Sensors, vision and networks: From video surveillance to activity recognition and health monitoring

Andrea Prati [a,*], Caifeng Shan [b] and Kevin I-Kai Wang [c]

[a] *Department of Engineering and Architecture, University of Parma, Italy*
E-mail: andrea.prati@unipr.it
[b] *Philips Research, Eindhoven, The Netherlands*
E-mail: caifeng.shan@philips.com
[c] *Faculty of Engineering, The University of Auckland, New Zealand*
E-mail: kevin.wang@auckland.ac.nz

Abstract. This paper presents an overview of the state of the art of three different fields with the shared characteristics of making use of a network of sensors, with the possible application of computer vision, signal processing, and machine learning algorithms. Namely, the paper first reports the state of the art and possible future directions for Intelligent Video Surveillance (IVS) applications, by recaping the history of the field in terms of hardware and algorithmic progresses. Then, the existing technologies of Wireless Sensor Networks (WSNs) are compared and described. Their applications to human activity recognition (HAR), both from a single or multiple sensors perspectives, are described and classified, followed by the current research trends and challenges. Finally, recent advances on camera-based health monitoring (including vision-based Ambient Assisted Living and patient monitoring, and camera-based physiological measurements) are described in full details, with the challenges faced.

Keywords: Intelligent video surveillance, wireless sensor networks, human activity recognition, camera-based health monitoring

1. Brief introduction

The recent advances in sensor development, embedded systems, wireless networks, and computer vision have brought to an increasing interest in corresponding applications of such enabling technologies in different fields of Ambient Intelligence. Among them, this paper will present the main components, current state of the art and future directions in three distinct (but intersecting) fields. In particular, Section 2 will present an overview of the history of *Intelligent Video Surveillance* (IVS) applications along two main directions: on the one hand, the progresses made in the last years in the hardware equipments used for IVS, from a single fixed camera, up to egocentric vision; on the other hand, and mostly in parallel, the progresses made in algorithm development and applications, from background suppression to high-level scene understanding. Section 3 will introduce the main wireless communication technologies at the base of *Wireless Sensor Networks* (WSN), while Section 4 will present research works using both single sensor and multiple sensors applications for *Human Activity Recognition* (HAR) over the last two decades, followed by the challenges ahead. Finally, Section 5 will describe recent advances on camera-based *health monitoring*, where vision-based Ambient Assisted Living, vision-based patient monitoring, and camera-based physiological measurements are reviewed in details.

---

*Corresponding author. E-mail: andrea.prati@unipr.it.

2. Intelligent video surveillance applications

Among the different fields where networks of sensors and cameras are applicable, intelligent video surveillance (IVS, hereinafter) has played an important role in the last decades. Sadly, one of the main driving reasons that contributed to its success is the increase of interest (and funding) that followed major worldwide security accidents, such as the crime diffusion and terrorist attacks. The increase of funding (both at European and international level) has brought many research groups to focus their activities to the development of systems and techniques for making automated video surveillance more intelligent and advanced.

It is hard to identify the exact year when IVS was born. First papers on this topic appeared around 1996–1998 [17,22,49,134]. However, one crucial milestone is represented by the publication of the special section on "Video Surveillance" on IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE T-PAMI) journal in August 2000 [36]. At that time, it was argued that existing video surveillance systems were used mainly for offline forensic analysis ("after the fact") and that there was an urgent demand for real-time, online automated analysis of video streams for alerting security officers to a burglary in progress or to suspicious behaviours, while there is still time to prevent the crime [36].

After this initial special section, there has been a large number of conferences (such as, IEEE AVSS – International Conference on Advanced Video and Signal-based Surveillance), workshops (such as, ACM VSSN – Video Surveillance and Sensor Networks, or IEEE PETS – International Workshop on Performance Evaluation of Tracking and Surveillance), journal special issues and papers, which addressed, under different perspectives, the many challenges that IVS had (and partially still has). A complete coverage of the state of the art in this field goes beyond the scope of this paper, but we aim at grouping the advances in the last 15+ years along two main directions.

The first direction (detailed in Section 2.1) can be named as *"hardware direction"*, not referring to the advances on hardware development (new cameras, powerful PCs/devices, etc.), but more on the different usage of hardware (mainly acquisition devices) which has driven new applications and, as a consequence, new algorithms. The second direction (detailed in Section 2.2) will be called *"algorithm direction"* and will include the progresses in terms of algorithms to address new applications (also coming from the usage of

new acquisition devices), with a constant increase on the level of complexity along the years.

2.1. History of hardware progresses in IVS

As said above, this section will briefly browse the temporal changes along the last years in terms of hardware progresses in IVS systems.

At the beginning, also due to cost and availability issues, most of the works in IVS tackled the *single fixed camera* scenario. The video feed provided by a single fixed camera was processed to automatically analyze the scene [19,22,37,55,118,134]. However, the single fixed viewpoint provided by a single camera has some limitations, in terms of both the amount of area covered and the robustness to occluding objects. The former limitation does not allow the system to monitor large areas (usual in security applications) and prevents the detection and tracking of objects when they move away from the field of view of the camera. The latter limitation, instead, makes the system unreliable when the targets are (partially) occluded by other moving/non-moving objects.

In order to solve these limitations, around 2002 some works on IVS started looking at different types of cameras. For instance, *PTZ* (Pan-Tilt-Zoom) or active cameras have been used to dynamically change the viewpoint of the single fixed camera in order to get rid of occlusions and to expand the monitored area [23]. One limitation of PTZ cameras is that a single field of view is available at a given time, and proper scheduling algorithms need to be designed to have a proper (and continuous) observations of the targets [39]. One possible alternative consists in using special cameras, such as the *omnidirectional cameras* (see, for instance, [87]) which provide a 360-degrees simultaneous view of the scene thanks to an hemispherical mirror on the top of the camera lens.

Despite the advantages of using cameras with variable (PTZ) or large (omnidirectional) fields of view, they still have limited applicability on large-area surveillance and are still prone to the occlusion problem, due to their fixed optical center. Therefore, from the very beginning, there has been an ever-increasing interest in using *multiple cameras* (either fixed or active, or both), composing, as a matter of fact, a network of (hybrid) sensors. Although some earlier works (such as [71,78,117]) already considered the use of multiple cameras, the milestone for a more-or-less concrete system with multiple cameras can be considered the work of Khan and Shah in 2003 [68]. This work can be con-

sidered among the first ones to address the problem of fusing multiple camera tracking outputs to construct a multi-camera (multi-object) tracking system. In fact, it proposed an algorithm to the *consistent labeling* problem, i.e. assigning the same (consistent) label (ID) to an object detected in different cameras. This labeling is the initial step to obtain a long-term multi-camera tracking of objects within a large area, followed by analysis of the object trajectories and thus of its behavior in the scene. Several approaches have been proposed for multiple camera tracking, classified depending on whether the fields of view of the cameras are partially overlapped or completely disjoint. In the former case, geometric as well as appearance-based features can be used to solve the data association task between the detections in two views, while in the latter only pure appearance-based features can be used [25], together with some reasoning about the travelling time between adjacent cameras.

There is still a very active research community working in this field, with an increasing emphasis on the size of the network (i.e., the number of cameras), the variety of sensors (hybrid systems), the complexity of the scene (especially, with non-overlapping fields of view), and the final application (see next section with regards to people re-identification topic). Interested readers may refer to the 2009 book [2] or to the 2013 review reported in [131].

While the research on multi-camera IVS was progressing, another active area of research has come up around 2004 [20]. In fact, thanks to the advances in electronics, the so-called *smart cameras* have started to be affordable at the beginning of 21th century. Smart cameras are not only acquisition devices, but also include processing, storage and communication modules. These modules allow the IVS system to preprocess the acquired video feeds on board and transmit only relevant information to the server(s) through the communication network. Besides the obvious relief of the communication and processing load, smart cameras are also more privacy respectful, by retaining and transmitting only data which do not violate the privacy laws.

These two last research areas (multi-camera systems and smart cameras) have been often fused together by building hybrid IVS systems that exploit the advantages of smart cameras within a distributed networks of sensors (cameras) [57].

One further step in the hardware progresses for IVS relates to the general concept of *mobile vision*. This term was first introduced in [59] and refers to the development of computer vision and image processing techniques for camera-equipped consumer devices, such as the latest generation smartphones and tablets. Mobile vision shares and builds upon prior work on active vision, which deals with computer vision algorithms for processing imagery captured by non-stationary cameras. Similar to cameras mounted on a robot or a vehicle, camera-equipped consumer devices that are carried by humans are non-stationary. Furthermore, it is oftentimes more difficult to model the motion of a camera embedded in a hand-held consumer device than that of a camera mounted on a robot. Mobile vision also poses a whole set of unique challenges stemming from the fact that these devices have limited computational, memory and bandwidth resources [103]. There are several reasons why mobile vision is an emerging field of research. First of all, the latest generation of consumer devices – smart phones and tablets – have the computational resources needed to support "sophisticated" computer vision algorithms. Many of these devices also boast multiple sensors, including a number of non-imaging sensors, such as compass, accelerometer, and gyroscope. Readings from these sensors can be incorporated into the image processing pipeline to support a variety of applications, e.g., augmented reality, geo-localization, etc. Furthermore, advances in communication technologies have resulted in a new generation of high-speed, low-latency wireless communication protocols. This suggests that it is now possible to leverage cloud-based computational resources and lower the onboard computational burden on consumer devices. Even more importantly, the last 30+ years of research on computer vision algorithms has produced highly-accurate and efficient algorithms that can now be deployed on consumer devices. All these factors have contributed to the development of a multitude of mobile vision projects and products, such as visual object (landmarks, logos, goods) recognition, augmented reality, gesture recognition, road sign recognition, lane detection, automatic cruise control, collision avoidance, etc. [103].

One last advance in terms of hardware used in IVS has been introduced around 2012. In fact, the advances in consumer electronics has made available in the market the so-called *egocentric cameras* [66]. At the end, these are normal wearable cameras that provide a "first-person" perspective of the scene. A part from the technical challenges (common to the above-mentioned ones for mobile or robot vision), egocentric vision has also opened a brand new set of applications related, among the others, to IVS: face recognition,

(b) Timeline of hardware progresses in IVS.

