


# A review on applications of activity recognition systems with regard to performance and evaluation

International Journal of Distributed  
Sensor Networks  
2016, Vol. 12(8)  
© The Author(s) 2016  
DOI: 10.1177/1550147716665520  
ijdsn.sagepub.com  


Suneth Ranasinghe, Fadi Al Machot and Heinrich C Mayr

## Abstract

Activity recognition systems are a large field of research and development, currently with a focus on advanced machine learning algorithms, innovations in the field of hardware architecture, and on decreasing the costs of monitoring while increasing safety. This article concentrates on the applications of activity recognition systems and surveys their state of the art. We categorize such applications into active and assisted living systems for smart homes, healthcare monitoring applications, monitoring and surveillance systems for indoor and outdoor activities, and tele-immersion applications. Within these categories, the applications are classified according to the methodology used for recognizing human behavior, namely, based on visual, non-visual, and multimodal sensor technology. We provide an overview of these applications and discuss the advantages and limitations of each approach. Additionally, we illustrate public data sets that are designed for the evaluation of such recognition systems. The article concludes with a comparison of the existing methodologies which, when applied to real-world scenarios, allow to formulate research questions for future approaches.

## Keywords

Human activity recognition, active and assisted living, sensor networks, smart systems

Date received: 9 March 2016; accepted: 11 July 2016

Academic Editor: José Molina

## Introduction

Human activity recognition (HAR) is a highly dynamic and challenging research topic. It aims at determining the activities of a person or a group of persons based on sensor and/or video observation data, as well as on knowledge about the context within which the observed activities take place. In the ideal case, an activity is recognized regardless of the environment it is performed in or the performing person.

In general, the HAR process involves several steps—from collecting information on human behavior out of raw sensor data to the final conclusion about the currently performed activity. These steps are as follows: (1) *pre-processing* of the raw data from sensor streams for handling incompleteness, eliminating noise and redundancy, and performing data aggregation and normalization; (2) *segmentation*—identifying the most

significant data segments; (3) *feature extraction*—extracting the main characteristics of features (e.g. temporal and spatial information) from the segmented data using, for example, statistical moments; (4) *dimensionality reduction*—decreasing the number of features to increase their quality and reduce the computational effort needed for the classification; and (5) *classification, the core machine learning and reasoning*—determining the given activity.<sup>1</sup>

Application Engineering Research Group, Alpen-Adria-Universität  
Klagenfurt, Klagenfurt, Austria

### Corresponding author:

Suneth Ranasinghe, Application Engineering Research Group, Alpen-Adria-Universität Klagenfurt, Klagenfurt 9020, Austria.  
Email: suneth.ranasinghe@aau.at



Creative Commons CC-BY: This article is distributed under the terms of the Creative Commons Attribution 3.0 License

(<http://www.creativecommons.org/licenses/by/3.0/>) which permits any use, reproduction and distribution of the work without

further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<http://www.uk.sagepub.com/aboutus/openaccess.htm>).

The main goals of HAR systems are to observe and analyze human activities and to interpret ongoing events successfully. Using visual and non-visual sensory data, HAR systems retrieve and process contextual (environmental, spatial, temporal, etc.) data to understand the human behavior. There are several application domains where HAR concepts are investigated and the systems are developed. We divide them roughly into four categories: *active and assisted living (AAL) systems for smart homes, healthcare monitoring applications, monitoring and surveillance systems for indoor and outdoor activities, and tele-immersion (TI) applications.*

Traditionally, the task of observing and analyzing human activities was carried out by human operators, for example, in security and surveillance processes or the processes of monitoring a patients' health condition. With the increasing number of camera views and technical monitoring devices, however, this task becomes not only more challenging for the operators but also increasingly cost-intensive, in particular, since it requests around-the-clock operation. In practice, for the case of home care, personnel deployment for such tasks often cannot be financially feasible.

Moreover, HAR systems within these fields are able to support or even replace human operators in order to enhance the efficiency and effectiveness of the observation and analysis process. As an example, with the help of sensory devices, an HAR system can keep track of the health condition of a patient and notify the health personnel in case of an urgent need.

On the other hand, scientific and technical progress continuously improved the living conditions of humans. This causes a dramatic societal change as it comes with decreasing birth rates and increasing life expectancy, which together turn the age pyramid upside down.<sup>2</sup> Intensive research and development in the field of Active and Assisted Living (AAL)<sup>3</sup> focuses on mastering one of the consequences of this change: the increasing need of care and support for older people. The goal of AAL systems, therefore, is to provide appropriate unobtrusive technical support enabling people to live as independent as possible for as long as possible in their homes. To be able to provide such support, an AAL system needs to know about a person's behavior; that is, it depends on powerful HAR systems for obtaining, collecting, compiling, and analyzing such knowledge. Similarly, TI systems also make use of HAR systems to track and simulate human behaviors in a virtual environment in order to build attractive game interfaces or to enhance the existing communication methods.

This survey article focuses on current activity recognition (AR) projects applied within these fields, their achievements, issues, and challenges. We do not focus on the classification (machine learning) approaches of HAR systems, as there is a rich body of previously published work in this area.<sup>4-6</sup>

The organization of the article is as follows: section "Notion of 'activity'" shortly discusses the concept of "activity" as understood within this study. Section "Applications of AR systems" presents a broad selection of state-of-the-art systems representing the above categories: active and assisted living systems for smart homes, healthcare monitoring applications, monitoring and surveillance systems for indoor and outdoor activities, and TI applications. In section "AR systems and approaches," these systems are classified based on different types of sensors, namely, video-based, non-video-based, or multimodal sensors. Section "Popular data sets" outlines popular AR data sets that are used as universal data sets to evaluate such systems. Based hereon, section "Discussion of HAR approaches" discusses the limitations of the existing HAR approaches and presents challenges for future research.

## Notion of "activity"

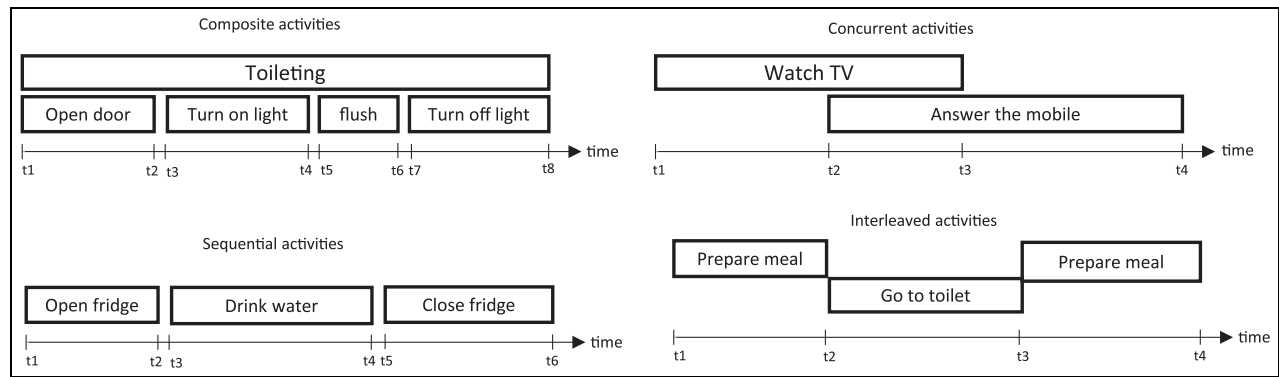
So far, no unique ontological definition of the notion/concept of "activity" has been proposed. However, common to the usage of this notion in the various application domains is its refinement into granularity levels. An example of such refinement can be found in the activity theory<sup>7,8</sup> which assumes activities (e.g. "preparing coffee") to be hierarchically structured into *actions* (e.g. "enter the kitchen," "fill the water container") that again are composed out of *operations*, the latter being understood as atomic steps implementing the action (like "push the door handle," "open the water tap"). Activities are understood as aggregations of actions, which again are understood as aggregations of atomic operations.

Within the AAL domain, taxonomies have been proposed in, for example, Katz<sup>9</sup> and Lawton and Brody<sup>10</sup> for the so-called (instrumental) activities of daily living (ADLs). In general, the taxonomies depend on the living scenarios, such as the ADLs and activities related to medical or other domains.

Usually, an activity is supposed to be performed within a certain time span, which is calculated based on the durations of the composing sub-units.<sup>11</sup> Considering the aspects of performance, duration, and order of these sub-units, activities can be divided into composite activities, sequential activities, concurrent activities, and interleaved activities (Figure 1).<sup>12</sup> Based hereon, actions and activities can be derived from a well-defined information hierarchy of events (based on sensor inputs) fused with additional context information.

## Applications of AR systems

During the last decade, there was a significant growth of the number of publications in the field of AR; in



**Figure 1.** Relations between actions and general structures of ADLs.<sup>12</sup>

**Table 1.** References, classified along application categories and observation methodology.

Application type	Visual-based systems	Non-visual-based systems	Multimodal systems
Active and assisted living (AAL) applications for smart homes	13–18	11,19,20	9,10,12,21–32
Healthcare monitoring applications	33–36	37,38	39–45
Security and surveillance applications	46–60		58,61–63
Tele-immersion (TI) applications	64–75		

particular, many researchers have proposed application domains to identify specific activity types or behaviors to reach specific goals in these domains. This section focuses on state-of-the-art applications that use HAR methodologies to assist humans. This review, in particular, discusses the application of the current AR approaches to AAL for smart homes, healthcare monitoring solutions, security and surveillance applications, and TI applications; these approaches are further classified along the observation methodology used for recognizing human behavior, namely, into approaches based on visual, non-visual, and multimodal sensor technology. Table 1 shows the references considered, each related to the respective category and class.

### Active and assisted living applications for smart homes

Advances in modern technologies have provided innovative ways to enhance the quality of independent living of elderly or disabled people. Active and assisted living systems use AR techniques to monitor and assist residents to secure their safety and well-being. According to Demirir et al.,<sup>21</sup> a smart home is an environment equipped with sensors that enhance the safety of residents and monitor their health conditions. Thus, smart homes improve the level of independency and the quality of life of the people who need support for physical and cognitive functions. In general, inside a smart

home, the behavior of the residents and their interaction with the environment are monitored by analyzing the data collected from sensors. Table 2 shows the state-of-the-art AAL systems using HAR components.

Recently, research has been conducted aimed at using the combination of wearable sensors and sensors implanted into the environment, which, together with audio–video-based solutions, allow for a gentle assistance of the elderly people. The GER’HOME project (<http://www-sop.inria.fr/members/Francois.Bremond/topicsText/gerhomeProject.html>) used multi-sensor analysis for monitoring the activity of the elderly people to improve their living conditions and to reduce the costs of long hospitalizations. GER’HOME is equipped with a network of cameras and contact sensors to track people and recognize their activities.

Similarly, HERMES (<http://www.fp7-hermes.eu/>) aimed at providing cognitive care for people who are suffering from mild memory problems. This is achieved by combining the functional skills of a particular person with his or her age-related cognitive incapacibilities and assist them when necessary. HERMES used visual and audio processing techniques, speech recognition, speaker identification, and face detection techniques to guide the people. In a similar manner, the universAAL (<http://www.universaal.org/index.php/en/>)<sup>22</sup> open platform and reference specification for AAL was introduced to technically produce cheaper products that are simple, configurable, and easily deployable at a smart home to provide useful “AAL services.”

