

*promoting access to White Rose research papers*



## **Universities of Leeds, Sheffield and York**

**<http://eprints.whiterose.ac.uk/>**

---

This is an author produced version of a paper published in **Journal of Ambient Intelligence and Smart Environments (JAISE)**.

White Rose Research Online URL for this paper:

**<http://eprints.whiterose.ac.uk/42983/>**

---

### **Published paper**

Fabien, Cardinaux, Deepayan , Bhowmik, Charith, Abhayaratne and Mark S., Hawley (2011) *Video Based Technology for Ambient Assisted Living: A review of the literature*. Journal of Ambient Intelligence and Smart Environments (JAISE) . ISSN 1876-1364 (In Press)

---

# Video Based Technology for Ambient Assisted Living: A review of the literature

Fabien Cardinaux, Deepayan Bhowmik, Charith Abhayaratne and Mark S.Hawley

School of Health and Related Research, University of Sheffield, UK

E-mail: {fabien.cardinaux,mark.hawley}@sheffield.ac.uk

Dept of Electronic and Electrical Engineering, University of Sheffield, UK

E-mail: {d.bhowmik, c.abhayaratne}@sheffield.ac.uk

## Abstract

Ambient assisted living (AAL) has the ambitious goal of improving the quality of life and maintaining independence of older and vulnerable people through the use of technology. Most of the western world will see a very large increase in the number of older people within the next 50 years with limited resources to care for them. AAL is seen as a promising alternative to the current care models and consequently has attracted lots of attention. Recently, a number of researchers have developed solutions based on video cameras and computer vision systems with promising results. However, for the domain to reach maturity, several challenges need to be faced, including the development of systems that are robust in the real-world and are accepted by users, carers and society. In this literature review paper we present a comprehensive survey of the scope of the domain, the existing technical solutions and the challenges to be faced.

## 1 Introduction

According to the European Union (EU) commission's projection, the number of older people will increase three-fold between 2008 and 2060. This envisages great challenges towards care for older people with limited available resources. On the other hand, research suggests that as many as 35% of older people who live in a care home could be supported to live at home or in extra care housing schemes through the use of technology [46]. As an al-

ternative to current care models, enriched assistive living based solutions can offer the possibility for many older and vulnerable people to live independently at home.

No universal definition of ambient assisted living (AAL) has been adopted, but it can be described as information and communication technology based products, services and systems to provide older and vulnerable people with a secure environment, improve their quality of life and reduce the costs of health and social care. According to the technical road map [1], outlined by the EU commission, the stakeholders' needs are categorised as: 1) AAL for persons (home and mobile), 2)AAL in the community and 3)AAL at work. However within the scope of this work we restrict our discussion to AAL for personal living in the home. State-of-the-art AAL technologies include the use of home embedded wireless sensor networks and body worn sensors. More recently, the research community, public institutions and industry have shown a growing interest in video and computer vision based solutions for AAL (VAAL) applications. The technology is rapidly reaching a stage of maturity with the first products being commercialised [53].

One of the main aims of AAL based systems is to provide personal safety to older people. Falling is a major risk faced by older people with one-third of people over 65 falling each year. Family carers are also impacted by an increased burden of care, anxiety and costs of residential care. The associated cost of falls is almost £1bn annually for the UK alone. With such severe consequences for the individual and society, it is not surprising that the vast majority of VAAL research focuses on fall detection.

Due to the multi-disciplinary nature of the topic, research reports tend to be scattered across conferences and journals specialised in diverse domains ranging from health and telecare to engineering and computer vision. As a consequence, the scientific community might experience difficulties in following recent advances in the domain. This survey aims to summarise the state-of-the-art of video technology currently used in AAL, to explore the challenges to be faced and to investigate potential future developments in the domain.

The rest of this paper is organised as follows. An overview of the application domains and system requirements of VAAL is given in Section 2. Section 3 covers current computer vision approaches and Section 4 discusses issues related to acceptability and protection of privacy. Methodologies for performance evaluation are discussed in Section 5. Finally, conclusions and future areas of research are given in Section 6.

## 2 Applications and Requirements

### 2.1 Application Domains and Scope

This section presents a list of AAL areas [1] along with potential applications for video technology. Note that this survey is concerned only with the use of automatic video analysis, therefore, other applications of video cameras such as tele-communication (videoconferencing) or surveillance of video by an operator are not included in this survey.

#### 2.1.1 Health, rehabilitation and care

People with long-term health conditions often experience a reduced quality of life and dependence upon health and social care services. Video technology can be used in this context to recognise and monitor physiological data, as recently proposed by Poh *et al.* [48] who presented a methodology for a contactless measurement of cardiac pulse rate using a video camera.

Physiological data can, however, not always be measured from embedded sensors and a popular alternative solution is to monitor users' activities with the premise that change in activities of daily living is a good indicator of declining health. Video technology can be used in this

context to recognise users' activities. For example Mihailidis *et al.* [38] propose a health monitoring scheme which tracks the location and motion of the user at home, builds probabilistic models of patterns of activity and then detects deviations in these patterns. With a similar assumption that activities are a good indicator of health, Nait-Charif and McKenna [40] introduced a method to summarise activity by tracking the user's position using an overhead camera.

Diet is also an important factor affecting health and Kim *et al.* [28] proposed a design concept of an image-based dietary assessment tool implemented in a mobile phone. Food images are recorded using a built-in camera on a mobile phone and sent to a server where the portion size of the meal is estimated. This information is subsequently used to keep personal dietary records as a means of monitoring energy and nutrient intakes.

The information related to activity, physiological data, diet, etc. can be subsequently fed back to users to promote health behaviour changes or sent to carers. As suggested in [42], these technologies can be part of a mechanism to support self-management of people suffering from chronic problems such as stroke and provide them with a better insight into how and how often they may perform their activity and rehabilitation exercises.

#### 2.1.2 Personal and home safety and security

Older and vulnerable people are particularly exposed to safety and security threats. Feeling safe and secure is thus essential to extend autonomous daily living. Video cameras can be used, for example to automatically detect intrusion by a burglar [19]. Such security applications have been the subject of a number of research papers in the domain of automatic video surveillance and will not be discussed in detail in this survey. For a recent overview of automatic video surveillance, see [30].

Currently fall detection is a well researched topic and sensor based solutions are available and commercialised. The majority of such solutions are based on accelerometers, but one of the drawbacks of these systems are that users always need to wear the sensor. The system will not work if the user forgets to wear the device. Alternative solutions are suggested, such as passive fall detection which uses floor vibration sensors, sound [3, 34, 72] or video based monitoring [7, 18, 26, 32, 33, 35, 51, 52, 57, 58, 60–

62]. Video based monitoring algorithms are discussed in section 3.

People with dementia are particularly vulnerable with most exhibiting behavioural disturbances. Long-term dementia care facilities face the risk of the patients eloping and putting themselves at risk. In [8], Chen *et al.* propose to improve safety of patients in a dementia unit with a video elopement detector. The system is based on multiple cameras covering the exit of the unit and demonstrates that it can accurately differentiate elopements from routine exit of the building.

### 2.1.3 Biorobotic systems

Biorobotic systems have been considered for a long time to improve the quality of life of older and vulnerable people. In the last few years, a growing number of research studies have investigated the development of companion robots for the care of older people. These robots can provide physical, cognitive and affective support to the users. A systematic literature review of the effectiveness of assistive robots has been conducted by Broekens *et al.* [5] in 2009 which shows the growing importance and promising preliminary results of assistive robotics. Broekens *et al.* differentiate two types of functionalities achieved by assistive robots: (1) Robots which support independent living by supporting basic activities and mobility and (2) Robots which aim to enhance health and wellbeing of elderly users by providing companionship. For both functionalities, robots must have a good understanding of the world around them and video cameras along with other sensors associated with appropriate processing can play this role. Mataric *et al.* [37] are developing a socially assistive robot equipped with a video camera that makes communicative actions such as looking at or away from the user. Other robots equipped with a camera include the Aibo [20] developed and produced by Sony.