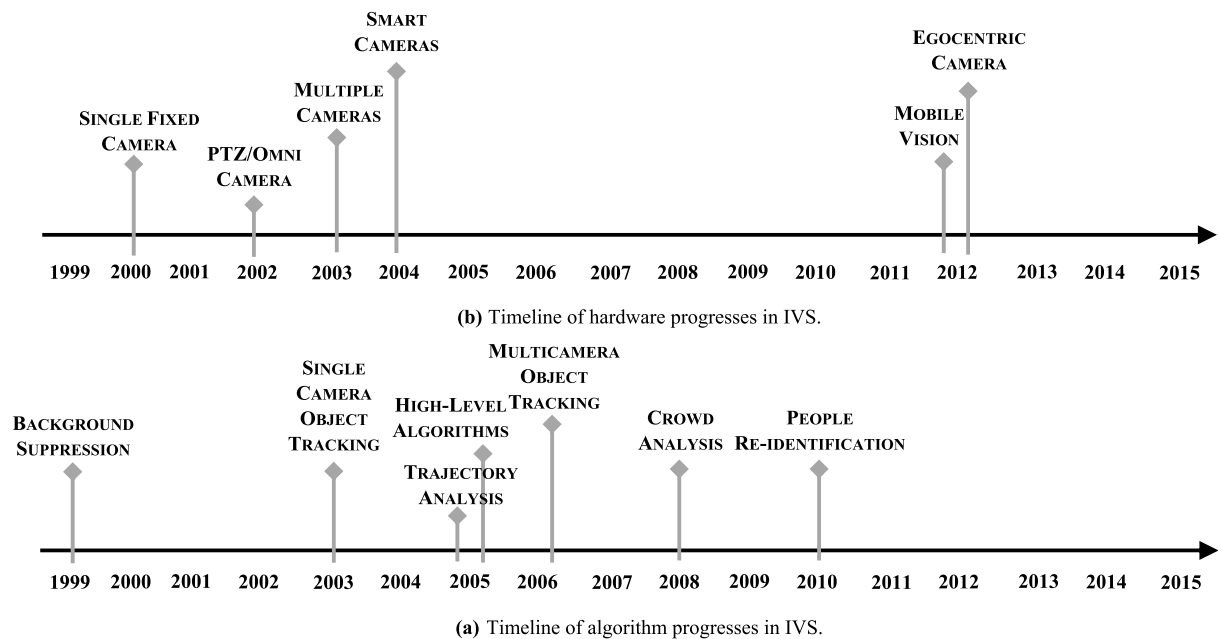**(a)** Timeline of algorithm progresses in IVS.

Fig. 1. (a) Approximated timeline of hardware progresses for IVS. (b) Approximated timeline of algorithm progresses for IVS.

just to mention the most common, has been brought to a new level thanks to egocentric vision. Other applications are related to object recognition, as well as to personal storytelling from visual lifelogging [18].

Figure 1(a) summarizes the approximated timeline of the major steps in hardware progresses for IVS. It is quite evident that there has been a gap between 2004 and 2012 in the hardware progresses: this can be explained by, on the one hand, the focus of the scientific community on the algorithms (see next section) and, on the other hand, by the time needed to develop affordable and powerful mobile devices and egocentric cameras.

2.2. History of algorithm progresses in IVS

This section will briefly outline the progresses in the IVS field in terms of algorithms. As a premise, it is quite straightforward to notice that these progresses are tightly related to those in the hardware used (described in the previous section).

As in the case of other fields of computer vision (and computer science in general), the algorithm progresses have followed a low-to-high-level path. As a consequence, the first set of papers had "object/target detection" as primary goal. Since, at the beginning, single fixed cameras were used, the first efforts can be classified as the development of effective (and efficient)

background suppression techniques. Background suppression refers to an algorithm (or set of algorithms) aiming at separating (at pixel- or object-level) the stationary parts of the scene (i.e., the background) from those which are moving (i.e., the foreground). A seminal work on this topic is dated back in 1999 [124]. Despite its age, this topic is still heavily researched in the literature, especially considering more complex scenes with slowly-changing backgrounds, distractors (e.g., waving trees, which are moving but should be classified as background) and moving cameras (where moving parts should be distinguished between actual background and moving objects). Some additional references (among the numerous existing in the literature) on this topic are [34,42,45,98,116].

Once the (moving) objects in the scene have been detected, the next step typically involves tracking objects, i.e., consistently associating the same identifier to an object, along time. There have been several papers addressing the *object tracking* problem. Although some earlier works are dated back to 1999 [24], a milestone for this topic can be set to 2003 when the work by Comaniciu *et al.* [38] has been published. Object tracking algorithms are often classified based on how objects are represented and on the data association phase [137]. Point-based trackers employ either a deterministic or probabilistic (based on Kalman filter or JPDAF – Joint Probability Data Association Filter) approach to

establish correspondence between keypoints on the objects in successive frames. Kernel-based trackers, instead, use some more-or-less complex kernel and a classifier for data association. Finally, some trackers exploit the information about the silhouette of the objects to find correspondences. Interested readers may also refer to the experimental survey reported in [115].

As a natural extension, once the objects are tracked in each single camera of a network of multiple cameras, there has been a significant number of papers addressing the *multicamera object tracking* problem. The extension of object tracking to multiple cameras requires to take into account, when deciding the object representation and the data association strategy, the differences in viewpoint, resolution, illumination conditions and colors of the different cameras. Also in this case, it is rather subjective to set a starting date for this topic, but with the work of 2006 by Kim and Davis [69] (and others that appeared in that period of time), the proposed solutions can be considered mature enough. By considering multiple cameras, also issues related to their coordination/cooperation, as well as to the communication between them, should be taken into consideration [121]. Finally, this is still a very active area of research [108,122].

Most of the existing multi-object (multi-camera) tracking algorithms are capable of tracking objects in a low-medium-density situation. Therefore, since approximately 2008 [139], several researchers have studied specific solutions to the *crowd analysis* problem. This refers to situations where the density of people/targets is high or very high. The proposed solutions are mainly based on flow dynamics approaches [5] or on social force models [86].

Tracking objects on a large surveilled area is per se useful, but it is typically followed by algorithms that exploit the temporal information about the objects (appearance, position, speed, etc.) for higher-level tasks. Among them, in the recent past, object trajectories (as sequence of positions in time) have been analyzed as a mean for extracting information about the object behavior. *Trajectory analysis* has been a flourishing area of research, starting from the seminal work of 2005 [81]. The approaches to trajectory analysis are based on a specific representation of the trajectories followed by a clustering algorithm. Each obtained cluster of trajectories represents a common behavior, whereas small clusters identify anomalies (and, thus, potentially suspicious behaviors) [26]. Interested readers can refer to the survey in [88].

There are also several research topics under the IVS umbrella aiming at exploiting low- and mid-level algorithms (object detection and tracking, trajectory analysis, etc.) to design *high-level algorithms*, such as posture analysis [41,89], fall detection [91], detection of abandoned packages [16], action/activity recognition [30,101] and others. At even higher level, some papers studied how to model and classify interactions both between targets (mainly people) and with the scene, for instance for assisted living applications [29]. In particular, group detection and recognition of social interactions have been hot topics in recent years [8,40]. The main approach for these tasks is based on understanding the spatio-temporal personal space of the targets and using social force models (SFMs) to understand the interactions among them.

Finally, a special attention must be given to *person re-identification* topic, since it is still a very hot topic in the scientific community [47,129,142]. On the one hand, this can be treated as a special case of multi-camera tracking (i.e., the goal is to assign a consistent label to a person when it disappears in one camera and re-appears in another), but, on the other hand, it is also a borderline topic with the general image retrieval topic, and most of the techniques used are borrowed by that topic.

Figure 1(b) summarizes the approximated timeline of the major steps in algorithm progresses for IVS, and it can be noticed that it well integrates with the progresses in hardware.

2.3. Future trends in IVS

Most of the progresses listed in the previous sections are still more-or-less active topics in IVS. However, generally speaking and without claim to impose or address the future directions of research in IVS, there are some topics which might be considered pivotal for IVS in the near future:

– *Exploitation of deep learning techniques*: with the recent diffusion of deep learning, several applications of IVS are going to be rethought, by replacing traditional approaches with some variants of Convolutionary Neural Networks (CNNs). Some recent examples are person re-identification [35] or action recognition [127].

– *Large-scale/city-scale IVS*: the availability of cameras installed in our cities nowadays makes reasonable to have access in the upcoming years to thousands of (hybrid) cameras, including drones/

UAVs [125]. This will require not only special computational and storage capabilities, but also specific and efficient techniques for handling different problems, such as communication, camera coordination, failure resilience, etc.

– *Higher and higher-level algorithms*: it is foreseeable that, thanks to the ever-increasing computational power and the availability of sophisticated tools (e.g., deep learning), the researchers in IVS will develop more than ever algorithms of high level for complex scene understanding. One interesting direction will be to move towards *prediction* more than *recognition* of events.

3. Wireless sensor networks

Different to the video surveillance systems, another important technology underpinning the modern ubiquitous computing, ambient intelligence (AmI), and the Internet of Things (IoT) is the wireless sensor networks (WSNs). This type of networks typically has the features of: 1) limited power source, 2) limited computational power, 3) limited communication bandwidth, and 4) preferably miniaturized physical size so that it is physically unobtrusive. These limitations are the main obstacles of WSNs in practical applications and have been the main research focuses over the last two decades.

The current state-of-the-art WSNs technology can be separated into five categories as shown in Table 1. The first category is the traditional WSNs that feature low power consumption at mW level during transmission, low data rate at 256 kbps, and short range between 20–100 meters. Popular technologies such as Zigbee [15] and 6LoWPAN [90], which are typically based on the IEEE 802.14.5 standard [54], fall into this category. There are also other proprietary networks such as Z-Wave [138] which is promoted by a group of industry stakeholders (i.e. the Z-Wave Alliance), targeting specific niche applications such as home automation. Over the last decade, this type of WSNs had

been widely deployed in various consumer electronics for ambient and wearable devices in various applications, such as home automation [52], agriculture [10] and healthcare wearable devices [9]. This type of WSNs is very often referred to as the Body Area Network (BAN).

Different to the 1st category, the 2nd category WSN technologies such as WiFi and Bluetooth, support higher data rate that can be used to exchange more sophisticated data, such as audio and video streams. As a compromise, this type of WSNs consumes more power than the previous type. Therefore, except a few of the early researches, technologies such as WiFi are typically used as a backhaul link for category 1 WSNs to provide connectivity to the Internet with permanent power support. However, it is worth mentioning that, different to Wi-Fi, the state-of-the-art Bluetooth is capable of providing relatively high data rate (up to 4 Mbps) while also being as power efficient as Category 1 type WSNs.

