The Knowledge Engineering Review, Vol. 33, e4, 1–25. © Cambridge University Press, 2018
doi:10.1017/S0269888918000048

A survey and taxonomy on intelligent surveillance from a system perspective

RADU-CASIAN MIHAILESCU, PAUL DAVIDSSON, ULRIK EKLUND and
JAN A. PERSSON

Department of Computer Science, Malmö University, Internet of Things and People Research Center, 205 06 Malmö, Sweden;
e-mail: radu.c.mihailescu@mah.se, paul.davidsson@mah.se, ulrik.eklund@mah.se, jan.a.persson@mah.se

Abstract

Recent proliferation of surveillance systems is mostly attributed to advances in both image-processing techniques and hardware enhancement of smart cameras, as well as the ubiquity of sensor-driven architectures. Owing to these capabilities, new aspects are coming to the forefront. This paper addresses the current state-of-the-art and provides researchers with an overview of existing surveillance solutions, analyzing their properties as a system and drawing attention to relevant challenges when developing, deploying and managing them. Also, some of the more prominent application domains are highlighted here. In an effort to understand the development of the advanced solutions, based on their most distinctive characteristics, we propose a taxonomy for surveillance systems to help classify them and reveal gaps in existing research. We conclude by identifying promising future research lines.

1 Introduction

Recent advances, in particular in the area of smart cameras, equipped with video sensing, processing and communication all on a single embedded device, have directed research toward the development of a new generation of surveillance systems, cf. Nilsson (2008), Qian *et al.* (2011). In broad terms, a surveillance system can be perceived as a system of connected sensors, deployed for real-time monitoring of persistent and transient objects. General characteristics include the ability to aggregate, store and interpret sensor data by performing primitive tasks such as object motion detection and recognition, as well as more high-level tasks such as tracking, behavioral analysis and situation awareness, in order to provide value added services to users, while also preventing the system operator from having to perform continuous supervision. As this technology matures, interesting research questions arise at the intersection of several disciplines from signal processing and Internet of things (IoT) to software engineering and artificial intelligence or even sociological studies.

In this paper we review existing surveillance solutions from a system perspective, being mainly concerned with architectural design in the context of large-scale surveillance networks that encompass a wide variety of sensing capabilities. Notably, scale and heterogeneity have a pronounced impact in increasing the system's complexity, making human operators unable to simultaneously review and interpret many streams of data. Hence, the system design, which represents the scope of this review, becomes of crucial importance in introducing an intelligence layer capable of automating the coordination and control of the surveillance infrastructure. The accompanying machine vision techniques employed in surveillance are out of scope for this paper and have already been extensively addressed in several literature reviews (i.e. Hu *et al.*, 2004; Brutzer *et al.*, 2011; Weinland *et al.*, 2011), so we do not involve with them deeply. In particular, we draw attention to Valera and Velastin (2005), which closely relates to our work and provides an overview of the evolving generations of surveillance systems. However, this study is limited to visual surveillance and is mainly focusing on image processing techniques such as object detection, recognition, tracking and behavior analysis. Whereas, this work provides a comprehensive survey, from a system perspective, on multi-modal sensing in surveillance applications. Additionally, we focus our survey on identifying the breadth of existing solutions, defining a taxonomy for classifying surveillance systems and identifying topics with high research potential. Moreover, the proposed framework is also intended to guide design decisions when engineering such systems. It is interesting to remark that the majority of works in this space are either solely concerned with, or evaluated for indoor environments. Commercial solutions are largely omitted from this survey due to the lack of basic details regarding the internal functioning of their systems.

The evolution of surveillance systems over time is considered to have gone through three major phases, according to Valera and Velastin (2005). Closed-circuit television cameras constitute the first generation of analogue surveillance systems. Essentially, they represent a set of cameras directly connected to a series of monitors under the supervision of a human operator, while no form of information processing is involved. Clearly, this proved to be an unsustainable approach, impracticable to scale due to its high dependency on the attention capacity of the human operator.

The second generation of surveillance systems marked the transition to digital video communication. This enabled reduced amounts of bandwidth and less storage space owing to digital compression, as well as a more robust transmission of the signal. Additionally, semi-automatic procedures for supporting the human user have been devised to generate alarms based on predefined event triggers. However, the inaccuracy of vision algorithms created a high number of false alarms. The systems deployed denote a centralized approach of processing data.

