Review of computer vision in intelligent environment design

REVIEW OF COMPUTER VISION IN INTELLIGENT ENVIRONMENTS DESIGN

Chenghua Lin[1], *Kecheng Liu*[1], *Hong Wei*[2]

[1]Informatics Research Centre

[2]School of System Engineering

{chenghua.lin, k.liu, h.wei}@reading.ac.uk

ABSTRACT

This paper discusses and compares the use of vision based and non-vision based technologies in developing intelligent environments. By reviewing the related projects that use vision based techniques in intelligent environment design, the achieved functions, technical issues and drawbacks of those projects are discussed and summarized, and the potential solutions for future improvement are proposed, which leads to the prospective direction of my PhD research.

1. INTRODUCTION

The term *Intelligent Building* is getting more and more popular nowadays and this concept has attracted a great deal of research interests during the last decade. So far there still is no standard definition of intelligent buildings, but it is generally agreed that intelligent buildings [1, 2] should be able to improve occupants' performance and productivity by autonomously managing the building environment, which optimise energy consumption, safety, monitoring functions and occupant's well-being. It should be noticed that although intelligent buildings could bring lots of benefits to the occupants and the environment, its rate of adoption is still slow in both commercial sector and home sector buildings. There are two major reasons for this. Firstly, as indicated by the home automation and control market research group at ABI, most consumers are unaware of the intelligent building technology and the benefits it can provide, let alone know how it is implemented [3]. Secondly, currently available intelligent building solutions are often highly sophisticated which arise from the techniques of commercial building management system, thus are consequently expensive especially in terms of hardware and installation [4]. Owing to the advance of technologies, especially in the fields of computer vision, agent technology and wireless sensor network, the technical challenges mentioned above are now feasible and enhance the service for occupants within the building.

A wide variety of sensing technologies have been deployed in developing intelligent buildings and intelligent environments for retrieving the information of the building and its occupants. According to the sensing types, these different kinds of sensing technologies can be divided into two categories: vision based and non-vision based, where the former is heavily relied on the visual data obtained from the camera while the latter could take advantage of various sensory information such as the temperature and light-level sensors, movement or occupancy sensors e.g. passive IR, pressure pads, smoke or gas detectors as well as active badge for detecting the location of specific individuals. Although the use of vision faces a lot of challenges including the technical issues of developing and optimising the image processing algorithms, along with the ethical issues concerned with the privacy of the users, it is believed that vision technology will have great potential in the intelligent environment design and could overcome many limitations of the non-vision based techniques, such as intrusive and lack of in-depth information about the environment. In this paper, we will briefly review the vision technology that has been used in the previous intelligent environment and system design, and identify the problems and weakness which need to be addressed in the future.

The rest of the paper is organised as following: section 2 compares computer vision with other techniques in developing intelligent environments. Section 3 reviews the related projects that utilise computer vision techniques. Section 4 discusses the algorithms and technical details of the projects mentioned in section 3. Section 5 summarises the paper and draws the problems that need to be solved in the future.

2. RATIONALE FOR USING VISION

The goal of pervasive computing is to integrate information and computing into everyday physical world, so that this technology is available to everyone in any context [5]. The concept of pervasive computing has gained increasing impact on intelligent environment design, especially when the environment needs to be aware of the activities performed by the occupants and then make considered decisions about the type of assistance to provide, such as environmental control and medical assist.

When deciding the type of sensing technology to use in an intelligent environment, it is vital to understand the user and its requirements. Only with an in depth understanding of the requirements can we better determine the type of functions that need to be performed within the intelligent environment, and hence the type of data to be collected.

