the proposed algorithm using tools classically dedicated to object detection in still images in a video analysis framework. We then present various extensions of the project including its potential ability to recognize people activity. Finally, some conclusions and perspectives are given.

2. Presence sensor

In this section, we describe the human detection system proposed for the *CAPTHOM* project. It is based on video analysis obtained from a camera. This method has three main steps: change detection, moving objects tracking and classification. Fig. 1 sums up the defined process for human detection.

2.1. Change detection

Working with static camera, the search space can be reduced by detecting regions of interest in the image where there is a high probability to find a person. This first step is achieved through a background subtraction: from an environment model and an observation, we try to only detect what has changed.

According to the results obtained in the comparative study of background subtraction algorithms [8], we decided to model each pixel of the background by a Gaussian probability density function [9]. This quite simple model is a good compromise between quality of detection, computation time and memory requirement.

Since the scene is never completely static, the model must be designed to adapt to different environment changes, such as:

1. slow variation in lighting conditions, caused for example by natural change of daylights,
2. sudden and significant variation due for example to artificial extra lighting adding, and
3. addition or removal of static objects.

The background model is consequently updated at three different levels: the pixel level updating each one with a temporal filter