**Table 2.** State-of-the-art active and assisted living applications for smart homes.

Author	Keywords	Sensor information used
Ricardo Costa et al. <sup>47</sup> TLM Van Kasteren et al. <sup>48</sup>	Ambient intelligence, assisted living Activity recognition, machine learning, wireless sensor networks (WSNs)	Domestic smart sensors WSN
Chen Wu et al. <sup>49</sup>	Multi-view activity recognition, decision fusion methods, smart home	Multi-camera views
Parisa Rashidi et al. <sup>64</sup>	Activity recognition, data mining, sequence mining, clustering, smart homes	Accelerometer, state-change sensor, motion sensors, and so on
Can Tunca et al. <sup>65</sup>	Ambient assisted living, WSNs, human activity recognition	WSN
Rubén Blasco et al. <sup>66</sup>	Ambient assisted living, ambient intelligence, smart homes, context and user awareness, distributed sensor networks	Radio-frequency identification (RFID) technology, WSNs, and so on
Saisakul Chernbumroong et al. <sup>67</sup>	Assisted living systems, activities of daily living (ADLs), wrist-worn multi-sensors, elderly care, feature selection, and classification	Accelerometers, gyroscope, magnetometer, bio-sensors, and so on
Nirmalya Roy et al. <sup>24</sup>	Multimodal sensing, context recognition	Body-worn sensors, motion sensors, and so on
M Blumendorf and S Albayrak <sup>25</sup>	Smart environments, multimodal interaction, model-based user interface development, ambient assisted living, multi-access service platform	Smart home sensors
A Dohr et al. <sup>13</sup>	eHealth, pervasive healthcare, telemedicine, Near Field Communication (NFC), RFID	RFID and NFC technology
Jaime Lloret et al. <sup>26</sup>	Ambient assisted living (AAL), WSNs, sensors and actuators, elderly people	Microphones, accelerometer, presence sensors, mobile phone sensors, cameras, and so on

Moreover, smart home system proposed by the Center for Advanced Studies in Adaptive Systems (CASAS)<sup>23</sup> used machine learning and data mining techniques to analyze the daily activity patterns of a resident to generate automation policies based on the identified patterns to support the residents. Automation policies were used to assist elder individuals in their urgent needs.

Another example of an intelligent home environment is the Mobiserv (<http://www.mobiserv.info/>) project.<sup>19</sup> The aim of this project is to offer health, medical, and well-being services to older adults. The support was provided based on the understanding of the user context data of their indoor daily living situations.

The SWEET-HOME project (2009–2013)<sup>39,76</sup> proposed by the French National Research Agency aimed at establishing a smart home solution which is capable to interact with old people or disabled individuals using an audio-based interaction technology. Initially, a multimodal sound corpus was recorded using the interactions of healthy individuals. Later on, this audio data were used to train the smart home's speech and sound recognition systems.<sup>40</sup>

Finally, the Human Behavior Monitoring and Support (HBMS) project<sup>46</sup> (this work was funded by the Klaus Tschira Stiftung gGmbH, Heidelberg) was initiated to support old or disabled individuals with cognitive

impairments to live autonomously in their familiar environments. In the first phase, a person's activities of daily living are observed by the *HBMS observation engine* to create a *Human Cognitive Model (HCM)* using the *Human Cognitive Modeling Language HCM-L* (<http://austria.omilab.org/psm/content/hcml/download?view=download>). Then, these HCMs are used by the *HBMS support engine* that applies reasoning mechanisms based on Answer Set Programming (ASP) to assist individuals in a smart and unobtrusive way.

### Healthcare monitoring applications

The development of medical science and technology considerably increased the life quality of patients. As stated by Goldstone,<sup>20</sup> life expectancy rates will increase dramatically in 2050, and approximately 30% of Americans, Canadians, Chinese, and Europeans will be over the age of 60 years. This will lead to higher demands for medical personnel which may be impossible to be supplied in the near future. Hence, researchers try to enhance the existing healthcare monitoring approaches that would handle urgent medical situations and shorten the hospital stay and regular medical visits of a patient. Table 3 shows the state-of-the-art healthcare monitoring applications.

**Table 3.** State-of-the-art healthcare monitoring applications.

Author	Keywords	Sensor information used
Yaniv Zigel et al. <sup>29</sup>	Acoustic signal processing, fall detector, feature extraction, pattern recognition	Floor vibration and sound sensing
Qiang Li et al. <sup>30</sup>	Fast fall detection, activities of daily living (ADL)	Accelerometers and gyroscopes
Derek Anderson et al. <sup>31</sup>	Activity analysis, fuzzy logic, fall detection, eldercare	Multi-camera views
Maarit Kangas et al. <sup>32</sup>	Elderly, independent living, movement analysis	Accelerometers
M Lustrek and B Kaluza <sup>41</sup>	Activity recognition, machine learning, body part	Video, infrared motion capture sensors, and so on
AK Bourke and GM Lyons <sup>42</sup>	Falls in the elderly, fall detection, gyroscope, ADL	Bi-axial gyroscope sensor
Chao Wang et al. <sup>33</sup>	Healthcare monitoring, near-threshold operation, reconfigurable computing	Healthcare sensors
Minh-Thanh Vo et al. <sup>43</sup>	Wireless sensor network (WSN), healthcare monitoring, simulation	Light-to-Frequency (LTF), infrared (IR) sensors, healthcare monitoring wireless sensors, and so on

Basically, healthcare monitoring systems are designed based on the combination of one or more AR components such as fall detection, human tracking, security alarm and cognitive assistance components. Most of the healthcare systems use body-worn and contextual sensors that are placed on patients' bodies and in their environment. Once help is needed, the system notifies the relevant parties (i.e. medical personnel) about the situation to assist the patient quickly. The E-safe fall detection and notification system<sup>27</sup> has used the zigbee-based wearable sensor system to automatically detect fall situations and notify the in-house correspondents via zigbee technology. Then the external correspondents are also notified via Short Message Service (SMS) and email.

The smart assisted living (SAIL) system was introduced in Zhu et al.<sup>28</sup> using human-robot interaction (HRI) to monitor the health condition of an elder or disabled individual. SAIL consists of a body sensor network, a companion robot, a smartphone, and a remote health provider. Based on the sensor data, the robot assists the human or the help is provided by a remote health provider contacted through a smartphone gateway.

CAALYX (<http://www.ij-healthgeographics.com/content/6/1/9>), a European Union (EU)-funded healthcare project, aims at assisting elderly people using a wearable device that is capable of measuring vital signs and fall detection events and of notifying care providers automatically in an emergency situation. Most importantly, the CAALYX is able to report the current medical status of the patient together with his or her current location that helps the emergency team to provide immediate assistance.

### Security and surveillance applications

Traditional surveillance systems are monitored by human operators. They should be continuously aware

of the human activities that are observed via the camera views. An increasing number of camera deployments and views makes the operators' work more stressful and, as a result, leads to decreasing their productivity levels. As a result, security firms are seeking help from vision-based technologies to automate the human operator processes and detect anomalies in camera views. Table 4 shows the state-of-the-art security and surveillance system applications.

Most of the traditional object recognition methods depend on the object's shape, but it is quite challenging to apply those approaches to a surveillance system which consists of cluttered, wide-angle, and low-detailed views.<sup>44</sup> Therefore, the proposed applications should be capable of addressing the environmental and contextual issues such as noise, occlusions, and shadows. The facts that affect an HAR system in an indoor environment are different from an outdoor environment. As an example, an approach used inside a bank may not be applicable to a crowded place such as a metro and an airport. Most importantly, those approaches should be robust and capable of working under real-world application conditions.

The surveillance system introduced in Brémond et al.<sup>37</sup> is based on the Video Surveillance Interpretation Platform (VSIP) and is able to recognize human behavior such as fighting and vandalism events occurring in a metro system using one or several camera views.<sup>44</sup> Additionally, as shown in Chang et al.,<sup>38</sup> this system was able to detect and predict the suspicious and aggressive behaviors of a group of individuals in a prison. The researchers used multiple camera views to detect situations such as loitering, distinct groups, and aggression scenarios in real time and in a crowded environment. The airport surveillance system proposed by Fusier et al.<sup>45</sup> is able to recognize 50 types of events including complex activities such as baggage unloading, aircraft arrival preparation, and refueling operation.

**Table 4.** State-of-the-art security and surveillance applications.

Author	Keywords	Sensor information used
Jun-Wei Hsieh et al. <sup>50</sup>	Behavior analysis, event detection, string matching	Camera views
Umut Akdemir et al. <sup>51</sup>	Activity ontologies, visual surveillance	Camera views
C Fookes et al. <sup>52</sup>	Supervised learning, multi-camera network, event detection	Multi-camera views
Frdric Dufaux et al. <sup>53</sup>	Privacy, selective encryption, surveillance, video processing	Camera views
Nils Krahnstoeve et al. <sup>54</sup>	Multi-camera view, real time, detection, and tracking algorithms	Multi-camera views
Shih-Chia Huang <sup>55</sup>	Background model, entropy, morphology, motion detection, video surveillance	Camera views
L Maddalena and A Petrosino <sup>56</sup>	Background subtraction, motion detection, neural network, self-organization, visual surveillance	Camera views
Liyuan Li et al. <sup>57</sup>	Mean-shift tracking, multi-object tracking, occlusion, video surveillance	Camera views
Donato Di Paola et al. <sup>58</sup>	Autonomous mobile robot, RFID tags, surveillance applications	Video, RFID tags, laser scene change detector
Zheng Xu et al. <sup>59</sup>	Video-structured description, surveillance videos, public security	Video
Evgeny Belyaev et al. <sup>60</sup>	Vehicular communication, video surveillance, 3D discrete wavelet transform	Camera views

RFID: radio-frequency identification; 3D: three-dimensional.

### TI applications

TI is an approach that allows users to share their presence in a virtual environment and interact with each other in real time such as being present in the same physical but in different geographical environments. These applications require a higher amount of computer processing power and generate a large amount of data that need to be transferred through a network in real time.

Usually, compression methods are applied to reduce the amount of data to be transferred. For example, the multi-camera TI system<sup>68</sup> can extract the kinematic parameters of a human body in each frame from cloud data using motion estimation, and thus, significantly minimize the network transfer between the remote sites.

Furthermore, three-dimensional (3D) videoconferencing applications successfully address the hardware bottlenecks emerging from complex computational algorithms in real-time 3D video conferences.<sup>69,70</sup> Similarly, the i3DPost project<sup>71</sup> used TI techniques to enhance the existing appearance of two-dimensional (2D) layouts (<http://www.i3dpost.eu/>) by converting 2D computer-aided draft (CAD) planning) into full color motion-rendered pictures.

Table 5 shows the state-of-the-art TI applications.

### AR systems and approaches

Advances in visual and sensor technology enabled AR systems to be widely used in daily life. During the last decade, scientists have taken various approaches to recognize the human behavior in many application domains. In this review, we categorize the AR systems based on their design methodology, mainly taking into account the data collection process. As a result, this section is divided into

subsections dedicated to visual sensor-based systems, non-visual sensor-based systems, and multimodal systems. Each approach is then discussed with regard to its usage under different categories.

#### Visual systems for AR

Identification of human activities using visual sensor networks is one of the most popular approaches in the computer vision research community. In early stages of visual recognition, systems were categorized into groups such as hand gesture recognition for human-computer interfaces, facial expression recognition, and human behavior recognition for video surveillance.<sup>81</sup>

The major difference between visual sensors and other sensor types is the way of perceiving the information in an environment. Most sensors provide the data as a one-dimensional data signal, whereas the visual sensors provide the data as a 2D set of data which is seen as images.<sup>82</sup> There are various types of visual-based AR approaches. Therefore, we organize this subsection along four categories: (1) visual AR systems for active and assisted living and smart home systems, (2) visual AR systems for healthcare monitoring systems, (3) visual AR systems for security and surveillance systems in public areas, and (4) visual AR systems in sports and outdoors.

*Visual systems for active and assisted living and smart home systems.* Given the visual sensor usage in indoor environments, active and assisted living systems provide residents with supervision and assistance to ensure their well-being. AAL systems should be easily deployable, robust systems which are capable of assisting users whenever required.

**Table 5.** State-of-the-art tele-immersion applications.

Author	Keywords	Used sensor information
G Kurillo and R Bajcsy <sup>72</sup>	3D video, 3D tele-immersion, human–computer interaction, remote collaboration, telepresence	16–48 VGA cameras (640 × 480 pixels)
C Zhang et al. <sup>73</sup>	Teleconferencing, 3D video processing, 3D video rendering, 3D audio processing	Three infrared (IR) cameras, three color XGA cameras, and one Kinect
B Petit et al. <sup>74</sup>	Multi-camera, real time, 3D modeling, telepresence	Eight 1-megapixel cameras
Z Huang et al. <sup>75</sup>	3D tele-immersion, synchronization	3D video camera
Kurillo et al. <sup>77</sup>	Tele-immersion, rehabilitation, tele-rehabilitation, lower extremities	3D video stream
Chun-Han Lin et al. <sup>78</sup>	Virtual collaboration, motion-sensing techniques, gesture-collaborative games	3D motion sensor data
Yunpeng Liu et al. <sup>79</sup>	Depth map, compression, 3D tele-immersion	Depth image streams
H Kim et al. <sup>80</sup>	Big data management, multimodal data registration, film production	2D/3D video data

2D: two-dimensional; 3D: three-dimensional.