The prior research identifies some important functions that an intelligent environment needs to perform in order to support the occupants and to be useable by a wide range of users, which include occupants' well-being, energy conservation and enhanced safety. A number of approaches have been investigated over the last couple of years for realizing the functions mentioned above, and progress has been made in many aspects of this issue. Davidsson and Boman [6, 7] presented a multi-agent system for intelligent building control, where the agents are resided on PDAs and Bluetooth access points and communicate via Bluetooth. The aims of the system are to achieve both energy conservation and occupants' satisfaction via value added service. Energy saving is realised by automatic control of lighting and heating devices, while occupants' satisfaction is realized by adapting the room temperature and light intensity according to occupants' preferences. Although the system is possible to detect which persons are in which room at any moment via the Bluetooth network, it is not capable to distinguish the actions taken by different occupants, which is necessary for learning occupants' behaviours. Due to the establishing of a Bluetooth connection is too slow, severe constrains have been placed on many applications. Other projects have also achieved similar functions in similar scenarios. For example, in the research of Sharples et al. [8], a multi-agent architecture consists of parallel distributed embedded agents is proposed for intelligent building control, with the aims to perform various tasks related to user comfort, energy saving and safety. A variety type of sensors are used in the system for collecting the status about the building and the occupants, such as the temperature detectors, movement sensors etc, while different occupants are distinguished by using a tagging system. Another example is the iDorm project [9], where an intelligent dormitory is developed as testbed for a multiuse ubiquitous computing environment. A myriad of sensors and effectors as well as several different communication networks are hidden in the floor level trunking and a false ceiling.

The projects in [6-9] utilise non-vision based sensors and demonstrate certain learning and autonomous behaviours. Although non-vision based sensors can provide reliable and wide range data, some issues still haven't been addressed. As more information need to be collected, more sensors would be required, which often results in a cluttered and obtrusive environment, like the one in [9]. When learning user's behaviour, it is very important to obtain rich source of information about user's activities. However, sensors used in those projects often result in incomplete information about the user and the environment. For example, a motion detector can detect that a person is in a room, but cannot tell where in the room the person is located, let alone whether the person is sitting, talking, standing or who the person is. More complex and complete information is needed to better describe the user's context and determine user's activities, especially when the system is required to predict inhabitant actions and provide a lot of interactions between the user and the system. In addition, since these types of sensors cannot determine the identity of the person, they are unreliable when there are more than one person lives in the environment. Other devices like RFID (Radio-frequency identification devices) can address this problem by being able to identify different users and track their actions. In the Intelligent Airport (TINA) project [10], a cellular network of combined RFID tag location units and high resolution panoramic video cameras are used to provide seamless tracking of the passengers, along with identification and location service. However, the use of RFID tags presents the same limitations as other on-person hardware, as they require the user continually wear the sensor, which could be obtrusive and inappropriate.

Vision can provide richer information than any other type of sensory information and it is believed that computer vision can be used to develop intelligent environment that could overcome the limitations of non-vision based sensing hardware. Vision can be used not only to obtain positional data, but also to collect more in-depth information in order for the system to build more accurate models of the user and the environment. For instance, vision can be used to determine more accurately the person's location within the environment, any unusual actions, and distinguish the user of interest from other people or pets sharing the same space. Furthermore, a vision-based system can be used to not only track gross movements of the person, but to also track the fine motor movements (including gestures), which is necessary for learning user's behaviours, predicting their actions and hence provide proper assistance. However, issues surrounding the use of vision must be addressed in order for such system to be become more widely accepted and used, such as technical issue of developing more robust and optimised image processing algorithm, and ethical issues concerning the privacy of the users.

3. RELATED WORK USING VISION

In order to explore the potential benefits of using a vision system, some research projects that developed using computer vision technique, especially in the area of intelligent environment, are investigated.

An intelligent room [11] was developed by the MIT Artificial Intelligence Laboratory based on a normal room with minimal decorations, which aims to provide ready at hand computation and information services for the people using the room. The system has a three-layer architecture, which utilises combined techniques of computer vision, robotics, speech understanding and natural language processing. At the bottom level, vision systems track people and identify their activities and gestures like shaking hands, sitting, talking, and through word spotting decide whether people in the room are talking to each other or the room itself. The second level provides a uniform agent-based interface to everything that is installed in the room, where each agent is responsible for a certain type of operation. At the highest level applications are written for providing specialized services to the occupants of the room.