Current research efforts tend to focus on dealing with a heterogeneous mix of sensors, including, though not being confined to smart cameras, audio sensors, radio-frequency identification (RFID) tags, proximity sensors, smart cards or even three-dimensional (3D) sensing technologies (e.g. time-of-flight, stereo vision, structured light). The challenge for this third generation of surveillance systems falls on handling a large number of geographically distributed sensors, autonomously coordinating and distributing video communication, as well as processing capabilities and resources over the network. Moreover, while a massive amount of data are being generated by this sensing infrastructure, efficient ways of analyzing and interpreting it are imperative in order to provide high-quality services to end users. For instance, smart buildings provide an edifying example, where a wide range of data can be employed for deriving energy-efficiency recommendations that are both relevant for the user and actionable. This third generation of surveillance systems represent the focus of this work.

The rest of this paper is organized as follows. In Section 2 we review the key enabling technologies that provide significant enhancements for the third generation of surveillance systems. In Section 3 we propose a taxonomy that classifies surveillance systems based on their most distinctive characteristics. Next, Section 4 describes some of the most prominent applications to surveillance. In Section 5 we summarise the findings of our analysis. Section 6 identifies several gaps in the existing literature and discusses promising future research directions. Finally, conclusion is given in Section 7.

2 Enabling technologies

In this section, we point toward several key technologies that are driving the development of the third generation of surveillance systems. Namely, we are interested in technologies that can support engineering large-scale surveillance systems consisting of heterogeneous sensors and devices, and that can facilitate orchestrating the system components, such that the system can exhibit a goal-directed behavior and take appropriate actions according to its design objectives. It is out of scope here to give a comprehensive survey of each technology, but rather to point to representative examples that emphasize the role they could play in surveillance (Table 1).

2.1 Internet of things and sensor networks

The IoT paradigm has emerged as means for levering the growing ubiquity of sensor devices, primarily by using the Internet to enable data communication in such a way that these devices are seamlessly

Table 1 List of representative works for different strategies

Ref. No.	Data fusion	Petri nets	Bayesian information	Rule engine	Greedy optimization	BDI	Semantics	Context	DWT	Kalman	Plan
Park *et al.* (2017)								X			
Mihailescu *et al.* (2016)	X						X				
Brown *et al.* (2015)									X		
Fernández-De-Alba *et al.* (2015)								X			
Meinel *et al.* (2014)										X	
Bicocchi *et al.* (2014)								X			
Patti *et al.* (2014)				X							
Gascuena *et al.* (2014)							X	X			
Compton *et al.* (2012)							X				
Silva *et al.* (2012)		X									
Castanedo *et al.* (2008)	X					X					
Biswas *et al.* (2008)	X										
Monari *et al.* (2008)			X								
Soldatos *et al.* (2007)								X			
Town (2007)	X		X								
Pavon *et al.* (2007)											X
Jovanovic and Rinner (2007)				X							
Collins *et al.* (2001)	X				X						

BDI = belief–intentions–desires; DWT = discrete wavelet transform.

integrated into the information network. Although the wide spread of IoT technologies and sensor networks in general is still in its infancy and significant research efforts are still required in this area, there are clear overlaps and synergies with the domain of surveillance systems. Several of such instances are described hereafter.

VigilNet, proposed by Vicaire *et al.* (2009), is a large scale, outdoor surveillance system composed of tiny power-constrained sensors performing intruder detection tasks. The goal of the research of Vicaire *et al.* (2009) is to enable the sensor devices to collaboratively increase the system lifetime by employing different energy-efficient strategies. A so-called *tripwire service* is implemented, consisting of a sensor partitioning phase and a duty cycle scheduling. Results from an actual deployment conclude that the power management techniques proposed are able to increase the lifetime of the network from 4 to 200 days. However, the trade-off between network longevity and performance is an aspect that is currently prescribed during network configuration and needs to be carefully adapted depending on the conditions of the environment.

Li *et al.* (2011) describe an IoT architecture, where a multi-hop network of smart homes are organized into a so-called smart community. Distributed community environment surveillance is realized through their *neighborhood watch* application, based on a classic wireless *ad hoc* network. Given the static nature of this domain, the global network topology is supposed to be known *a priori* and loaded to each of the homes' gateways. Due to the fact that the interconnected homes use wireless communication standards, which are generally highly susceptible to security or reliability faults, the authors suggest some ways to alleviate these shortcomings. For example, detecting unreliable nodes, such as home gateways, could be achieved by having neighboring nodes monitor each other's retransmissions. Also, filtering for false network traffic may be essential in avoiding false data to be propagated in the community or inefficient bandwidth usage.

