



## State of the art of smart homes

Liyanage C. De Silva<sup>a,c,\*</sup>, Chamin Morikawa<sup>b</sup>, Iskandar M. Petra<sup>a</sup>

<sup>a</sup> Faculty of Science, Applied Physics Group, University of Brunei Darussalam, Brunei Darussalam

<sup>b</sup> Interfaculty Initiative in Information Studies, The University of Tokyo, Japan

<sup>c</sup> School of Engineering and Advanced Technology, Massey University, Palmerston North, New Zealand

### ARTICLE INFO

#### Article history:

Received 5 January 2012

Received in revised form

10 April 2012

Accepted 2 May 2012

Available online 12 June 2012

#### Keywords:

Smart monitoring

Audiovisual event detection

Eldercare

Energy efficient homes

Video summarization

Smart home

### ABSTRACT

In this paper we present a review of the state of the art of smart homes. We will first look at the research work related to smart homes from various view points; first in the view point of specific techniques such as smart homes that utilize computer vision based techniques, smart homes that utilize audio-based techniques and then smart homes that utilize multimodal techniques. Then we look at it from the view point of specific applications of smart homes such as eldercare and childcare applications, energy efficiency applications and finally in the research directions of multimedia retrieval for ubiquitous environments. We will summarize the smart homes based research into these two categories. In the survey we found out that some well-known smart home applications like video based security applications has seen the maturity in terms of new research directions while some topics like smart homes for energy efficiency and video summarization are gaining momentum.

© 2012 Elsevier Ltd. All rights reserved.

## 1. Introduction

Humans interact with the environment that surrounds them, in numerous ways. They perceive the environmental conditions and act, react or adjust accordingly. If the environment can be made to reciprocate this behavior and respond to human behavior, it will lead to several advantages. Such behavior can automate various tasks that humans have to perform manually, and also provide novel services and facilities. A smart home is a home-like environment that possesses ambient intelligence and automatic control, which allow it to respond to the behavior of residents and provide them with various facilities.

The standard approach for building smart homes is to *computerize* them. A set of sensors gather different types of data, regarding the residents and utility consumption of the home. Computers or devices with computing power (e.g., micro-controllers) analyze these data to identify *actions* of residents or *events*. They then respond to these actions and events by *controlling* certain mechanisms that are built in to the home. A simple example for such smart behavior is turning the lights on when a person enters a room. However, more complicated tasks such as detecting if an elderly resident is alone and not feeling well are also desired.

Smart homes have been researched for nearly a couple of decades. The pioneering work in this area are the Smart Rooms implemented by the MIT Media Lab (Pentland, 1996). Thereafter, several researches have investigated this topic with a wide range of prospective applications. At the current state, there are many types of smart homes with three major application categories.

The first category aims at providing services to the residents by detecting and recognizing their actions or by detecting their health conditions. Such smart homes act as information collection testbeds to support the wellbeing of the residents of the home. These smart homes can be further divided into three types; smart homes that provide eldercare, smart homes that provide healthcare and smart homes that provide childcare.

The second category of smart homes aims at storing and retrieving of multi-media captured within the smart home, in different levels from photos to experiences. One might argue that the issue of privacy of such type of information collection, but it will be a matter of acceptance in to one's lifestyle with time.

The third category is surveillance, where the data captured in the environment are processed to obtain information that can help to raise alarms, in order to protect the home and the residents from burglaries, theft and natural disasters like flood etc. A few researches attempted to combine these functions into one smart home.

Apart from the 3 types of smart homes we have discussed there is an emerging trend of a special type of smart homes which can help the occupants to reduce the energy consumption of the house by monitoring and controlling of the devices and rescheduling their operating time according to the energy demand and supply.

\* Corresponding author at: Faculty of Science, Applied Physics Group, University of Brunei Darussalam, Brunei Darussalam. Tel.: +673 8180458.

E-mail addresses: [liyanagecd@yahoo.co.nz](mailto:liyanagecd@yahoo.co.nz) (L.C. De Silva), [chamin@me.com](mailto:chamin@me.com) (C. Morikawa), [iskandar.petra@ubd.edu.bn](mailto:iskandar.petra@ubd.edu.bn) (I.M. Petra).

With recent advances in electronics and computing, sensing technologies and computing power required to implement a smart home is now available in small sizes, low prices and energy efficiency. However, providing the ambient intelligence that is required to make decisions for smart behavior is still a challenging task. Human behavior at home is highly unstructured. Multiple sensory modalities are required to sense such behavior. Advance pattern recognition techniques are required to recognize the behavior of multiple residents. Privacy becomes an important issue once the systems store the data. Due to such challenges, smart room technologies at the current state are far from being matured.

This paper surveys the state of the art of smart home technologies. It organizes related researches as follows. Before going into specific application categories we will first look at different techniques used in smart homes namely; video based techniques, audio based techniques and multimodal techniques. Section 2 outlines video based techniques for human activity detection in smart environments. Section 3 covers audio-based techniques. Section 4 investigates how to combine multiple sensory modalities to recognize actions and events that take place in a smart home. Thereafter we will look at specific applications of smart homes. First in Section 5 we will look at smart homes applications for eldercare and childcare. Then in Section 6 we look at the energy efficient smart homes and Section 7 investigates research directions on Multimedia Retrieval for Ubiquitous Environments. We conclude the paper with a brief discussion on the state of the art, and suggestions for future research directions.

## 2. Video-based techniques in smart homes

Video is highly prospective as an input modality for smart homes, due to its non-intrusive nature and rich information content. However, its use in a home-like environment has been sometimes questioned and argued about, due to privacy concerns. Despite such concerns, research on vision-based activity detection in smart environments has been growing rapidly.

Many types of vision-based systems for surveillance and monitoring of closed environments have been described and built over the past (Harper, 2003). Henry Tan and De Silva (2003), Tan et al. (2003) has proposed a simple technique for human activity recognition using head movement detection. *Smart environments* are an immediate application of human activity detection. Alex Pentland's research group at MIT Media Laboratory designed a *smart room* in 1991 (Pentland). This has evolved from its initial design to its matured state of five networked smart rooms in the United States, Japan and the United Kingdom. These rooms use several machines, none more powerful than a personal computer, to identify the location, identity, facial expression and hand gestures of the persons in the room. Few more related research can be found in (Bobick; MIT). Systems based on single or multiple cameras, both stationary and moving, have been designed and implemented for automatic detection, tracking and recognition of humans and their actions (Vadakkepat et al., 2008).

In the paper (Utsumi et al., 1998a,b) the authors propose a system that analyses image sequences from multiple stationary cameras, acquired from a particular scene, to detect humans and their actions, and index the sequence according to these activities. The image sequences are indexed using the results for faster searching. *Key frames* are extracted from the image sequences for each entry in the index, to facilitate visual inspection without browsing the image sequence. In addition to the index, visualizations of motion paths for humans in the scene are created to provide a faster way of tracking human movements in the scene.