Following the Intelligent Room Project, Hanssens et al. [12] further discussed the issues in building agent-based intelligent workspaces, such as distributed software systems need to allow large numbers of software components to locate and communicate with one another; interfaces between humans and computers need to enable natural, unencumbered interaction; environments need to be aware of what users are trying to do so that they can offer appropriate assistance. By presenting the developed concepts, technologies and applications, they proposed how to address those needs, ranging from low-level communication infrastructure to distributed applications with multi-modal interfaces.

Mihailidis et al. [5] proposed a pervasive healthcare system aiming to support and guide older adults with dementia to the complete the task of handwashing. The sensing equipment of this system is a colour video camera, which is mounted directly above the bath room vanity to monitor user's handwashing process. Three agents are designed in the system: sensing agent, planning agent and prompting agent. Sensing agent is for monitoring user's actions by determining the spatial coordinates of user's hands and body, based on which the planning agent will determine which plan the user is attempting and whether the step being completed is appropriate. If system detects that the user has made an error, the prompting agent will guide the user with customized prompt.

In [13], a surveillance system that uses a network with different kind of sensors for localizing and tracking people in an office environment is presented. The people localization is based on a Bayesian framework which allows the integration of evidence of multiple sensor sources including video cameras, infrared tag readers and a fingerprint reader. The system architecture is comprised of three layers. The bottom layer deals with real-time image acquisition and feature extraction via the agents run on several computers. The middle layer consists of a set of application agents which use the features extracted from the bottom layer, and the top layer uses the results of the middle layer applications to provide a set of meta-applications.

Yang et al. [14] present a multi-camera indoor surveillance system for real-time people detection and tracking. The foreground segmentation is based on a feature level subtraction approach without any post processing module, which is claimed work well even under complex environment conditions. A greedy search based approach is used for tracking multiple people through occlusion and a multi-camera handoff mechanism is provided for associating individual trajectories in adjacent cameras.

Pinhanez and Bobick [15] built an intelligent robotic cameras system which is able to automatically frame subjects and objects in a TV studio upon verbal request from a TV director. An architecture based on two levels of representation is used to cope with the problem of relating visual imagery to symbolic knowledge about the scene. The high level world models roughly describe the objects and occurring actions while low level view representations are obtained by vision routines selected according to the present state of the world.

The researchers from the MIT Media Laboratory [16] described a real-time computer vision and machine learning system for modelling and recognizing human behaviours in a visual surveillance task. This system combines top-down with bottom-up information in a closed feedback loop, with both components employing a statistical Bayesian approach, and is particularly concerned with detecting when interactions between people occur and classifying the type of interaction. Examples of interaction behaviours include following one person, altering one's path to meet another and so forth.

Rui et al. [17] present an automated camera management system for capturing lectures. The major functions of the system including lecturer tracking, audience tracking and video editing, which are all fully automated. Some predefined knowledge is used to assist lecturer tracking: 1) the lecturer is usually moving or gesturing during the lecture which is an important tracking cue; 2) the lecturer's moving space is usually confined to podium area, which can be used to predefine a tracking region to help distinguish lecturer's movement from that of the audience. The audience tracking is based on microphone array technique which estimates the sound source direction, due to that the computer vision based techniques are not able to cope with the audience environment which is usually dark and the audience members sit quite close to each other.

4. SUMMARY OF VISION TECHNIQUES

In this section, some of the computer vision techniques that used in the projects discussed in Section III will be summarized.

4.1. Blob Detection

The notion of *blobs* as a representation for image features has a long history in computer vision and has had many different mathematical definitions. Some refer *blob detection* to visual modules that are aimed at detecting points and/or regions in the image that are either brighter or darker than the surrounding [18]. In

[16], blob is defined as a compact set of pixels that share some visual properties that are not shared by the surrounding pixels. These properties could be colour, texture, brightness, motion, shading, a combination of these, or any other salient spatio-temporal property derived from the image sequence. An example of blob detection is shown in Fig. 1.