2.2 Multi-agent systems (MASs)

The allure of agent technologies for the design and control of surveillance systems can be attributed to several key MAS characteristics, which appear most adequate for this application domain. MAS allows to address challenges of autonomous and decentralized decision-making in a flexible manner, by decomposing complex tasks and assigning subproblems to loosely coupled components that interact and coordinate autonomously to solve system-level design goals (Woolridge, 2009). A significant part of the work surveyed in this paper applies some form of MAS implementation and below we present some aspects which are more particular to these systems.

Pavon *et al.* (2007) demonstrate the use of INGENIAS, an agent-based model-driven methodology to develop a multi-sensor surveillance system. Essentially, INGENIAS provides a modeling language to describe the system, based on five viewpoints from which a MAS can be regarded: organization, agent, goals-tasks, interactions and environment. The organization level deals with the definition of roles as functional responsibilities of agents. The agent behavior is described using so-called mental states for managing goals, which need to be mapped to system tasks. Workflows are used to define collaborative plans and specify the interactions between agents. Lastly, the environment description contains details about the data collected by sensors and the possible actions to be performed by actuators. A security guard usecase, where different kinds of alarms can be specified, is depicted using this framework.

A multi-agent architecture for visual sensor networks is presented by Castanedo *et al.* (2008). The authors adopt the well-known *belief–intentions–desires* (BDI) model (Rao & Georgeff, 1995) to represent the internal states of the agents. The solution is demonstrated in an indoor tracking application and is implemented using the JADEX toolkit (Braubach *et al.*, 2003). In this context, the agent's beliefs correspond to the information the agent acquires from the environment and the other agents. Desires express the motivation of the agent, in this case to perform tracking, while intentions are the basic steps chosen by the agent to achieve its desire. Additionally, a so-called fusion agent gathers track data from all the agents and determines a performance score for each agent based on the coherence with the rest of the data. This information is feedbacked to the agent, allowing it to use it as external information in order to reason about the confidence of its findings.

Similarly, Bustamante *et al.* (2014) used the BDI paradigm to model an active camera, agent-based surveillance system. The system proposes a three-layer architecture: a sensor management tier, which provides distributed active camera control and management, a data fusion tier for data processing and a user interface tier, where the operator can monitor the system. The focus lies here in using data fusion as an integral part of the architectural design. Regarding the BDI agent model, beliefs are used to represent information about the environment, desires express the goal of the system (e.g. tracking targets), while intentions are used to implement these goals (e.g. controlling the camera). For implementation purposes, the authors define a scenario-specific agent ontology. Implementation is carried out using the JADEX framework. This prototype is evaluated in an indoor tracking application with a limited number of visual sensors. However, the architecture cannot offer support for integrating heterogeneous sensors.

2.3 Service-oriented and specialized middleware

Middleware systems and service-oriented architectures (SOA) are gaining traction as various kinds of sensor networks and their applications increase in complexity. Middleware is designed to abstract hardware-specific issues and facilitate interoperability of various components, such as *ad hoc* communication, discovery of new devices or establishing new communication links. Essentially, this simplifies development at the application level and exempts developers from handling implementation at the lower layers. SOA have emerged as a popular approach in terms of providing a middleware with standard protocols, common interfaces and well-defined components.

Memory Jog, introduced by Soldatos *et al.* (2007), is a non-intrusive digital assistant for meetings, prototyped using an agent-based SOA middleware implemented in JADE (Bellifemine *et al.*, 2008). The goal of the service is to provide pertinent information facilitating humans to accomplish tasks during different types of events (i.e. meetings), that are occurring in a room setting equipped with a rich sensing infrastructure. Prototyping the service is done using a middleware designed to boost rapid application development. Whenever a new device is installed, the system instantiates a proxy agent for controlling this device. The room agent is the gateway for accessing all services available in that room. The procedure for a client to request a particular service is straightforward: the client sends a request to the gateway, which searches for a proxy for that service, which in turn finds an available service provider. Moreover, situation models can be defined, specifying the entities, properties and relations to be observed for triggering services. Situation models are expressed in XML format and parsed by a designated watching agent. Defining situation models can nevertheless be a tedious and error-prone task in case of more complex scenarios.