While the previous two types of WSNs can already cover a majority of real-world applications, coverage and scalability can become big issues with certain applications such as climate monitoring [79] and smart metering [102] that require very large geographical coverage. Although it is possible to achieve a wide coverage using the previous two types of WSNs, the required deployment cost, and sophisticated network topology and management, will make it infeasible in actual practice. Therefore, a new category of WSNs has emerged in recent years, called Low Power Wide Area Network (LPWAN), that targets specifically long distance and low data rate monitoring applications, such as smart metering. Technologies such as LoRAWAN [1,133] and Sigfox [74] fall in this category and are rapidly penetrating the market.

While the low power and long range communication feature makes LoRAWAN and Sigfox very popular, their highly-constrained bandwidth deemed them impractical for many smart cities applications, such as smart surveillance and intelligent transportation. To mitigate the bandwidth issue, a slightly different type

Table 1
Comparison of different categories of wireless sensor networks

Category	Typical technology	Data rate	Transmission range	Power consumption
1	Zigbee, Z-Wave, 6LoWPAN (IEEE 802.15.4)	256 kbps	20 to 100 m	Low
2	WiFi, Bluetooth	up to 600 Mbps	50–400 m	High
3	LoRAWAN, Sigfox	100 bps–50 kbps	2–20 km	Low
4	NB-IoT, LTE Cat M.	200 kbps–1 Mbps	1–10 km	Medium
5	3G/4G	50 Mbps or more	A few to up to 40 km range	High

of LPWANs exists that offers higher data rate at a compromise of being more power hungry. The state-of-the-art NB-IoT [106,114] and LTE Cat. M [97] fall into this category.

The last category, strictly speaking, may not be considered as WSNs, but, rather, as the backhaul link (e.g. cellular networks) that interconnects specific WSNs with the Internet and exchange data/commands with each other. However, such network technologies have been applied in different types of wireless sensing applications.

4. Wearable sensors for human activities recognition (HAR)

Central to the modern concepts of IoT and WSNs is the wireless sensor nodes (which will be referred to as nodes in the rest of text). The ease of deployment and unobtrusiveness nature of nodes make them the ideal candidates for many real-world applications from environmental monitoring, smart cities, healthcare, industrial automation, to military [3,4]. A typical node consists of three main subsystems, that is the sensing, computing and wireless communication subsystems. Depending on the application requirements, different sensors, wireless transceivers, computing platform, and necessary power source will be selected. In this paper, we will focus on the use of ambient and wearable wireless sensor nodes for context-aware systems, or more specifically human activity recognition (HAR) systems, which are crucial for many applications such as healthcare and military [73,96].

4.1. Single accelerometer approach

Understanding the physical context such as body position, posture, gait, vital signs and physical activities of a person can offer great amount of information for context-aware applications such as smart home, elderly care, rehabilitation, and sports science. In this paper, we focus on the identification of the so-called Activities of Daily Living (ADL). Enabling or assisting elderlies and disabled patients to be able to perform ADL is critical in maintaining their quality of life. Majority of the ADL can be classified into three categories namely static, dynamic, and transitional activities. Static activities consist of activities such as standing, sitting, and lying, whereas dynamic activities consist of activities such as walking, running, jumping. In addition, there are also transitional activities that sit in between static and dynamic activities, such as sit-to-stand and lie-to-sit, which are of critical importance in inferring the correct physical context.

Naturally, most of these ADL are determined by human locomotion and therefore inertial sensors, or more specifically the accelerometer, are the most popular choice. Table 2 covers some of the prominent research works over the last two decades that use only a single accelerometer for recognizing human ADL.

Table 2

Human activity recognition (HAR) systems that use a single accelerometer

Sensor locations	Activity	Techniques	Accuracy	Year
3D Acc (waist)	Walk continuously along a corridor, then up a stairway and along another corridor, and then down a stairway	Threshold-based	98.8	2000 [112]
3D Acc (back, on the vertebrate)	Standing, sitting down with lowering subjects head, sitting down and leaning against, lying down straight, lying upside down, walking, going up/down stairs, running	Threshold-based	95.1	2003 [76]
3D Acc (worn near the pelvic region)	Standing, walking, running, going up/down stairs, sit-ups, vacuuming, brushing teeth	Naïve Bayes, k-NN, SVM, Binary decision	46.3–99.3	2005 [107]
3D Acc (waist)	Sitting, standing, lying, sit-to-stand, stand-to-sit, lie-to-stand, stand-to-lie, walking	Gaussian Mixture Model (GMM)	91.3	2006 [6]
3D Acc (hip)	Sit-to-stand, stand-to-sit, lying, lying-to-sit, sit-to-lying, walking (slow, normal, fast), fall (active, inactive, chair), circuit	Binary decision	90.8	2006 [67]
3D Acc in the trouser pocket (not fixed)	Jumping, still, walking, running	Wavelet Autoregressive Model	95.45	2010 [56]
3D Acc (right waist)	Walking, standing, sitting, lying face-up/down, stand-to-sit, sit-to-stand, sit-to-lie, lie-to-sit, falling	Threshold-based	96.5	2017 [94]
2D Acc	Walking, running, sitting, walking up/down stairs, standing	ANN	95	2000 [105]

The main advantage of single accelerometer systems is their simplicity and unobtrusive nature that cause less impediment to users' daily activities. Since there is only one accelerometer, sensor placement is important in achieving a more accurate recognition. While different sensor locations may lead to varying recognition accuracies of different activities, it can be observed that the most popular sensor placement location is around the pelvic region. The locomotion of the pelvic region is relatively steady, while still capable of providing discriminating features to separate various ADL.

HAR systems are typically considered as a classification problem where various types of feature selection, feature extraction, and machine learning based classification approaches are developed. While most of the developed techniques are able to achieve reasonably high accuracies, such results must be interpreted with care. Due to the computation power and energy constraints, most of the existing works are still considering the deployed WSNs as data acquisition tools and analyses are very often offline with emulated activities. Therefore, the real-time performance with real life data can be a key challenge for many existing works.

4.2. Multiple sensor fusion approach

While single-accelerometer-based systems have been proved to be very effective and less obtrusive to users' daily lives, multiple sensor fusion approaches are still heavily investigated for various reasons (see Table 3 for some relevant examples). In the early days, a common alternative approach is to use multiple one-dimensional or two-dimensional accelerometers for HAR, as opposed to a single triaxial accelerometer.

Table 3

Human activity recognition (HAR) systems that use multiple sensors fusion

Sensor locations	Activity	Techniques	Accuracy	Year
2x 1D Acc (waist belt and left thigh)	Sitting, standing, lying, locomotion	–	87.7	1998 [126]
4x 1D Acc (sternum, wrist, thigh, and lower leg)	Sitting, standing, lying supine, sitting and talking, sitting and operating PC keyboard, walking, stairs up/down, cycling	Template matching	95.8	1999 [48]
2x 3D Acc (left and right hip)	Start/Stop walking, level walking, up/down stairs	ANN	83–90	2001 [83]
1x 2D Acc, 1x digital compass sensor, 1x angular velocity sensor	Standing, 2 steps north, 40 steps east, 3 steps south, and then 6 steps west	Threshold-based	92.9–95.9	2002 [75]
1x 2D Acc + Gyro, front chest	Sitting, lying, standing, walking, sit-to-stand, stand-to-sit	Threshold-based	>90	2003 [92]
2x 3D Acc (hip and wrist) + GPS	lying down, sitting and standing, walking, running, cycling, rowing, playing football, Nordic walking, and cycling with a regular bike	Decision tree	89	2008 [46]
5x 2D Acc (hip, wrist, arm, ankle, and thigh)	Lying, cycling, climbing, walking, running, sitting, standing	HMM	92.2–98.5	2010 [82]
4x 3D Acc (left ankle, right ankle, sternum)	Fall detection	Linear Discriminant Analysis (LDA)	89	2013 [12]
3D Acc at the waist & ECG on the front chest	Lying, sitting, standing, walking, walking up/down stairs, running	Linear Discriminant Analysis (LDA) & Relevance Vector Machines (RVM)	99.57	2013 [62]
3D Acc, 3D Gyro, hold in the left hand	Stationary, walking, running, going up/down stairs, falling down	Extreme Learning Machine (ELM)	66.9–78.1	2014 [58]
3D Acc, 3D Gyro (right thigh)	Running, standing, sitting, walking, up/down stairs	Nonlinear Kernel Discriminant Analysis (KDA) & Extreme Learning Machine (ELM)	99.81	2015 [135]
1x 3D Acc at knee & 2x Force Sensitive Resistors (FSRs) insole	Walking gait detection	Threshold-based & Finite State Machine Modelling	98.9	2018 [113]

The construction of such systems are usually more difficult even with wireless communication capability.

Another typical fusion is to integrate accelerometer and gyroscope into the so-called *Inertial Measurement Unit (IMU)*. Gyroscope provides more accurate detection on body orientation and posture which complements well with the accelerometer. However, as it can be observed in Table 3, fusing multiple sensor data does not necessary offer better accuracy. How to better fuse heterogeneous sensor data to achieve more accurate classification remains an important research topic.

Last but not least, accelerometers or IMUs are also commonly fused with other ambient sensors (e.g. RFID, GPS) to better infer more complex human activities/behaviours, or vital sign monitoring sensors (e.g. ECG, heart rate) to detect specific types of activities such as fall detection and gait analysis.

4.3. Fine-grained behaviour recognition

As mentioned previously, many different machine learning techniques had been applied for classifying simple ADL. It is worth noting that most of those approaches for classifying human physical activities are data driven (or learning-based). Data driven approaches are preferred for their abilities of handling noise and uncertainties, which are of critical importance when dealing with raw sensor data. On the other hand, data driven approaches also often require large amount of training/testing data in order to build a reasonably accurate classification model. The necessary preparation work such as collecting and labelling data can be very time consuming, which limits the applicability of these approaches [32,136]. In addition, recognition of simple physical activities may not provide sufficient context information to differentiate complex behaviours. For example, knowing the physical activity of standing may not be sufficient to differentiate more complex behaviours, such as cooking or brushing teeth.

In order to determine more complex human activities, the so-called knowledge driven (or specification-based) approaches [93] is more popular as opposed to the data driven approaches. These types of approaches offer the ability to construct high level semantic model and support reasoning/inferencing using the model with multimodal input sensor data. In addition, knowledge-based approaches offer formal structures that can provide semantic meanings to sensor data that enable the construction of high-level semantic activity model to recognize more complex human activities/behaviours.

Among various knowledge-based approaches, ontology-based approaches are ones of the most popular, which are capable of providing a formal specification representing domain knowledge using a set of axioms and constraints [33,51,95,109]. The main drawback of ontology-based approaches is the computationally intensive reasoning process and its lack of support for temporal relations. Further, traditional ontology-based approaches are not able to handle missing sensor data. Some recent research works have been focusing on associating probability and confidence level with ontology to support certain level of uncertainties in sensor data.