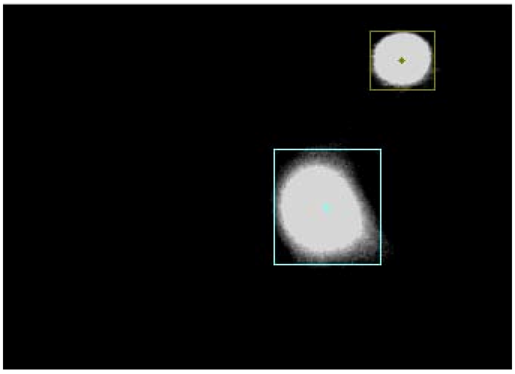

Figure 1. Blob detection

Oliver et al. [16] used 2D blob features for modelling pedestrian in their experiment, and the motion is the main cue for clustering the pixels into blobs based on the concern that they have a static background moving objects. Mihailidis et al. [5] use blob analysis algorithm to identify the skin-coloured objects and ignores all but the four largest blobs that are greater than 200 pixels in size. The location and orientation information of each blob is then used to determine which pair of possible blobs correspond the user's hands regardless the size. Other projects like the *Multiple-Sensor Indoor Surveillance System* [13] uses blob extraction and separation techniques to separate the foreground pixels, and the different part of the blob is used to model different part of the human body, i.e. top 15% of the blob represents the head and bottom 20% represents the feet.

4.2. Object Tracking

4.2.1. Kalman Filter

A Kalman filter is used to estimate the state of a linear system where the state is assumed to be distributed by a Gaussian [19]. Kalman filtering is composed of two steps, prediction and correction. The prediction step uses the state model to predict the new state of the variables:

$$\overline{X}^t = \mathbf{D}X^{t-1} + W,$$

$$\overline{\Sigma}^t = \mathbf{D}\Sigma^{t-1}\mathbf{D}^T + Q^t,$$

where $\overline{X}^t$ and $\overline{\Sigma}^t$ are the state and the covariance predictions at time *t*. **D** is the state transition matrix which defines the relation between the state variables at time *t* and *t*-1. *Q* is the covariance of the noise *W*. Similarly, the correction step uses the current observations Z^t to update the object's state:

$$K^t = \overline{\Sigma}^t \mathbf{M}^t [\mathbf{M}\overline{\Sigma}\,\mathbf{M}^T + R^t]^{-1}, \qquad (1)$$

$$X^t = \overline{X}^t + K^t \underbrace{[Z^t - \mathbf{M}\overline{X}^t]}_{v}, \qquad (2)$$

$$\Sigma^t = \overline{\Sigma}^t + K^t\mathbf{M}\overline{\Sigma}^t,$$

where *v* is called the innovation, **M** is the measurement matrix, *K* is the Kalman gain, which is the Riccati Equation (1) used for propagation of the state models. Note that the updated state X^t is still distributed by a Gaussian.

The Kalman filter has been extensively used in the vision domain for tracking. Oliver et al. [16] use Kalman filter to track objects' location, coarse shape, colour pattern and velocity, along with predicting object's position and velocity in the next frame, by associating a first order Kalman filter to the trajectory of each object blob. Broida and Chellappa [20] used the Kalman filter to track points in noisy image. Rosales and Sclaroff [21] use the extended Kalman filter to estimate 3D trajectory of an object from 2D motion. However, Kalman filter is restricted to situations where the probability distribution of the state parameters is unimodal [22]. In the presence of occlusion, cluttered background resembling the tracked objects, and complex dynamics, the distribution is likely to be multimodal and alternative algorithm need to be used, e.g. condensation algorithm.

4.2.2. Colour

Colour has always been useful information for image processing. Generally, colour is accepted to be made up of two components: *Chrominance* and *Luminance*. The Chrominance value of an object identifies the colouring property of that object whereas luminance is a property of the environment where the object is. A colour space [23] is a mathematical model describing the way in which colour can be represented as numbers. The most common used colour spaces are RGB, gray colour and HSI space, where each of which has three or four components.