Aguilar-Ponce *et al.* (2007) present an architecture for automated scene surveillance using wireless visual sensor networks. Again, the multi-agent framework is used to model the system components, which can be classified into object processing units (OPUs), scene processing units (SPUs), region agents (RAs) and object agents (OAs). Several OPUs are grouped to form a cluster, where each one contains a cluster head, responsible for collecting the information coming from the other members of the cluster and sending it to the SPU for further analysis. RAs are associated to cameras, while OAs are instantiated once an object is detected. OAs and RAs reside at the OPU level. Communication between OPU and SPU occurs by passing through the network only the object parameters. Finally, once this information reaches the SPU, a decision is taken regarding whether the scene contains a threat or not.

Piette *et al.* (2015) propose a method for automatically deploying and configuring applications onto a heterogeneous hardware infrastructure, by taking into account the hardware properties and characteristics. In particular, the authors are focused on modeling ambient systems and ubiquitous applications. Both the hardware infrastructure and the application functionality are represented using graphs. A graph-matching algorithms determines the correspondence between the available hardware entities and the application requirements.

A middleware focused on dynamic reconfiguration in distributed camera networks is introduced by Jovanovic and Rinner (2007). Here, dynamic reconfiguration is interpreted in the sense of modifying the system functionality at run-time via exchanging software tasks between the system's components or altering the Quality of Service (QoS)-levels of tasks. Reconfiguration is achieved using policies, which represent in effect a set of rules, specified based on the Houdini language (Hull *et al.*, 2005) and triggered when changes in the internal or external state of the system occur. The authors argue that a multi-agent implementation can benefit from flexibility and scalability, but fail to flesh-out the details of this implementation, nor do they provide rigorous experimental results to back their claims.

The concept of a task-oriented architecture for collaborative video surveillance is presented by Monari *et al.* (2008). Although object tracking is the problem under investigation, the authors claim that the architecture is not application specific, but could also accommodate different tasks. On a similar note with previous works, a cluster-based structure is proposed. However, clusters are dynamically instantiated for each tracking task and not predetermined at design time. This approach has the advantage of overcoming the issues of object handover between sensors. Without going into extensive detail, the authors mention that data obtained from the cluster's sensors is aggregated using Bayesian fusion and that a cluster manager is used to receive tracking jobs from a human user. The solution was experimentally tested at the Fraunhofer institute using IP-cameras, yet an interesting challenge remains in adapting the architecture to make use of heterogeneous sensors.

2.4 Data fusion in distributed systems

Especially in the case of large-scale systems consisting of heterogeneous nodes, with limited or complementary sensing capabilities, collaboration among the sensor nodes becomes essential in order to improve the reliability of the analysis and to estimate correctly the state of the environment. In contrast, relying on a single node to receive and process all data is becoming increasingly impractical.

Snidaro *et al.* (2015) provide a comprehensive survey on the importance of context in information fusion processes. The authors distinguish different levels of context representation that can serve as input to the data fusion process. Namely, contextual information can take the form of static data sets (e.g. geographic information system databases), contextual variables from direct observations (e.g. weather conditions) or inferred relationships between entities and the environment. Additionally, context can act as a constraint during the data fusion process or it can bring new problem dimensions, that can influence the possible interpretations of a situation. Overall, this work emphasizes the lack of reference architectures for the next generation of data fusion systems, which can integrate context in a meaningful way.

Collaborative information processing is achieved by Biswas *et al.* (2008) using a mobile agent-based framework. The idea proposed is that of employing mobile agents that migrate from node to node following a certain itinerary, which could be predetermined or determined on-the-fly, to fuse data locally at each node. Mobile agents are used in an attempt to reduce communication cost and minimize energy consumption across the sensor network. Communication is carried out in a purely peer-to-peer fashion. An agent transfer from one node to another is initiated by the so-called source node. The agent code alongside its state is sent to the new location, where the agent is restored and resumes its activity. Notably, experiments are conducted evaluating the performance of this approach against the traditional client/server architecture. The comparison is focusing on energy consumption and execution time as the main metrics. The authors advocate their solution in the context of large-scale deployments of sensor networks, where the communication overhead increases significantly due to the amount of sensor nodes. Results show that under large-scale network conditions (i.e. large values of the following parameters: size of the network, raw data size and file access overhead) the mobile agent-based approach outperforms the client/server implementation. Namely, given the simulation setup in the paper, MAS performed better both in terms of execution time and energy consumption when the number of nodes in the network exceeded 16. Similarly, the same occurs when the ratio between message size and the mobile agent size exceeds 20. However, to be noted that mobile agent solutions may introduce additional issues such as replication faults, security problems or planning the agents' itinerary.