## 2.2 Video Based vs Sensor Based Solutions

Applications of AAL described in the previous section are typically based on solutions employing sensors either embedded in the environment or body worn [47]. The most widely used sensors for AAL are reported in Table 1 along with their domains of application. To reach the level of analysis required for AAL applications (*e.g.*

recognising activity), a large network of embedded sensors is generally required, therefore systems are usually costly to maintain, are relatively obtrusive (*e.g.* sensors set up on every cupboard door) and are highly sensitive to the performance of the sensors. Another approach is to use body worn sensors which are usually quite effective when the user actually wears them, however, it is recognised that user compliance with wearable systems is poor.

Recently a number of researchers and companies have been looking at developing solutions based on video cameras and computer vision approaches. A monitoring system based on video cameras has potential advantages. In principle a single camera in a room could pick up most of the activities performed in the room and, consequently, could replace a large number of sensors.

Remarkably, the last five years have shown a large interest in video based solutions. One of the major reasons for this gain of interest is the cost of video cameras. The price of video cameras has drastically reduced within the last decade and good quality video cameras are now available at very low cost. Another important factor in favour of video technology is the maturity of computer vision technology. Latest research in computer vision, motivated by needs in domains such as security and surveillance, provides new means of interpreting rich information provided by video cameras.

On the other hand, video based AAL has several obstacles to overcome to reach full maturity. The barriers include users' acceptability and risk of loss of privacy. These are discussed in detail in section 4.

## 2.3 Multi-modal Approaches

Although a growing popularity has been noted towards video based monitoring in AAL, other sensor based solutions are currently being used in many real life projects. At the same time video based sensors can not replace many traditional sensors. More recently many multi-modal approaches using a network of sensor arrays have been adopted and used. For example, GER'HOME [49] uses a multi-sensor based analysis for everyday elderly activity monitoring; audio-video based solutions are proposed in the HERMES project [50] to facilitate AAL applications for older people using technologies such as: speech recognition, speaker identification, face detection,

Table 1: Commonly used sensors and their domains of application

Sensor	Application domains	Embedded/Body-worn
Motion detector	Fall detection, Activity monitoring , Security	Embedded
Door open/close	Activity monitoring, Security	Embedded
Electrical appliance	Activity monitoring, Safety	Embedded
Microphone	Fall detection	Embedded
Accelerometer	Fall detection, Activity monitoring	Body worn
RFID	Activity monitoring	Body worn/Embedded
Temperature/CO2/Smoke sensor	Safety	Embedded
Vital sign/physiological	Health care	Body worn/Embedded
<b>Camera</b>	<b>Fall detection, Activity monitoring, Security, Safety, Health care</b>	<b>Embedded</b>

Table 2: Sensors used in multi-modal approaches

Sensor	Project	GER'HOME	HERMES	UbiMon	Toreyin <i>et al.</i> [62, 63]
Light sensor	X				
Humidity sensor	X				
Temperature sensor	X		X		
Presence sensor	X			X	
Pressure sensor	X				
Water sensor	X				
Electrical sensor	X				
Door and window sensor	X				
Open / close sensor	X				
ECG			X		
Accelerometer			X		
Blood pressure monitoring sensor			X		
Audio sensor		X	X	X	
<b>Video sensor</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>	

person tracking and face identification. The UbiMon project [44] proposed a framework to collect information from wearable and implantable sensors to a local processing unit such as: PDA or smart phone through a body network, before transmitting to a remote database. Toreyin *et al.* [62] proposed to use a similar algorithm based on hidden Markov models to detect fall from video [62], audio, passive infra-red and vibration sensors [63]. A list of sensors used in these multi-modal approaches is shown in Table 2. In conclusion, additional sensors, such as, audio sensors can offer complimentary feature parameters and robustness to existing vision based behaviour monitoring systems while bio-medical sensors provide different types of information to the complete AAL system.

## 2.4 General Structure of Video Based AAL

A video based monitoring typically has the following system blocks as shown in Fig. 1.a):

1. video data capture,
2. a local video multiplexer and transmitter which collects video data from different camera hardware (wired or wireless connection) and transmits to a central server,
3. a central server which collects and stores data from many local transmitters at remote locations,
4. a video processing system block which analyses video data for action or activity recognition and offers output alerts or decisions,
5. operator to monitor the video data before or after processing.

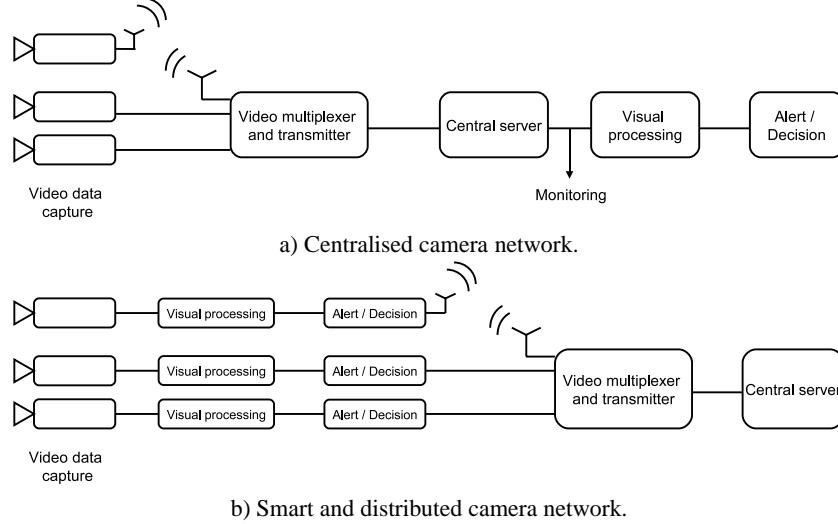


Figure 1: Video monitoring system blocks within AAL.

Such a centralised camera network often requires high bandwidth for video data transmission and an alternative solution is suggested in [17] that uses a smart and distributed camera network which processes the video data locally and sends the alert or decision output to the central server as shown in Fig. 1.b). Distributed smart camera solutions also have advantages in terms of confidentiality (see Section 4). In another approach addressing the bandwidth issue in centralised camera networks, compressed video transmission is suggested in [33].

During the video data capture, traditionally a network of fixed cameras fitted on the walls are used to cover the target areas. In such a scenario, the viewing angle of the individual's movement causes a significant difference in the action recognition processes. Viewing angles of the individual's movement are categorised as follows [33]:

- Object motion towards camera,
- Object motion away from camera and
- Object movement horizontal to camera. Sideways camera location.

In order to avoid such view angle problems, view invariant computer vision algorithms [51] using a single camera or image fusion based multi camera setup [17] or



Figure 2: Omnidirectional camera to capture 360° view [66].

an omnidirectional camera [66] are suggested. In the case of an omnidirectional camera, the system can capture a 360° viewing angle of video data and process accordingly as shown in Fig. 2.

### 3 Visual Processing

In this section we first give details of the computer vision algorithms that can be used to recognise short term actions and higher level activities for behaviour monitoring, then we provide an overview of approaches for health measurements from video or images. Then we discuss video processing challenges.

#### 3.1 Action Recognition Algorithms

The processing block is one of the most important blocks in video monitoring systems (see Fig. 1) and uses various computer vision approaches [2, 23, 39, 64, 65]. This section focuses on visual processing for action recognition. Approaches to link actions into activities are summarised in section 3.2.

Fig. 3 represents the taxonomy of computer vision approaches for action recognition in AAL. Actions are usually defined either by the individual postural features (e.g. the person is bending over, lying, squatting, etc.) or by the individual ambulatory information obtained through motion tracking. In turn postural features can be established using static approaches which analyse the posture of an individual from a single image and dynamic approaches which include the temporal information in movement detection. The postural feature recognition process often includes a preliminary step called *background subtraction* (section 3.1.1) followed by *posture* (section 3.1.2) or *movement recognition* (section 3.1.3). Motion tracking based approaches are discussed in section 3.1.4. Table 3 summarises the action recognition algorithms used in the literature.