A camera based detection approach is given in Kainka (2003) where it involves a single camera tracking a number of people. The system works by extracting points and identifying feature

points from an image, creates a path and clusters them and finally each of these clusters corresponds to a person. The  $W^4$ : Who? When? Where? What? (Haritaoglu et al., 1998) technique relies on the system to solely identify a combination of shapes and sizes from the image segmentation of the monochromatic imagery to identify a subject's presence and its interaction and time. The system in (Utsumi et al., 1998a,b) uses multiple cameras to detect human motion by selecting the best viewpoints of the images to extract a maximum amount of information on the individual or multiple amounts of individuals. The results of the system are reconstructions of the human position, normal axis and body size.

### 2.1. Video sensor-based event recognition

In the paper (De Silva et al., 2006) the authors use a state-based approach to recognize events and actions. The state diagram in Fig. 1 shows the transitions between states defined for a tracked human in the image sequence. Table 1 summarizes the rules for the state transitions.

In order to execute the various actions in the above state diagram the authors detect and track humans entering the environment. Each human entering the smart room is modelled using the detected location and motion parameters. These parameters are buffered over a window of 10 seconds and validated against a set of decision rules to detect state transitions (Table 1).

### 2.2. Performance evaluation of tracking using video sensors

Although tracking is an intermediate step in any activity detection system, there is no immediate method for evaluation. In the paper (De Silva et al., 2006) the authors use 20 image sequences for evaluating the performance of tracking. The results are compared with ground truth for quantitatively estimating the accuracy. On the other hand, (de Silva, 2003) qualitatively evaluate their system based on user studies by residents.

### 2.3. Detection of unusual events and actions

It is possible that actions or events that are important to be recorded, but not specified by the rules, take place. For example, there can be a situation where a person trying to block the camera. In the paper (De Silva et al., 2006) the authors keep an index to such an action as “unrecognized” to facilitate human observation to recognize the action at a later stage. If the amount of scene change occurring between two frames is substantial and the action/event cannot be recognized, the scenario is identified as an unrecognized event. Key frames showing the scene change are extracted from the image sequence.

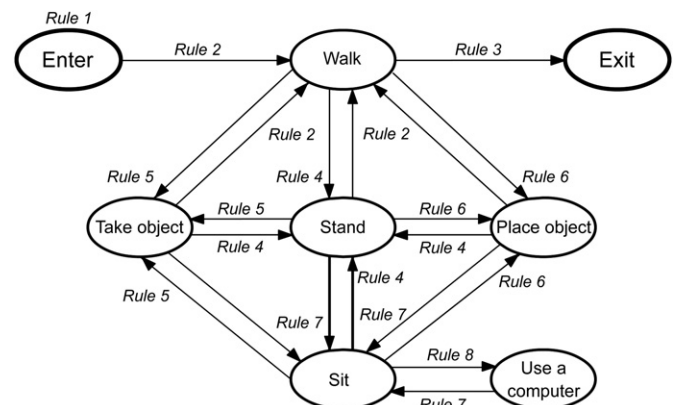


Fig. 1. State transitions for video sensor based detection (as of De Silva et al. (2006)).

**Table 1**  
Rules for state transitions.

Rule ID	State moved to	Specification
1	Enter (initial state)	New human model instance detected near the entrance region for 10 frames
2	Walk	Coordinates of the centroid of the head-shoulder region changes gradually over 20 frames
3	Exit (final state)	Human model instance located near the entrance region for more than 10 frames, and could not be tracked thereafter
4	Stand	Y coordinate of the head centroid increases by more than 2 head-heights, with less than 10% change in head-height
5	Take object	Location of object changes together with the location of the person
6	Place object	New object detected on the table region for more than 15 frames
7	Sit	Y coordinate of the head centroid decreases by more than 2 head-heights, with less than 10% change in head-height
8	Use a computer	Person sitting near computer with only small movements below shoulder region

Falling is an another unusual event that need to be investigated in a smart home that cater for elderly since a sudden fall can lead the person into life threatening situations. In the paper (De Silva, 2008e) the author describes the method of detection of such an event using a video sequence (Fig. 2).

### 3. Audio-based techniques in smart homes

Video sensor based event detection approach has some shortfalls like it fails to cover the entire room and also event non-detection due to occlusion. Hence some authors have proposed audio sensor based methods for actions detection (De Silva, 2008e). In his paper for the separation of cough, cry and shout (vocal track generated sounds) from walk, door-open and fall (non-vocal sounds), they use pitch contour detection. This is due to the fact that vocal track generated audio signals consists of its inherent formant frequency component. They proposed a method of detecting walking action using audio signals. Events like door open and walking had been separated by using the audio intensity profiles. According to the authors of the paper walking has its inherent property of gradual increase of the intensity profile till the steps are getting closer to the microphone and then gradual decrease when walks past the microphones while the door opening is one off high pitch isolated and localized sound. Also they combined the falling audio activity with the video based falling detection to get a higher accuracy.

Different types of events can be recognized by processing audio signals captured within a smart home. Smeaton and McHugh (2005) demonstrated that a network of microphones can detect audio events that can significantly summarize a video collection from multiple cameras. The inference of this statement is that in order to extract vital information for video summarization, audio events can be used as they are connected in the time domain up to some extent. Hence it is somewhat easier, for example, to follow the stepping sound of a child in a house to extract the video clips where the child is walking from room to room. In order to track the child only using video information the field of view of cameras should be overlapping. But with the use of network of microphones one can bridge the gap very easily. The Human Speech Home Project of MIT (Roy et al., 2006) investigated how continuously captured audio in a smart home can be used for media retrieval, early language learning, and several other applications.

A home is an environment with a number of rooms that are connected to each other, directly or indirectly. The degree of connectivity can change with opening and closing of doors. These conditions make sound source localization a prerequisite for multiple source audio analysis in a smart home. Different researches employed different techniques for this task (Rui and Florencio, 2003); Microphone arrays (Chen et al., 2005) and beam-forming (Hoshuyama et al., 1999) are common approaches. Bian et al. (2004) employed sound source localization followed by

classification to infer activities inside a smart home. Liu and Wan (2001) demonstrated the use of time-domain audio features with supervised learning for multi-class audio classification Fig. 3.

### 4. Multimodal-based techniques in smart homes

In their paper (Vadakkepat et al., 2008) addressed a scenario where a robot tracks and follows a human using multimodal techniques. The use of multimodality is to improve the accuracy of detection and tracking. A neural network is utilized to learn the skin and non-skin colors of the human in the room. The skin-color probability map is utilized for skin classification and morphology-based preprocessing. In addition to the visual cues, the tracking process considers 16 sonar scan and tactile sensor readings from the robot to generate a robust measure of the person's distance from the robot. The robot thus decides an appropriate action, namely, to follow the human subject and perform obstacle avoidance. Heuristic rule is used for face-ratio analysis and Bayesian cost analysis for label classification.

Let us look at the research work related to audio-visual based approaches for human activity detection in smart homes. There are some audio-visual based attempts for human emotion recognition (De Silva et al., 1997; De Silva and Miyasato, 1998; De Silva, 2004) which can support smart home activities as a background processing mechanism. An excellent survey of audio and video based multimodal affect recognition can be found in Zeng et al. (2009). There have also been some attempts to use audio-visual sensors for human activity detection in the context of smart homes (De Silva, 2008e).

Some researches combined data from multiple sensors other than audio or video sensors, to improve the accuracy (Demongeot et al., 2002; Brooks et al., 2003; Beigl et al., 2004; Petrushin et al., 2006; Chana et al., 2008).