Fosty et al.<sup>83</sup> presented an RGB-D camera monitoring system for event recognition which uses a hierarchical model-based approach. The aim of the approach is to recognize physical tasks that are evaluated depending on the patients with dementia. The extraction of complex events from video sequences is carried out by combining RGB-D data stream with the corresponding tracking information, the contextual objects (zones or pieces of equipment), and the event models. Experimental results indicated 95.4% accuracy rate for the events such as balance test, walking test, repeated transfer test, and up-and-go tests.

Xia et al.<sup>84</sup> presented a human action recognition approach for an indoor environment based on histograms of 3D joint locations (HOJ3D), as a compact representation of postures. The researchers extracted the 3D skeletal joint locations from Kinect depth maps using Shotton et al.'s<sup>85</sup> method. Then, they re-projected the computed HOJ3D from the action depth sequences using Linear Discriminant Analysis (LDA) and clustered them into *k* posture visual words which represent the prototypical poses of actions. The temporal evolutions of those visual words are modeled by discrete hidden Markov models (HMMs). The special data set has been collected which consists of 10 types of human actions (i.e. walk, sit down, and stand up) in an indoor setting. Experimental results show 90.92% overall accuracy rate for action types such as walk, sit down, stand up, pick up, carry, throw, push, pull, wave, and clap hands.

With the purpose of handling uncertainty, Romdhane et al.<sup>86</sup> described a complex event recognition approach with probabilistic reasoning. The estimation of a primitive state's probability is based on the Bayesian process model and computes the confidence of a complex event as Markov process considering the probability of the event at a previous time. The proposed event recognition algorithm uses the tracked mobile objects as input (extracted by vision algorithms,

segmentation, detection, and tracking), a priori knowledge of the scene, and predefined event models. After calculating the probability of an event, the system can make a recognition decision by accepting events with a probability above a threshold or rejecting them. Experimental results present a higher accuracy for recognizing indoor events: 92.59% for up-and-go event, 100% for beginning guided test, 100% for interacting with a chair, 93.3% for staying at kitchen, and 87.5% for preparing a meal event.

Hartmann et al.<sup>14</sup> proposed a robust and intelligent vision system which detects a person who is staying or lying on the floor. The system uses only one fisheye camera which is located in the middle of the room. The solution is based on image segmentation using Gaussian mixture models to detecting moving residents and then analyzing their main and ideal orientation using image moments. It has a low latency and a detection rate of 88%.

**Visual systems for healthcare monitoring systems.** Although visual sensor-based approaches are not popular in healthcare monitoring systems, these techniques are widely used for implementing fall detection systems, particularly to take care of patients who suffer from diseases such as dementia and Alzheimer. Having in mind the visual-based healthcare monitoring approaches, Foroughi et al.<sup>87</sup> proposed a fall detection system combined with integrated time motion images (ITMI) and eigenspace techniques. The proposed system was able to detect a wide range of daily life activities including abnormal and unusual behaviors. The researchers extracted the eigenmotion space and classified the motion using multilayer perceptron (MLP) neural network to decide on a fall event. This system showed a reliable average recognition rate of 89.99% as their final result.

An intelligent monitoring system proposed by Chen et al.<sup>88</sup> monitors the “elopement” events of dementia

units and is able to automatically detect such events and alert the caregivers. The monitoring system uses a camera network to collect the audio/visual records of daily activities. Using an HMM-based algorithm, the authors were able to detect elopement events from the collected data. Experimental results demonstrate that the proposed system was able to successfully detect elopement events with almost 100% accuracy.

*Visual systems for security and surveillance systems in public areas.* Security and surveillance systems have used visual processing approaches extensively to track human behaviors in public environments. Visual sensor-based techniques are the most suitable approaches for implementing such systems because of the valid evidential proofs which can be provided by videos and images due to their nature.

Zaidenberg et al.<sup>89</sup> proposed an approach to Scenario Recognition based on knowledge (ScReK) framework model to automatically detect the behavior of a group of people in a video surveillance application. It keeps track of individuals moving together by maintaining spatial and temporal group coherence in a video stream. The proposed framework models the semantic knowledge of the objects of interest and scenarios to recognize events associated with the detected group based on spatiotemporal constraints. At the beginning, people are individually detected and tracked. Then, their trajectories are analyzed over a temporal window and clustered using the mean-shift algorithm. The obtained coherence value describes the activity performed by the group. Furthermore, the researchers have proposed a description language for formalizing events. The group event recognition approach has been successfully validated using three data sets collected from four cameras (an airport, a subway, a shopping center corridor, and an entrance hall): group events such as fighting, split up, joining, entering and exiting the shop, browsing, and getting off a train have been successfully detected with low false positive and false negative rates.

In the context of visual surveillance of metro scenes, Cupillard et al.<sup>90</sup> proposed an approach for recognizing isolated individuals, groups of people, or crowded environments using multiple cameras. The functionality of the proposed vision module is composed of three tasks: (1) motion detection and frame-to-frame tracking, (2) combining multiple cameras, and (3) long-term tracking of individuals, groups of people, and crowd evolving in the scene. For each tracked actor, the behavior recognition module performs reasoning on three levels: states, events, and scenarios. Also, the authors defined a general framework to combine and tune various recognition methods (e.g. automaton, Bayesian network, or AND/OR tree) that are dedicated to analyzing specific situations (e.g. mono/multi-actor activities,

numerical/symbolic actions, or temporal scenarios). This method was able to successfully recognize the scenarios like “Fraud” 6/6 (6 times out of 6), “Vandalism” 4/4, “Fighting” 20/24, “blocking” 13/13, and the scenario “overcrowding” 2/2.

Nievas et al.<sup>91</sup> proposed bag-of-words framework to detect fight events using Space–Time Interest Points (STIP) and Motion SIFT (MoSIFT) action descriptors. They have introduced a new video database containing 1000 sequences that are grouped as fights and non-fights for evaluation purposes. Experimental results with this video database detected fight events from action movies with an accuracy of nearly 90%.

*Visual systems in sports and outdoors.* Computer vision techniques can also be used to recognize sport activities to enhance the performance of players and analyze the game plan. Direkoglu and O’Connor<sup>92</sup> proposed a method to analyze an entire playground of team activities of a handball game. Frame differencing and optical flow methods have been used to extract the motion features and recognize the sequence of position distribution of the team players. The proposed approach was evaluated using the European handball data set and is able to successfully identify six different team activities in a handball game, namely, slowly going into offense, offense against setup defense, offense fast break, fast returning into defense to prevent fast break, slowly returning into defense, and basic defense.

Action bank presented by Sadanand and Corso<sup>93</sup> combined a large set of action detectors that represent a high-dimensional “action-space” with a linear classifier to arrange a semantically rich representation for AR. The action bank is inspired by the object bank method<sup>94</sup> which mines a high level of human action in a video. The authors have tested the action bank with popular data sets and achieved improved performances for various data sets: 98.2% for the Kungliga Tekniska Högskolan (KTH) Royal Institute of Technology, data set (better by 3.3%), 95.0% for the University of Central Florida (UCF), Sports data set (better by 3.7%), 57.9% for the UCF50 data set (baseline is 47.9%), and 26.9% for the HMDB51 data set (baseline is 23.2%).

Recognizing activities in noisy videos, that is, videos consisting of blurred motions, occlusions, missing observations, and dynamic backgrounds, is quite challenging. Using probabilistic event logic (PEL) interpretation, Brendel et al.<sup>95</sup> were able to understand events and time intervals of a basketball video which consisted of noisy data. They have successfully identified the basketball events, namely, dribbling, jumping, shooting, passing, catching, bouncing, and so on.

Tang et al.<sup>96</sup> proposed a method to identify complex events in video streams by utilizing a conditional model. The model is trained in a max-margin



framework to automatically detect discriminative and interesting segments in a video. It competitively achieved a higher accuracy for detecting and recognizing difficult tasks. The latent variables over the video frames have been introduced to discover and assign the most discriminative sequence of states for the event. This model is based on a HMM which tracks the model durations of states in addition to the transitions between states. Accordingly, experimental results acquired using the Multimedia Event Detection (MED) Transparent Development (DEV-T) data set have showed a 15.44% rate for attempting a board trick, 3.55% for feeding an animal, 14.02% for landing a fish, 15.09% for a wedding ceremony, and 8.17% for working on a woodworking project.

Crispim and Bremond<sup>97</sup> proposed a probabilistic framework to handle the uncertainty of a constraint-based ontology framework for event detection. The uncertainty modeling constraint-based framework is monitored by a RGB-D sensor (Kinect®, Microsoft©) vision system. The RGB-D monitoring system is composed of three main steps: people detection, people tracking, and event detection. The people detection step is performed by a depth-based algorithm proposed by Nghiem et al.<sup>98</sup> The detection is evaluated by a multi-feature tracking algorithm proposed by Chau et al.<sup>99</sup> The event detection step uses a set of tracked people generated in the previous step and a priori knowledge of the scene provided by a domain expert. Experimental results showed that the uncertainty modeling improves the detection of elementary scenarios in recall (e.g. in zone “phone”: 85%–100%), the precision indices (e.g. in zone “reading”: 54.71%–73.15%), and the recall of complex scenarios.

Touati and Mignotte<sup>100</sup> introduced a set of prototypes that are generated from different viewpoints of the video sequence data cube to recognize human actions. The prototypes are generated using a multidimensional scaling (MDS) based on a nonlinear dimensionality reduction technique. This strategy aims at modeling each human action in a low-dimensional space as a trajectory of points or a specific curve for each viewpoint of the video cube. Then, a k-Nearest Neighbors (K-NN) classifier is used to classify the prototype, for a given viewpoint, associated with each action to be recognized. Fusion of classification results of each viewpoint has been used to improve the recognition rate performance. The overall performance showed a 92.3% accuracy rate to recognize activities such as walking, running, skipping, jacking, jumping, siding, and bending.

### AR systems using non-visual sensors

Besides visual-based analyzing, researchers have made many other attempts such as analyzing human voice streams and sensor-based detection to automatically detect

human behavior. There are various types of sensors that can be used for AR, ranging from simple sensors such as ball switches and radio-frequency identification (RFID) tag readers to accelerometers which are used for complex audio processing and computer vision sensing. Also, other sensors such as fiber optical sensors for posture measuring, foam pressure sensors for respiration rate measuring, and physiological sensors such as oximetry sensors, skin conductivity sensors, electrocardiographs, and body temperature sensors can be included.<sup>101</sup>

*Non-visual AR systems in active and assisted living and smart home systems.* Smart home technologies use various types of sensors which provide light, sound, contact, motion, and state-change information of the residents to determine their behaviors. Fleury et al.<sup>102</sup> discussed a methodology regarding the classification of daily living activities in a smart home using support vector machines. The proposed system was able to identify seven activities, namely, hygiene, toilet use, eating, resting, sleeping, communication, and dressing/undressing, using location sensors, microphones, wearable sensors, temperature, and hygrometry sensors.

Challenges of most smart home approaches are the inconsistency and unreliability of the sensors which misread the actual information. Hong et al.<sup>103</sup> introduced a framework to deal with such uncertainty by combining the reliability level of each sensor with the overall decision-making process.

Fleury et al.<sup>102</sup> have presented an ADL (smart home) classification which is based on support vector machines. They have deployed sensors such as infrared presence sensors, door contacts, temperature and hygrometry sensors, and microphones in a smart home and collected the data of the residents (young individuals) to identify the residents’ daily living activities. The researchers were able to successfully identify seven activities: hygiene, toilet use, eating, resting, sleeping, communication, and dressing/undressing with a 75% of classification rate for the polynomial kernel and 86% classification rate for the Gaussian kernel.

Most of the AAL and smart home approaches made use of the annotated data labeled with the corresponding activities. Szewczyk et al.<sup>104</sup> have proposed a mechanism to annotate sensor data with correspondent activity labels; thus, they were able to achieve a higher accuracy in recognizing daily activities such as sleeping, eating, personal hygiene, preparing a meal, working at computer, watching TV, and others. The average accuracy of the models was increased from 66.35% to 75.15%. In general, annotating sensor data with activities is time-consuming and may require the assistance of the smart home residents.