##### 3.1.1 Background subtraction

Background subtraction is popularly used to separate moving objects (human shape here) in the scene by segmenting into background and foreground [64] and many methods are proposed in the literature [9, 14, 27, 54]. In its simplest form the background modelling considers a stationary camera with a stationary or nearly stationary background to detect the relative movement of human activity in the scene.

A comparison of the intensities within the frames is done over time to estimate the static pixels which is then

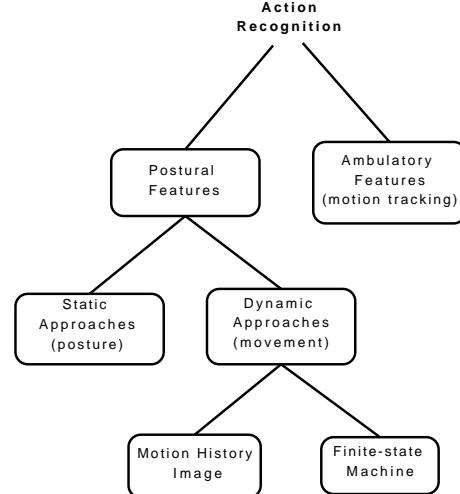


Figure 3: Taxonomic scheme of computer vision approaches for action recognition in AAL

modelled as background and the pixels can be represented by multivariate Gaussian values in HSV (hue, saturation and value) colour space [57].

However, in a real scenario, many other parameters, such as, lighting condition, long-term scene change, small cluttered movement, ghost and shadow, can affect the background modelling. To address such problems, Stauffer and Grimson [54] proposed an adaptive backgrounding to reliably deal with lighting changes, repetitive motions from clutter and long-term scene changes; Cucchiara *et al.* [9] used a statistical approach using the knowledge of moving objects, shadows and ghost objects from the previous frames to update the background; Elgammal *et al.* [14] proposed a non-parametric model to handle clutters and small motions in the background by estimating the probability of observing pixel intensity values.

In [15], Eng *et al.* propose a Bayesian framework to deal with background subtraction under challenging conditions. This framework could provide more robust background subtraction when the foreground object is matches the background (e.g. the individual having clothes similar to the background or if the scene is very dark) or if two objects in the foreground are not separable due to occlusion.

Table 3: Different approaches for action recognition. The silhouette is normally extracted using background subtraction (BS). Posture can be represented by the bounding box features (BoundingBox), a 5 dimensional vector (5D vector) representing region of interest features or projected histograms (Hist) and recognised using k-nearest neighbour (KNN). Motion can be represented by motion history image (MHI) and recognised using motion quantification (MQ), distance between histograms (HistDist), Multi-Layer Perceptron (MLP), finite state model (FSM), hidden Markov model (HMM) or layered HMM (LHMM).

Reference	Silhouette Extraction	Posture Posture	Movement Movement	Motion Tracking	Detected Actions
Hezelhoff <i>et al.</i> [43]	BS	BoundingBox + KNN	-	-	Standing, Sitting, Bending, Lying, Lying Toward
Rougier <i>et al.</i> [52]	BS	BoundingBox	MHI+MQ	-	Fall detection
Foroughi <i>et al.</i> [18]	BS	-	MHI+MLP	-	Walk, Run, Stumble, Limp
Tao <i>et al.</i> [58]	BS	BoundingBox	FSM	-	Forward Fall, Backward Fall, Sideway Fall, Bend down, Sit down, Lie down
Töreyin <i>et al.</i> [62]	BS	BoundingBox	HMM	-	Sideway/Backward/Forward Fall detection
Thome <i>et al.</i> [61]	BS	BoundingBox	LHMM	-	Walking, Falling
Chen <i>et al.</i> [8]	BS	5D vector	HMM	-	Walking, Falling
Lin and Ling [33]	BS	Hist	HistDist	-	Elopement
Nait-Charif and McKenna [40]	-	-	-	Overhead tracking	Walking, Squatting, Falling, Moving toward / away / horizontally with regard to the camera
Rougier <i>et al.</i> [51]	-	-	-	3D head tracking	Activity recognition, Fall detection
					Fall detection
					Fall detection

sion.

### 3.1.2 Static Approaches

Posture is thought to be a good indicator of a person's activity and events such as falls. In static posture analysis the video sequence is analysed using frame-wise image processing. Typically, the first stage consists of extracting the silhouette from the image using background subtraction.

*Aspect ratio measurement* is one of the simplest and quickest methods in this category and used in many video based AAL algorithms, particularly in *fall detection* [17, 33, 43, 58]. Aspect ratios are calculated by detecting the bounding box of the foreground segmented moving object (silhouette) and constructing approximated ellipses, rectangles etc. The bounding box is estimated by finding the furthest (leftmost, rightmost, topmost and bottommost) foreground pixels. However sometimes noises and shadows can introduce error (larger size of human object) in the bounding box calculation. To avoid such error, the foreground pixels are projected onto x and y axes by

calculating the number of changed pixels row-wise and column-wise to build horizontal and vertical projection histograms. The bounding box can be identified by applying a threshold on generated histograms.

Images can be classified into different postures using the bounding box aspect ratios [58]. Emmanuel [43] proposed to classify postures using the k-Nearest Neighbour (KNN) algorithm. The KNN algorithm compares histograms of the stored posture images with histograms of the current input frame to find the most similar posture. The authors proposed an additional bounding box angle test to filter out false detection between a bending posture and lying towards camera.

### 3.1.3 Dynamic Approaches

Static approaches have strong limitations. For example, static approaches alone cannot differentiate a lying posture due to a fall or intentional lying. The static approaches are often restricted to a range of movement patterns as it does not consider the temporal information during posture detection. To overcome such shortcom-

ings dynamic approaches have been proposed considering movement information within the postural feature extraction stage.

As a first attempt to include temporal information into posture classification, in [43], the time of the last standing position is preserved in order to differentiate between normal *lying* and a real *fall*. The algorithm checks the difference between time of lying ( $t_{lying}$ ) and last standing ( $t_{standing}$ ) and a *fall* is declared if  $(t_{lying} - t_{standing})$  is less than a specified threshold.

Other ways to model the dynamic nature of a movement include 1) methods which combine several frames into one single image representing the silhouette at different points in time [18, 52] and 2) methods which use finite state models (FSM) to represent the sequence of posture information [8, 58, 61, 62].

Rougier *et al.* [52] suggested a Motion History Image (MHI) based solution. The MHI is a special representation of a foreground object where the intensity of the pixels are derived according to recency of the motion within the image sequence and gives the most recent movement of a person during action. However, a repeated motion in the same position results in a similar MHI and Foroughi *et al.* [18] proposed an alternative solution to overcome this issue. A spatio-temporal template is proposed to store occurrence time of motion emphasizing the final motion. Finally, in both cases, the derived dynamic images are converted into a scalar valued image using a motion quantification coefficient [52]. In [18], Foroughi *et al.* use eigenspace to reduce the dimensionality of the MHI. The resulting vector is classified into a set of postures using a Multi-layer Perceptron (MLP).

Tao *et al.* [58] proposed to model the sequence of bounding box ratio using a two state finite state machine (FSM). The state transitions indicate changes between seating, standing and lying posture. Similarly, Töreyin *et al.* [62] used combined audio and video features as inputs to a hidden Markov model to propose a robust *fall* detection. In a dementia patient elopement monitoring scheme, Chen *et al.* [8] outlined the region of interest (ROI) at possible exit points of the unit and extracted 5 dimensional feature vectors corresponding to number of foreground pixels, geometric mean and variance of the ROI. An HMM is used here comprising 5 hidden states to model elopement sequences. Finally, Thome *et al.* [61] propose a system based on a Layered HMM (LHMM)

structure which allow to recognise actions (described by posture) as well as higher level behaviour patterns (activities) within the same model. The LHMM segments the problem into distinct layers that operate at different temporal granularities. The proposed approach implement a semantic of activities, other algorithms to translate actions into activities are presented in section 3.2.