In the next three sections we will look at specific applications of smart homes. Either one or a couple of the three techniques we have discussed in the previous three sections, audio based, video based or multi-modal are being used to achieve these specific application goals.

### 5. Smart homes for eldercare

Smart homes for eldercare applications are growing at a very fast phase in all parts of the world. One main requirement of such applications is the human detection and activity classification. The necessity for the development of human detection methods in the field of modern Home-care and security systems has become very popular and essential. There are many techniques currently being used for human detection using many different kinds of sensory information. It is necessary to detect the presence of the human in advance before processing the human activities such as falling, standing or walking etc. (Kainka, 2003).

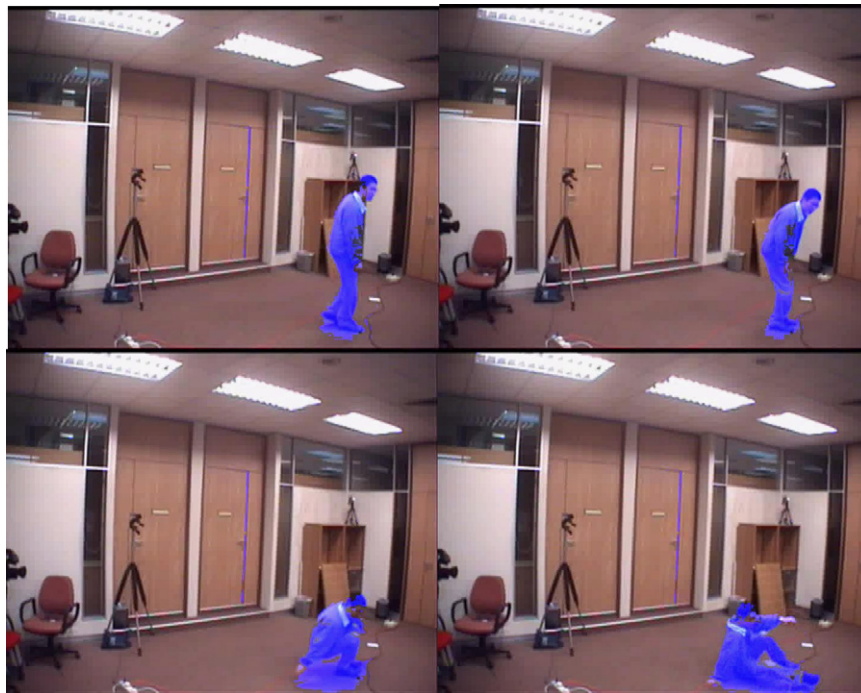


Fig. 2. Fall Detection Results (as of De Silva 2008e).

In Elliott et al. (2009) the authors described an experiment that extends the distributive sensing approach to identify the three-dimensional location of an object in constant motion that can be applied to human activity detection in a smart home. Distributive sensing has previously been used in the identification of size and location of statically placed objects. Here, in their paper they propose a novel system to measure balance or sway of a human. They presented an experimental set-up which consisted of a pendulum structure positioned on a supported steel plate. Three low-cost deflection sensors were positioned under the plate with the resulting signals used as inputs to a neural network implemented on a field-programmable gate array. The results showed that the embedded system can accurately track the pendulum position in real time with a mean tracking error of around 6% in all three dimensions. This evidence indicated that their technique is sufficiently sensitive and could be implemented in a pragmatic configuration for discriminating between balance and sway, which can be used to detect fall of an elderly in a smart home environment just before it happen.

There are numerous Smart Home projects that intend to make daily life comfortable (Demongeot et al., 2002; Ogawa et al., 2002; Pereira et al., 2006) Aware Home Project (Abowd et al., 2002) for supporting elderly residents. Basic activities such as opening and closing of doors was recorded using switch-based sensors (Ogawa and Togawa, 2000). Numerous types of sensors are used for tracking and detection of the persons and recognize their activities. Use of cameras and image analysis for this purpose is common. In Easy Living Project (Brumitt, 2000; Krumm et al., 2000) and Intelligent Space (Lee et al., 1999), the positions of humans are detected using multiple cameras. However, alternative methods such as Radio Frequency Identification (RFID) tags (Juels, 2006), optical tags (Kainka, 2003) and Infra-red based motion sensors (Moore, 1999) have been used where image acquisition and analysis is not possible due to issues such as privacy, disk space, and computational cost.

Human detection techniques at present can be either video based or any other sensor based. Sensor based detections are such as (Ogawa and Togawa, 2000; Ogawa et al., 2001; Lim and

Kriegman, 2004) where infrared sensors and carbon dioxide sensors are used to detect motion and magnetic sensors are utilized to detect the opening and closing of doors. An illumination sensor is a type of sensor where once the subject is present, the sensor relies on changes in the environment caused by the subject to trigger a chain of events in the circuit. A more fascinating approach is a system called Cenwits (Huang and Mishra, 2005). Connection-less Sensor-Based Tracking Using Witnesses. This is a mobile system that emits a signal from time to time using RF communication. When two of these mobile sensors are close to each other, information is gathered such as time and location at that time of the subject carrying the sensor and finally all information is dumped at an access point. This system would be useful for application in a large area where it being necessary to keep track of individuals.

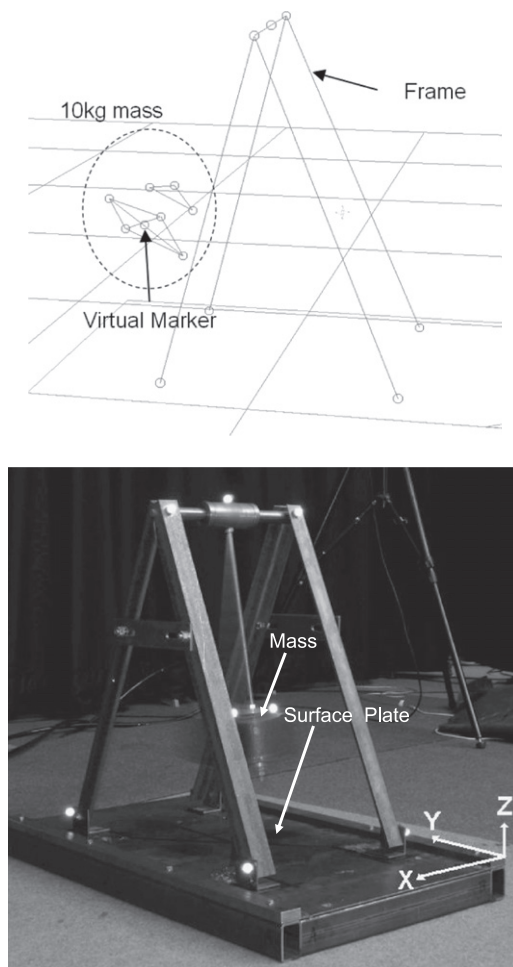
The paper by You et al. (2008) has two goals, one is elder care and the other is childcare support. AIST Japan under their Digital Human project had built a Smart home for child accident prevention (AIST).