*Non-visual AR systems in other indoor environments.* Apart from smart home systems, other sensor-based indoor

systems have been proposed. As an example, Viani et al.<sup>105</sup> presented a Learning-by-Example (LBE) approach which is able to track and localize passive targets of an indoor non-infrastructure environment. This approach used the interaction between targets and wireless links, so that the proposed strategy does not need any additional use of radio devices or specific sensors to track the objects. Furthermore, Viani et al.<sup>106</sup> introduced a system to detect theft attempts in an indoor museum. This system was able to monitor the artworks inside the museum and to estimate the visitor behaviors using its wireless sensor network (WSN). Multi-sensors available in a WSN gathered all the information in a central control unit and processed the data in real time to assist the museum authorities, which enables them to give immediate feedback when required.

Wang et al.<sup>107</sup> proposed a CSI-based human Activity Recognition and Monitoring (CARM) system which consists of CSI-speed model (to measure the correlation between CSI value dynamics and human movement speeds) and CSI-activity model (to measure the correlation between the movement speeds of different human body parts and a specific human activity). CARM was able to detect human activities, namely, running, walking, sitting down, opening refrigerator, falling, boxing, pushing one hand, brushing teeth, and empty (i.e. no activity) with 96% average accuracy. Also, CARM has been evaluated in different environments, namely, lab, lobby, office, and apartment, and achieved an average accuracy of 90%, 93%, 83%, and 80%, respectively.

### *Multimodal sensors for AR approaches*

Multimodal AR approaches are becoming popular in the last decade. They use visual and non-visual sensors at the same time to recognize human activities. Depending on the user requirements, the type of the model and the usage of sensors may differ. For example, one camera would be able to cover a wide area of a particular context, but it may not be enough to analyze sensitive data such as temperature of the environment, humidity, and user information like the heart rate. Clearly, systems using a single modality sensor approach would not perform well in situations where such different kinds of input data are needed. To overcome these limitations,<sup>108</sup> multi-sensor modality is introduced by exploiting various kinds of sensors in the same recognition system using, for example, multi-sensor/camera networks combined with body-worn sensors and mobile devices.

*Multimodal systems in the AAL and smart home domain.* Considering the privacy concerns of residents, sensor-based technologies are more widely used in AAL domains as compared to video-based approaches.

Vacher et al.<sup>76</sup> address the problem of learning and recognizing ADLs in smart homes using (hierarchical) hidden semi-Markov models. They have introduced a model using typical duration patterns and inherently hierarchical structures that can learn the behavior of a resident during the day. Captured activities are classified and segmented to detect the abnormal behaviors of that person.

Chen et al.<sup>109</sup> introduced a knowledge-driven approach to assist smart homes using multi-sensor data streams. This approach has been evaluated using various ADLs and user scenarios. A 94.44% average recognition rate has been achieved for recognizing activities such as making tea, brushing teeth, making coffee, having bath, watching TV, making chocolate, making pasta, and washing hands. The average time for recognizing an operation was 2.5 s.

Brdiczka et al.<sup>110</sup> presented an approach for a smart home environment capable of learning and recognizing human behaviors from multimodal observation data. A 3D video tracking system has been used for tracking the entities (persons) of a scene and a speech activity detector to analyze audio streams of each entity. The system was able to identify basic individual activities such as “walking” and “interacting with table.” Based on this individual information, group situations such as aperitif, presentation, and siesta were identified.

Chahuara et al.<sup>111</sup> presented an application to recognize ADLs in a smart home using Markov Logic Networks (MLN). Sensor raw data from non-visual and non-wearable sensors are used to create the classification model. The usage of a formal domain knowledge description and a logic-based recognition method led to a higher re-usability of the trained model from one home to another. MLN achieved 85.3% overall accuracy for the events such as eating, tidying up, hygiene, communicating, dressing up, sleeping, and resting.

Vacher et al.<sup>76</sup> presented SWEET-HOME, a project that aims at a new smart home system based on audio technology.<sup>112</sup> The developed system detects the distress situations of a person and provides assistance via natural man-machine interaction (voice and tactile commands), and security reassurance anytime, anywhere in the house. The system uses the Dedicated Markov Logic Network (DMLN) for domain knowledge representation to handle the sensor data uncertainty. During the experiment, the participants took part into four different scenarios: feeding (preparing and having a meal), sleeping, communicating (with specialized e-lío device; <http://www.technosens.fr/>), and resting (listening to the radio). The experimental results showed an accuracy of 70% for all four scenarios.

*Multimodal systems in the healthcare domain.* Multimodal sensors are deployed or used as observation tools to

monitor the health condition of patients. Maurer et al.<sup>113</sup> introduced ewatch, a multi-sensor platform that monitors and recognizes activities of a person using sensory devices deployed at different points of the body. It identifies user activities in real time and records these classification results during the day. Then, the multiple time domain feature sets and sampling rates are compared to analyze the trade-off between recognition accuracy and computational complexity. Classification accuracy is calculated for every activity collected from six different body points (wrist, pocket, bag, necklace, shirt, and belt). The data from all subjects are finally combined to train the general classifier. As a result of the performed evaluation, the collected data indicated that all six points were good for detecting walking, standing, sitting, and running activities. Based on these results, the authors have implemented a decision tree classifier that runs inside ewatch.

Similarly, Vacher et al.<sup>114</sup> proposed a multi-processing system to monitor tasks such as signal detection and channel selection, sound/speech classification, life sound classification, and speech recognition tasks of a patient. Once the system detects an emergency, it transfers the information to the remote medical monitoring application through the network.

**Multimodal systems in outdoor environments.** Multimodal sensors are also used by researchers to detect human activities in outdoor environments. Al Machot et al.<sup>61,62</sup> presented a Smart Resource Aware Multi-Sensor Network (SRSnet) to automatically detect complex events using ASP. The SRSnet system is based on audio and video processing components. The system detects complex events by aggregating simple events from audio and video data processing. Information about the detected events and PTZ configuration form the input for the network subsystem. Detected complex and simple events are stored in a multimedia data warehouse. Experimental results of SRSnet showed over 94% detection ratio for the complex events, namely, group running, fighting, and people running in different directions.

Theekakul et al.<sup>63</sup> presented a rule-based framework for human activity classification. It consists of two main components: rule learning and rule-based inference. Domain-specific knowledge is used together with sensor data to construct the classification rules. The knowledge base includes assumptions of activity characteristics, constraints of device packaging design, and so on. This approach illustrates how the domain-specific knowledge and feature space observation data can be used for rule construction. The orientation and baseline rules were applied for training and testing processes. Test data sets successfully detected lying, sitting and standing, walking, running, and jumping activities.

The detection was performed with an accuracy of 76.43% and 74.46% for the training and test data sets, respectively.

**Mobile-based multimodal systems.** Mobile devices have recently become sophisticated, and most of the today's mobile devices are equipped with powerful sensors, such as Global Positioning System (GPS), audio (i.e. microphones), image (i.e. camera), temperature, direction (i.e. compasses), and acceleration, and provide a significant computing power. Consequently, they open a wide area for researchers to find innovative solutions for HAR.

Lara et al.<sup>115</sup> presented Centinela, a mobile platform that combines mobile acceleration sensor data with vital signs (e.g. heart rate and respiration rate). This platform consists of a single sensing device in a mobile phone, which makes it a portable and unobtrusive real-time data collection platform. It uses both statistical and structural detectors while introducing two new features, trend and magnitude, to detect the differentiation of vital sign stabilization during periods of activities. Centinela was able to recognize five different activities, namely, walking, running, sitting, ascending, and descending activities.

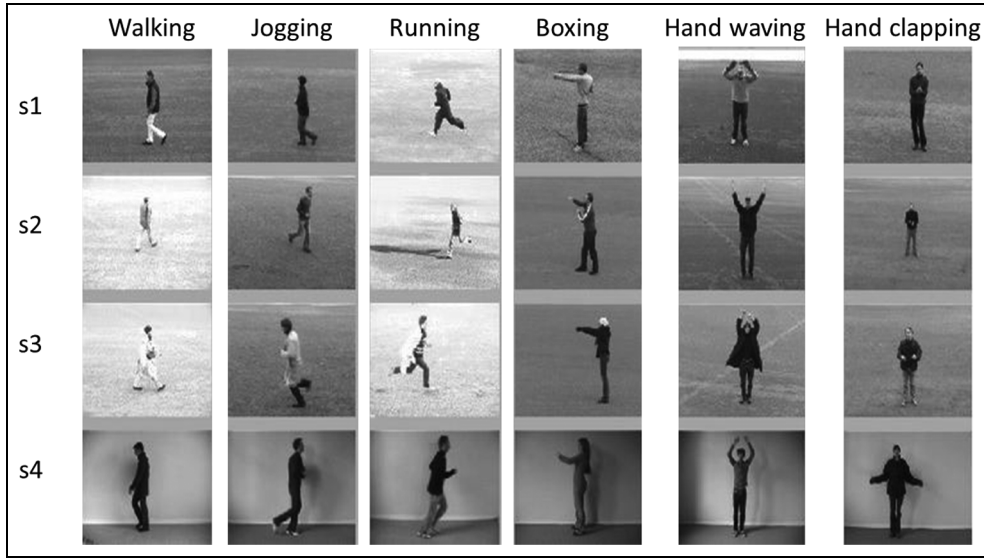
Vigilante,<sup>116</sup> an HAR application on Android, was also developed by Lara et al. based on Centinela. Vigilante uses a library called MECLA to exploit multiple sensing devices which are integrated into a body area network (BAN). Evaluation results showed that the application could run up to 12.5 h continuously in a mobile phone and achieved 92.6% of overall accuracy to identify walking, running, and sitting activities.

Kwapisz et al.<sup>117</sup> introduced a system which uses a phone-based accelerometer to recognize human activities. This system is reported to be able to collect labeled accelerometer data, namely, walking, jogging, climbing stairs, sitting, and standing events. The results are used to train a predictive model for recognizing activities. The experimental results showed recognition rates over 90%.

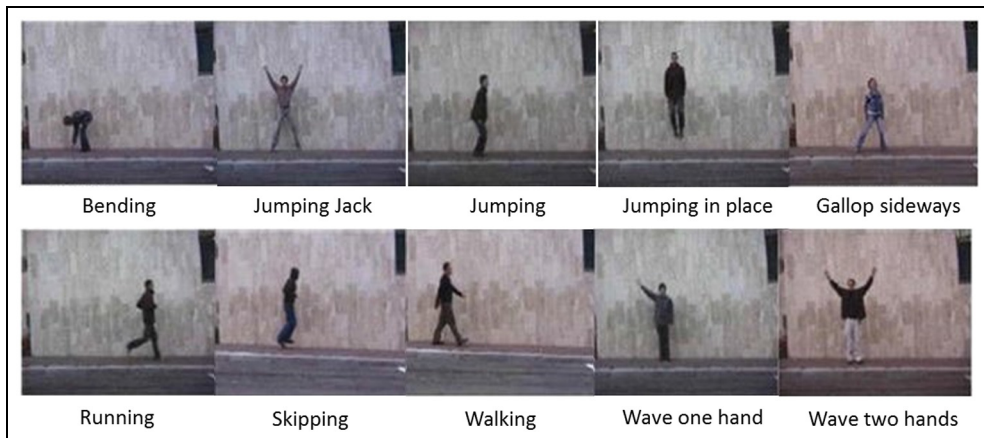
Similarly, Riboni et al.<sup>118</sup> presented COSAR, a framework built on Android for context-aware AR. This framework uses ontologies and ontological reasoning methods, combined with statistical inferencing. The structured symbolic knowledge about the environment surrounding the user allows the system to successfully identify a particular user activity. Integrating ontological reasoning with statistical methods demonstrated an overall accuracy of 92.64% which is higher than that of other pure statistical methods.

## Popular data sets

In this section, we outline some publicly available data sets currently used in the HAR research community for



**Figure 2.** Examples of KTH data set.<sup>119</sup>



**Figure 3.** Examples of Weizmann data set.<sup>120</sup>

evaluation and comparison purposes. We divide the section into subsections describing video-based and non-video-based data sets.