### 3.1.4 Motion Tracking

Apart from postural information, other important features that can be used to recognise activity are based on tracking of positional information of the individual.

In one of the initial approaches of tracking based algorithms, Nait-Charif and McKenna [40] proposed an overhead tracking scheme, where the target person is tracked using an ellipse representation so that the state at time  $t$  can be described as  $\mathbf{e}_t = (x_t, y_t, \psi_t, s_t, e_t)$  where  $(x_t, y_t)$  is the ellipse center,  $\psi_t$  is orientation,  $s_t$  is scale and  $e_t$  is the eccentricity. The tracking of this ellipse is then performed by applying an Iterated Likelihood Weighting particle filter [41] to provide trajectories in 5D ellipse parameter space followed by a temporal smoothing operation. These smoothed trajectories and the person's speed (estimated at every point) provide a compact representation of the person's global motion. The global motions are useful to annotate the activities and make a decision in case of an event. However, this scheme is based on 2D vertical velocity estimation and may not distinguish between similar motion and posture information, such as a real fall and a person sitting down. In addressing these issues Rougier *et al.* [51] used a 3D head tracking model which reduces the effect of perspective foreshortening. The scheme used the POSIT algorithm [10] to estimate the relative position of a 3D head ellipsoid in the camera coordinate system and the 3D tracking is performed by using a conditional density propagation based particle filter [24]. At every position of the ellipse, a 3D pose of the head is calculated to obtain the 3D trajectory which is analyzed to detect a *fall*.

### 3.1.5 Combination of algorithms:

The algorithms presented above do not always achieve the degree of sensitivity required for real life applications, therefore, some researchers have proposed to combine some classifiers to achieve better results.

For example, Emmanuel [43] observed that their approach based on projected histograms and KNN sometimes classifies other postures wrongly as 'lying toward' and propose a bonding box angle test to remove uncertainty. Similarly, Lin and Ling [33] proposed to reduce the number of false alarms by using a three stage approach which sequentially apply three different algorithms to detect a fall:

1. Detection of motion and duration of an event (typical fall time 0.4s~0.8s),
2. Human shape detection, location and change of rate of the centroid of the shape, and
3. Vertical projection histogram calculation and posture classification, *i.e.*, fall, walk and squat.

Rougier *et al.* [52] propose a sequential scheme for fall detection which uses both static and dynamic analysis of the scene. First, the system detects the motion of falling, then a static posture recognition algorithm is used to detect if the individual is lying or standing and finally, the system checks if the individual is on the ground without moving. If the three sequential tests are positive, the fall is detected.

Note that the three mentioned systems are based on a combination of static and dynamic posture recognition. To our knowledge, no action recognition system for VAAAL has been proposed which combines postural features with motion tracking features. These two features are, however, likely to be complementary and could potentially improve general action recognition performance. Furthermore, the three approaches combine the results of multiple classifiers using rather simplistic AND and OR rules to take the final decision. The final decision could also be made by weighting the results of the different classifiers or use a more advanced theoretical framework [29] for classifier combination.

### 3.2 Monitoring Behaviour

Most video based AAL applications are concerned with recognising and monitoring human behaviour. Exceptions are the methods which take health measurement from video data discussed in Section 3.3.

Currently the majority of the literature focuses on fall detection [18, 32, 33, 35, 51, 52, 57, 61, 62], while some

efforts have been made to detect elopement of patients from a dementia unit [8]. However the scope of video based monitoring is much wider and yet to be explored fully. Table 4 shows a list of behaviour that can potentially be monitored through video based systems and the corresponding domain of application.

**Table 4: List of behaviours monitored in AAL**

Activities of Daily Living (ADL):
<i>Bathing, Dressing, Grooming, Walking, Climbing stairs, Eating, Cooking, Managing Medication, Using Phone, Housework, Doing laundry, etc.</i>
Sleep/Awake pattern
Immobility/Activity pattern
Elopement
Fall

Although *action* and *activity* often refer to a similar concept, *action* is more related to a simple motion pattern such as opening a door, turning the oven on or falling whereas *activity* refers to complex sequences such as cleaning a room or cooking. Actions are the primitive components used to define an activity. For some applications such as fall detection, the task consists in detecting particular actions, however, in other cases the application requires to recognise higher level activities. Activities are composed of primitive actions linked in a spatio-temporal structure. Note that activities may be composed of several sub-activities which can eventually be broken down into actions. Behavioural monitoring usually involves a hierarchical structure. The lower level consists of recognising simple actions using the computer vision techniques presented in section 3. The output of the action detectors provides the input stream for a higher level activity recognition component.

Inspired by research in natural language processing, stochastic context free grammar parsing is a popular approach to recognising semantically defined activities. With this approach, for each activity, a grammar is defined specifying expected orders of sub-activities and actions. Parsing mechanisms are used to translate the action sequence into activities description. Note that apart from the actions ordering, no notion of temporality is defined in the grammar. In [25], Ivanov *et al.* propose such a system which has been demonstrated to recognise car activities in a car park. Such a grammar representation and parsing mechanism is well suited to modelling activities for AAL.

More recently researchers have addressed this problem by using dynamic Bayesian networks (DBN) to model activities. DBN also encodes expected actions and the ordering constraints for each considered activity, however DBN approaches can enforce temporal constraints that provide powerful contextual cues to recognising activities. Laxton *et al.* [31] propose a DBN approach which model the time period between actions. A limitation of both stochastic context free grammars and DBNs is that the semantic must be specified *a priori*. In the context of AAL, each individual could perform a particular activity in their own way and it is unlikely that the same grammar definition could fit all users. It would thus be beneficial if the system could automatically learn the grammar or topology for each user using recorded data which is not the case for existing solutions.

### 3.3 Physiological monitoring

Another very promising application is the contact free measurement of physiological data through the use of video cameras. The ability to monitor a patient's physiological measurements has been demonstrated by several research papers [16, 21, 48, 56].

Fei and Pavlidis [16] proposed a method to recover the breathing waveform of patients using thermal video. First, the nostril region is segmented and tracked over time, then the breathing information is extracted through wavelet analysis. Garbey *et al.* [21] also used thermal imaging, developing a method to measure cardiac pulse based on a thermal signal emitted from major superficial vessels of the face. Takano and Ohta [56] proposed a low cost solution to measure both the respiratory and pulse rate from a standard camera. The brightness of a region of the face is measured for 30s. Series of operations are applied to the generated signal of brightness: the derivative, a low pass filter, and spectral analysis. The peaks in the spectral signal are then used to determine the respiratory and the pulse rate.

Poh *et al.* [48] have taken a step further by proposing a high accuracy cardiac pulse measurement from a basic webcam even in the presence of head movements. Moreover they show that the approach can be extended to simultaneous measurement of multiple people. First an automatic face tracker detects faces in the video, then the extracted face image is separated into the three RGB

channels. The three signals are normalised and decomposed into three independent components using independent component analysis (ICA). The power spectrum is then extracted from one of the components and the pulse frequency is designated as the frequency that corresponds to the peak within an operational range of frequency. This approach has proven to be effective even in the presence of head movements with a mean bias of only 0.64 bpm.

### 3.4 Discussion and Challenges

Table 3 summarises the action recognition algorithms used in the literature as well as the intended recognition task. Most of the proposed systems are devoted to detecting falls or to differentiate falling from walking. Other papers present systems to classify a larger number of actions such as squatting, bending or eloping or different types of falls (forward, backward or sideways) and Nait-Charif and McKenna [40] recognise ambulatory information to summarise activities and detect falls. VAAL is a relatively new domain and, supported by the success of achieving these tasks, the rich content of video sensing has the potential to recognise more detailed actions such as 'opening a cupboard' or 'picking up an object' which would be very useful information for AAL applications.