## 6. Smart homes for energy efficiency

In recent days we can find there are a growing number of new research proposals and findings in related to new and alternative energy technologies. However there are many easy and cheap ways to reduce energy use at our homes by efficient energy management. Most of these simply require a change in behavior of the occupants of the home. In the paper (Lach et al., 2007) the authors proposed an automatic monitoring system to reduce the energy usage of a typical home by using WIFI technology enabled smart switches. Fig. 4 shows the GUI of the presented prototype system. In their project they were looking ahead to enhance this technology by adding different types of sensors to enhance the automatic monitoring and control of the environment according to the user preference based on their profiling.

In the paper by Reinisch et al. (2011) in their Think Home (Fig. 5) they looked into the use of a multi-agent technique to





**Fig. 3.** Configuration of the markers on a swing designed to mimic the body balancing patterns of an elderly which can be sensed by the platform underneath before a fall (as of Elliott et al. (2009)).

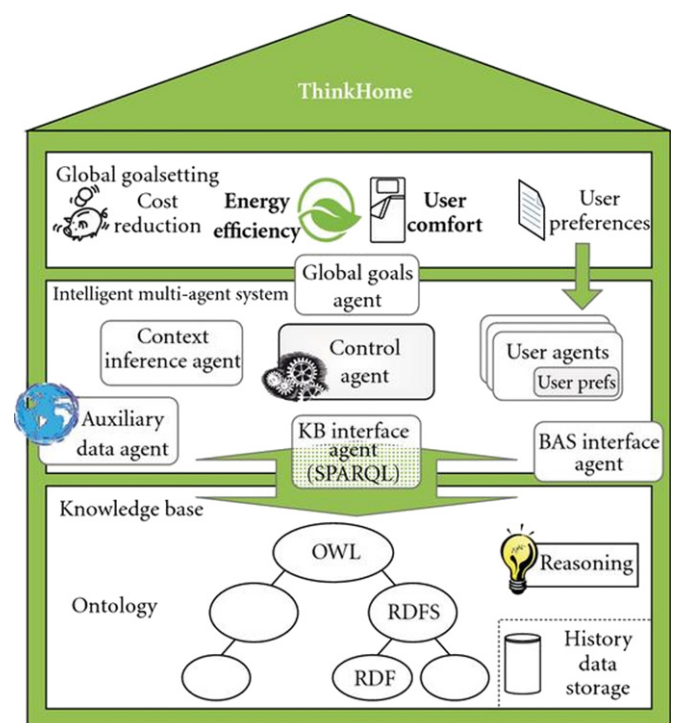
reduce the energy consumption in a house. Their work puts its focus on alleviating the current problems by proposing a comprehensive system concept that ensures the smart homes can keep their promise in the future. The system operates on an extensive knowledge base that stores all information needed to fulfill the goals of energy efficiency and user comfort. Its intelligence is implemented as and within a multi-agent system that also caters for the system's openness to the outside world. They propose several agents such as Global goal agent, Control agent, User agent, Auxiliary data agent, and Interface agent etc. to execute the system. Different agents contain and provide intelligence regarding aspects such as user preferences, auxiliary data, and control parameters.

## 7. Multimedia retrieval for ubiquitous environments

Automated capture of experiences taking place at home is interesting owing to a number of reasons. Home is an environment where a variety of important events and experiences take place. Some of these, such as the first footsteps of a child, provide no opportunity for manual capture. Some others are so important that humans do not want to keep themselves out of the experience to shoot photos or video. A corpus of interactions and experiences at home can provide valuable information for studies related to the



**Fig. 4.** Energy efficient smart homes – GUI of the control system for energy efficiency – as of Lach et al. (2007).



**Fig. 5.** Think home – an energy efficient smart home – as of Reinisch et al. (2011).

design of better housing, human behavior, etc. Other prospective applications include assistance for elderly residents and aiding recollection of things that were forgotten.

Both capture and retrieval of experiences in a home-like environment is extremely difficult due to a number of reasons. Even the simplest and the smallest of the houses are partitioned into a number of rooms or regions, making it necessary to have a large number of cameras and a fair number of microphones for complete data capture. Continuous recording of data from these devices, to ensure the capture of all important experiences, results in a very large amount of data. The level of privacy differs

at different places of a house, and sometimes certain regions are shared only among certain residents.

The most difficult problems, however, arise during retrieval and summarization of the captured data. Content captured at home is much less structured compared to that from any other environment. Queries for retrieval could be at very different levels of complexity, and the results can be in various levels of granularity. Some examples are shown below (as of de Silva (2007)):

- Show the video from the camera near the entrance to the living room, from 8:30 pm to 9:00 pm, on the 1st of February, 2005.

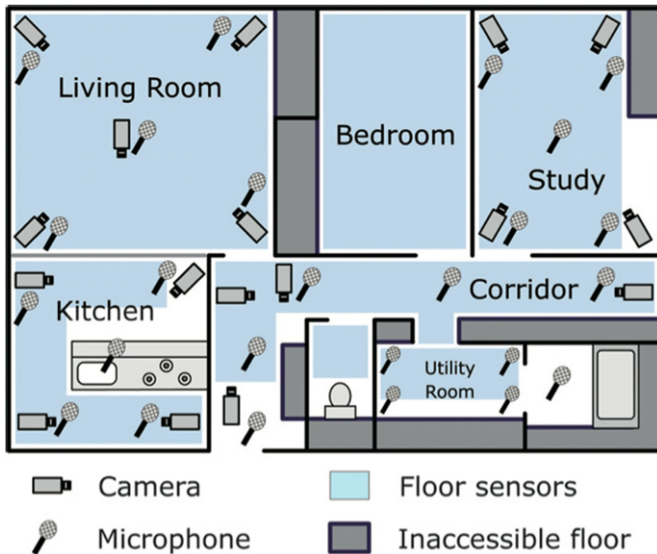


Fig. 6. The sensor arrangement of the ubiquitous smart home (as of de Silva et al. (2008c,d)).

- What was our child doing between 5:30 and 6:30 pm yesterday?
- On which date did Jeff visit us last month?
- How did the strawberry jam that I bought last week finish in 4 days?

Given the large content and the state of the art of content processing algorithms, multimedia retrieval for ubiquitous environments based solely on content analysis is neither efficient nor accurate. Therefore, it is desirable to make use of supplementary data from other sensors for easier retrieval. For example, proximity sensors that get activated by human presence will remove the burden of image analysis for human detection. Since ubiquitous environments are built with infrastructure to support cameras and microphones for capture, it is relatively easy to add additional sensors to acquire such data. Domain knowledge, such

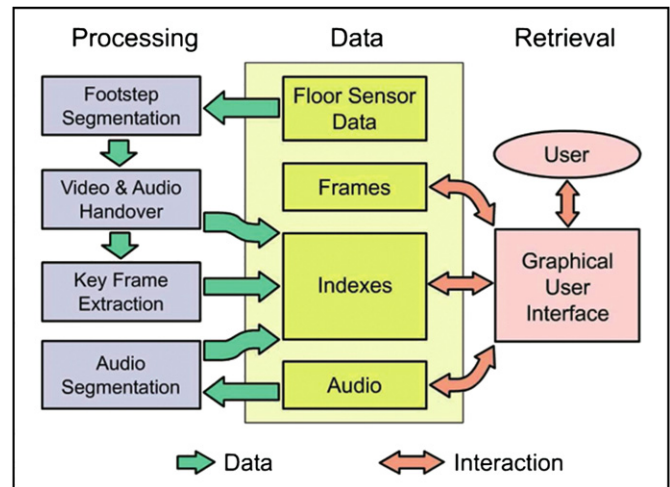


Fig. 8. The user interaction used in the retrieval process of the summary of movements inside the ubiquitous smart home (as of de Silva et al. (2008c)).