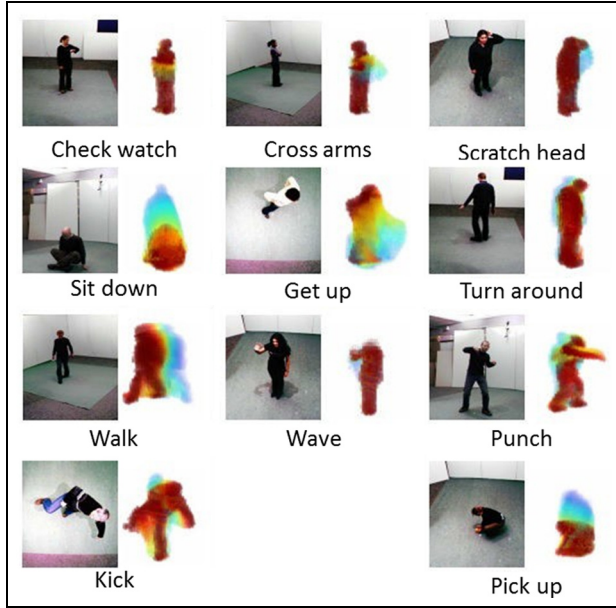
#### Video-based data sets

**KTH data set.** KTH video database<sup>119</sup> consists of clips showing sequences like walking, running, jogging, hand-waving, boxing, and hand-clapping actions and activities (Figure 2). These video sequences are performed by 25 persons in different locations and settings, for example, outdoor, outdoor with scale variation, outdoors with different clothes, and indoors. The current KTH database consists of 2391 sequences; all sequences were collected in similar backgrounds with a 25 fps frame rate.

**Weizmann data set.** The Weizmann data set<sup>120</sup> consists of 10 action classes, namely, walk, run, jump, gallop

sideways, one hand wave, two hands wave, bend, skip, jump in place, and jumping jack (Figure 3). The data sequences were captured in similar outdoor backgrounds that consist of irregular versions (with dog, occluded, with bag, etc.) to be used in robustness experiments for the “walk” activity.

**INRIA Xmas Motion Acquisition Sequences multi-view data set.** INRIA Xmas Motion Acquisition Sequences (IXMAS) data set was introduced by Weinland et al.<sup>121</sup> It contains 14 actions, namely, check watch, cross arms, scratch head, sit down, get up, turn around, walk, wave, punch, kick, point, pick up, throw overhead, and throw from bottom up (see Figure 4). Actions were captured using five cameras. These actions were performed by 11 persons and captured three times from each person. The camera viewpoints were fixed and set up with static background and illumination settings.



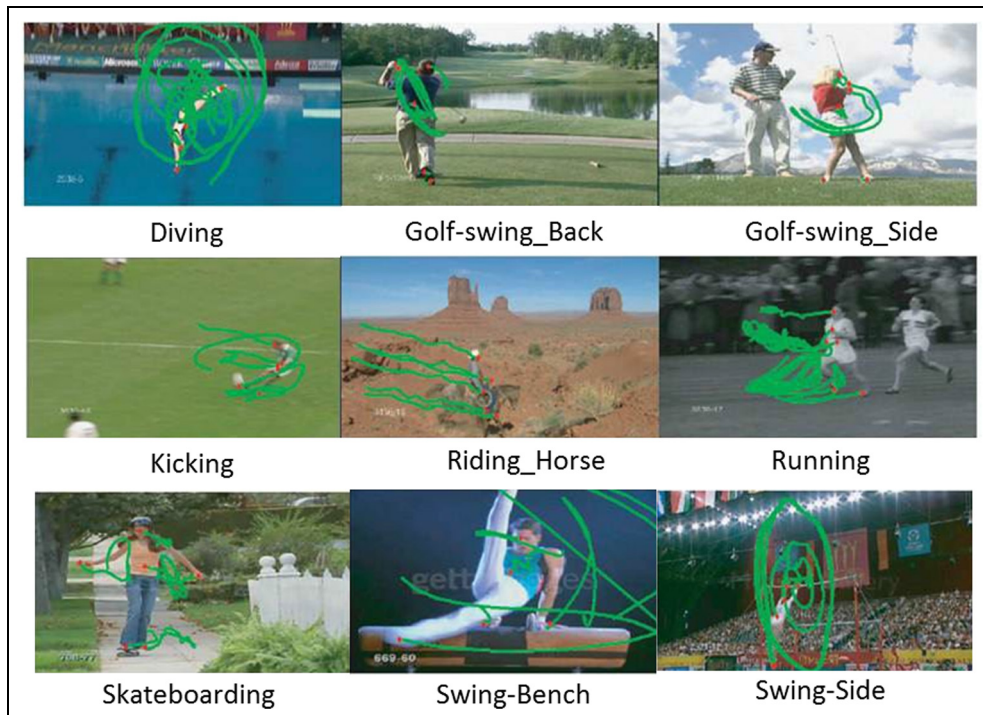
**Figure 4.** Examples of IXMAS multi-view data set.<sup>121</sup>

**UCF sport data set.** The UCF sport data set<sup>122</sup> consists of nearly 200 sports action videos that are collected from broadcast television channels at a resolution of  $720 \times 480$  and show scenes like diving, golf swinging, lifting, kicking, horseback riding, skating, running, swinging, and walking (Figure 5).

**YouTube action data set.** This data set consists of the videos divided into 11 action categories, namely, basketball shooting, biking/cycling, diving, golf swinging, horseback riding, soccer juggling, swinging, tennis swinging, trampoline jumping, volleyball spiking, and walking with a dog (Figure 6). These videos include large variations of camera motions, for example, object scale and viewpoints, object appearance and poses, cluttered backgrounds, and illumination conditions.<sup>123</sup>

**i3DPost multi-view human action data sets.** This data set was created in the framework of the i3DPost project and contains synchronized/uncompressed-HD, 8-view image sequences of 13 actions performed by 8 people (104 in total). It consists of scenes of the following activities: walking, running, jumping, hand-waving, jumping in place, sitting-standing up, running-falling, walking-sitting, running-jumping-walking, handshaking, pulling, and performing facial expression actions (see Figure 7). Additionally, it provides background images for camera calibration parameters to reconstruct 3D mesh models.<sup>71</sup>

**MOBISERV-AIIA database.** This data set is specialized to train machine learning models for recognizing “having a meal” activities, including eating and drinking (see Figure 8). MOBISERV video sequences were recorded with a resolution of  $640 \times 480$  pixels using 12 persons (6 females and 6 males) aging between 22 and 39 years



**Figure 5.** Examples of UCF sport data set.<sup>122</sup>





**Figure 6.** Examples of YouTube action data set.<sup>123</sup>



**Figure 7.** Examples of i3DPost multi-view human action data set.<sup>71</sup>

with different facial characteristics, for example, related to eyes, glasses, and beard. In this data set, eight videos were recorded with four distinct “having a meal” sessions with different cloths for each person. In total, the database consists of 384 video sequences.<sup>124</sup>

**IMPART data sets.** The IMPART multimodal/multi-view data set<sup>125</sup> consists of multimodal data footage and

3D reconstructions of various indoor/outdoor scenes, for example, LiDAR scans, digital snapshots of reconstructed 3D models, spherical camera scans, and reconstructed 3D models. Additionally, it provides multi-view video sequences of actions in indoor and outdoor environments using different types of cameras, for example, fixed multiple HD camera sequences (360°/120° setup), free-moving principal HD camera,



**Figure 8.** Examples of MOBISERV-AIIA data set.<sup>124</sup>

nodal cameras, and multi-view facial expression captures. The provided facial expressions are categorized as, for example, Neutral—Anger—Fear—Happiness—Sadness—Surprise (see Figure 9).

### Non-video-based data sets

**CASAS.** The CASAS data set<sup>126</sup> has been introduced by the Washington State University, CASAS, as a part of the CASAS smart home project. Five sequences of activities, namely, using telephone, washing hands, preparing and eating meals, using medication, and cleaning were asked to be performed by participants, and relevant sensor information was collected using motion, temperature, water, burner, phone usage (for completed calls), and item sensor readings (see Figure 10).

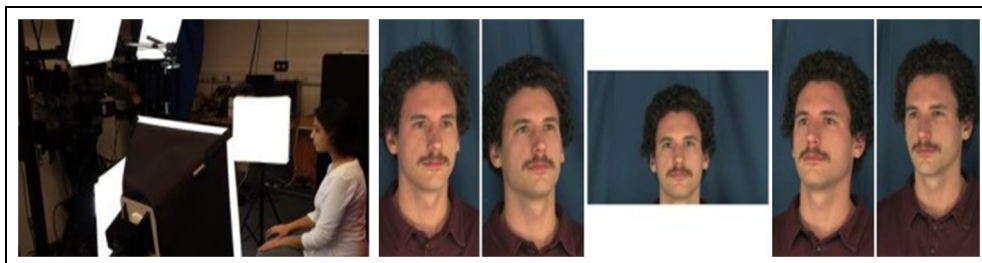
**Benchmark (Van Kasteren) data set.** This data set<sup>127</sup> consists of four data sets which record several weeks of human behaviors inside their homes. Reed switches, pressure mats, mercury contacts, passive infrared, and float sensors were used to collect the data. Activities such as leaving, toileting, showering, brushing teeth, sleeping, having breakfast, preparing dinner, preparing snacks, and drinking were annotated for each data set.

**Sweet-Home data sets.** The SWEET-HOME multimodal corpus data set<sup>76</sup> was recorded by observing individuals performing ADLs in a fully equipped smart home using microphones and home automation sensors. This multimodal corpus consists of three data sets, namely, *multimodal subset*, *home automation speech subset*, and *interaction subset*. Initially, the model was trained using the data collected from 16 persons (who were healthy and had no disabilities) to recognize the human behaviors automatically. Then, the automatic voice command recognition system was developed using the home automation speech subset (audio) data. The audio data set was recorded using microphones which were placed at distant locations, for example, on ceilings (no body-worn microphones had been used). Finally, the interaction data subset was recorded based on the observations of user interactions of 11 persons (6 elder persons and 5 visually impaired persons) during the Sweet-Home system evaluation.

### Discussion of HAR approaches

In the last decades, the enthusiasm for the vision-based technology has emerged rapidly. Recent advances in computer vision and sensor device technologies have assisted researchers to address various real-world HAR problems, but there are still important issues that need to be further investigated.

As it becomes clear from this survey, there is a high demand for existing and upcoming AR solutions. Scientists have investigated different types of sensors depending on available user contexts and research requirements. For example, most of the security and surveillance solutions were developed using video cameras because of the obvious advantage that camera observations can be immediately used as valid evidential proofs. On the other hand, considering a patient monitoring system, it would be more effective and guarantee higher privacy to use body-worn sensors over a visual-based approach, since such sensors provide more accurate physiological data than can be obtained from visual data. However, forcing an old or disabled person to wear such body-disturbing and



**Figure 9.** Examples of multimodal/multi-view data set—multi-view facial expression capture.<sup>125</sup>



**Figure 10.** Resident “washing hands” (left). This activity triggers motion sensor ON/OFF events and water flow sensor values (right).<sup>126</sup>

intrusive sensors is still a controversial issue. According to the survey by Ni et al.,<sup>12</sup> the acceptance does not only rely on people’s capabilities and limitations, it also depends on the personal, socioeconomic, and cultural contexts as well. To avoid such barriers, the authors suggested technology designers to pay attention to human factor-related functionalities before starting the development of such devices. Introducing design patterns and low-level design guidelines would be helpful for the developers to easily customize the devices for each specific application, person, and context. Most of the emergency or fall detection systems are designed to trigger a response to abnormal user behavior. For avoiding the triggering of false alarms, Young and Mihailidis<sup>128</sup> developed an interface, the *Personal Emergency Response System (PERS)*, which aimed at recognizing the keywords and seniors’ speech. PERS was able to identify high-risk emergencies successfully while ignoring false alarm situations.

A study of Rialle et al.<sup>129</sup> carried out with 270 families, and dealing with their user perceptions of 14 innovative technologies, showed that smart home technologies allow caregivers to leave the nursing home without compromising the patients’ safety. This fact was highly appreciated by both, patients and caregivers. Consequently, the design of smart home technologies has to take user aspects comprehensively into account to become acceptable for the future.

TI systems and gaming interface applications use multiple camera views to identify human interactions and simulate them within a virtual environment. Although TI systems require higher processing power due to their nature, cameras can hardly be replaced by other sensory devices.