Cho *et al.* (2010) introduces a homogeneous architecture with equally powered visual sensor nodes. The goal of the surveillance system is to reason about the occupancy of certain areas. Additionally, a locality structure is used to decrease the computational complexity using clusters. Clustering is performed taking into account the topology of the camera network and more specifically, according to overlapping fields of view (FOVs). Data processing within the cluster entails that all data are collected to a sink node where computation occurs. This brings about the challenge of determining an efficient way to pass data such that the path from leafs to the root is minimized. The solution proposed by the authors to tackle this problem resembles a simplification of the distributed constraint optimization approach of Meisels (2007), whereby the topology graph is reduced to a spanning tree using a breadth first search and values are propagated from leafs to a root node. However, this only captures part of the problem because attaching many child nodes to one node can reduce the path, but at the same time may create bottlenecks in the case of resource limited nodes. The trade-off between path length and enabling parallelized processing needs to be further explored.

SPIRIT, introduced by Town (2007), is a surveillance system based on an ultrasonic location infrastructure to track people and devices in an office building, also relaying on information coming from cameras. A number of vision techniques used in this project include background and object appearance modeling, face detection, segmentation, and tracking. The system has been deployed at Cambridge University. Data fusion is realized via a so-called shared world model, which incorporates static information, such as positioning of sensors, as well as dynamic information, like the location of people or devices. Sensor readings are updated to the model such that it remains consistent and accurate. In order to do so, Bayesian networks are used to perform inference and newly derived hypotheses compete for observations with existing hypotheses. The disadvantage of this approach lies in the fact that people are supposed to carry on them a bat device which can be triggered by radio to emit an ultrasonic pulse, based on which the location of a person can be determined.

2.5 Image processing for computer vision

Computer vision is one of the most active research topics with direct application to surveillance systems. Processing video scenes involves a number of steps of increasing complexity, initiated by modeling the environment, detecting motion, followed by object segmentation, classification and tracking, while more advanced solutions have the capability to recognize activities and understand certain behaviors (Ko, 2008). As already mentioned, an in-depth analysis of computer vision techniques is outside the scope of this paper, thus we only describe briefly several notable elements and contributions of this area.

Although motion detection can be achieved in a number of ways, statistical methods, such as Gaussian mixture model (GMM) (Power & Schoonees, 2002) have the benefit of robustness to noise, shadows or changes in lighting conditions. GMM describes each pixel as a mixture of Gaussians distributions and classifies it as background or foreground. The limitation of this approach is its computational complexity, which can be alleviated by using specialized hardware (i.e. graphics processing unit, field programmable gate array).

Object classification can essentially be carried out using either a shape-based or a motion-based approach. The former uses some form of shape information (i.e. blobs, silhouettes) to distinguish objects, while the later identifies periodic properties of moving entities to categorize them. Tracking methods can also be divided into several categories: region-based tracking, active contour-based tracking, feature-based tracking, and model-based tracking. In practice, exploiting contextual information has proved a helpful approach in improving the accuracy, by constraining the object recognition and tracking problems in the context of a particular application (i.e. vehicles are constrained to roads as opposed to the sky). For an extensive survey on tracking methods we refer the reader to Yilmaz *et al.* (2006).

Finally, behavioral analysis is concerned with the generation of high-level descriptions of actions and interactions for the identified objects and represents an open research challenge. Current techniques range from template-based approaches, such as dynamic time warping (Veeraraghavan *et al.*, 2006), that try to match time-varying data to predefined reference patterns, to state-based methods. Hidden Markov models (Zhang *et al.*, 2006) is a state-based techniques which has successfully been used to recognize particular classes of activities and which generated a number of variations to the initial model (i.e. PHMM (Vogler & Metaxas, 1999), CHMM (Oliver *et al.*, 2000), LHMM (Oliver *et al.*, 2002)). Also various supervised machine learning techniques for classification have been applied for this task, however, in order to obtain a high statistical confidence, large amounts of training data are required. For instance, time-delay neural networks, which essentially append delay units to a general static network, have been applied for hand gesture recognition and lip reading. Other examples include the use of Bayesian networks and context-free grammars (Ryoo & Aggarwal, 2008) for describing human actions by recognize composite activities.