Some issues are, however, still to be addressed fully such as dealing with occlusions. Indeed, in home environment filled with furniture and various objects, it is impossible to cover the totality of a room using a single camera. Hardware solutions to this problem include the use of an overhead camera. Alternatively, the solution proposed in [8] make use of multi-views of the scene by combining features from multiple cameras covering the same area.

Another challenge to be addressed is the multi-occupancy of the accommodation. While most AAL solutions are targeted for people living alone, they are still likely to receive visitors, therefore algorithms for identity recognition (e.g. face recognition) and tracking of multiple people should be included in VAAL systems.

## 4 Acceptability and Privacy

As seen in previous sections, AAL technology can potentially provide a wide range of solutions to support older people in their own homes and improve their quality of

life. However, to be applicable to a large scale the technology must be accepted by potential users and their families. Some users show some anxiety toward new technologies or are concerned with anaesthetic and invasive sensors (or cameras) fitted in their home. However, the main concern over the adoption of AAL in general and VAAL in particular is the loss of privacy. Maintaining privacy is thus a very important consideration when developing VAAL technology.

In the first part of this section the technical challenges to protect privacy are discussed, while the second part is an overview of acceptability studies.

#### 4.1 Data and Privacy Protection

In the current literature only a handful of papers [17, 69, 70] discussed the issue of data protection in AAL. Sensor based telecare information security is discussed in [70], whereas video based privacy solutions are proposed in [17] and [69].

Data protection, including medical data, is crucial in the context of AAL. The data should be stored securely in a centralised server and access is controlled by various authenticity measures such as password protection. The main threats involved are reply attacks (a network attack to send data repeatedly), on-line / off-line password guessing attacks, impersonation attacks (providing false data to legitimate user) or stolen-verifier attacks (stealing passwords by insider member). Increased security solutions are proposed by Wu *et al.* [70] by using a two stage authentication process which involves a cryptographical session key followed by password authentication. In conclusion, preserving video or sensor based AAL data is an *information security* problem and can be addressed by solutions from database management, cryptography or data mining related fields.

When dealing with video recordings in users' homes, the privacy protection issues are predominant. In most VAAL systems, some authorised people (referred to as carers) have access to a level of information recorded and processed by the system. The most sensitive level being the live video or recorded footage and the less sensitive level being the automatically interpreted information (e.g. a fall detected). Fig. 4 shows a classification of information of VAAL systems. Different carers may have access to different levels of information and access to a level of

information could be open to a carer only in specific situations. For example, when a critical situation is detected by the system, clinical professionals may need to access the live or recorded footage, therefore a hierarchical access control can address such issues.

#### Sensitivity of information

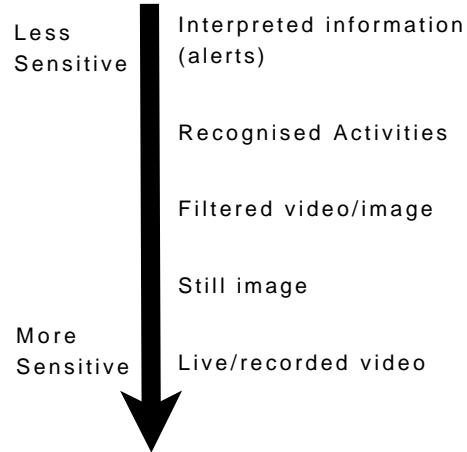


Figure 4: Sensitivity of information of VAAL systems.

In an attempt towards providing security and privacy in video based monitoring for AAL, Fleck and Strasser [17] discussed various aspects and proposed a solution using a smart camera network. In a traditional centralised monitoring system, a carer can have direct access to the privacy sensitive footage (Fig. 5,(1)); or a network intruder (during channel transmission or in the server) can fetch the data illegally (Fig. 5,(2)); or the direct camera hardware can be targeted (Fig. 5,(3)). Using a smart camera, part of the processing is performed locally and the more sensitive information is not transmitted to the server or a carer, thus reducing the vulnerability of the system. Smart cameras can either process the video locally and transmits decision information as in [71] or they can filter the video to obscure individual identities or actions as in [69] and [17]. Several computer vision approaches have been proposed to remove private information from the video data such as using augmented symbols of the target or silhouette of the region of interest as shown in Fig. 6. Table 5 summarises

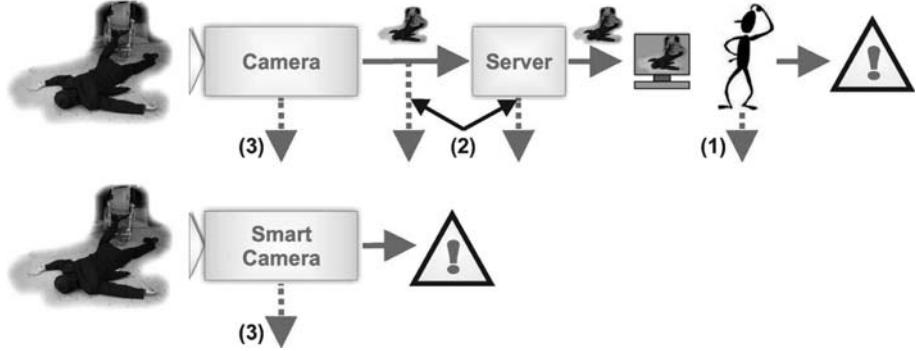


Figure 5: Centralised & distributed camera network and privacy issues. Figure is adopted from [17].

several privacy filter solutions.

Table 5: Privacy filter solutions

Ref	Filter
Fleck <i>et al.</i> [17]	Augmented symbols of the target or silhouette of the region of interest
Winkler <i>et al.</i> [69]	Edges of the region of interest
Dufaux <i>et al.</i> [12]	Scrambling regions of interest
Gross <i>et al.</i> [22]	Model-Based Face De-Identification



Figure 6: Example of individual privacy: silhouette of region of interest. Figure is adopted from [17].

A similar approach is taken in [69] to protect privacy and the proposed solution focuses on four different aspects:

**Integrity:** Any manipulation in video transmission must be detected, *e.g.*, using checksum or digital signature.

**Authenticity:** Evidence of source of data, *i.e.*, camera identification etc.

**Confidentiality:** Privacy protection by encrypting the region of interest.

**Multi-level user control:** Different level of protection on confidential data using user access control.

## 4.2 User Requirements and Acceptability

The Royal Academy of Engineering discusses the dilemmas of privacy and surveillance in [45]. The dilemmas arise from the impossibility of weighing up the benefits of surveillance systems and the potential loss of privacy. In AAL technology, only the end-user perspective can help to define the right balance.

Marquis-Faulkes *et al.* [36] present a qualitative user requirement study for the use of video technology for fall detection and activity monitoring. Four focus groups were conducted with a total of 37 participants. Three groups were composed of older people with different levels of dependence and the last group comprised sheltered housing wardens. One outcome of the study was that the participants were generally content to have a video based monitoring device provided that the video footage was analysed only by computer and that no-one would actually watch them. The participants in the user group with the highest level of independence expressed concerns about carers having access to too much information about their

activities but were favourable to a system that would alert carers in case of fall or abnormal activity detected.

Turgeon Londei *et al.* [35] explore the receptivity and requirements of older people regarding the introduction of an intelligent video monitoring system in Canada. Interviews were conducted with 25 older people with a history of falls. Both qualitative and quantitative data were reported and analysed. The perception toward video monitoring was generally good with 96% of the participants at least partially favourable to the intelligent video monitoring system and 88% accepted the sending of an image to a designated carer. The participants were presented with different types of images that could be sent to a designated carer. The proposed images were the original image and images processed with blurring, pixelisation or silhouette extraction. Interestingly, nearly all the participants chose to send the original image rather than modified images. Although 19 participants originally refused to be videotaped in one or more room, 8 of them would accept if the original image was modified. Note that only half of the participants who were favourable to the system did not agree to use it, claiming that the system was too complex, that they were not old or ill enough or that they did not need more protection.