A single camera can replace multiple sensor deployments in an AAL environment as it tracks a wide angle of an environment and captures all the environmental information at once. Nevertheless, using camera observations may not be the most suitable option for AAL environments considering the user privacy and

acceptance issues. Also, it is quite challenging to identify complex user activities using visual-based approaches considering the need of higher computer processing power for the analysis. Non-visual fall detection experiments within the AAL context showed higher recognition rates than experiments using visual-based approaches. In general, for AAL purposes, non-visual-based HAR approaches are preferable.<sup>130,131</sup>

A large set of sensor network can track the information of simple human activities in detail. This information can later be combined using computer fusion algorithms to detect the complex behavior of a particular AAL resident. In order to achieve highly accurate results, it is quite essential to deploy a sufficient amount of sensors inside the desired environment, for example, sensor placed in every door at a smart home and every daily used equipment. Otherwise, final predictions may mislead to false results. On the other hand, deploying and maintaining such a sensor network is quite challenging and expensive. Therefore, multimodal approaches might be a good choice, when the obtrusive components are embedded and used carefully and consciously; for example, body-worn sensors and mobile phone sensors have been used to detect such situations and showed higher success results.<sup>132,133</sup>

Patient monitoring systems benefit from using sensor-based solutions; however, acceptance aspects have to be carefully considered; moreover, wearing a sensor kit for a long time may not be comfortable for the user.

Most of the visual sensor-based equipment is sensitive to light/brightness factors of the environment; higher brightness or lower brightness image streams could possibly lack information sufficient for the classifiers to identify human behavior. Similarly, sensor-based approaches suffer from sensor robustness issues, that is, some sensors may not automatically get activated by the predefined user behavior due to a malfunction. Also, some of the sensors may automatically be activated even without human interaction due to weather conditions such as lightning, wind, and rain.

Considering the performance factors, visual sensor-based approaches require a higher amount of computer processing power to process the data compared to other approaches due to the complexity of the vision algorithms and the data volumes. In contrast to this, non-visual sensor approaches perform comparatively faster and consume less energy. Most of the multimodal but non-visual algorithms can even run inside an embedded platform, consuming a low amount of energy. Table 6 shows a comparison.

## Conclusion and further research

Most of the existing AR systems are designed and tested under laboratory conditions. Thus, using such a system in a real-time scenario would not give a better



**Table 6.** Comparison of HAR approaches.

	Advantages	Disadvantages	Performance
Visual sensor-based approaches	Single camera can track a wide angle of an environment Able to replace many sensory devices with one camera Easy to operate Provide reliable data Suitable for security and surveillance systems and tele-immersion systems	Privacy issues Track only specific details of the environment High-sensitive video cameras are comparatively expensive Computer processing power is too high Sensitive for the light/brightness factors Higher power consumption to operate Existing training data sets are not simulating realistic environments Require more processing time Acceptance issues	Need a higher amount of computer processing power to perform Processing time is comparatively high
Sensor-based (non-visual) approaches	Large set of sensor-based network can track every detail of human behavior including human body information Secure the privacy aspects compared to video-based solutions Comparatively less computer processing power is required Comparatively low power consumption to operate Suitable for healthcare and AAL systems Low cost	Need a large set of sensors, specifically to track each behavior Provide unreliable data Accuracy issues One sensor malfunction may lead for false detections Acceptance issues	Able to perform faster even with a smaller amount of computer processing power Processing time is comparatively low
Multimodal approaches	Suitable for detecting complex activities Light-weight sensors Comparatively higher accuracy rate Comparatively low power consumption to operate Suitable for healthcare and AAL systems	Most of the time, target persons need to carry or wear the smart kit Require multiple sensors to capture full body movements Intrusiveness of wearing single or multiple sensors Acceptance issues Data fusion algorithms may lead to false predictions	Able to perform faster even with a smaller amount of computer processing power Processing time is comparatively low

performance if the system is not adapted to the new environment. However, deploying such an application in a real environment will not be appropriate unless the application is tested under real-world conditions which consist of noise, occlusions, shadows, and other factors.

Human behavior is spontaneous; in particular, humans tend to do several activities at the same time, for example, cooking while talking to friends, and watching TV while eating. Therefore, future HAR systems should recognize such concurrent activities rather than focusing only on a single activity at a given time. Furthermore, some of the human activities may be interleaved, for example, when watching TV and phoning with a friend in parallel, there could be moments where you concentrate on the TV program and thus delay phone answers to your friend. Future HAR systems should be designed to successfully recognize such interleaved situations. Additionally, HAR systems should strengthen the handling of uncertainty, that is, avoid

ambiguous behavior interpretations; as an example, the action *opening a refrigerator* can belong to several activities such as cooking and cleaning.

Most of the recently reported experiments were based on non-visual sensor data collected from a single user activity. But when considering real-life scenarios, activities can be performed by multiple users concurrently. Recognizing multi-user activities is challenging; it should incorporate an appropriate amount of sensors, suitable methods to model multi-user interactions, and filtering useful information from the obtained data. Carrying out sensor data fusion for such settings to achieve sufficient accuracy for activity monitoring is still an open research issue.<sup>12</sup>

Ziefle and Wilkowska<sup>134</sup> showed that people's willingness to use medical technology in a case of need outweighs negative feedback. However, they state that most of representatives of the older generation ("early technical generation") expressed higher levels of

aloofness and distrust against innovative medical solutions. Nevertheless, it seems to be a common belief that older and disabled people should use the upcoming medical devices in order to be autonomous in their homes, if there are no alternatives at affordable cost.

Several research groups have published data sets which were produced by the observation of different individuals, for different activity classes, and used for different evaluation methods, as was discussed in section “Popular data set.” When evaluating a new approach, it is quite essential to use such a publicly open data set and use them as a benchmark for the evaluation.

### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was funded by the Klaus Tschira Stiftung gGmbH, Heidelberg.

### References

1. Krishnan NC, Juillard C, Colbry D, et al. Recognition of hand movements using wearable accelerometers. *J Ambient Intell Smart Environ* 2009; 1(2): 143–155.
2. Michael J, Grieser A, Strobl T, et al. Cognitive modeling and support for ambient assistance. In: Mayr HC, Kop C, Liddle S, et al. (eds) *Information systems: methods, models, and applications*. Berlin and Heidelberg: Springer 2012, pp.96–107.
3. Menschner P, Prinz A, Koene P, et al. Reaching into patients’ homes—participatory designed AAL services. *Electron Market* 2012; 21(1): 63–76.
4. Turaga P, Chellappa R, Subrahmanian VS, et al. Machine recognition of human activities: a survey. *IEEE T Circ Syst Vid* 2008; 18(11): 1473–1488.
5. Aggarwal JK and Ryoo MS. Human activity analysis: a review. *ACM Comput Surv* 2011; 43(3): 16.
6. Bashir F, Khokhar A and Schonfeld D. Object trajectory-based activity classification and recognition using hidden Markov models. *IEEE T Image Process* 2007; 16(7): 1912–1919.
7. Leontev AN. *Activity, consciousness, and personality*. Englewood Cliffs, NJ: Prentice Hall, 1978.
8. Mayr HC, Al Machot F, Michael J, et al. HCM-L: domain specific modeling for active and assisted living. In: Karagiannis D, Mayr HC and Mylopoulos J (eds) *Domain-specific conceptual modeling—concepts, methods and tools*. Berlin: Springer, 2016, pp.527–552.
9. Katz S. Assessing self-maintenance: activities of daily living, mobility, and instrumental activities of daily living. *J Am Geriatr Soc* 1983; 31: 721–727.
10. Lawton MP and Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist* 1969; 9: 179–186.
11. Okeyo G, Chen L and Wang H. Combining ontological and temporal formalisms for composite activity modeling and recognition in smart homes. *Future Gener Comp Sy* 2014; 39: 29–43.
12. Ni Q, García Hernando AB and de la Cruz IP. The elderly’s independent living in smart homes: a characterization of activities and sensing infrastructure survey to facilitate services development. *Sensors* 2015; 15(5): 11312–11362.
13. Dohr A, Modre-Opsrian R, Drobits M, et al. The Internet of things for ambient assisted living. In: *2010 seventh international conference on information technology*, Las Vegas, NV, 12–14 April 2010, pp.804–809. New York: IEEE.
14. Hartmann R, Al Machot F, Mahr P, et al. Camera-based system for tracking and position estimation of humans. In: *2010 conference on design and architectures for signal and image processing (DASIP)*, Edinburgh, 26–28 October 2010, pp.62–67. New York: IEEE.
15. Chaaraoui AA, Padilla-López JR, Ferrández-Pastor FJ, et al. A vision-based system for intelligent monitoring: human behaviour analysis and privacy by context. *Sensors* 2014; 14(5): 8895–8925.
16. Romdhane R, Mulin E, Derreumeaux A, et al. Automatic video monitoring system for assessment of Alzheimer’s disease symptoms. *J Nutr Health Aging* 2012; 16(3): 213–218.
17. Negin F, Cosar S, Koperski M, et al. Generating unsupervised models for online long-term daily living activity recognition. In: *Asian conference on pattern recognition (ACPR 2015)*, November 2015, <https://hal.inria.fr/hal-01233494/document>
18. Bilinski P, Corvee E, Bak S, et al. Relative dense tracklets for human action recognition. In: *2013 10th IEEE international conference and workshops automatic face and gesture recognition (FG)*, Shanghai, China, 22–26 April 2013, pp.1–7. New York: IEEE.
19. Van den Heuvel H, Huijnen C, Caleb-Solly P, et al. Mobiserv: a service robot and intelligent home environment for the Provision of health, nutrition and safety services to older adults. *Gerontechnology* 2012; 11(2): 373.
20. Goldstone JA. The new population bomb: the four megatrends that will change the world. *Foreign Affairs* 2010, pp.31–43, <https://www.foreignaffairs.com/articles/2010-01-01/new-population-bomb>
21. Demiris G, Hensel BK, Skubic M, et al. Senior residents perceived need of and preferences for smart home sensor technologies. *Int J Technol Assess Health Care* 2008; 24(1): 120–124.
22. Ram R, Furfari F, Girolami M, et al. UniversAAL: provisioning platform for AAL services. In: Van Berlo A, Hallenborg K, Corchado Rodríguez JM, et al. (eds) *Ambient intelligence-software and applications*. Berlin: Springer International Publishing, 2013, pp.105–112.
23. Rashidi P and Cook DJ. Keeping the resident in the loop: adapting the smart home to the user. *IEEE T Syst Man Cy A* 2009; 39(5): 949–959.
24. Roy N, Misra A and Cook D. Ambient and smartphone sensor assisted ADL recognition in multi-inhabitant smart environments. *J Ambient Intell Human Comput* 2016; 7(1): 1–19.