4.4. Challenges of WSN-based context-aware systems

As the WSNs technologies are becoming more mature in satisfying various communication requirements, the challenges of WSN-based context-aware systems are shifting towards addressing the utility of the system.

4.4.1. Interoperability

With the increasing number of embedded ambient devices and wearable devices, the issue of interoperability is becoming more prominent. As indicated previously, fusing multimodal sensor data are generally beneficial for finding different varieties of context. Fusion will not be possible if individual networks, or even, individual devices are running in isolation. The issue of interoperability must be addressed to support a wider uptake and realization of intelligent devices and context-aware systems. Open standards and open architectures may offer possible solutions but remain a challenge.

4.4.2. Scalability

In addition to interoperability, as the number of ambient and wearable devices are increasing, interactions between users and devices must also change to support a scalable deployment. Specifically, how can users interpret the data, and further, act accordingly to make intelligent decisions.

4.4.3. Reliability

To further improve the utility of context-aware systems, the issue of reliability must be addressed. With the envisaged trend of IoT and WSN, billions of ambient and wearable devices will be deployed in the next decade or two. However, with these infrastructures and

potential redundancy in data collection, identifying the correctness and quality of data remains a challenging research topic.

4.4.4. Power consumption

Obviously, in order for ambient and wearable devices to be more useful, continuous operation is the key for ultimate utility. Power saving approaches at different layers (i.e. hardware and communication layers), different harvesting sources, advanced battery technologies are still ongoing research challenges that require more research efforts.

5. Camera-based health monitoring

As a further application of network of cameras with computer vision algorithms, this section will present recent advances on camera-based health monitoring. Compared to video surveillance where video analytics is used for security and safety, in this field camera sensing and monitoring are used for health and care of people. Camera-based methods offer an unobtrusive solution for the monitoring and diagnosis of subjects. With the low-cost cameras being pervasive in our lives, this field has received more and more attention, considering people are taking a more proactive approach to their health and wellness and the older population continues to grow worldwide.

We will particularly focus on three key topics: vision-based Ambient Assisted Living (AAL) at home, vision-based patient monitoring in clinical environments, and camera-based physiological measurement. In the end the challenges and future trends are briefly discussed.

5.1. Vision-based Ambient Assisted Living (AAL)

Ambient Assisted Living (AAL) is to improve the quality of life and maintain independence of older and vulnerable people through the use of technology [27]. State-of-the-art AAL technologies are typically based on solutions employing sensors either embedded in the environment or body worn. The systems are usually costly to maintain, and relatively obtrusive (e.g. sensors set up on every cupboard door) and are highly sensitive to the performance of the sensors. The body worn sensors are quite effective when the user actually wears them, however, it is recognized that user compliance with wearable systems is poor. More recently, the research community and industry have shown a grow-

ing interest in video and computer vision based solutions for AAL applications [14,21,43,44,80].

One of the main aims of AAL systems is to provide personal safety to older people. Falling is a major risk faced by older people with one-third of people over 65 falling each year. Therefore the vast majority of vision-based AAL research focuses on fall detection [110]. Zhang *et al.* [141] present a comprehensive survey of recent proposed fall detection methods, where they divide the existing vision-based fall detection into three categories: fall detection using a single RGB camera, 3D-based methods using multiple cameras, and 3D-based methods using depth cameras. Fall detections using a single RGB camera have been extensively studied, as the systems are easy to set up and are inexpensive. Shape related features, inactivity detection and human motion analysis are the most commonly used clues for detecting falls. The disadvantage of using a single RGB camera is that most of them lack flexibility. Fall detectors using a single RGB camera are often case specific and viewpoint-dependent. Moving a camera to a different viewpoint (especially a different height from the floor) would require the collection of new training data for that specific viewpoint. Single RGB camera based methods also suffer from occlusion problems where part of the human body is occluded by a certain object, like a bed. Another category is 3D-based methods using multiple RGB cameras. The calibrated multi-camera systems allow 3D re-construction of the object but require a careful and time-consuming calibration process. Depth camera based systems [44] share the same advantages with the calibrated multi-camera systems and at the same time do not need a time-consuming calibration step. Depth cameras enable researchers to model fall events under a camera-independent world coordinate system and reconstruct the 3D information of the object.

In a recent paper [80], using a dataset collected from an assisted-living facility for seniors, Luo *et al.* presented a computer vision-based approach that leverages non-intrusive, privacy-compliant multimodal sensors and state-of-the-art computer vision techniques to continuously detect seniors' activities and provide the corresponding long-term descriptive analytics. This analytics includes both qualitative and quantitative descriptions of senior daily activity patterns, which can be further interpreted by caregivers. They developed multimodal, privacy-preserving, and easy-to-install sensors and deployed them in a senior home facility. Particularly they utilized the depth imaging

and thermal imaging to detect and monitor senior daily activities. Both modalities are privacy-compliant and thus prevent identification of the person in the images. They recorded long-term continuous video data and annotated the data, and showed that, using state-of-the-art Convolutional Neural Networks-based activity classification and detection models, their approach can classify the clinically-relevant activities in real-time as well as to perform temporal activity detection in the long term. Six types of activities: sitting, standing, walking, sleeping, using bedside commode, getting assistance were considered in this work.

5.2. Vision-based patient monitoring

Patient monitoring system is one of the most important devices used in hospitals, which allows the caregivers to regularly observe patients, mainly their vital signs and trends. It is believed that camera-based monitoring can provide more useful information (beyond vital signs) to the caregivers. For example, in clinical environments, abnormal activities in wards such as unbalanced walking, bed climbing, and irregular body posing can lead to serious consequences, for instance falling, pains or injuries. These information cannot be monitored with the existing patient monitor, but can be well detected in the videos. In [31], cameras were installed at intensive care units for observing agitation in the sedated patients. Kittipanya-Ngam *et al.* [70] studied the use of intelligent video analytic system in monitoring patients in hospital wards. Cameras are utilized for monitoring the activities of the patients and performing better assessments. The proposed system can provide alarming features to draw attentions from the caregivers for detecting falls, risky movements, and irregular actions of the patients. Furthermore, information of the monitored activities is recorded for further studies. They installed cameras in the patient wards with two different views, Bed View and Room View. The Bed View is to capture the activities of the patient on the bed, while for Room View, the purpose is to monitor the activities happen in the common area. The camera of Bed View can help medical staffs to have a better understanding in patient conditions by monitoring the patient 24/7 and providing other information such as bed occupancy and extubation. Apart from activities around their own beds, patients also have activities on the ward corridors. The camera of Room View will provide the data such as accidents and activities inside or outside the wards.

Falls occurring in hospital settings can cause severe emotional and physical injury to patients as well as increase healthcare costs for the hospitals and the patients. Banerjee *et al.* [13] presented an approach for patient activity recognition in hospital rooms using depth data collected using a Kinect sensor. Depth sensors such as the Kinect ensure that activity segmentation is possible during day time as well as night while addressing the privacy concerns of patients. With the depth sensor, their approach detects the presence of a patient in the bed as a means to reduce false alarms from an existing fall detection algorithm. Their system can reduce false alerts of fall detection such as pillows falling off the bed or equipment movement.

Monitoring neonatal movements may be exploited to detect symptoms of specific disorders. In fact, some severe diseases are characterized by the presence or absence of rhythmic movements of one or multiple body parts. Clonic seizures and apneas can be identified by sudden periodic movements of specific body parts or by the absence of periodic breathing movements, respectively. In [28], motion signals from multiple digital cameras or depth-sensor devices (e.g., Kinect) are extracted and properly processed to estimate the periodicity of pathological movements using the Maximum Likelihood (ML) criterion. Their non-invasive system can help to monitoring 24/7 all the newborns present in a NICU, or even at home. Infants are particularly vulnerable to the effects of pain and discomfort, which can lead to abnormal brain development, yielding long-term adverse neurodevelopmental outcomes. In [119], Sun *et al.* proposed a video-based method for automated detection of the infants' discomfort.

5.3. Camera-based physiological measurement

The health condition of a subject is normally assessed by measuring certain physiological parameters of the body, such as body temperature, blood pressure, heart rate (HR), respiration rate (RR), oxygen saturation (SpO2), and so on. These physiological measurements traditionally are performed by attaching sensors in contact with the body (skin), for example, the widely used pulse oximeter. In recent years, camera-based contactless physiological measurement has received much attention because of its easy-to-use and unobtrusiveness.

A number of new methods have been presented [85], where the greatest amount of research has focused on remote photopletysmography (PPG). Photopletysmography is the measurement of light reflected from,

or transmitted through, the body, which captures the volumetric flow of blood within the vessels [7,63]. The standard PPG measures the blood volume changes in the microvascular tissue bed beneath the skin in a "contact" manner. Non-contact PPG remotely measures the PPG signals using a camera [120]. Compared to the standard PPG, non-contact PPG have several advantages: firstly, the contact force between probe and skin is not present anymore, avoiding the deformation of vessels walls; furthermore, it is possible to acquire data from larger areas of the skin simultaneously, which allows evaluating the blood circulation in a large area or multi-sites. Verkruysse *et al.* [128] presented the first example of PPG measurement using a regular, red-green-blue (RGB) camera and ambient light. However, the accuracy of the recovered wave was not rigorously quantified. Poh *et al.* [99] showed that the PPG signal could be recovered from videos recorded with an ordinary webcam using ambient light.

Different algorithms have been proposed for contactless PPG measurement. The most straightforward method is the spatial averaging; by averaging the pixel intensity values in the region of interest (ROI), the signal to noise ratio (SNR) of PPG can be increased [128, 132,143]. A weighted averaging method was proposed in [72], where a higher weight is assigned to pixels that contain a stronger PPG signal and a lower weight to those with weaker signal. Blind Source Separation (BSS) algorithms have been utilized to extract PPG signal from different color channels, for example, Poh *et al.* used the Independent Component Analysis (ICA) [85,99,100] to extract the pulse signal. To increase the reliability, in [84] they considered a camera with five color channels, red, green, blue, cyan and orange. Similar to ICA, Lewandowska *et al.* used Principle Component Analysis (PCA) for pulse signal extraction [77]. As a understanding of the algorithm principle behind the non-contact PPG, Wang *et al.* presented a mathematical model that incorporates the pertinent optical and physiological properties of skin reflection [130]. With this model, a plane orthogonal to the skin-tone can be defined in the temporally normalized RGB space for pulse extraction.