Both studies [35, 36] suggest that VAAL would be accepted in principle by users if they feel that the technology would make a real difference to their security and quality of life. However, while all the participants of the study from Marquis-Faulkes *et al.* [36] would not allow anyone to see the video footage, most participants from the study from Turgeon Londei *et al.* [35] would accept sending original images to a designated carer in the case of a critical situation detected by the system. It could be hypothesised that, in the former study the participants did not discuss the possibility of a still image sent to the carer only in the case of a critical situation but were referring to a constant surveillance by an operator. The second study included only participants with a history of falls, therefore, potentially with a greater need for a fall detector.

### 4.3 Discussion

Acceptability and privacy are key elements to consider in the design of new VAAL systems. As seen previously, the technology exists to protect users' privacy against access by unauthorised individuals. The question still to be an-

swered is: What level of information could be accessed by designated carers? To answer this question, the right balance between efficacy and loss of privacy should be found. The requirement studies presented in the previous section provide good indicators, however the privacy issues of video based monitoring have not been fully explored in line with the requirements from the users at both ends of AAL systems, *i.e.*, patients and clinical professionals. It is also believed that by improving the technology, the VAAL systems can become more reliable and consequently depend less on human operations, thus improving confidentiality.

It is also apparent from the requirement studies that the design of the system might not be fully understood by potential users. It is reported in [35] that some participants did not make any distinction between the image analysis executed by a computer and the images seen by authorised carers. As a consequence, it can be expected that some users can be reluctant to adopt a system which involves a camera even if the video is never transmitted or stored and only abstracted information is sent to a server as proposed by Yang *et al.* [71].

Both acceptability studies [35, 36] were conducted with a relatively small number of participants and it would be interesting to study the user requirements at a larger scale. As mentioned by Broadbent *et al.* in a review of acceptance of health care robots for older population [4], research studies show cultural differences in attitudes towards robots. It is likely that similar cross-cultural differences could be observed in acceptability of VAAL and large scale cross-cultural acceptability studies would benefit the domain.

Social implications of AAL should also be considered, for example Sun *et al.* [55] observe that the transfer of the dependence from human side to technology also reduces the social connections of the assisted people.

## 5 Evaluation Methodology

### 5.1 System Evaluation

VAAL systems can be evaluated at multiple levels:

1. **Technical performance evaluation** consists in measuring the accuracy to which the proposed system

achieves the designed tasks. Technical performance evaluation strategies are discussed in section 5.2.

2. **Outcome evaluation** aims to determine the effect of the system on users, carers and society. While dependent on the technical performances, outcome evaluation is strongly linked to the way carers, clinicians and emergency services respond to the information provided by the system. A typically used methodology to evaluate the effectiveness of health services is to 1) recruit users fulfilling the eligibility criteria, 2) select randomly a group of users who will be provided the technology and associated services and 3) compare the outcome measures for the group receiving the technology and the group receiving traditional care. To our knowledge, no VAAL has been evaluated on a large scale clinical trial yet, however, for an example of clinical evaluation of AAL based on a network of sensors, see Brownsell *et al.* [6].
3. **Privacy protection validation** aims to insure that the system fulfils the required level of security. Wu *et al.* [] compare telecare information systems according to their vulnerability to several types of attacks (Replay attacks, Password guessing attacks, etc.). Some international standards for computer security certification also exist such as the *Common Criteria for Information Technology Security Evaluation* [68].
4. **User acceptability evaluation** aims to analyse how the system is perceived by potential users. In [11], Demiris *et al.* developed a questionnaire to assess patients' impressions of the risks and benefits of home telecare.

## 5.2 Technical Performance Evaluation

Many approaches have been recently proposed to tackle VAAL, however the community is lacking a consistent methodology for evaluating model and system technical performance. Key for the field to progress and to be adopted, is experimental validation and presentation of experimental results.

Technical performance evaluation aims to characterise how well the recognised activities match the true sequence of activities (the groundtruth). Although, activ-

ity and action recognition can be a multi-class problem, the performance evaluation is typically done a class at a time which leads to the performance evaluation of a binary classification.

For action recognition tasks such as fall detection, the technical evaluation typically estimates the detection rate ( $TPR = \text{true positive rate}$ ) and the false alarm rate ( $FPR = \text{false positive rate}$ ). Most algorithms can be adjusted to reduce the number of missed detections with the effect of increasing the number of false alarms. A popular technique to compare, select and visualise performance of such systems is to plot receiver operating characteristic (ROC) curves which depict the tradeoffs between  $FPR$  and  $TPR$  [13].

The evaluation of performance of activity recognition is more complex and no common methodology has been adopted yet. Indeed, the characterisation of activity recognition performances using  $FPR$  and  $TPR$  does not take into account other types of errors such as fragmentation, merging and timing offset. Fig. 7 shows examples of the different types of error generated by activity recognition systems.

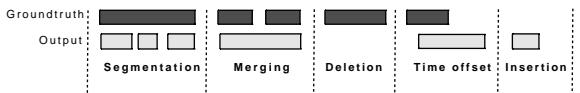


Figure 7: Examples of different types of error which can be generated from activity recognition.

Ward *et al.* [67] have recently proposed a comprehensive overview and critique of existing performance metrics for activity recognition. These metrics typically use one of two basic units, namely *frames* and *events*. The frame based scoring decomposes the time in frames of fixed-length unit. For each frame the groundtruth and recognised activity are matched making it possible to find a  $FPR$  and  $TPR$ . The scoring based on events compares groundtruth and recognised sequences of activities defined by the start time and stop time. A recognised activity can then be scored as either correctly detected or falsely inserted or deleted. We note that a definition of *correct* detection should be provided. For example Tapia *et al.* [59] propose to label an activity as correctly detected if it is detected within a time interval around the end of the groundtruth activity.

As stated in [67], both scoring approaches suffer from limitations. The frame based scoring does not distinguish between the types of errors which means that an undetected activity (i.e. deletion) could have less effect on the error rate than a timing offset in an otherwise correctly detected activity. On the other hand, event based scoring can differentiate time offset, deletion and insertion but is unable to account for fragmented and merged activities. Moreover, there is no clear definition of what is a *correct* activity and each researcher seems to provide their own definition. In light of the problems with existing approaches, Ward *et al.* [67] propose a novel set of performance metrics for activity recognition which can characterise insertion, merging, time offset (overfill and underfill), deletion and fragmentation.

In the literature, results are often presented using video sequences of simulated falls, often recorded by the researchers themselves. Unfortunately, each proposed approach is evaluated on a separate dataset which make comparisons impossible. We would thus recommend that a multidisciplinary group of researchers join efforts to build a large and realistic database of video sequences of falls with an associated evaluation protocol that could be put together and used by the research community. For obvious ethical reasons, it might not be possible to collect a dataset with actual falls of older people but the simulated falls should be made as realistic as possible by consulting specialist clinicians.

## 6 Conclusion

Video and computer vision technologies are developed to aid ambient assisted living with particular emphasis in the detection of falls. However the scope of the domain is rapidly expanding towards new applications related to rehabilitation and care of older people or personal activity management. In our view, three main challenges need to be addressed for the domain to reach maturity:

- 1. The Technical Challenge:** The technological solutions proposed for VAAL in the last few years, particularly to tackle fall detection, are very promising but robustness to real-world conditions needs to be addressed. The rich information provided by video solutions could also be used to solve problems currently tackled by conventional sensor based systems.

Efforts should be made to explore further new applications.