2.6 Semantic technologies

Ontology-based approaches are used to provide a shared conceptualization of a given domain of interest. In this manner, meaning can be encoded separately and decoupled from the implementation code or the data being streamed through the system. This allows a machine (but also human) interpretable model, encapsulating the relationships between the terms of the ontology. Because of this principle, ontologies can be further augmented and modified at execution time without the need of re-engineering the application code. Among the most prevalent technologies, we note the RDF (Lassila & Swick, 1999) and WSDL (Christensen *et al.*, 2001) formats, the SPARQL (Prud'hommeaux & Seaborne, 2006) query language and the OWL (W3C OWL Working Group, 2009) computational logic-based language, which enables inference capabilities.

In the surveillance domain, there are a number of works explicitly concerned with developing ontologies built on a specific use case. A forensic domain ontology is proposed by Sobhani *et al.* (2015) to support an automated visual surveillance system. The authors use a top-down approach to construct the ontology by describing activities, entities, relationships, resources and consequences. Notably, the ontology is built to capture outdoor riot scenarios, subject to the data provided by the London Metropolitan Police during the 2011 riots in United Kingdom. Another interesting example is the BOnSAI ontology introduced by Stavropoulos *et al.* (2012), which is specialized for incorporating ambient intelligence concepts for smart building scenarios. It provides the capability to model explicitly functionality, QoS, hardware, context, as well as the user. Hachem *et al.* (2011) propose an ontology designed to capture the heterogeneity in terms of hardware sensors and devices in the IoT. In addition, they include a representation of real-world objects by modeling them in the ontology as physical concepts. Similarly, the Semantic Sensor Network ontology by Compton *et al.* (2012) can describe sensors in terms of capabilities, measurement processes, observations and deployments. These ontologies could prove instrumental in terms of managing complex surveillance infrastructures with heterogeneous sensing equipment.

Mihailescu *et al.* (2016) propose an IoT system design, based on an internal information representation for sensor-enabled environments. New sensing tasks can be injected into the system, while a semantics-empowered search heuristic is responsible for dynamically assigning data sources for each task.

The solution consists of instantiating virtual sensors, which can exploit the existing sensing infrastructure in order to support new value-added services. Solution updates are generated on-the-fly in response to changes in the contextual data. Moreover, the system is demonstrated in a smart building application concerning activity monitoring. For the same task of activity monitoring in office spaces, Nguyen *et al.* (2014) propose a solution which uses ontological modeling, representation and reasoning. The ontology models activities, spacial contexts, artifactual contexts and the relations between them. Artifacts represents smart things in the environment. The system monitors the state of all artifacts in the environment and reasons with the ontology to recognize which activity is taking place. A time synchronization mechanism is applied in order to generate coordinated sensor readings. The system can handle multiple users, as well as areas.