Most of the existing works focus on, as substitute contact PPG, measuring vital sign parameters, e.g., heart rate and respiration rate [11], blood oxygen saturation (SpO$_2$) [61,132], pulse rate variability [100], and blood pressure [60,140]. A few works have focused on PPG imaging to measure the PPG signals in everything location for, e.g., perfusion monitoring. Hu *et al.* compared reflection and transmission mode on PPG perfusion maps [144], by acquiring images at wavelengths of 660, 870 and 940 nm on the finger tips. Rubins *et al.* conducted experiments on elaborating two-dimensional mapping of the dermal perfusion [111]. Kamshilin and colleagues introduced a method for extracting high spatial resolution photoplethysmographic images using the lock-in amplification [64]. They showed qualitatively that it is possible to detect perfusion changes due to scratching of the skin or occlusion of the vascular bed [65].

Thermal imaging is another method for measuring physiological information [104]. It is based on the information contained in the thermal signal emitted from major superficial vessels. This signal is acquired through a highly sensitive thermal imaging system. Temperature of the vessel is modulated by pulsative blood flow. The method is based on the assumption that temperature modulation due to pulsating blood flow produces the strongest thermal signal in a superficial vessel. This signal is affected by physiological and environmental thermal phenomena. Therefore, the resulting thermal signal that is being sensed by the infrared camera is a composite signal, with the pulse being only one of its components. Garbey *et al.* [50] developed a method to measure cardiac pulse based on a thermal signal emitted from major superficial vessels of the face. Human faces convey not only information about physiological parameters, but also information regarding mental condition and disease symptoms. Certain medical conditions alter the expression or appearance of the face due to physiological or behavioral responses. Thevenot *et al.* [123] presented a general overview of different works in the literature on using computer vision for health diagnosis from faces.

Most of the car accidents happen because of people's unsafe behaviors [104], and many aspects of the driver's conditions may affect the safe driving, for example sleepiness and sleep disorder, clinical depression, alcohol consumption, emergencies (myocardial dysfunction), and even their emotions (anger, excitement, etc.). By continuously monitoring the physiological parameters of the driver, the driving safety could be further improved. Therefore, driver monitoring systems have drawn a lot of attention from the research community and the industry [104]. The use of remote, non-contact camera imaging has been investigated in recent years for in-car health monitoring. For example, Guo *et al.* [53] presented a camera-based non-contact approach to monitor driver's Heart Rate Variability (HRV) continuously under real world driving circumstances. Rahman *et al.* [104] discussed the lat-

est developments in the literature on monitoring phys-iological parameters using camera and investigate the possibility of using this in driver monitoring domain.

5.4. Challenges and future directions

Nations around the world face a rising demand for costly health care for the ageing population. With the low cost of camera and the maturity of computer vision technology, camera monitoring for healthcare will be further developed with more interests. Although it presents significant promise, camera-based technology has several obstacles to overcome to reach full maturity. The main barriers include users' acceptability and risk of loss of privacy. The use of cameras would be acceptable in public facilities, i.e. nursing centres and hospitals, and during specific activities or events, such as tele-rehabilitation or safety assessment. However, when used in private spaces (e.g. bedrooms, bathrooms), privacy is a big concern. The use of cameras in private spaces without any privacy protection mechanisms yields in low acceptance by the end user. Hence, maintaining privacy as much as possible is a very important consideration when developing camera-based health monitoring solutions; privacy needs to be protected either by design, or by explicitly taking appropriate measures. One way is to use, instead of normal color camera, the depth sensor and/or thermal sensor, as did recently in [80].

Although we can develop more sophisticated vision-based techniques, it is also worth emphasizing that the vision-based methods are not necessary to be a standalone system for healthcare monitoring. It can be combined with other sensors to improve the robustness and accuracy. For instance, in some AAL applications, audio-video based solutions were proposed, where sound processing and speech recognition are included to help improve the robustness. More recently many multimodal approaches using a network of sensor arrays have been adopted and used.

6. Conclusions

This paper aims at presenting some insights of the state-of-the-art research works in the fields of Intelligent Video Surveillance (IVS), Wireless Sensor Network (WSN)-based human activity recognition, and camera-based health monitoring, with indications of future directions and challenges of these three prominent fields of Ambient Intelligence. Without preten-

sion of completeness, this paper has the main objective to describe the historical trend over the last two decades, the current status in these fields, and draw some possible directions of research that tackle the main challenges for each field.

References

[1] F. Adelantado, X. Vilajosana, P. Tuset-Peiro, B. Martinez, J. Melia-Segui and T. Watteyne, Understanding the limits of LoRaWAN, *IEEE Communications Magazine* **55**(9) (2017), 34–40, ISSN 0163-6804. doi:10.1109/MCOM.2017. 1600613.

[2] H. Aghajan and A. Cavallaro, *Multi-Camera Networks: Principles and Applications*, Academic Press, 2009.

[3] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam and E. Cayirci, Wireless sensor networks: A survey, *Computer Networks* **38**(4) (2002), 393–422. ISSN 1389-1286. http://www. sciencedirect.com/science/article/pii/S1389128601003024. doi:10.1016/S1389-1286(01)00302-4.

[4] H. Alemdar and C. Ersoy, Wireless sensor networks for healthcare: A survey, *Computer Networks* **54**(15) (2010), 2688–2710, ISSN 1389-1286. http://www.sciencedirect. com/science/article/pii/S1389128610001398. doi:10.1016/j. comnet.2010.05.003.

[5] S. Ali and M. Shah, A Lagrangian particle dynamics approach for crowd flow segmentation and stability analysis, in: *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, IEEE, 2007, pp. 1–6.

[6] F.R. Allen, E. Ambikairajah, N.H. Lovell and B.G. Celler, Classification of a known sequence of motions and postures from accelerometry data using adapted Gaussian mixture models, *Physiological Measurement* **27**(10) (2006), 935, http://stacks.iop.org/0967-3334/27/i=10/a=001. doi:10.1088/ 0967-3334/27/10/001.

[7] J. Allen, Photoplethysmography and its application in clinical physiological measurement, *Physiological measurement* **28**(3) (2007), 1. doi:10.1088/0967-3334/28/3/R01.

[8] S. Alletto, G. Serra, S. Calderara and R. Cucchiara, Understanding social relationships in egocentric vision, *Pattern Recognition* **48**(12) (2015), 4082–4096. doi:10.1016/j.patcog. 2015.06.006.

[9] J. Andreu-Perez, D.R. Leff, H.M.D. Ip and G.Z. Yang, From wearable sensors to smart implants-toward pervasive and personalized healthcare, *IEEE Transactions on Biomedical Engineering* **62**(12) (2015), 2750–2762, ISSN 0018-9294. doi:10. 1109/TBME.2015.2422751.

[10] Aqeel-ur-Rehman, A.Z. Abbasi, N. Islam and Z.A. Shaikh, A review of wireless sensors and networks' applications in agriculture, *Computer Standards & Interfaces* **36**(2) (2014), 263–270, ISSN 0920-5489. http://www.sciencedirect. com/science/article/pii/S0920548911000353. doi:10.1016/j. csi.2011.03.004.

[11] A. Aridarma, T. Mengko and S. Soegijoko, Personal medical assistant: Future exploration, in: *Electrical Engineering and Informatics (ICEEI), 2011 International Conference on*, IEEE, 2011, pp. 1–6.

[12] O. Aziz, E.J. Park, G. Mori and S.N. Robinovitch, Distinguishing the causes of falls in humans using an array of wearable tri-axial accelerometers, *Gait & Posture* **39**(1) (2014), 506–512, ISSN 0966-6362. http://www.sciencedirect.com/science/article/pii/S0966636213005924. doi:10.1016/j.gaitpost.2013.08.034.

[13] T. Banerjee, M. Enayati, J.M. Keller, M. Skubic, M. Popescu and M. Rantz, Monitoring patients in hospital beds using unobtrusive depth sensors, in: *Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS'14*, 2014.

[14] T. Banerjee, M. Yefimova, J.M. Keller, M. Skubic, D.L. Woods and M. Rantz, Exploratory analysis of older adults' sedentary behavior in the primary living area using kinect depth data, *Journal of Ambient Intelligence and Smart Environments* **9**(2) (2017), 163–179. doi:10.3233/AIS-170428.

[15] P. Baronti, P. Pillai, V.W.C. Chook, S. Chessa, A. Gotta and Y.F. Hu, Wireless sensor networks: A survey on the state of the art and the 802.15.4 and ZigBee standards, *Computer Communications* **30**(7) (2007), 1655–1695, http://www.sciencedirect.com/science/article/pii/S0140366406004749. doi:10.1016/j.comcom.2006.12.020.

[16] M.D. Beynon, D.J. Van Hook, M. Seibert, A. Peacock and D. Dudgeon, Detecting abandoned packages in a multi-camera video surveillance system, in: *Advanced Video and Signal Based Surveillance, 2003. Proceedings. IEEE Conference on*, IEEE, 2003, pp. 221–228. doi:10.1109/AVSS.2003.1217925.

[17] A. Bobick and J. Davis, Real-time recognition of activity using temporal templates, in: *Applications of Computer Vision, 1996. WACV'96, Proceedings 3rd IEEE Workshop on*, IEEE, 1996, pp. 39–42.

[18] M. Bolanos, M. Dimiccoli and P. Radeva, Toward storytelling from visual lifelogging: An overview, *IEEE Transactions on Human–Machine Systems* **47**(1) (2017), 77–90.

[19] T. Boult, R. Micheals, A. Erkan, P. Lewis, C. Powers, C. Qian and W. Yin, Frame-rate multi-body tracking for surveillance, in: *Proc. DARPA Image Understanding Workshop*, Citeseer, 1998, pp. 305–308.

[20] M. Bramberger, J. Brunner, B. Rinner and H. Schwabach, Real-time video analysis on an embedded smart camera for traffic surveillance, in: *Real-Time and Embedded Technology and Applications Symposium, 2004. Proceedings. RTAS 2004. 10th IEEE*, IEEE, 2004, pp. 174–181. doi:10.1109/RTTAS.2004.1317262.

[21] A. Braun and T. Dutz, Low-cost indoor localization using cameras – evaluating AmbiTrack and its applications in Ambient Assisted Living, *Journal of Ambient Intelligence and Smart Environments* **8**(3) (2016), 243–258. doi:10.3233/AIS-160377.

[22] C. Bregler and J. Malik, Tracking people with twists and exponential maps, in: *Computer Vision and Pattern Recognition, 1998. Proceedings. 1998 IEEE Computer Society Conference on*, IEEE, 1998, pp. 8–15.

[23] T. Brodsky, R. Cohen, E. Cohen-Solal, S. Gutta, D. Lyons, V. Philomin and M. Trajkovic, Visual surveillance in retail stores and in the home, in: *Video-Based Surveillance Systems*, Springer, 2002, pp. 51–61. doi:10.1007/978-1-4615-0913-4_4.