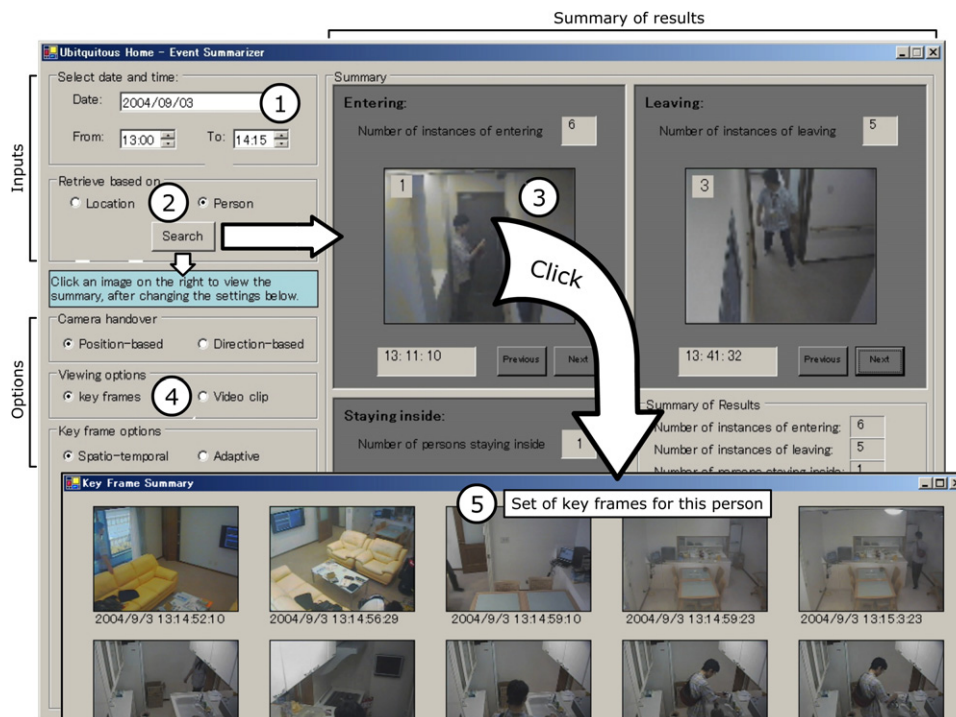


Fig. 7. The functional overview of the proposed systems as of de Silva et al. (2008c,d). Here they have used different sensory modalities for indexing actions and events.

**Table 2**

Comparison of smart home research by application areas.

Type of smart home	References
1 Smart homes for security	(de Silva, 2003; Tan and De Silva, 2003; Tan et al., 2003; De Silva, 2008e)
2 Smart homes for eldercare	(De Silva et al., 1997; De Silva and Miyasato, 1998; Abowd et al., 2002; De Silva, 2004)
3 Smart homes for healthcare	(Yong and Lee, 2002; Cai et al., 2005)
4 Smart homes for childcare	(AIST; You et al., 2008)
5 Smart homes for energy efficiency	(De Silva et al., 2009; De Silva et al., 2010)
6 Smart homes for better life (music, entertainment etc.)	(Panasonic; Brumitt, 2000; de Silva et al., 2005a, b; de Silva et al., 2005a, b; de Silva et al., 2007; de Silva et al., 2008c, d; de Silva et al., 2008c, d)

**Table 3**

Comparison of smart home research by activity detection methods.

Technique used for Activity detection	Papers
1 Audio based techniques	(Bush, 1945; Hoshuyama et al., 1999; Rui and Florencio, 2003; Vacher et al., 2003; Bian et al., 2004; Cai et al., 2005; Chen et al., 2005; Kraft et al., 2005; Smeaton and McHugh, 2005; de Silva et al., 2008c, d)
2 Video based techniques	(Pentland, 1996; Segen and Pingali, 1996; De Silva and Miyasato, 1998; Haritaoglu et al., 1998; Healey and Picard, 1998; Utsumi et al., 1998a, b; Utsumi et al., 1998a, b; Ivanov and Bobick, 1999; Krumm et al., 2000; Pentland and Choudhury, 2000; de Silva, 2001; Petkovic and Jonker, 2002; Tjondronegoro et al., 2002; Viola and Jones, 2002; Yong and Lee, 2002; de Silva, 2003; Kainka, 2003; Tan and De Silva, 2003; Tan et al., 2003; Tan et al., 2003; de Silva et al., 2004; Gandhi and Trivedi, 2004; Lam and Chiu, 2004; Lim and Kriegman, 2004; Wood et al., 2004; Duan et al., 2005; Huayong and Hui, 2005; Snoek and Worring, 2005; Valera and Velastin, 2005; Wang and Parameswaran, 2005; Anavi, 2006; Gemmell et al., 2006; Girgensohn et al., 2006; Pereira et al., 2006; De Silva et al., 2008a, b; De Silva et al., 2008a, b)
3 Audio-visual techniques	(De Silva et al., 1997; Pye et al., 1998; Wang et al., 2000; De Silva, 2004; Cucchiara, 2005; de Silva et al., 2005a, b; Kim et al., 2005; de Silva, 2007; de Silva et al., 2008c, d; De Silva, 2008e; You et al., 2008)
4 Sensor based techniques	(Ogawa and Togawa, 2000; Collins et al., 2001; Ogawa et al., 2001; Lee and Mase, 2002; Ogawa et al., 2002; Brooks et al., 2003; Beigl et al., 2004; Mori et al., 2004; Arampatzis et al., 2005; Huang and Mishra, 2005; Juels, 2006; Petrushin et al., 2006; De Silva et al., 2009; Elliott et al., 2009; De Silva et al., 2010)
5 Mixture of audio, video and sensor based techniques	(Matsuoka and Fukushima, 2004; de Silva et al., 2005a, b; Vadakkepat et al., 2008)

as the purpose of use for each room, is also helpful in the design of algorithms for retrieval.

### 7.1. Motivation

Investigation in to automated retrieval of experiences at home can be useful in several other aspects, in addition to the significances mentioned above. This topic encompasses the general research areas of multimedia retrieval and ubiquitous environments. However, a home is much less controlled compared to the other ubiquitous environments used in related research. Video captured at home are unstructured content, marking a significant contrast from news, sports or instructional video which are the common inputs for automated retrieval. Therefore, the selected topic will pose several research challenges, with prospects of significant contributions to these areas. The outcomes of this research are applicable in areas with practical significance, such as automated surveillance, elder care, and automated video summarization.

### 7.2. Related research

Research on multimedia retrieval for smart homes and other smart environments has become possible due to the recent developments in storage technologies facilitating recording large amounts of data. There are several ongoing projects that work on this topic. Applications in this category include meeting video retrieval and summarization of instructional video. Some of the projects, such as CHIL (Waibel, 2005), attempted to combine both the above directions by supporting user interaction real-time and using retrieval for long term support. The college of computing, Georgia Institute of Technology (Bian et al., 2004), has constructed several smart classrooms. These rooms are equipped with multiple data projectors, cameras and active white boards,

to facilitate capturing of lectures for later review by students. The classrooms are also equipped with stylus based tablets for the use of the students. Xerox PARC uses infrared beacons to provide improved user interfaces for smart rooms (Want and Borriello, 2000). A graphical user interface is used to control equipment in the room, with the aid of these beacons.