2.7 *Self-organizing and self-adaptive mechanisms*

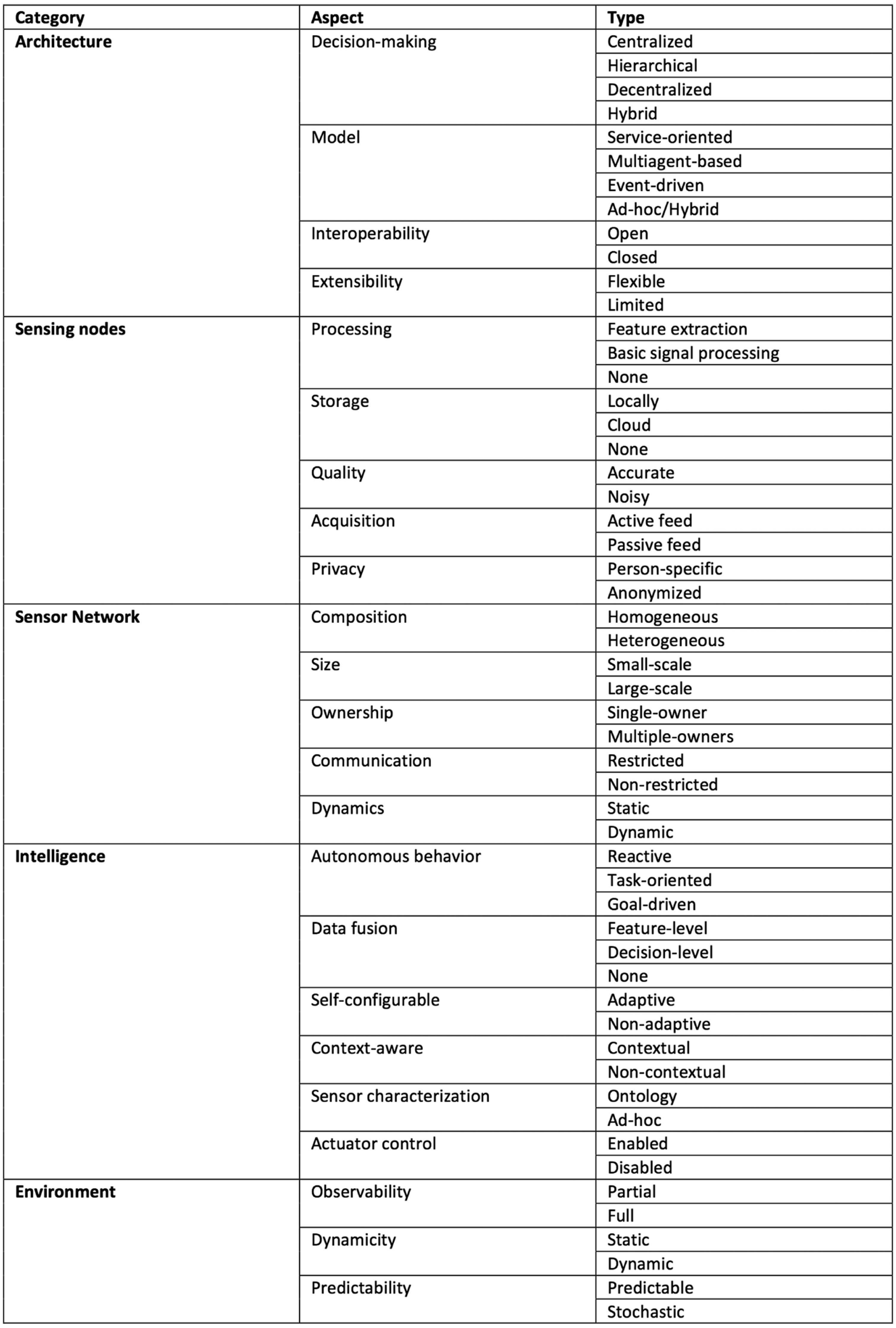
The goal of applying self-organizing processes to surveillance systems is that configuration aspects can be handled autonomously at run-time, such that the system can adapt as the situation demands. Much of the work on surveillance system falls under the assumption of having the network configuration already in place and thus do not deal with this problem. Hoffmann *et al.* (2008) are concerned with such an aspect, namely, spatially partitioning an area of interest given a set of cameras. They elaborate a self-organizing mechanism, where cameras communicate with their neighbors and independently calculate their FOV. Maintaining a low-communication traffic and configuring networks containing a larger number of cameras are also the objectives of this work. Essentially, cameras locally adjust their FOV and exchange state information with neighboring cameras in order to minimize FOV overlap and thereby maximizing surveillance coverage. To reduce communication overhead, only improvements over a certain threshold are considered. Yet, the authors do not account for the fact that small local changes could lead to major global improvements. This means that the solution could in fact be arbitrary far from optimal.

Self-adaptation is yet another important feature that can offer surveillance systems the flexibility to provide services in accordance to the situational characteristics of the monitored environment. Such an example is given by Byun and Park (2011), where a surveillance system is used for providing control and energy saving services in buildings. From an architecture standpoint, the system consists of self-adapting gateways and sensors. The gateway provides location-based services and can control a set of devices, while also having information about users, their associated devices and their profile. A user is recognized based on his MAC address and a list of manageable devices becomes available on the user's smartphone once he enters the room. Hence, specific services are provided to users under given situations, such as building energy monitoring and control services. A test bed has been developed and experiments were conducted showing a moderate reduction in energy consumption.

Also, nowadays, especially in the case of smart buildings, surveillance systems extend beyond a network that is owned and operated by a single stakeholder and instead consist of nodes which are property of multiple owners. In order to accommodate interoperability and heterogeneity between services of increasing diversity, service discovery is essential. Moreover, in a decentralized setting where there is no central repository responsible for the management of services, self-organizing mechanisms can play a central role. Zhu *et al.* (2005) survey a number of key service discovery protocols and point out the fact that service discovery design needs to address the challenge of operating in unfamiliar, dynamic contexts without trusting the environment. Thus, self-organization can be instrumental in establishing the structural relations between the system's components and adapting to changes in the environment in order to improve the system performance.