[24] Q. Cai and J.K. Aggarwal, Tracking human motion in structured environments using a distributed-camera system, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **21**(11) (1999), 1241–1247. doi:10.1109/34.809119.

[25] S. Calderara, R. Cucchiara and A. Prati, Bayesian-competitive consistent labeling for people surveillance, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **30**(2) (2008).

[26] S. Calderara, U. Heinemann, A. Prati, R. Cucchiara and N. Tishby, Detecting anomalies in people's trajectories using spectral graph analysis, *Computer Vision and Image Understanding* **115**(8) (2011), 1099–1111. doi:10.1016/j.cviu.2011.03.003.

[27] F. Cardinaux, D. Bhowmik, C. Abhayaratne and M.S. Hawley, Video based technology for ambient assisted living: A review of the literature, *Journal of Ambient Intelligence and Smart Environments* **3**(3) (2011), 253–269.

[28] L. Cattani, D. Alinovi, G. Ferrari, R. Raheli, E. Pavlidis, C. Spagnoli and F. Pisani, Monitoring infants by automatic video processing: A unified approach to motion analysis, *Computers in Biology and Medicine* **80**(1) (2017), 158–165. doi:10.1016/j.compbiomed.2016.11.010.

[29] A.A. Chaaraoui, P. Climent-Pérez and F. Flórez-Revuelta, A review on vision techniques applied to human behaviour analysis for ambient-assisted living, *Expert Systems with Applications* **39**(12) (2012), 10873–10888. doi:10.1016/j.eswa.2012.03.005.

[30] J.M. Chaquet, E.J. Carmona and A. Fernández-Caballero, A survey of video datasets for human action and activity recognition, *Computer Vision and Image Understanding* **117**(6) (2013), 633–659. doi:10.1016/j.cviu.2013.01.013.

[31] J.G. Chase, F. Agogue, C. Starfinger, Z. Lam, G.M. Shaw, A.D. Rudge and H. Sirisena, Quantifying agitation in sedated ICU patients using digital imaging, *Computer Methods and Programs in Biomedicine* **76**(2) (2004), 131–141. doi:10.1016/j.cmpb.2004.03.005.

[32] L. Chen, J. Hoey, C.D. Nugent, D.J. Cook and Z. Yu, Sensor-based activity recognition, *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* **42**(6) (2012), 790–808, ISSN 1094-6977. doi:10.1109/TSMCC.2012.2198883.

[33] L. Chen, C. Nugent and G. Okeyo, An ontology-based hybrid approach to activity modeling for smart homes, *IEEE Transactions on Human–Machine Systems* **44**(1) (2014), 92–105, ISSN 2168-2291. doi:10.1109/THMS.2013.2293714.

[34] M. Chen, X. Wei, Q. Yang, Q. Li, G. Wang and M.-H. Yang, Spatiotemporal GMM for background subtraction with superpixel hierarchy, in: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.

[35] Y. Chen, X. Zhu, S. Gong et al., 2018, Person re-identification by deep learning multi-scale representations.

[36] R.T. Collins, A.J. Lipton and T. Kanade, Introduction to the special section on video surveillance, *IEEE Transactions on pattern analysis and machine intelligence* **22**(8) (2000), 745–746. doi:10.1109/TPAMI.2000.868676.

[37] R.T. Collins, A.J. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, O. Hasegawa, P. Burt et al., A system for video surveillance and monitoring, *VSAM final report* (2000), 1–68.

[38] D. Comaniciu, V. Ramesh and P. Meer, Kernel-based object tracking, *IEEE Transactions on pattern analysis and machine intelligence* **25**(5) (2003), 564–577. doi:10.1109/TPAMI.2003.1195991.

[39] C.J. Costello, C.P. Diehl, A. Banerjee and H. Fisher, Scheduling an active camera to observe people, in: *Proceedings of the ACM 2nd International Workshop on Video Surveillance & Sensor Networks*, ACM, 2004, pp. 39–45. doi:10.1145/1026799.1026808.

[40] M. Cristani, L. Bazzani, G. Paggetti, A. Fossati, D. Tosato, A. Del Bue, G. Menegaz and V. Murino, Social interaction discovery by statistical analysis of F-formations, in: *BMVC*, Vol. 2, 2011, p. 4.

[41] R. Cucchiara, C. Grana, A. Prati and R. Vezzani, Probabilistic posture classification for human-behavior analysis, *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans* **35**(1) (2005), 42–54. doi:10.1109/TSMCA.2004.838501.

[42] R. Cucchiara, A. Prati and R. Vezzani, Real-time motion segmentation from moving cameras, *Real-Time Imaging* **10**(3) (2004), 127–143. doi:10.1016/j.rti.2004.03.002.

[43] P. Dadlani, T. Gritti, C. Shan, B.D. Ruyter and P. Markopoulos, SoPresent: An awareness system for connecting remote households, in: *Proceedings of the European Conference on Ambient Intelligence, ECAI'14*, 2014.

[44] A. Dubois and F. Charpillet, Measuring frailty and detecting falls for elderly home care using depth camera, *Journal of Ambient Intelligence and Smart Environments* **9**(4) (2017), 469–481. doi:10.3233/AIS-170444.

[45] A. Elgammal, Background subtraction: Theory and practice, *Synthesis Lectures on Computer Vision* **5**(1) (2014), 1–83. doi:10.2200/S00613ED1V01Y201411COV006.

[46] M. Ermes, J. Pärkkä, J. Mäntyjärvi and I. Korhonen, Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions, *IEEE Transactions on Information Technology in Biomedicine* **12**(1) (2008), 20–26, ISSN 1089-7771. doi:10.1109/TITB.2007.899496.

[47] M. Farenzena, L. Bazzani, A. Perina, V. Murino and M. Cristani, Person re-identification by symmetry-driven accumulation of local features, in: *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, IEEE, 2010, pp. 2360–2367. doi:10.1109/CVPR.2010.5539926.

[48] F. Foerster, M. Smeja and J. Fahrenberg, Detection of posture and motion by accelerometry: A validation study in ambulatory monitoring, *Computers in Human Behavior* **15**(5) (1999), 571–583, ISSN 0747-5632. http://www.sciencedirect.com/science/article/pii/S0747563299000370. doi:10.1016/S0747-5632(99)00037-0.

[49] N. Friedman and S. Russell, Image segmentation in video sequences: A probabilistic approach, in: *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann Publishers Inc., 1997, pp. 175–181.

[50] M. Garbey, S. Nanfei, A. Merla and I. Pavlidis, Contact-free measurement of cardiac pulse based on the analysis of thermal imagery, *IEEE Transactions on Biomedical Engineering* **54**(8) (2007), 1418–1426. doi:10.1109/TBME.2007.891930.

[51] K.S. Gayathri, K.S. Easwarakumar and S. Elias, Probabilistic ontology based activity recognition in smart homes using Markov logic network, *Knowledge-Based Systems* **121** (2017), 173–184, ISSN 0950-7051. http://www.sciencedirect.com/science/article/pii/S0950705117300370. doi:10.1016/j.knosys.2017.01.025.

[52] C. Gomez and J. Paradells, Wireless home automation networks: A survey of architectures and technologies, *IEEE Communications Magazine* **48**(6) (2010), 92–101, ISSN 0163-6804. doi:10.1109/MCOM.2010.5473869.

[53] Z. Guo, Z. Jane and Z. Shen, Physiological parameter monitoring of drivers based on video data and independent vector analysis, in: *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing, ICASSP'14*, 2014.

[54] J.A. Gutierrez, M. Naeve, E. Callaway, M. Bourgeois, V. Mitter and B. Heile, IEEE 802.15.4: A developing standard for low-power low-cost wireless personal area networks, *IEEE Network* **15**(5) (2001), 12–19, ISSN 0890-8044. doi:10.1109/65.953229.

[55] I. Haritaoglu, D. Harwood and L.S. Davis, W/sup 4: Real-time surveillance of people and their activities, *IEEE Transactions on pattern analysis and machine intelligence* **22**(8) (2000), 809–830. doi:10.1109/34.868683.

[56] Z. He, Activity recognition from accelerometer signals based on wavelet-AR model, in: *2010 IEEE International Conference on Progress in Informatics and Computing*, Vol. 1, 2010, pp. 499–502. doi:10.1109/PIC.2010.5687572.

[57] S. Hengstler, D. Prashanth, S. Fong and H. Aghajan, Mesh-Eye: A hybrid-resolution smart camera mote for applications in distributed intelligent surveillance, in: *Proceedings of the 6th International Conference on Information Processing in Sensor Networks*, ACM, 2007, pp. 360–369.

[58] L. Hu, Y. Chen, S. Wang and Z. Chen, b-COELM: A fast, lightweight and accurate activity recognition model for mini-wearable devices, *Pervasive and Mobile Computing* **15** (2014), 200–214, Special Issue on Information Management in Mobile Applications Special Issue on Data Mining in Pervasive Environments, ISSN 1574-1192, http://www.sciencedirect.com/science/article/pii/S157411921400090X. doi:10.1016/j.pmcj.2014.06.002.

[59] G. Hua, Y. Fu, M. Turk, M. Pollefeys and Z. Zhang, Introduction to the special issue on mobile vision, *International Journal of Computer Vision* **96**(3) (2012), 277–279. doi:10.1007/s11263-011-0506-3.

[60] S.-C. Huang, P.-H. Hung, C.-H. Hong and H.-M. Wang, A new image blood pressure sensor based on PPG, RRT, BPTT, and harmonic balancing, *IEEE Sensors Journal* **14**(10) (2014), 3685–3692. doi:10.1109/JSEN.2014.2329676.

[61] K. Humphreys, T. Ward and C. Markham, *A CMOS Camera-Based Pulse Oximetry Imaging System*, IEEE, 2006, pp. 3494–3497.

[62] R. Jia and B. Liu, Human daily activity recognition by fusing accelerometer and multi-lead ECG data, in: *2013 IEEE International Conference on Signal Processing, Communication and Computing (ICSPCC 2013)*, 2013, pp. 1–4. doi:10.1109/ICSPCC.2013.6664056.

[63] A. Kamal, J. Harness, G. Irving and A. Mearns, Skin photoplethysmography – a review, *Computer methods and programs in biomedicine* **28**(4) (1989), 257–269. doi:10.1016/0169-2607(89)90159-4.

[64] A.A. Kamshilin, S. Miridonov, V. Teplov, R. Saarenheimo and E. Nippolainen, Photoplethysmographic imaging of high spatial resolution, *Biomedical optics express* **2**(4) (2011), 996–1006. doi:10.1364/BOE.2.000996.