- 2. The Acceptability Challenge:** To be successful, VAAL systems must be accepted by users, carers and society. The existing acceptability studies suggest that users would accept VAAL in principle with the condition that they feel that it would make a real difference to their security, health and quality of life. However, the studies also show that only a small proportion of the interviewees would actually use the proposed systems. The gap between the acceptability in principle and the consent to adopt the technology can only be reduced by developing the technology together with the study of user needs.
- 3. The Integration Challenge:** An integration of technical solutions to build an AAL based society is still far from reality. Steps need to be taken towards the integration of VAAL including a standardisation performance evaluation, as seen in section 5, and the integration of VAAL with existing telecare services. One of the great advantages of video technology for AAL is that the same hardware setup can be used for multiple applications to help older people to live independently (see Table 1). The integration of multiple applications in a single system could make the solution more attractive.

## References

- [1] AALIANC. Ambient assisted living roadmap, 2010. <http://www.aaliance.eu/public/documents/aaliance-roadmap>.
- [2] J. K. Aggarwal and S. Park. Human motion: modeling and recognition of actions and interactions. In *Proc. 3D Data Processing, Visualization and Transmission 3DPVT*, pages 640 – 647, 2004.
- [3] M. Alwan, P. J. Rajendran, S. Kell, D. Mack, S. Dalal, M. Wolfe, and R. Felder. A smart and passive floor-vibration based fall detector for elderly. In *Proc. Information and Communication Technologies (ICTTA)*, volume 1, pages 1003 –1007, 2006.

- [4] E. Broadbent, R. Stafford, and B. MacDonald. Acceptance of healthcare robots for the older population: Review and future directions. *International Journal of Social Robotics*, 1:319–330, 2009.
- [5] J. Broekens, M. Heerink, and H. Rosendal. Assistive social robots in elderly care: a review. *Gerontechnology*, 8:94–103, 2009.
- [6] S. Brownsell, S. Blackburn, and M. S. Hawley. An evaluation of second and third generation telecare services in older people’s housing. *Journal of Telemedicine and Telecare*, 1:8–12, 2008.
- [7] G. Cambul, M. R. Sahin, and M. E. Karshgil. Automatic fall detection in real time video based applications. In *Proc. IEEE Signal Processing and Communications Applications*, pages 1–4, 2006.
- [8] D. Chen, A. J. Bharucha, and H. D. Wactlar. Intelligent video monitoring to improve safety of older persons. In *Proc. IEEE International Conference on Engineering in Medicine and Biology Society (EMBS)*, pages 3814–3817, 2007.
- [9] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati. Detecting moving objects, ghosts, and shadows in video streams. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 25(10):1337 – 1342, Oct. 2003.
- [10] D. F. DeMenthon and L. S. Davis. Model-based object pose in 25 lines of code. *International Journal of Computer Vision*, 15:123–141, 1995.
- [11] G. Demiris, S. Speedie, and S. Finkelstein. A questionnaire for the assessment of patients’ impressions of the risks and benefits of home telecare. *Journal of Telemedicine and Telecare*, 6:278–284, 2000.
- [12] F. Dufaux and T. Ebrahimi. Scrambling for video surveillance with privacy. In *Proc. 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW)*, page 160, 2006.
- [13] J. P. Egan. Signal detection theory and ROC analysis. *Series in Cognition and Perception*, 1975.
- [14] A. M. Elgammal, D. Harwood, and L. S. Davis. Non-parametric model for background subtraction. In *Proc. 6th European Conference on Computer Vision-Part II (ECCV)*, pages 751–767, 2000.
- [15] How-Lung Eng, Junxian Wang, A.H. Kam, and Wei-Yun Yau. A bayesian framework for robust human detection and occlusion handling human shape model. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 2, pages 257 – 260 Vol.2, 2004.
- [16] J. Fei and I. Pavlidis. Thermistor at a distance: Unobtrusive measurement of breathing. *IEEE Transactions on Biomedical Engineering*, 57(4):988 –998, Apr. 2010.
- [17] S. Fleck and W. Strasser. Smart camera based monitoring system and its application to assisted living. *Proceedings of the IEEE*, 96(10):1698 –1714, Oct. 2008.
- [18] H. Foroughi, A. Naseri, A. Saberi, and H. S. Yazdi. An eigenspace-based approach for human fall detection using integrated time motion image and neural network. In *Proc. International Conference on Signal Processing (ICSP)*, pages 1499 –1503, 2008.
- [19] J. A. Freer, B. J. Beggs, H. L. Fernandez-Canque, F. Chevrier, and A. Goryashko. Automatic intruder detection incorporating intelligent scene monitoring with video surveillance. In *Proc. European Conference on Security and Detection (ECOS)*, pages 109 –113, 1997.
- [20] M. Fujita. AIBO: Toward the era of digital creatures. *International Journal on Robotic Research*, 20(10):781–794, 2001.
- [21] M. Garbey, S. Nanfei, A. Merla, and I. Pavlidis. Contact-free measurement of cardiac pulse based on the analysis of thermal imagery. *IEEE Transactions on Biomedical Engineering*, 54(8):1418 – 1426, Aug. 2007.
- [22] R. Gross, L. Sweeney, F. D. T. Frade, and S. Baker. Model-based face de-identification. In *Proc. 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW)*, pages 161 – 168, 2006.

- [23] S. Hongeng, R. Nevatia, and F. Bremond. Video-based event recognition: activity representation and probabilistic recognition methods. *Computer Vision and Image Understanding*, 96(2):129 – 162, 2004.
- [24] M. Isard and A. Blake. CONDENSATION - conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29:5–28, 1998.
- [25] Y. A. Ivanov and A. F. Bobick. Recognition of visual activities and interactions by stochastic parsing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):852 –872, Aug. 2000.
- [26] B. Jansen and R. Deklerck. Context aware inactivity recognition for visual fall detection. In *Proc. Pervasive Health Conference and Workshops*, pages 1 –4, 2006.
- [27] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis. Real-time foreground-background segmentation using codebook model. *Real-Time Imaging*, 11(3):172 – 185, 2005. Special Issue on Video Object Processing.
- [28] S. Kim, T. Schap, M. Bosch, R. Maciejewski, E. Delp, D. Ebert, and C. Boushey. Development of a mobile user interface for image-based dietary assessment. In *Proc. 9th International Conference on Mobile and Ubiquitous Multimedia*, 2010.
- [29] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):226 –239, Mar 1998.
- [30] T. Ko. A survey on behavior analysis in video surveillance for homeland security applications. In *Proc. IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, pages 1 –8, 2008.
- [31] B. Laxton, L. Jongwoo, and D. Kriegman. Leveraging temporal, contextual and ordering constraints for recognizing complex activities in video. In *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 1 –8, 2007.
- [32] T. Lee and A. Mihailidis. An intelligent emergency response system: preliminary development and testing of automated fall detection. *Journal of Telemedicine and Telecare*, 11(4):194–198, 2005.
- [33] C. W. Lin and Z. H. Ling. Automatic fall incident detection in compressed video for intelligent homecare. In *Proc. International Conference on Computer Communications and Networks (ICCCN)*, pages 1172 –1177, 2007.
- [34] D. Litvak, I. Gannot, and Y. Y. Zigel. Detection of falls at home using floor vibrations and sound. In *Proc. IEEE Convention of Electrical and Electronics Engineers in Israel (IEEEI)*, pages 514 –518, 2008.
- [35] S. T. Londei, J. Rousseau, F. Ducharme, A. St-Arnaud, J. Meunier, J. Saint-Arnaud, and F. Giroux. An intelligent videomonitoring system for fall detection at home: perceptions of elderly people. *Journal of Telemed Telecare*, 15(8):383–390, 2009.
- [36] F. Marquis-Faulkes, S. J. McKenna, A. F. Newell, and P. Gregor. Gathering the requirements for a fall monitor using drama and video with older people. *Journal Technology and Disability*, 17(4):227–236, 2005.
- [37] M. Mataric, J. Eriksson, D. Feil-Seifer, and C. Winstein. Socially assistive robotics for post-stroke rehabilitation. *Journal of NeuroEngineering and Rehabilitation*, 4(1):5, 2007.
- [38] A. Mihailidis, T. Tam, M. McLean, and T. Lee. An intelligent health monitoring and emergency response system. In *Proc. International Conference on Smart Homes and Health Telematics (ICOST)*, pages 272–281, 2005.
- [39] T. B. Moeslund, A. Hilton, and V. Krüger. A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, 104:90–126, Nov. 2006.
- [40] H. Nait-Charif and S. J. McKenna. Activity summarisation and fall detection in a supportive home environment. In *Proc. IEEE International Conference on Pattern Recognition (ICPR)*, volume 4, pages 323 – 326, 2004.