The RGB colour space [24] consists of the three additive primaries: red, green and blue. Spectral components of these colours are combined additively to produce a resultant colour. Although the RGB colour space simplifies the design of computer graphics systems, it is not ideal for all applications due to the red, green and blue colour components are highly correlated. This makes it difficult to execute some image processing algorithms such as histogram equalization. For a gray colour, it is in fact one in which the red, green and blue components all have equal intensity in RGB space. So it is only necessary to specify a single intensity value for each pixel, whereas three intensities are needed to specify each pixel in a full colour image. Gray colour space has been widely used for the reasons that less information needs to be provided for each pixel and it is entirely sufficient for many tasks. In terms of the HSI colour model [8], three properties, hue, saturation and intensity are used to describe colour. There is no need to know what percentage of blue or green is required to produce a colour in the HIS space, as the required colour can be obtained simply by adjusting the hue. To change a deep red to pink, adjust the saturation. To make it darker or lighter, alter the intensity. As a result, this model is an ideal tool for developing image processing algorithms based on colour descriptions that are natural and intuitive to humans, such as histogram operations, intensity transformations and convolutions [5]. An example of the RGB and HIS colour space is shown in Fig. 2.

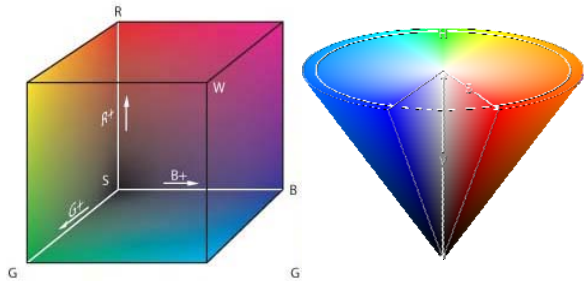

Figure 1. RGB colour space (L); HSI colour space (R)

Among various object tracking techniques, using colour for tracking has the advantages of being orientation and size invariant while being fast to profess. Mihailidis et al. [5] use skin colour as a non-obtrusive marker for real-time hand tracking, as the skin tones have some distinguishing natural feature that they differ primarily in their intensity while their chromaticity or hue remains relatively unchanged. For achieving effective skin segmentation, the normalized colour coordinate (NCC) model is used to locate all skin-coloured pixels. Oliver et al. [16] use Gaussian probability distribution function (PDF) in RGB colour space to model the appearance of each detected blob, in order to handle the occlusions as well as to solve the correspondence between blobs over time. Other projects like the *Multiple-Sensor Indoor Surveillance System* [13] use colour histograms in different colour spaces as major features for distinguishing people by their clothes, based on the concern that at any given day, a person usually wears the same clothes, and thus the colour of the person's clothes is consistent and good discriminator.

4.3. Background modelling

Many computer vision tasks require robust segmentation of foreground objects from dynamic scenes, such as people detection, visual surveillance, vehicle tracking and so forth. And the very first and fundamental step of handling those tasks is the background modelling. In [19], some of the popular background modelling techniques are summarised: 1) Mixture of Gaussians [25]; 2) Eigenbackground [16]; 3) Wall flower; 4) Dynamic texture background.

Stauffer and Grimson [25] present a popular method for real-time video tracking by using adaptive background mixture models, which is capable to deal with lighting changes, repetitive motions from clutter, and long-term scene changes. They model each pixel of the sequence frame as a mixture of Gaussians and use an on-line approximation to update the model. The Gaussian distributions of the adaptive mixture model are then evaluated to determine which are most likely to result from a background process. Each pixel is classified based on whether the Gaussian distribution which represents it most effectively is considered part of the background model.