[65] A.A. Kamshilin, V. Teplov, E. Nippolainen, S. Miridonov and R. Giniatullin, Variability of microcirculation detected by blood pulsation imaging, *PloS one* **8**(2) (2013), 57117. doi:10.1371/journal.pone.0057117.

[66] T. Kanade and M. Hebert, First-person vision, *Proceedings of the IEEE* **100**(8) (2012), 2442–2453. doi:10.1109/JPROC.2012.2200554.

[67] D.M. Karantonis, M.R. Narayanan, M. Mathie, N.H. Lovell and B.G. Celler, Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring, *IEEE Transactions on Information Technology in Biomedicine* **10**(1) (2006), 156–167, ISSN 1089–7771. doi:10.1109/TITB.2005.856864.

[68] S. Khan and M. Shah, Consistent labeling of tracked objects in multiple cameras with overlapping fields of view, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25**(10) (2003), 1355–1360. doi:10.1109/TPAMI.2003.1233912.

[69] K. Kim and L.S. Davis, Multi-camera tracking and segmentation of occluded people on ground plane using search-guided particle filtering, in: *European Conference on Computer Vision*, Springer, 2006, pp. 98–109.

[70] P. Kittipanya-Ngam, O.S. Guat and E.H. Lung, Computer vision applications for patients monitoring system, in: *Proceedings of 15th International Conference on Information Fusion*, 2012.

[71] J. Krumm, S. Harris, B. Meyers, B. Brumitt, M. Hale and S. Shafer, Multi-camera multi-person tracking for easyliving, in: *Visual Surveillance, 2000. Proceedings*, Third IEEE International Workshop on, IEEE, 2000, pp. 3–10.

[72] M. Kumar, A. Veeraraghavan and A. Sabharwal, DistancePPG: Robust non-contact vital signs monitoring using a camera, *Biomedical optics express* **6**(5) (2015), 1565–1588. doi:10.1364/BOE.6.001565.

[73] O.D. Lara and M.A. Labrador, A survey on human activity recognition using wearable sensors, *IEEE Communications Surveys Tutorials* **15**(3) (2013), 1192–1209, ISSN 1553-877X. doi:10.1109/SURV.2012.110112.00192.

[74] M. Lauridsen, B. Vejlgaard, I.Z. Kovacs, H. Nguyen and P. Mogensen, Interference measurements in the European 868 MHz ISM band with focus on LoRa and SigFox, in: *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, 2017, pp. 1–6. doi:10.1109/WCNC.2017.7925650.

[75] S.-W. Lee and K. Mase, Activity and location recognition using wearable sensors, *IEEE Pervasive Computing* **1**(3) (2002), 24–32, ISSN 1536-1268. doi:10.1109/MPRV.2002.1037719.

[76] S.H. Lee, H.D. Park, S.Y. Hong, K.J. Lee and Y.H. Kim, A study on the activity classification using a triaxial accelerometer, in: *Proceedings of the 25th Annual, International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No.03CH37439)*, Vol. 3, 2003, pp. 2941–29433, ISSN 1094-687X. doi:10.1109/IEMBS.2003.1280534.

[77] M. Lewandowska and J. Nowak, Measuring pulse rate with a webcam, *Journal of Medical Imaging and Health Informatics* **2**(1) (2012), 87–92. doi:10.1166/jmihi.2012.1064.

[78] A. Lipton, H. Fujiyoshi and R. Patil, Moving target detection and classification from real-time video, in: *Proc. IEEE Workshop Application of Computer Vision*, 1998.

[79] Lukas, W.A. Tanumihardja and E. Gunawan, On the application of IoT: Monitoring of troughs water level using WSN, in: *2015 IEEE Conference on Wireless Sensors (ICWiSe)*, 2015, pp. 58–62. doi:10.1109/ICWISE.2015.7380354.

[80] Z. Luo, J.-T. Hsieh, N. Balachandar, S. Yeung, G. Pusiol, J. Luxenberg, G. Li, L.-J. Li, N.L. Downing, A. Milstein and L. Fei-Fei, Computer vision-based descriptive analytics of seniors' daily activities for long-term health monitoring, in: *Proceedings of the Machine Learning for Healthcare (MLHC) Conference, MLHC '18*, 2018, pp. 1–18.

[81] D. Makris and T. Ellis, Learning semantic scene models from observing activity in visual surveillance, *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* **35**(3) (2005), 397–408. doi:10.1109/TSMCB.2005.846652.

[82] A. Mannini and A.M. Sabatini, Machine learning methods for classifying human physical activity from on-body accelerometers, *Sensors* **10**(2) (2010), 1154–1175, ISSN 1424-8220. http://www.mdpi.com/1424-8220/10/2/1154. doi:10.3390/s100201154.

[83] J. Mantyjarvi, J. Himberg and T. Seppanen, Recognizing human motion with multiple acceleration sensors, in: *2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace (Cat.No.01CH37236)*, Vol. 2, 2001, pp. 747–7522, ISSN 1062-922X. doi:10.1109/ICSMC.2001.973004.

[84] D. McDuff, S. Gontarek and R.W. Picard, Improvements in remote cardiopulmonary measurement using a five band digital camera, *IEEE Transactions on Biomedical Engineering* **61**(10) (2014), 2593–2601. doi:10.1109/TBME.2014.2323695.

[85] D.J. McDuff, J.R. Estepp, A.M. Piasecki and E.B. Blackford, A survey of remote optical photoplethysmographic imaging methods, in: *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, IEEE, 2015, pp. 6398–6404.

[86] R. Mehran, A. Oyama and M. Shah, Abnormal crowd behavior detection using social force model, in: *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, IEEE, 2009, pp. 935–942. doi:10.1109/CVPR.2009.5206641.

[87] S. Morita, K. Yamazawa and N. Yokoya, Networked video surveillance using multiple omnidirectional cameras, in: *Computational Intelligence in Robotics and Automation, 2003. Proceedings. 2003 IEEE International Symposium on*, Vol. 3, IEEE, 2003, pp. 1245–1250.

[88] B.T. Morris and M.M. Trivedi, A survey of vision-based trajectory learning and analysis for surveillance, *IEEE transactions on circuits and systems for video technology* **18**(8) (2008), 1114–1127. doi:10.1109/TCSVT.2008.927109.

[89] S. Mota and R.W. Picard, Automated posture analysis for detecting learner's interest level, in: *Computer Vision and Pattern Recognition Workshop, 2003. CVPRW'03. Conference on*, Vol. 5, IEEE, 2003, p. 49. doi:10.1109/CVPRW.2003.10047.

[90] G. Mulligan, The 6LoWPAN architecture, in: *Proceedings of the 4th Workshop on Embedded Networked Sensors, EmNets '07*, ACM, New York, NY, USA, 2007, pp. 78–82. ISBN 978-1-59593-694-3. doi:10.1145/1278972.1278992.

[91] H. Nait-Charif and S.J. McKenna, Activity summarisation and fall detection in a supportive home environment, in: *Pattern Recognition, 2004. ICPR 2004, Proceedings of the 17th*

International Conference on, Vol. 4, IEEE, 2004, pp. 323–326.

[92] B. Najafi, K. Aminian, A. Paraschiv-Ionescu, F. Loew, C.J. Bula and P. Robert, Ambulatory system for human motion analysis using a kinematic sensor: Monitoring of daily physical activity in the elderly, *IEEE Transactions on Biomedical Engineering* **50**(6) (2003), 711–723, ISSN 0018-9294. doi:10.1109/TBME.2003.812189.

[93] M.H.M. Noor, Z. Salcic and K.I.-K. Wang, Enhancing onto-logical reasoning with uncertainty handling for activity recog-nition, *Knowledge-Based Systems* **114** (2016), 47–60, ISSN 0950-7051. http://www.sciencedirect.com/science/article/pii/S0950705116303604. doi:10.1016/j.knosys.2016.09.028.

[94] M.H.M. Noor, Z. Salcic and K.I.-K. Wang, Adap-tive sliding window segmentation for physical activity recognition using a single tri-axial accelerometer, *Per-vasive and Mobile Computing* **38** (2017), 41–59, ISSN 1574-1192. http://www.sciencedirect.com/science/article/pii/S1574119216302280. doi:10.1016/j.pmcj.2016.09.009.

[95] M.H.M. Noor, Z. Salcic and K.I.-K. Wang, Ontology-based sensor fusion activity recognition, *Journal of Ambient In-telligence and Humanized Computing* (2018), ISSN 1868-5145, https://doi.org/10.1007/s12652-017-0668-0. doi:10.1007/s12652-017-0668-0.

[96] V. Osmani, S. Balasubramaniam and D. Botvich, Hu-man activity recognition in pervasive health-care: Sup-porting efficient remote collaboration, *Journal of Network and Computer Applications* **31**(4) (2008), 628–655, ISSN 1084-8045. http://www.sciencedirect.com/science/article/pii/S1084804507000719. doi:10.1016/j.jnca.2007.11.002.

[97] J. Petäjäjärvi, K. Mikhaylov, M. Hämäläinen and J. Iinatti, Evaluation of LoRa LPWAN technology for remote health and wellbeing monitoring, in: *2016 10th International Sym-posium on Medical Information and Communication Tech-nology (ISMICT)*, 2016, pp. 1–5. doi:10.1109/ISMICT.2016.7498898.

[98] M. Piccardi, Background subtraction techniques: A review, in: *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, Vol. 4, IEEE, 2004, pp. 3099–3104.

[99] M.-Z. Poh, D.J. McDuff and R.W. Picard, Non-contact, au-tomated cardiac pulse measurements using video imaging and blind source separation, *Optics express* **18**(10) (2010), 10762–10774. doi:10.1364/OE.18.010762.

[100] M.-Z. Poh, D.J. McDuff and R.W. Picard, Advancements in noncontact, multiparameter physiological measurements us-ing a webcam, *IEEE transactions on biomedical engineering* **58**(1) (2011), 7–11. doi:10.1109/TBME.2010.2086456.

[101] R. Poppe, A survey on vision-based human action recog-nition, *Image and vision computing* **28**(6) (2010), 976–990. doi:10.1016/j.imavis.2009.11.014.

[102] P. Prasad, Recent trend in wireless sensor network and its applications: A survey, *Sensor Review* **35**(2) (2015), 229–236, https://doi.org/10.1108/SR-08-2014-683. doi:10.1108/SR-08-2014-683.

[103] A. Prati and F.Z. Qureshi, Integrating consumer smart cam-eras into camera networks: Opportunities and obstacles, *Com-puter* **47**(5) (2014), 45–51. doi:10.1109/MC.2014.125.