- [41] H. Nait-charif and S. J. McKenna. Tracking poorly modelled motion using particle filters with iterated likelihood weighting. In *Proc. Asian conference on computer vision (ACCV)*, pages 156–161, 2004.
- [42] N. Nasr, S. Torsi, S. Mawson, P. Wright, and G. Mountain. Self management of stroke supported by assistive technology. In *Proc. Virtual Rehabilitation International Conference*, page 193, 2009.
- [43] A. H. Nasution and S. Emmanuel. Intelligent video surveillance for monitoring elderly in home environments. In *Proc. IEEE Workshop on Multimedia Signal Processing (MMSP)*, pages 203 –206, 2007.
- [44] J. W. P. Ng, B. P. L. Lo, O. Wells, M. Sloman, N. Peters, A. Darzi, C. Toumazou, and G. Yang. Ubiquitous monitoring environment for wearable and implantable sensors (ubimon). In *Proc. 6th International Conference on Ubiquitous Computing (UBICOMP 04)*, 2004.
- [45] The Royal Academy of Engineering. Dilemmas of privacy and surveillance: Challenges of technological change. March 2007. [http://www.raeng.org.uk/societygov/policy/current\\_issues/privacy\\_surveillance/pdf/dilemmas\\_of\\_privacy\\_and\\_surveillance\\_report.pdf](http://www.raeng.org.uk/societygov/policy/current_issues/privacy_surveillance/pdf/dilemmas_of_privacy_and_surveillance_report.pdf).
- [46] Department of Health Publication. Efficiencies in telecare. Feb. 2010. [http://www.csed.dh.gov.uk/\\_library/Resources/CSED/CSEDProduct/Telecare\\_North\\_Yorks\\_Case\\_Study\\_webversionvs\\_1.pdf](http://www.csed.dh.gov.uk/_library/Resources/CSED/CSEDProduct/Telecare_North_Yorks_Case_Study_webversionvs_1.pdf).
- [47] E. J. Pauwels, A. A. Salah, and R. Tavenard. Sensor networks for ambient intelligence. In *Proc. IEEE Workshop on Multimedia Signal Processing (MMSP)*, pages 13 –16, 2007.
- [48] M. Z. Poh, D. J. McDuff, and R. W. Picard. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Opt. Express*, 18(10):10762–10774, May 2010.
- [49] GER’HOME Project. <http://gerhome.cstb.fr/>.
- [50] HERMES Project. <http://www.fp7-hermes.eu/>.
- [51] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau. Monocular 3d head tracking to detect falls of elderly people. In *Proc. IEEE International Conference on Engineering in Medicine and Biology Society (EMBS)*, pages 6384 –6387, 2006.
- [52] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau. Fall detection from human shape and motion history using video surveillance. In *Proc. International Conference on Advanced Information Networking and Applications Workshops (AINAW)*, volume 2, pages 875 –880, 2007.
- [53] Link Care Services. [http://www.linkcareservices.com/lcs/LCS\\_Presentation\\_Guide.pdf](http://www.linkcareservices.com/lcs/LCS_Presentation_Guide.pdf).
- [54] C. Stauffer and W. E. L. Grimson. Learning patterns of activity using real-time tracking. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 22:747–757, Aug. 2000.
- [55] H. Sun, V. De Florio, N. Gui, and C. Blondia. Promises and challenges of ambient assisted living systems. In *Proc. International Conference on Information Technology: New Generations (ITNG)*, pages 1201 –1207, 2009.
- [56] C. Takano and Y. Ohta. Heart rate measurement based on a time-lapse image. *Medical Engineering and Physics*, 29(8):853 – 857, 2007.
- [57] J. Tao, T. Mukherjee, M. F. Wong, M. Wang, and Y. P. Tan. Fall incidents detection for intelligent video surveillance. In *Proc. International Conference on Information, Communications and Signal Processing*, pages 1590 –1594, 2005.
- [58] J. Tao, M. Turjo, and Y. P. Tan. Quickest change detection for health-care video surveillance. In *Proc. IEEE International Symposium on Circuits and Systems (ISCAS)*, page 4 pages, 2006.
- [59] E. M. Tapia, S. S. Intille, and K. Larson. Activity recognition in the home using simple and ubiquitous sensors. *Pervasive Computing*, 3001:158–175, 2004.

- [60] N. Thome and S. Migue. A HHMM-based approach for robust fall detection. In *Proc. International Conference on Control, Automation, Robotics and Vision (ICARCV)*, pages 1–8, 2006.
- [61] N. Thome, S. Migue, and S. Ambellouis. A real-time, multiview fall detection system: A LHMM-based approach. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(11):1522 –1532, Nov. 2008.
- [62] B. U. Töreyin, Y. Dedeoglu, and A. E. Çetin. HMM based falling person detection using both audio and video. *Computer Vision in Human-Computer Interaction*, 3766:211–220, 2005.
- [63] B. U. Töreyin, E. Birey Soyer, Ibrahim Onaran, and A. E. Çetin. Falling person detection using multi-sensor signal processing. *EURASIP Journal on Advances in Signal Processing*, 2008, 2008.
- [64] P. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea. Machine recognition of human activities: A survey. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(11):1473 –1488, Nov. 2008.
- [65] L. Wang and D. Suter. Learning and matching of dynamic shape manifolds for human action recognition. *IEEE Transactions on Image Processing*, 16(6):1646 –1661, June 2007.
- [66] M. L. Wang, C. C. Huang, and H. Y. Lin. An intelligent surveillance system based on an omnidirectional vision sensor. In *Proc. IEEE Cybernetics and Intelligent Systems*, pages 1–6, 2006.
- [67] J. A. Ward, P. Lukowicz, and H. W. Gellersen. Performance metrics for activity recognition. *ACM Trans. Intell. Syst. Technol.*, 2:6:1–6:23, Jan. 2011.
- [68] A. Warren, R. Bayley, C. Bennett, A. J. Charlesworth, R. A. Clarke, and C. Oppenheim. Privacy impact assessments: The UK experience. In *Proc. 31st International Conference of Data Protection and Privacy Commissioners*. SSRN, 2009.
- [69] T. Winkler and B. Rinner. TrustCAM: Security and privacy-protection for an embedded smart camera based on trusted computing. In *Proc. IEEE Conference on Advanced Video and Signal Based Surveillance*, volume 0, pages 593–600, 2010.
- [70] Z. Y. Wu, Y. C. Lee, F. Lai, H. C. Lee, and Y. Chung. A secure authentication scheme for tele-care medicine information systems. *Journal of Medical Systems*, pages 1–7, 2010.
- [71] G. Z. Yang, B. Lo, L. Wang, M. Rans, S. Thiemjarus, J. Ng, P. Garner, S. Brown, B. Majeed, and I. Neid. From sensor networks to behaviour profiling: A homecare perspective of intelligent building. In *Proc. IEE Seminar for Intelligent Buildings*, 2004.
- [72] Y. Zigel, D. Litvak, and I. Gannot. A method for automatic fall detection of elderly people using floor vibrations and sound – proof of concept on human mimicking doll falls. *IEEE Transactions on Biomedical Engineering*, 56(12):2858 –2867, Dec. 2009.