The Ubiquitous Sensor Room is an environment that captures data from both wearable and ubiquitous sensors to retrieve video diaries related to experiences of each person in the room. Jaimes et al. (2004) utilize graphical representations of important memory cues for interactive video retrieval from a ubiquitous environment. The Sensing Room is a ubiquitous sensing environment equipped with cameras, floor sensors and RFID sensors for long-term analysis of daily human behavior. Video and sensor data are segmented into 10-min intervals and the activity in the room during each segment is recognized using a Hidden Markov Model. Matsuoka and Fukushima (2004) attempted to understand and support daily activities in a house, using a single camera installed in each room and sensors attached to the floor, furniture and household appliances.

In their paper (de Silva and Yamasaki et al., 2008c,d) presented their techniques for experience retrieval in a smart home. The smart home they have analysed was equipped with 19 cameras, 25 microphones and pressure-based sensors mounted on the house floor. Hierarchical clustering of pressure sensor data followed by video handover automatically created videos tracking residents as they walked to and from different rooms. Video summarization reduced these videos to sets of key frames, for faster viewing. Sound source localization followed by supervised machine learning facilitated video indexing by audio events. A user interface based on hierarchical media segmentation combined these results to enable residents to retrieve their life experiences in a fast and effective manner. The system was evaluated by a family who stayed in the smart home and used the system to retrieve their experiences six months after their



stay. Fig. 6 shows the sensor arrangement of their smart home. The functional overview of their proposed methodology is summarized in Fig. 7. A block diagram of user interaction for retrieving a summary of movement is shown in Fig. 8.

## 8. Conclusions

In this paper we looked at the research work related to smart homes from various view points; first in the view point of specific techniques and then in the last 3 sections in the view point of specific applications. Specifically we looked at computer vision based techniques in smart homes, audio-based techniques in smart homes, multimodal techniques in smart homes at first. Then when we looked from the view point of applications of smart homes we looked at applications like eldercare and childcare, energy efficiency and finally in the research directions of multimedia retrieval for ubiquitous environments. Table 2 and Table 3 summarize the past research works in to some specific sub topics based on application centered and technique centered.

Here we observed that some of the smart home research areas are getting saturated and some new emerging areas are taking place. For example, video based person identification and activity detection for security applications has almost come to maturity but new areas like multiple sensor integration based techniques for energy efficient applications are taking the lead.

In the future we can envisage that more and more computer power given to smart sensors researchers will make use of them in home area distributed sensor networks. Each sensor will either report in real time to the host or they will keep the information in the memory for offline processing. In order to address complex situations in smart rooms, multiple agents based intelligent and distributed software/hardware frameworks have been proposed recently. Some top software companies are building application specific software targeted for smart homes so that existing infrastructure can be upgraded with added intelligence and decision making support. In the future our homes will not be the same. A simple example is our mobile phones. About 10 years ago it was just a phone that can make calls. Now the number of functions it has are countless. This analogy can be easily applied to our homes. One day it will be a robot inside out. The house will look at us from many directions to protect us from potential dangers due to our forgetfulness or due to other physical disabilities. Then we will always have a friend to live in with us.

## References

Abowd, G.D., Bobick, I., et al., 2002. The aware home: developing technologies for successful aging. *Proceedings of American Association of Artificial Intelligence (AAAI 2002)*.

AIST. AIST Digital home project. Retrieved 20/12/2011, 2011, from <<http://www.dh.aist.go.jp/en/research/enabling/>>.

Anavi, S., 2006. Information Processing in a Ubiquitous Home Using Image Analysis. Signal Processing Institute, School of Engineering, Swiss Federal Institute of Technology, Lausanne.

Arampatzis, T., Lygeros, J., et al., 2005. A survey of applications of wireless sensors and wireless sensor networks. *Proceedings of the 13th Mediterranean Conference on Control and Automation*. Limassol, Cyprus: 719–724.

Beigl, M., Krohn, A., et al., 2004. Typical sensors needed in ubiquitous and pervasive computing. *First International Workshop on Networked Sensing Systems (INSS 2004)*. Tokyo, Japan: 153–158.

Bian, X., Abowd, G.D., et al., 2004. Using sound source localization to monitor and infer activities in the home. *GVU Technical Report; GIT-GVU-04-20*, Georgia Institute of Technology.

Bobick, A. Kids' Room. Retrieved 2011/12/28, from <<http://vismod.media.mit.edu/vismod/demos/kidsroom/kidsroom.html>>.

Brooks, R.R., Ramanathan, P., et al., 2003. Distributed target classification and tracking in sensor networks. *Proc. IEEE* 91, 1163–1171.

Brumitt, B., 2000. Easy living: technologies for intelligent environments. *Proceedings of International Symposium on Handheld and Ubiquitous Computing*.

Bush, V., 1945. As we may think. *Atlantic Mon.* 176 (1), 101–108.

Cai, R., Lu, L., et al., 2005. Unsupervised content discovery in composite audio. *Proceedings of ACM Multimedia 2005*. Singapore, ACM: 628–637.

Chana, M., Estève, D., et al., 2008. A review of smart homes—present state and future challenges. *Comput. Methods Program. Biomed.* 91 (1).

Chen, J.F., Shue, L., et al., 2005. An adaptive microphone array with local acoustic sensitivity. *IEEE International Conference on Multimedia & Expo (ICME 2005)*. Amsterdam, The Netherlands.

Collins, R.T., Lipton, A.J., et al., 2001. Algorithms for cooperative multisensor surveillance. *Proc. IEEE* 89, 1456–1477.

Cucchiaro, R., 2005. Multimedia surveillance systems. *Proceedings of the Third ACM international Workshop on Video Surveillance & Sensor Networks (VSSN '05)*. Hilton, Singapore, ACM Press, New York, NY: 3–10.

De Silva, C.R., Ranganath, S., et al., 2008a. Cloud basis function network: a modified RBF network architecture for holistic facial expression recognition. *Elsevier Pattern Recognit. J.* 41 (4), 1241–1253.