25. Blumendorf M and Albayrak S. Towards a framework for the development of adaptive multimodal user interfaces for ambient assisted living environments. In: Stephanidis C (ed.) *Universal access in human-computer interaction, Intelligent and ubiquitous interaction environments*. Berlin and Heidelberg: Springer, 2009, pp.150–159.
26. Lloret J, Canovas A, Sendra S, et al. A smart communication architecture for ambient assisted living. *IEEE Commun Mag* 2015; 53(1): 26–33.
27. Gannapathy VR, Ibrahim AFBT, Zakaria ZB, et al. Zigbee based smart fall detection and notification system with wearable sensor (e-safe). *Int J Res Eng Technol* 2013; 2(8): 337–344.
28. Zhu C and Sheng W. Wearable sensor-based hand gesture and daily activity recognition for robot-assisted living. *IEEE T Syst Man Cy A* 2011; 41(3): 569–573.
29. Zigel Y, Litvak D and Gannot I. A method for automatic fall detection of elderly people using floor vibrations and soundproof of concept on human mimicking doll falls. *IEEE T Biomed Eng* 2009; 56(12): 2858–2867.
30. Li Q, Stankovic JA, Hanson MA, et al. Accurate, fast fall detection using gyroscopes and accelerometer-derived posture information. In: *BSN 2009, sixth international workshop wearable and implantable body sensor networks*, Berkeley, CA, 3–5 June 2009, pp.138–143. New York: IEEE.
31. Anderson D, Luke RH, Keller JM, et al. Linguistic summarization of video for fall detection using voxel person and fuzzy logic. *Comput Vis Image Und* 2009; 113(1): 80–89.
32. Kangas M, Konttila A, Lindgren P, et al. Comparison of low-complexity fall detection algorithms for body attached accelerometers. *Gait Posture* 2008; 28(2): 285–291.
33. Wang C, Zhou J, Liao L, et al. Near-threshold energy- and area-efficient reconfigurable DWPT/DWT processor for healthcare-monitoring applications. *IEEE T Circuits II* 2015; 62(1): 70–74.
34. Doulamis A, Doulamis N, Kalisperakis I, et al. A real-time single-camera approach for automatic fall detection. In: *International archives of photogrammetry, remote sensing and spatial information sciences, XXXVIII(5), commission V symposium*, 2010, pp.207–212, <http://www.isprs.org/proceedings/xxxviii/part5/papers/142.pdf>
35. Giroux S and Pigot H. An intelligent health monitoring and emergency response system. In: Giroux S and Pigot H (eds) *From smart homes to smart care: ICOST 2005, 3rd international conference on smart homes and health telematics*, vol. 15, Oxford: IOS Press, 2005, p.272.
36. Poh MZ, McDuff DJ and Picard RW. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Opt Express* 2010; 18(10): 10762–10774.
37. Brémont F, Thonnat M and Zuniga M. Video-understanding framework for automatic behavior recognition. *Behav Res Methods* 2006; 38(3): 416–426.
38. Chang M-C, Krahnstoever N, Lim S, et al. Group level activity recognition in crowded environments across multiple cameras. In: *2010 seventh IEEE international conference advanced video and signal based surveillance (AVSS)*, Boston, MA, 29 August–1 September 2010, pp.56–63. New York: IEEE.
39. Vacher M, Istrate D, Portet F, et al. The sweet-home project: audio technology in smart homes to improve well-being and reliance. *Conf Proc IEEE Eng Med Biol Soc* 2011; 2011: 5291–5294.
40. Vacher M, Caffiau S, Portet F, et al. Evaluation of a context-aware voice interface for Ambient Assisted Living: qualitative user study vs. quantitative system evaluation. *ACM T Access Comput* 2015; 7(2): 1–36.
41. Lustrek M and Kaluza B. Fall detection and activity recognition with machine learning. *Informatica* 2009; 33(2): 197–204.
42. Bourke AK and Lyons GM. A threshold-based fall-detection algorithm using a bi-axial gyroscope sensor. *Med Eng Phys* 2008; 30(1): 84–90.
43. Vo MT, Nghi TT, Tran VS, et al. Wireless sensor network for real time healthcare monitoring: network design and performance evaluation simulation. In: Toi VV and Lien Phuong TH (eds) *5th international conference on biomedical engineering in Vietnam*. Berlin: Springer International Publishing, 2015, pp.87–91.
44. Peursum P, West G and Venkatesh S. Combining image regions and human activity for indirect object recognition in indoor wide-angle views. In: *Tenth IEEE international conference computer vision, ICCV 2005, Beijing, China*, 17–21 October 2005, vol. 1, pp.82–89. New York: IEEE.
45. Fusier F, Valentin V, Bremond F, et al. Video understanding for complex activity recognition. *Mach Vision Appl* 2007; 18(3–4): 167–188.
46. Michael J and Mayr HC. Creating a domain specific modelling method for ambient assistance. In: *2015 fifteenth international conference advances in ICT for emerging regions (ICTer)*, Colombo, Sri Lanka, 24–26 August 2015, pp.119–124. New York: IEEE.
47. Costa R, Carneiro D, Novais P, et al. Ambient assisted living. In: Rodríguez C, Manuel J, Dante T, et al. (eds) *3rd symposium of ubiquitous computing and ambient intelligence 2008*. Berlin and Heidelberg: Springer, 2009, pp.86–94.
48. Van Kasteren TLM, Englebienne G and Krose BJ. An activity monitoring system for elderly care using generative and discriminative models. *Pers Ubiquit Comput* 2010; 14(6): 489–498.
49. Wu C, Khalili AH and Aghajan H. Multiview activity recognition in smart homes with spatio-temporal features. In: *Proceedings of the fourth ACM/IEEE international conference on distributed smart cameras*, Atlanta, GA, 31 August–4 September 2010, pp.142–149. New York: ACM.
50. Hsieh JW, Hsu YT, Liao HY, et al. Video-based human movement analysis and its application to surveillance systems. *IEEE T Multimedia* 2008; 10(3): 372–384.
51. Akdemir U, Turaga P and Chellappa R. An ontology based approach for activity recognition from video. In: *Proceedings of the 16th ACM international conference on multimedia*, Vancouver, BC, Canada, 26–31 October 2008, pp.709–712. New York: ACM.
52. Fookes C, Denman S, Lakemond R, et al. Semi-supervised intelligent surveillance system for secure environments. In: *2010 IEEE international symposium industrial*

- electronics (ISIE)*, Bari, 4–7 July 2010, pp.2815–2820. New York: IEEE.
53. Dufaux F and Ebrahimi T. Scrambling for privacy protection in video surveillance systems. *IEEE T Circ Syst Vid* 2008; 18(8): 1168–1174.
  54. Krahnstoeber N, Yu T, Lim SN, et al. Collaborative real-time control of active cameras in large scale surveillance systems. In: *Workshop on multi-camera and multi-modal sensor fusion algorithms and applications (M2SFA2 2008)*, 2008, <https://hal.inria.fr/inria-00326743/document>
  55. Huang SC. An advanced motion detection algorithm with video quality analysis for video surveillance systems. *IEEE T Circ Syst Vid* 2011; 21(1): 1–14.
  56. Maddalena L and Petrosino A. A self-organizing approach to background subtraction for visual surveillance applications. *IEEE T Image Process* 2008; 17(7): 1168–1177.
  57. Li L, Huang W, Gu IH, et al. An efficient sequential approach to tracking multiple objects through crowds for real-time intelligent CCTV systems. *IEEE T Syst Man Cy B* 2008; 38(5): 1254–1269.
  58. Di Paola D, Naso D, Milella A, et al. Multi-sensor surveillance of indoor environments by an autonomous mobile robot. *Int J Intell Syst* 2010; 8(1): 18–35.
  59. Xu Z, Hu C and Mei L. Video structured description technology based intelligence analysis of surveillance videos for public security applications. *Multimed Tools Appl* 2015; 1–18. DOI: 10.1007/s11042-015-3112-5.
  60. Belyaev E, Vinel A, Surak A, et al. Robust vehicle-to-infrastructure video transmission for road surveillance applications. *IEEE T Veh Technol* 2015; 64(7): 2991–3003.
  61. Al Machot F, Kyamakya K, Dieber B, et al. Real time complex event detection for resource-limited multimedia sensor networks. In: *8th IEEE international conference on advanced video and signal-based surveillance (AVSS)*, Klagenfurt, 30 August–2 September 2011, pp.468–473. New York: IEEE.
  62. Al Machot F, Kyamakya K, Dieber B, et al. Smart resource-aware multimedia sensor network for automatic detection of complex events. In: *8th IEEE international conference on advanced video and signal-based surveillance (AVSS)*, Klagenfurt, 30 August–2 September 2011, pp.402–407. New York: IEEE.
  63. Theekakul P, Thiemjarus S, Nantajeewarawat E, et al. A rule-based approach to activity recognition. In: Theeramunkong T, Kunifuji S, Sornlertlamvanich V, et al. (eds) *Knowledge, information, and creativity support systems*. Berlin and Heidelberg: Springer, 2011, pp.204–215.
  64. Rashidi P, Cook DJ, Holder LB, et al. Discovering activities to recognize and track in a smart environment. *IEEE T Knowl Data En* 2011; 23(4): 527–539.
  65. Tunca C, Alemdar H, Ertan H, et al. Multimodal wireless sensor network-based ambient assisted living in real homes with multiple residents. *Sensors* 2014; 14(6): 9692–9719.
  66. Blasco R, Marco Á, Casas R, et al. A smart kitchen for ambient assisted living. *Sensors* 2014; 14(1): 1629–1653.
  67. Chernbumroong S, Cang S, Atkins A, et al. Elderly activities recognition and classification for applications in assisted living. *Expert Syst Appl* 2013; 40(5): 1662–1674.
  68. Lien JM, Kurillo G and Bajcsy R. Multi-camera tele-immersion system with real-time model driven data compression. *Visual Comput* 2010; 26: 3–15.
  69. Feldmann I, Waizenegger W, Atzpadin N, et al. Real-time depth estimation for immersive 3D videoconferencing. In: *Proceedings of 3DTV-conference (3DTV-CON 10)*, Tampere, 7–9 June 2010, pp.1–4. New York: IEEE.
  70. Mekuria R, et al. A 3D tele-immersion system based on live captured mesh geometry. In: *Proceedings of the ACM multimedia systems conference (MMSys 13)*, Oslo, Norway, 27 February–1 March 2013, pp.24–35. New York: ACM.
  71. Gkalelis N, Kim H, Hilton A, et al. The i3DPost multi-view and 3D human action/interaction. In: *CVMP'09, conference for visual media production*, London, 12–13 November 2009, pp.159–168. New York: IEEE.
  72. Kurillo G and Bajcsy R. 3D teleimmersion for collaboration and interaction of geographically distributed users. *Virtual Real* 2013; 17(1): 29–43.
  73. Zhang C, Cai Q, Chou P, et al. *Viewport: a fully distributed immersive teleconferencing system with infrared dot pattern*. Technical report MSR-TR-2012-60, Microsoft Research, 1 April 2012, pp.1–11.
  74. Petit B, Lesage JD, Menier C, et al. Multicamera real-time 3D modeling for telepresence and remote collaboration. *Int J Digital Multimedia Broadcast* 2010; 2010: Article ID 247108 (12 pp.).
  75. Huang Z, Wu W, Nahrstedt K, et al. TSynC: a new synchronization framework for multi-site 3D tele-immersion. In: *Proceedings of the 20th international workshop on network and operating systems support for digital audio and video*, Amsterdam, 2–4 June 2010, pp.39–44. New York: ACM.
  76. Vacher M, Lecouteux B, Chahuara P, et al. The Sweet-Home speech and multimodal corpus for home automation interaction. In: *The 9th edition of the language resources and evaluation conference (LREC)*, Reykjavik, Iceland, 26–31 May 2014, pp.4499–4506, [http://www.lrec-conf.org/proceedings/lrec2014/pdf/118\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2014/pdf/118_Paper.pdf)
  77. Kurillo G, Koritnik T, Bajd T, et al. Real-time 3D avatars for tele-rehabilitation in virtual reality. In: *Proceedings of 18th medicine meets virtual reality (MMVR) conference*, Newport Beach, CA, February 2011, pp.290–296, <http://people.eecs.berkeley.edu/~gregorij/MMVR2011.pdf>
  78. Lin CH, Sun PY and Yu F. Space connection: a new 3D tele-immersion platform for web-based gesture-collaborative games and services. In: *2015 IEEE/ACM 4th international workshop games and software engineering (GAS)*, Florence, 18 May 2015, pp.22–28. New York: IEEE.
  79. Liu Y, Beck S, Wang R, et al. Hybrid lossless-lossy compression for real-time depth-sensor streams in 3D telepresence applications. In: Ho Y-S, Sang J, Ro YM, et al. (eds) *Advances in multimedia information processing—PCM 2015*. Berlin: Springer International Publishing, 2015, pp.442–452.
  80. Kim H, Pabst S, Sneddon J, et al. Multi-modal big-data management for film production. In: *2015 international conference on image processing (ICIP)*, Quebec City, QC, Canada, 27–30 September 2015, pp.4833–4837. New York: IEEE.