Oliver et al. [16] build an eigenspace that adaptively models the probability distribution function of the background for detecting the moving objects. This eigenspace model is formed by taking a sample of N images and computing both the mean background image and its covariance matrix, which describes the range of appearances that have been observed, including lighting variations over the day, weather variations and so on. The eigenspace could also be generated from a site model using standard computer graphics techniques.

4.4. Behaviour Modelling

One of the development trend of intelligent building or smart home is that environment should be able to predict the next low-level action, the next location, and the next high-level task that an inhabitant is likely to perform, which is crucial for providing assistance for home automation and adaptation to inhabitant's needs [26]. However, many of the proposed systems for human behaviour analysis using computer vision or machine learning are still limited to low or middle level such as detect and segment human object. Relative few efforts have been made to understand human behaviours that have substantial extent in time, particularly when they involve interactions between people [16].

An interesting piece of work is presented by Panduranga et al. [26], where the actions of inhabitant are modelled as states in a simple Markov model for predicting the next inhabitant action. Enhancement to the model has been made by categorising the actions into abstract tasks and using this information to make subsequent predictions. Other researchers have described a computer vision system using a statistical Bayesian approach for modelling human interactions [16]. They used and compared two different state-based learning architectures, HMMs and CHMMs, for modelling people's behaviour and interactions in a visual surveillance task. The examples of interesting interaction behaviours include following another person, altering one's path to meet another and so forth.

5. DISCUSSION AND FUTURE WORK

In summary, computer vision can have a significant role in the development of intelligent environments. When applying vision technology to accomplish certain application, object detection and tracking are always the technical issues that must be deal with and have great impacts on the overall system performance. Based on the papers that have been reviewed, some of the future directions of object tracking are summarised.

First of all, many of the assumptions used to solve the tracking problems are violated in realistic scenarios, e.g. smoothness of motion, minimal amount of occlusion, illumination constancy etc., which place limitation to a tracker's usefulness in many applications. Tracking multiple people in a small field of view with severe occlusion is another great challenge. Another important issue that has been neglected in the development of tracking algorithm is integration of contextual information. Research [19] has shown that a tracker designed to give the best average performance in a variety of scenarios can be less accurate for particular scene than a tracker that is attuned to the characterises of that scene. A wide range of feature selection algorithms have been investigated in the machine learning and pattern recognition communities. However, the drawbacks of these algorithms are they require offline training information about the target and/or the background which is not always available, and the discriminative features are not always the same as the object appearance or background varies. Therefore, there is a need for online selection of discriminative features. One of the promising directions to achieve this goal is the use of the online boosting methods for feature selection [19]. Similarly, most tracking algorithms use predefined models for object representation, which attracts lot of research inserts in unsupervised learning of object models for multiple nonrigid moving objects from a single camera. Probabilistic state-space methods like Kalman filters, HMMs and Dynamic Bayesian Networks (DBNs) have been extensively used to estimate object motion parameters. Among these methods, DBNs provide a principled framework for fusing information from different sources. However, there is a need for more efficient solutions for inference before DBNs are more commonly used in tracking applications.

Overall, regarding the future direction of the PhD research, one of the interesting directions will be investigating an approach that can exploit disparate sources of information in a general object tracker, in particular contextual and prior information, to make the general tracker being employed in a variety of scenarios possible. Another appealing direction will be using computer vision technique and statistic method to model human behaviours and predict their actions, which is vital for the intelligent environment to be aware of the activities performing within it and thus provide proper assistance to the occupants.

6. REFERENCES

[1] D. Clements-Croome, *Intelligent Buildings: Design, Management And Operation*: Thomas Telford, 2004.

[2] B. Qiao, K. Liu, and C. Guy, "A Multi-Agent System for Building Control," *Proceedings of the IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT 2006 Main Conference Proceedings)(IAT'06)-Volume 00*, pp. 653-659, 2006.

[3] P. Schuster, "Zigbee Enabled Home Control and Security," 2007.