De Silva, C.R., Ranganath, S., et al., 2008b. Cloud basis function neural network: a modified RBF network architecture for holistic facial expression recognition. *Pattern Recognit.* 41 (4), 1241–1253.

de Silva, G.C., 2001. Traffic Flow Measurement Using Video Image Sequences. Department of Computer Science and Engineering, Univ. of Moratuwa, Sri Lanka, MEng.

de Silva, G.C., 2003. Tracking and Indexing of Human Actions in Video Image Sequences. Dept. of Electrical and Computer Engineering, Singapore. National University of Singapore, MEng.

de Silva, G.C., 2007. Multimedia Experience Retrieval in Ubiquitous Home. Japan. University of Tokyo, PhD.

de Silva, G.C., De Silva, L.C., et al., 2004. Tracking and indexing of human actions in video image sequences. Published in the *Proceedings of the Annual Conference on Image and Vision Computing IVNCZ*.

de Silva, G.C., Yamasaki, T., et al., 2005a. Evaluation of video summarization for a large number of cameras in ubiquitous home. *Proceedings of ACM Multimedia 2005*: 820–828.

de Silva, G.C., Yamasaki, T., et al. (2007). An Interactive multimedia diary for the home. *IEEE Computer, Special Issue on Human Centered Computing*: 52–59.

de Silva, G.C., Yamasaki, T., et al., 2008c. Audio analysis for multimedia retrieval from a ubiquitous home. *Proceedings of MMM 2008*: 466–476.

de Silva, G.C., Yamasaki, T., et al., 2008d. Ubiquitous home: retrieval of experiences in a home environment. *IEICE Trans. Inf. Syst.* E91-D (2), 330–340.

de Silva, G.C., Yamasaki, T., et al., 2005b. Video handover for retrieval in a ubiquitous environment using floor sensor data. *Proceedings of ICME 2005*.

De Silva, L.C., 2004. Audiovisual emotion recognition. Invited Paper in the *Proceedings of IEEE International Conference on Systems, Man and Cybernetics (SMC2004)*, The Hague, The Netherlands, IEEE.

De Silva, L.C., 2008e. Audiovisual sensing of human movements for home-care and security in a smart environment. *Int. J. Smart Sens. Intell. Syst.* 1 (1), 220–245.

De Silva, L.C., Dewana, T., et al., 2010. Multiple sensor based autonomous monitoring and control for energy efficiency. Springer book series titled communications. In: Vadakkepat, P., Kim, J.H. (Eds.), *Computer and Information Science*, vol. 103. Springer, pp. 361–368.

De Silva, L.C., Iskandar, M.P., et al., 2009. Ambient intelligence in a smart home for energy efficiency and eldercare. Springer book series titled communications. In: J. H. K., et al. (Eds.), *Computer and Information Science*, vol. 44. Springer, pp. 187–194.

De Silva, L.C., Miyasato, T., 1998. Hierarchical Expression Recognition. Japanese National Patent (2967058).

De Silva, L.C., Miyasato, T., et al., 1997. Facial emotion recognition using multi-modal information. *Proceedings of IEEE International Conference on Information, Communications and Signal Processing (ICICSP1997)*, Singapore, IEEE.

De Silva, L.C., Pereira, N., et al., 2006. Video based event detection and tracking for eldercare and security in smart rooms. *Proceedings of Image Media Processing Symposium (IMPS2006)*. Shizuoka, Japan.

Demongeot, J., Virone, G., et al., 2002. Multi-sensors acquisition, data fusion, knowledge mining and alarm triggering in health smart homes for elderly people. *C.R. Biol.* 325 (6), 673–682.

Duan, L., Xu, M., et al., 2005. A unified framework for semantic shot classification in sports video. *IEEE Trans. Multimedia* 7 (6), 1066–1083.

Elliott, M.T., Petra, M.I., et al., 2009. Quantifying sway through surface deflection patterns: a novel approach using distributive tactile sensing. *Proc. Inst. Mech. Eng. Part H J. Eng. Med.* 223, 7.

Gandhi, T., Trivedi, M.M., 2004. Calibration of a reconfigurable array of omnidirectional cameras using a moving person. *Proceedings of the ACM Second International Workshop on Video Surveillance & Sensor Networks*. New York, NY, USA.

Gemmell, J., Bell, G., et al., 2006. MyLifeBits: a personal database for everything. *Commun. ACM* 49 (1), 88–95.

Girgensohn, F., Shipman, S., et al., 2006. Support for Effective Use of Multiple Video Streams in Security. *Proceedings of the Fourth ACM international Workshop on Video Surveillance & Sensor Networks (VSSN '06)*. ACM Press, New York, NY 19–26.

Haritaoglu, I., D., Harwood, et al., 1998. W4: Who? When? Where? What? A real time system for detecting and tracking people. *Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition*.

Harper, R., 2003. Inside the Smart Home: Ideas, Possibilities, and Methods. Springer-Verlag, UK.