81. Weinland D, Ronfard R and Boyer E. A survey of vision-based methods for action representation, segmentation and recognition. *Comput Vis Image Und* 2011; 115(2): 224–241.
82. Soro S and Heinzelman W. A survey of visual sensor networks. *Adv Multimedia* 2009; 2009: 1–22.
83. Fosty B, Crispim-Junior CF, Badie J, et al. Event recognition system for older people monitoring using an RGB-D camera. In: *ASROB-workshop on assistance and service robotics in a human environment*, 2013, <http://www-sop.inria.fr/members/Francois.Bremond/Postscript/baptiste-ASROB2013.pdf>
84. Xia L, Chen CC and Aggarwal J. View invariant human action recognition using histograms of 3D joints. In: *2012 IEEE computer society conference on computer vision and pattern recognition workshops (CVPRW)*, 2012, pp.20–27. IEEE, [http://cvrc.ece.utexas.edu/Publications/Xia\\_HAU3D12.pdf](http://cvrc.ece.utexas.edu/Publications/Xia_HAU3D12.pdf)
85. Shotton J, Sharp T, Kipman A, et al. Real-time human pose recognition in parts from single depth images. *Commun ACM* 2013; 56(1): 116–124.
86. Romdhane R, Boulay B, Bremond F, et al. Probabilistic recognition of complex event. In: Crowley JL, Draper BA and Thonnat M (eds) *Computer vision systems*. Berlin and Heidelberg: Springer, 2011, pp.122–131.
87. Foroughi H, Naseri A, Saberi A, et al. An eigenspace-based approach for human fall detection using integrated time motion image and neural network. In: *9th international conference on signal processing, ICSP*, Beijing, China, 26–29 October 2008, pp.1499–1503. New York: IEEE.
88. Chen D, Bharucha AJ and Wactlar HD. Intelligent video monitoring to improve safety of older persons. *Conf Proc IEEE Eng Med Biol Soc* 2007; 2007: 3814–3817.
89. Zaidenberg S, Boulay B and Bremond F. A generic framework for video understanding applied to group behavior recognition. In: *2012 IEEE ninth international conference on advanced video and signal-based surveillance (AVSS)*, 2012, pp.136–142, <https://hal.inria.fr/hal-00702179/file/avss.pdf>
90. Cupillard F, Bremond F and Thonnat M. Behaviour recognition for individuals, groups of people and crowd, 2003, <http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=1300131&url=http%3A%2F%2Fieeexplore.ieee.org%2Fiel5%2F9108%2F28886%2F01300131.pdf%3Farnumber%3D1300131>
91. Nievas EB, Suarez OD, Garcia GB, et al. Violence detection in video using computer vision techniques. In: Real P, Diaz-Pernil D, Molina-Abril H, et al. (eds) *Computer analysis of images and patterns*. Berlin and Heidelberg: Springer, 2011, pp.332–339.
92. Direkolu C and O'Connor NE. Team activity recognition in sports. In: Fitzgibbon A, Lazebnik S, Perona P, et al. (eds) *Computer vision—ECCV 2012*. Berlin and Heidelberg: Springer, 2012, pp.69–83.
93. Sadanand S and Corso JJ. Action bank: a high-level representation of activity in video. In: *2012 IEEE conference on computer vision and pattern recognition (CVPR)*, Providence, RI, 16–21 June 2012, pp.1234–1241. New York: IEEE.
94. Li LJ, Su H, Fei-Fei L, et al. Object bank: a high-level image representation for scene classification & semantic feature sparsification. In *Advances in neural information processing systems*, 2010, pp.1378–1386, <http://vision.stanford.edu/pdf/LiSuXingFeiFeiNIPS2010.pdf>
95. Brendel W, Fern A and Todorovic S. Probabilistic event logic for interval-based event recognition. In: *2011 IEEE conference on computer vision and pattern recognition (CVPR)*, Providence, RI, 20–25 June 2011, pp.3329–3336. New York: IEEE.
96. Tang K, Fei-Fei L and Koller D. Learning latent temporal structure for complex event detection. In: *2012 IEEE conference on computer vision and pattern recognition (CVPR)*, Providence, RI, 16–21 June 2012, pp.1250–1257. New York: IEEE.
97. Crispim C and Bremond F. Uncertainty modeling framework for constraint-based elementary scenario detection in vision system. In: Agapito L, Bronstein MM and Rother C (eds) *European conference on computer vision*. Berlin: Springer International Publishing, 2014, pp.269–282.
98. Nghiem AT, Auvinet E and Meunier J. Head detection using kinect camera and its application to fall detection. In: *2012 11th international conference on information science, signal processing and their applications (ISSPA)*, Montreal, QC, Canada, 2–5 July 2012, pp.164–169. New York: IEEE.
99. Chau DP, Bremond F and Thonnat M. A multi-feature tracking algorithm enabling adaptation to context variations. In: *4th international conference imaging for crime detection and prevention 2011 (ICDP 2011)*, London, UK, 3–4 November 2011, pp.1–6. IET.
100. Touati R and Mignotte M. MDS-based multi-axial dimensionality reduction model for human action recognition. In: *2014 Canadian conference on computer and robot vision (CRV)*, Montreal, QC, Canada, 6–9 May 2014, pp.262–267. New York: IEEE.
101. Huynh DTG. *Human activity recognition with wearable sensors*. PhD Thesis, TU Darmstadt, Darmstadt, 2008.
102. Fleury A, Vacher M and Noury N. SVM-based multi-modal classification of activities of daily living in health smart homes: sensors, algorithms, and first experimental results. *IEEE T Inf Technol B* 2010; 14(2): 274–283.
103. Hong X, Nugent C, Mulvenna M, et al. Evidential fusion of sensor data for activity recognition in smart homes. *Pervasive Mobile Comput* 2009; 5(3): 236–252.
104. Szewczyk S, Dwan K, Minor B, et al. Annotating smart environment sensor data for activity learning. *Technol Health Care* 2009; 17(3): 161–169.
105. Viani F, Martinelli M, Ioriatti L, et al. Real-time indoor localization and tracking of passive targets by means of wireless sensor networks. In: *Antennas and propagation society international symposium, APSURSI'09*, Charleston, SC, 1–5 June 2009, pp.1–4. New York: IEEE.
106. Viani F, Salucci M, Rocca P, et al. A multi-sensor WSN backbone for museum monitoring and surveillance. In: *2012 6th European conference on antennas and propagation (EUCAP)*, Prague, 26–30 March 2012, pp.51–52. New York: IEEE.
107. Wang W, Liu AX, Shahzad M, et al. Understanding and modeling of WiFi signal based human activity

- recognition. In: *Proceedings of the 21st annual international conference on mobile computing and networking*, Paris, 7–11 September 2015, pp.65–76. New York: ACM.
108. Chen C, Jafari R and Kehtarnavaz N. A survey of depth and inertial sensor fusion for human action recognition. *Multimed Tools Appl* 2015; 1–21. DOI: 10.1007/s11042-015-3177-1.
109. Chen L, Nugent CD and Wang H. A knowledge-driven approach to activity recognition in smart homes. *IEEE T Knowl Data En* 2012; 24(6): 961–974.
110. Brdiczka O, Langet M, Maisonnasse J, et al. Detecting human behavior models from multimodal observation in a smart home. *IEEE T Autom Sci Eng* 2009; 6(4): 588–597.
111. Chahuara P, Fleury A, Portet F, et al. Using Markov Logic Network for on-line activity recognition from non-visual home automation sensors. In: Paternò F, de Ruyter B, Markopoulos P, et al. (eds) *Ambient intelligence*. Berlin and Heidelberg: Springer, 2012, pp.177–192.
112. Vacher M. Projet SWEET-HOME—ANR, 2016, <http://sweet-home.imag.fr/> (accessed 5 May 2016).
113. Maurer U, Smailagic A, Siewiorek DP, et al. Activity recognition and monitoring using multiple sensors on different body positions. In: *BSN 2006, international workshop wearable and implantable body sensor networks*, Cambridge, MA, 3–5 April 2006, p.4. New York: IEEE.
114. Vacher M, Serignat JF, Chaillol S, et al. Speech and sound use in a remote monitoring system for health care. In: Sojka P, Kopeček I and Pala K (eds) *Text, speech and dialogue*. Berlin and Heidelberg: Springer, 2006, pp.711–718.
115. Lara OD, Perez AJ, Labrador MA, et al. Centinela: a human activity recognition system based on acceleration and vital sign data. *Pervasive Mobile Comput* 2012; 8(5): 717–729.
116. Lara OD and Labrador MA. A mobile platform for real-time human activity recognition. In: *2012 IEEE consumer communications and networking conference (CCNC)*, Las Vegas, NV, 14–17 January 2012, pp.667–671. New York: IEEE.
117. Kwapisz JR, Weiss GM and Moore SA. Activity recognition using cell phone accelerometers. *ACM SigKDD: Explorat Newsletter* 2011; 12(2): 74–82.
118. Riboni D and Bettini C. COSAR: hybrid reasoning for context-aware activity recognition. *Pers Ubiquit Comput* 2011; 15(3): 271–289.
119. Schuldt C, Laptev I and Caputo B. Recognizing human actions: a local SVM approach. In: *Proceedings of the 17th international conference on pattern recognition, ICPR 2004*, 2004, vol. 3, pp.32–36, [http://www.irisa.fr/vista/Papers/2004\\_icpr\\_schuldt.pdf](http://www.irisa.fr/vista/Papers/2004_icpr_schuldt.pdf)
120. Gorelick L, Blank M, Shechtman E, et al. Actions as space-time shapes. *IEEE Trans Pattern Anal Mach Intell* 2007; 29(12): 2247–2253.
121. Weinland D, Ronfard R and Boyer E. Free viewpoint action recognition using motion history volumes. *Comput Vis Image Und* 2006; 104(2): 249–257.
122. Mikel JA, Rodriguez D and Shah M. Action mach a spatio-temporal maximum average correlation height filter for action recognition. In: *IEEE conference on computer vision and pattern recognition, CVPR 2008*, Anchorage, AK, 23–28 June 2008, pp.1–8. New York: IEEE.
123. Liu J, Luo J and Shah M. Recognizing realistic actions from videos in the wild. In: *IEEE conference on computer vision and pattern recognition, CVPR 2009*, 2009, pp.1996–2003, [http://www.vision.eecs.ucf.edu/papers/cvpr2009/cvpr2009\\_liu1.pdf](http://www.vision.eecs.ucf.edu/papers/cvpr2009/cvpr2009_liu1.pdf)
124. Iosifidis A, Marami E, Tefas A, et al. The MOBISERV-AIIA eating and drinking multi-view database for vision-based assisted living. *J Inform Hiding Multimed Sig Process* 2015; 6(2): 254–273.
125. Kim H and Hilton A. Influence of colour and feature geometry on multi-modal 3D point clouds data registration. In: *2014 2nd international conference on 3D vision*, Tokyo, Japan, 8–11 December 2014, vol. 1, pp.202–209. New York: IEEE.
126. Cook D, Schmitter-Edgecombe M, Crandall A, et al. Collecting and disseminating smart home sensor data in the CASAS project. In: *Proceedings of the CHI workshop on developing shared home behavior datasets to advance HCI and ubiquitous computing research*, 2009, pp.1–7, <http://web.mit.edu/datasets/downloads/pp-cook.pdf>
127. Van Kasteren TLM, Englebienne G and Kröse B. Human activity recognition from wireless sensor network data: benchmark and software. In: Chen L, Nugent CD, Biswas J, et al. (eds) *Activity recognition in pervasive intelligent environments*. Paris: Atlantis Press, 2011, pp.165–186.
128. Young V and Mihailidis A. An automated, speech-based emergency response system for the older adult. *Geron-technology* 2010; 9(2): 261.
129. Rialle V, Ollivet C, Guigui C, et al. What do family caregivers of Alzheimer's disease patients desire in smart home technologies? Contrasted results of a wide survey. *Method Inform Med* 2008; 47(1): 63–69.
130. Shi G, Chan CS, Li WJ, et al. Mobile human airbag system for fall protection using MEMS sensors and embedded SVM classifier. *Sens J* 2009; 9(5): 495–503.
131. Rougier C, Meunier J, St-Arnaud A, et al. Robust video surveillance for fall detection based on human shape deformation. *IEEE T Circ Syst Vid* 2011; 21(5): 611–622.
132. Abbate S, Avvenuti M, Bonatesta F, et al. A smartphone-based fall detection system. *Pervasive Mobile Comput* 2012; 8(6): 883–899.
133. Madansingh S, Thrasher TA, Layne CS, et al. Smartphone based fall detection system. In: *2015 15th international conference on control, automation and systems (ICCAS)*, Busan, South Korea, 13–16 October 2015, pp.370–374. New York: IEEE.
134. Ziefle M and Wilkowska W. Technology acceptability for medical assistance. In: *2010 4th international conference on pervasive computing technologies for healthcare*, Munich, 22–25 March 2010, pp.1–9. New York: IEEE.