[4] B. E. Markwalter, S. K. Fitzpatrick, P. J. Hargaden, and S. C. A. A. S. C. Appling, "Design influences for the CEBus automation protocol," *Consumer Electronics, IEEE Transactions on*, vol. 37, pp. 145-153, 1991.

[5] A. Mihailidis, B. Carmichael, and J. Boger, "The use of computer vision in an intelligent environment to support aging-in-place, safety, and independence in the home," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 8, pp. 238-247, 2004.

[6] P. Davidsson and M. Boman, "Saving Energy and Providing Value Added Services in Intelligent Buildings: A MAS Approach," *Agent Systems, Mobile Agents, and Applications*, vol. 1882, p. 166?77.

[7] P. Davidsson and M. Boman, "Distributed monitoring and control of office buildings by embedded agents," *Information Sciences*, vol. 171, pp. 293-307, 2005.

[8] S. Sharples, V. Callaghan, and G. Clarke, "A Multi-Agent Architecture For Intelligent Building Sensing and Control," *International Sensor Review Journal*, vol. 19, pp. 13-14, 1999.

[9] A. Pounds-Cornish and A. Holmes, "The iDorm-a practical deployment of Grid technology," *Cluster Computing and the Grid 2nd IEEE/ACM International Symposium CCGRID2002*, pp. 433-438, 2002.

[10] I. White, R. Penty, J. Crowcroft, J. Elmirghani, A. Seeds, and P. Brennan, "The Intelligent Airport (TINA)."

[11] R. A. Brooks, "The Intelligent Room project," in *Cognitive Technology, 1997. 'Humanizing the Information Age'. Proceedings., Second International Conference on*, 1997, pp. 271-278.

[12] N. Hanssens, A. Kulkarni, R. Tuchinda, and T. Horton, "Building Agent-Based Intelligent Workspaces," *ABA Conference Proceedings*, p. 675?81, 2002.

[13] V. A. Petrushin, G. Wei, O. Shakil, D. Roqueiro, and V. Gershman, "Multiple-Sensor Indoor Surveillance System," *Proceedings of the 3rd Canadian Conference on Computer and Robot Vision (CRV)*, pp. 53-60, 2006.

[14] Y. Tao, F. Chen, D. Kimber, and J. A. V. J. Vaughan, "Robust People Detection and Tracking in a Multi-Camera Indoor Visual Surveillance System," in *Multimedia and Expo, 2007 IEEE International Conference on*, 2007, pp. 675-678.

[15] C. Pinhanez and A. Bobick, "Intelligent Studios: Using Computer Vision to Control TV Cameras."

[16] N. M. Oliver, B. Rosario, and A. P. Pentland, "A Bayesian Computer Vision System for Modeling Human Interactions," 2000.

[17] Y. Rui, L. He, A. Gupta, and Q. Liu, "Building an Intelligent Camera Management System."

[18] Wikipedia, "Blob detection."

[19] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Computing Surveys (CSUR)*, vol. 38, 2006.

[20] T. J. Broida and R. Chellappa, "Estimation of object motion parameters from noisy images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, pp. 90-99, 1986.

[21] R. Rosales and S. Sclaroff, "3 D trajectory recovery for tracking multiple objects and trajectory guided recognition of actions," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 117-123, 1999.

[22] T. B. Moeslund and E. Granum, "A survey of computer vision-based human motion capture," *Computer Vision and Image Understanding*, vol. 81, pp. 231-268, 2001.

[23] S. K. Singh, D. S. Chauhan, M. Vatsa, and R. Singh, "A robust skin color based face detection algorithm," *Tamkang Journal of Science and Engineering*, vol. 6, pp. 227-234, 2003.

[24] H. Rein-Lien, M. Abdel-Mottaleb, and A. K. Jain, "Face detection in color images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, pp. 696-706, 2002.

[25] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 246-252, 1999.

[26] S. Rao and D. J. Cook, "Predicting Inhabitant Actions Using Action and Task Models with Application to Smart Homes," *International Journal of Artificial Intelligence Tools*, vol. 13, pp. 81-100, 2004.