- Healey, J., Picard, R.W., 1998. Startle cam: a cybernetic wearable Camera. Proceedings of the Second IEEE International Symposium on Wearable Computers: 42.
- Hoshuyama, A., Sugiyama, A., et al., 1999. A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters. *IEEE Trans. Signal Process.* 47 (10), 2677–2684.
- Huang, J.H., Mishra, S., 2005. A sensor based tracking system using witnesses. Proceedings of the 25th IEEE International Conference on Distributed Computing System Workshops.
- Huayong, L., Hui, Z., 2005. A content-based broadcasted sports video retrieval system using multiple modalities: sportbr. Proceedings of the Fifth International Conference on Computer and Information Technology (CIT 2005). Washington, DC, IEEE Computer Society: 652–656.
- Ivanov, Y.A., Bobick, A.F., 1999. Recognition of multi-agent interactions in video surveillance. *IEEE Proceedings of the International Conference on Computer Vision*. Kerkyra, Greece. 1: 169–176.
- Jaimes, A., Omura, K., et al., 2004. Memory Cues for Meeting Video Retrieval. *CARPE 2004*: 74–85.
- Juels, A., 2006. RFID security and privacy: a research survey. *IEEE J. Sel. Areas Commun.* 24 (2), 381–394.
- Kainka, B., 2003. Passive-optical person detector. *Elector Electron.*, 2.
- Kim, S., Hwang, D.S., et al., 2005. An effective news anchorperson shot detection method based on adaptive audio/visual model generation. Proceedings of the Fourth International Conference in Image and Video Retrieval (CIVR 2005). Singapore: 276–285.
- Kraft, F., Malkin, R., et al., 2005. Temporal ICA for Classification of Acoustic Events in a Kitchen Environment. *INTERSPEECH*. Lisbon, Portugal.
- Krumm, H., Harris, S., et al., 2000. Multi-camera multi-person tracking for easy living. In: Proceedings of Third IEEE International Workshop on Visual Surveillance: 3–10.
- Lach, C., Punchihewa, A., et al., 2007. Smart home system operating remotely Via 802.11b/g wireless technology. Proceedings of the Fourth International Conference Computational Intelligence and Robotics and Autonomous Systems (CIRAS2007). Palmerston North, New Zealand.
- Lam, K., Chiu, K.H.C., 2004. Adaptive visual object surveillance with continuously moving panning camera. Proceedings of the ACM Second International Workshop on Video Surveillance & Sensor Networks. New York, NY, USA.
- Lee, J., Ando, N., et al., 1999. Design policy for intelligent space. Proceedings of IEEE SMC'99, IEEE.
- Lee, S., Mase, K., 2002. Activity and location recognition using wearable sensors. *Pervasive Computing* (July–September): 10–18.
- Lim, J., Kriegman, D., 2004. Tracking humans using prior and learned representations of shape and appearance. Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition: 869–874.
- Liu, M., Wan, C., 2001. A study on content-based classification and retrieval of audio database. Proceedings of the 2001 International Database Engineering and Applications Symposium. Grenoble, France: 339–345.
- Matsuoka, K., Fukushima, K., 2004. Understanding of living activity in a house for real-time life support. *SCIS & ISIS 2004*: 1–6.
- MIT. Massachusetts Institute of Technology Media Laboratory. Retrieved 2011/12/28, from <<http://www.media.mit.edu>>.
- Moore, D.J., 1999. Implementing phicons: combining computer vision with infra red technology for interactive physical icons. Proceedings of ACM UIST 1999. New York, ACM Press: 67–78.
- Mori, T., Noguchi, H., et al., 2004. Sensing room: distributed sensor environment for measurement of human daily behavior. First International Workshop on Networked Sensing Systems (INSS2004): 40–43.
- Ogawa, M., Ochiai, S., et al., 2001. Remote monitoring of daily activities and behaviors at home. Proceedings of the 23rd Annual EMBS International Conference. Istanbul, Turkey. 4: 3973–3976.
- Ogawa, M., Suzuki, R., et al., 2002. Long-term remote behavioural monitoring of the elderly using sensors installed in domestic houses. Proceedings of the Second Joint EMBS/BMES Conference. TX, USA. 3: 1853–1854.
- Ogawa, M., Togawa, T., 2000. Monitoring daily activities and behaviours at home by using brief sensors. First Annual International IEEE-EMBS Special Topic Conference on Microtechnologies in Medicine & Biology. Lyon, France: 611–614.
- Panasonic. Smart Life for an Entire Lifestyle. Retrieved 20/12/2011, from <<http://ch.panasonic.co.jp/global/contents/03270/>>.
- Pentland, A., 1996. Smart Rooms. *Scientific American*: 54–62.
- Pentland, A., Choudhury, T., 2000. Face recognition for smart environments. Computer, IEEE Press United Kingdom: 50–55.
- Pentland, A. P. Smart Rooms. Retrieved 2011/12/28, from <<http://vismod.www.media.mit.edu/vismod/demos/smartroom/ive.html>>.
- Pereira, N., De Silva, L.C., et al. (2006). Vision based human activity detection for eldercare and security. Proceedings of the Annual Conference on Image and Vision Computing (IVCNZ 2006). Great Barrier Island, New Zealand.
- Petkovic, M., Jonker, W., 2002. Cobra: a content-based video retrieval system (advances in database technology). Proceedings of the Eighth International Conference on Extending Database Technology (EDBT 2002). Prague, Czech Republic: 736–738.
- Petrushin, V.A., G. W., et al., 2006. Multiple-sensor indoor surveillance system. Proceedings of the Third Canadian Conference on Computer and Robot Vision (CCRV'06).
- Pye, D., Hollinghurst, N., et al., 1998. Audio-visual segmentation for content-based Retrieval. The International Conference on Spoken Language Processing (ICSLP'98). Sydney, Australia.
- Reinisch, C., Kofler, M.J., et al., 2011. Thinkhome energy efficiency in future smart homes. *EURASIP Journal on Embedded Systems* 2011: 19.
- Roy, D., Patel, R., et al., 2006. The human speechome project. Proceedings of the 28th Annual Cognitive Science Conference.
- Rui, Y., Florencio, D., 2003. New direct approaches to robust sound source localization. *IEEE International Conference on Multimedia & Expo (ICME 2003)*. Baltimore, USA.
- Segen, J., Pingali, S., 1996. A camera-based system for tracking people in real Time. Proceedings of the 13th International Conference on Pattern Recognition.
- Smeaton, A.F., McHugh, M., 2005. Towards event detection in an audio-based sensor network. Proceedings of the Third ACM Workshop on Video Surveillance and Sensor Networks (VSSN'05). Singapore: 87–94.
- Snoek, G.M., Worring, M., 2005. Multimedia event-based video indexing using time intervals. *IEEE Trans. Multimedia* 7 (4), 638–647.
- Tan, H.C.C., De Silva, L.C., 2003. Human activity recognition by head movement using Elman network and Neuro-Markovian hybrids. Proceedings of Image and Vision Computing New Zealand (IVCNZ 2003). Massey University, Palmerston North, New Zealand: 320–326.
- Tan, H.C.C., Janapriya, R.E.G., et al., 2003. An automatic system for multiple human tracking and actions recognition in office environment. Proceedings of IEEE the International Conference on Acoustics, Speech and Signal Processing (ICASSP 2003). Hong Kong, IEEE.
- Tan, H.C.C., Jia, K., et al., 2003. Human activities recognition by head movement using partial recurrent neural network. *Proc. SPIE* 5150, 2007.
- Tjondronegoro, D., Chen, Y.-P.P., et al., 2002. A framework for customizable sport video management and retrieval. Mining Multimedia and Complex Data: KDD Workshop MDM/KDD 2002, PAKDD Workshop KDMCD 2002 Revised Papers. O. Zaiane, S. Simoff and C. Djeraba, Springer: 248–265.
- Utsumi, A., Mori, H., et al., 1998a. Multiple-human tracking using multiple cameras. Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition: 498–503.
- Utsumi, A., Mori, H., et al., 1998b. Multiple view-based tracking of multiple humans. Proceedings of the 14th International Conference on Pattern Recognition. 1: 597–601.
- Vacher, M., Istrate, D., et al., 2003. Smart audio sensor for telemedicine. Smart Objects Conference 2003. Grenoble, France.
- Vadakkapat, P., Lim, P., et al., 2008. Multimodal approach to human-face detection and tracking. *IEEE Trans. Ind. Electron.* 55 (3), 1385–1393.
- Valera, M., Velastin, S.A., 2005. Intelligent distributed surveillance systems: a review. *IEE Proc. Image Signal Process.* 152, 192–204.
- Viola, P., Jones, M., 2002. Robust real-time object detection. *Int. J. Comput. Vision*, 1026–1033.
- Waibel, A., 2005. CHIL—Computers in the human interaction loop. Proceedings of Learntec 2005.
- Wang, R., Parameswaran, N., 2005. Survey of sports video analysis: research issues and applications. Proceedings of the Pan-Sydney area workshop on Visual information processing (VIP '05). M. Piccardi, T. Hintz, S. He, M. L. Huang and D. D. Feng. Darlinghurst, Australia, Australian Computer Society, Inc.: 87–90.
- Wang, Y., Liu, Z., et al., 2000. Multimedia content analysis using both audio and visual cues. *IEEE Signal Processing Magazine*: 12–36.
- Want, R., Borriello, G., 2000. Survey on information appliances. *IEEE Comput. Graphics Appl.* 20 (3), 24–31.
- Wood, K., Fleck, R., et al., 2004. Playing with SenseCam. Proceedings of Playing with Sensors: Exploring the Boundaries of Sensing for Playful Ubiquitous Computing (W3). UbiComp 2004.
- Yong, S.I., Lee, W.S., 2002. Content-based retrieval of video data with flexibly managed attributes. *Knowl. Inf. Syst.* 4 (4), 507–519.
- You, F., Luo, H., et al., 2008. Prototyping and evaluation for smart home controller based on chinese families behavior analysis. Proceedings of the Eighth Asia-Pacific conference on Computer-Human Interaction (APCHI '08). S. Lee, H. Choo, S. Ha and I. C. Shin: 437–445.
- Zeng, Z., Pantic, M., et al., 2009. A survey of affect recognition methods: audio, visual, and spontaneous expressions. *IEEE Trans. Pattern Anal. Mach. Intell.* 31 (1), 39–58.