[104] H. Rahman, S. Barua and S. Begum, Intelligent driver moni-toring based on physiological sensor signals: Application us-ing camera, in: *Proceedings of the 18th IEEE International Conference on Intelligent Transportation Systems*, 2015.

[105] C. Randell and H. Muller, Context awareness by analysing accelerometer data, in: *Digest of Papers. Fourth Interna-tional Symposium on Wearable Computers*, 2000, pp. 175–176. doi:10.1109/ISWC.2000.888488.

[106] R. Ratasuk, B. Vejlgaard, N. Mangalvedhe and A. Ghosh, NB-IoT system for M2M communication, in: *2016 IEEE Wireless Communications and Networking Conference*, 2016, pp. 1–5. doi:10.1109/WCNC.2016.7564708.

[107] N. Ravi, D. Nikhil, P. Mysore and M.L. Littman, Activity recognition from accelerometer data, in: *Proceedings of the Seventeenth Conference on Innovative Applications of Artifi-cial Intelligence (IAAI)*, 2005, pp. 1541–1546.

[108] E. Ristani and C. Tomasi, Features for Multi-Target Multi-Camera Tracking and Re-Identification, preprint, arXiv:1803.10859, To appear in CVPR 2018 (2018).

[109] N.D. Rodríguez, M.P. Cuéllar, J. Lilius and M.D. Calvo-Flores, A survey on ontologies for human behavior recog-nition, *ACM Comput. Surv.* **46**(4) (2014), 43:1–43:33, ISSN 0360-0300.

[110] C. Rougier, A. St-Arnaud, J. Rousseau and J. Meunier, Video surveillance for fall detection, in: *Video Surveillance*, W. Lin, ed., IntechOpen, Rijeka, 2011, Chap. 20, https://doi.org/10.5772/14609. doi:10.5772/14609.

[111] U. Rubins, R. Erts and V. Nikiforovs, The blood perfusion mapping in the human skin by photoplethysmography imag-ing, in: *XII Mediterranean Conference on Medical and Bi-ological Engineering and Computing 2010*, Springer, 2010, pp. 304–306. doi:10.1007/978-3-642-13039-7_76.

[112] M. Sekine, T. Tamura, T. Togawa and Y. Fukui, Classification of waist-acceleration signals in a continuous walking record, *Medical Engineering & Physics* **22**(4) (2000), 285–291, ISSN 1350-4533. http://www.sciencedirect.com/science/article/pii/S1350453300000412. doi:10.1016/S1350-4533(00)00041-2.

[113] M.F. Shaikh, Z. Salcic, K.I.-K. Wang and A.P. Hu, Bipedal gait model for precise gait recognition and optimal trigger-ing in foot drop stimulator: A proof of concept, Medical & Biological Engineering & Computing (2018), ISSN 1741-0444, https://doi.org/10.1007/s11517-018-1810-7. doi:10.1007/s11517-018-1810-7.

[114] R.S. Sinha, Y. Wei and S.-H. Hwang, A survey on LPWA technology: LoRa and NB-IoT, *ICT Express* **3**(1) (2017), 14–21, ISSN 2405-9595. http://www.sciencedirect.com/science/article/pii/S2405959517300061. doi:10.1016/j.icte.2017.03.004.

[115] A.W. Smeulders, D.M. Chu, R. Cucchiara, S. Calderara, A. Dehghan and M. Shah, Visual tracking: An experimental survey, *IEEE transactions on pattern analysis and machine intelligence* **36**(7) (2014), 1442–1468. doi:10.1109/TPAMI.2013.230.

[116] A. Sobral and A. Vacavant, A comprehensive review of back-ground subtraction algorithms evaluated with synthetic and real videos, *Computer Vision and Image Understanding* **122** (2014), 4–21. doi:10.1016/j.cviu.2013.12.005.

[117] C. Stauffer and W.E.L. Grimson, Adaptive background mix-ture models for real-time tracking, in: *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, Vol. 2, IEEE, 1999, pp. 246–252.

[118] C. Stauffer and W.E.L. Grimson, Learning patterns of activity using real-time tracking, *IEEE Transactions on pattern anal-ysis and machine intelligence* **22**(8) (2000), 747–757. doi:10.1109/34.868677.

[119] Y. Sun, C. Shan, T. Tan, X. Long, A. Pourtaherian, S. Zinger and P.H.N. de With, Video-based discomfort detection for infants, *Machine Vision and Applications* (2018). doi:10.1007/s00138-018-0968-1.

[120] Y. Sun and N. Thakor, Photoplethysmography revisited: From contact to noncontact, from point to imaging, *IEEE Transactions on Biomedical Engineering* **63**(3) (2016), 463–477. doi:10.1109/TBME.2015.2476337.

[121] M. Taj and A. Cavallaro, Multi-view multi-object detection and tracking, in: *Computer Vision*, Springer, 2010, pp. 263–280. doi:10.1007/978-3-642-12848-6_10.

[122] Y.T. Tesfaye, E. Zemene, M. Pelillo and A. Prati, Multi-object tracking using dominant sets, *IET computer vision* **10**(4) (2016), 289–298. doi:10.1049/iet-cvi.2015.0297.

[123] J. Thevenot, M.B. Lòpez and A. Hadid, A survey on computer vision for assistive medical diagnosis from faces, *IEEE Journal of Biomedical and Health Informatics* **22**(5) (2017), 1497–1511. doi:10.1109/JBHI.2017.2754861.

[124] K. Toyama, J. Krumm, B. Brumitt and B. Meyers, Wallflower: Principles and practice of background maintenance, in: *Computer Vision, the Proceedings of the Seventh IEEE International Conference on*, Vol. 1, IEEE, 1999, pp. 255–261.

[125] A. Trotta, F. D'Andreagiovanni, M. Di Felice, E. Natalizio and K.R. Chowdhury, When UAVs Ride A Bus: Towards Energy-efficient City-scale Video Surveillance, in: *IEEE International Conference on Computer Communications (INFOCOM) 2018*, 2018.

[126] M. Uiterwaal, E.B.C. Glerum, H.J. Busser and R.C.V. Lummel, Ambulatory monitoring of physical activity in working situations, a validation study, *Journal of Medical Engineering & Technology* **22**(4) (1998), 168–172. doi:10.3109/03091909809032535.

[127] A. Ullah, J. Ahmad, K. Muhammad, M. Sajjad and S.W. Baik, Action recognition in video sequences using deep Bi-directional LSTM with CNN features, *IEEE Access* **6** (2018), 1155–1166. doi:10.1109/ACCESS.2017.2778011.

[128] W. Verkruysse, L.O. Svaasand and J.S. Nelson, Remote plethysmographic imaging using ambient light, *Optics express* **16**(26) (2008), 21434–21445. doi:10.1364/OE.16.021434.

[129] R. Vezzani, D. Baltieri and R. Cucchiara, People reidentification in surveillance and forensics: A survey, *ACM Computing Surveys (CSUR)* **46**(2) (2013), 29. doi:10.1145/2543581.2543596.

[130] W. Wang, A.C. den Brinker, S. Stuijk and G. de Haan, Algorithmic principles of remote PPG, *IEEE Transactions on Biomedical Engineering* **64**(7) (2017), 1479–1491. doi:10.1109/TBME.2016.2609282.

[131] X. Wang, Intelligent multi-camera video surveillance: A review, *Pattern recognition letters* **34**(1) (2013), 3–19. doi:10.1016/j.patrec.2012.07.005.

[132] F.P. Wieringa, F. Mastik and A.F.v.d. Steen, Contactless multiple wavelength photoplethysmographic imaging: A first step toward "SpO 2 camera" technology, *Annals of biomedical engineering* **33**(8) (2005), 1034–1041. doi:10.1007/s10439-005-5763-2.

[133] A.J. Wixted, P. Kinnaird, H. Larijani, A. Tait, A. Ahmadinia and N. Strachan, Evaluation of LoRa and LoRaWAN for wireless sensor networks, in: *2016 IEEE SENSORS*, 2016, pp. 1–3. doi:10.1109/ICSENS.2016.7808712.

[134] C.R. Wren, A. Azarbayejani, T. Darrell and A.P. Pentland Pfinder, Real-time tracking of the human body, *IEEE Transactions on pattern analysis and machine intelligence* **19**(7) (1997), 780–785. doi:10.1109/34.598236.

[135] W. Xiao and Y. Lu, Daily Human Physical Activity Recognition Based on Kernel Discriminant Analysis and Extreme Learning Machine, *Mathematical Problems in Engineering* (2015).

[136] J. Ye, S. Dobson and S. McKeever, Situation identification techniques in pervasive computing: A review, *Pervasive and Mobile Computing* **8**(1) (2012), 36–66, ISSN 1574-1192. http://www.sciencedirect.com/science/article/pii/S1574119211000253. doi:10.1016/j.pmcj.2011.01.004.

[137] A. Yilmaz, O. Javed and M. Shah, Object tracking: A survey, *ACM computing surveys (CSUR)* **38**(4) (2006), 13. doi:10.1145/1177352.1177355.

[138] Z-Wave Alliance, 2015, http://z-wavealliance.org//, accessed 10 April 2015.

[139] B. Zhan, D.N. Monekosso, P. Remagnino, S.A. Velastin and L.-Q. Xu, Crowd analysis: A survey, *Machine Vision and Applications* **19**(5–6) (2008), 345–357. doi:10.1007/s00138-008-0132-4.

[140] G. Zhang, C. Shan, I. Kirenko, X. Long and R.M. Aarts, Hybrid optical unobtrusive blood pressure measurements, *Sensors* **17**(7) (2017), 1541. doi:10.3390/s17071541.

[141] Z. Zhang, C. Conly and V. Athitsos, A survey on vision-based fall detection, in: *Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments, PETRA '15*, ACM, New York, NY, USA, 2015, pp. 46:1–46:7. ISBN 978-1-4503-3452-5. doi:10.1145/2769493.2769540.

[142] R. Zhao, W. Oyang and X. Wang, Person re-identification by saliency learning, *IEEE transactions on pattern analysis and machine intelligence* **39**(2) (2017), 356–370. doi:10.1109/TPAMI.2016.2544310.

[143] J. Zheng and S. Hu, The preliminary investigation of imaging photoplethysmographic system, *Journal of Physics: Conference Series* **85** (2007), 012031.

[144] J. Zheng, S. Hu, V. Chouliaras and R. Summers, Feasibility of imaging photoplethysmography, in: *BioMedical Engineering and Informatics, BMEI 2008. International Conference on*, Vol. 2, IEEE, 2008, pp. 72–75.