3 A taxonomy for surveillance systems

In the following, we propose to characterize surveillance systems along several dimensions: architectural design, sensor nodes, sensing network, intelligence layer and application environment. The features of our proposed taxonomy are depicted in Figure 1 and below we give a brief characterization of what we consider to be the most salient aspects.



Category	Aspect	Type
Architecture	Decision-making	Centralized
		Hierarchical
		Decentralized
		Hybrid
	Model	Service-oriented
		Multiagent-based
		Event-driven
		Ad-hoc/Hybrid
	Interoperability	Open
		Closed
	Extensibility	Flexible
		Limited
Sensing nodes	Processing	Feature extraction
		Basic signal processing
		None
	Storage	Locally
		Cloud
		None
	Quality	Accurate
		Noisy
	Acquisition	Active feed
		Passive feed
	Privacy	Person-specific
		Anonymized
Sensor Network	Composition	Homogeneous
		Heterogeneous
	Size	Small-scale
		Large-scale
	Ownership	Single-owner
		Multiple-owners
	Communication	Restricted
		Non-restricted
	Dynamics	Static
		Dynamic
Intelligence	Autonomous behavior	Reactive
		Task-oriented
		Goal-driven
	Data fusion	Feature-level
		Decision-level
		None
	Self-configurable	Adaptive
		Non-adaptive
	Context-aware	Contextual
		Non-contextual
	Sensor characterization	Ontology
		Ad-hoc
	Actuator control	Enabled
		Disabled
Environment	Observability	Partial
		Full
	Dynamicity	Static
		Dynamic
	Predictability	Predictable
		Stochastic

Figure 1 A taxonomy for intelligent surveillance systems (part 1)

3.1 System architecture

We identify in the literature several prominent architectural designs that are typically employed and which we propose to characterize over four main dimensions:

- Decision-making: Refers to the organization and interaction of the system's components required in order to carry out the system objectives.
  - In a *centralized* setting the information collected throughout the system is concentrated at a single processing point that carries out the analysis. The *Server/Client* design pattern belongs to the class of

  centralized architectures and is usually encountered in traditional surveillance systems, where resources such as computing power, memory and storage are concentrated at the server-side.
  - In order to minimize communication and computation overhead the architecture can aim to distribute the processing load between the nodes. Distributed architectures are typically organized in a *hierarchical* manner, having low-level functions carried out as close as possible to the sensing nodes, which are often, in practice, structured into clusters, while only the results of these computations are passed to the upper layers.
  - In a fully *decentralized* architecture, each node retains full controllability of its processes and a system-wide *negotiation* is conducted to realize the desired goals. Interaction between the systems' components takes place in a *peer-to-peer* fashion, where nodes pool their resources together in order to implement some functionality at the system level.
  - Various combinations of the previous types of design can pose interesting alternatives leading to *hybrid* architectures.
- Model: Refers to the framework that underlies the system's design.
  - SOA represent a good approach to achieve interoperability between heterogeneous devices, making use of services to abstract device-specific functionalities and communication capabilities. A midddleware is then required to enable basic service management such as service discovery and service composition.
  - Surveillance systems can be regarded as an area with high potential for the application of *multi-agent-based* technologies, which provide a natural way to model a complex, distributed system by means of breaking it down into loosely coupled, autonomous components that coordinate to fulfill high-level tasks.
  - Several works propose events as the building blocks for describing surveillance systems. For *event-driven* architectures, detection and classification of events are the key features, while more high-level interpretations (i.e. scene understanding) can be further derived.
  - In practice, often times *ad hoc* solutions are devised to solve particular tasks that do not provide a generic framework, extendable to other functionalities.
- Interoperability: Concerns the capability of the system's components to exchange data without or with very little knowledge about the unique characteristics of the other components.
  - Closed: This is a limitation present in the majority of existing surveillance systems, which makes appending and upgrading the infrastructure or functionality very difficult, often times requiring to re-engineering the system at large.
  - Open: Engineering open system essentially means enabling interaction with other components whose presence could not be determined at design time and regardless of the programming languages used in their implementation, however, abiding to certain requirements or protocols.
- Extensibility: Refers to the level of effort required to implement the extension.
  - Flexible: Allows the addition of new functionality, while minimizing the impact to existing system functions.
  - Limited: Hinders systematic reuse in implementing extensions.

3.2 Sensing nodes

Sensing nodes represent the building blocks of the sensing infrastructure and can be primarily distinguished based on their capabilities, as well as the type of data being manipulated in the system. We propose to characterize them over the following dimensions:

- Processing: Sensors can be classified depending on their on-chip capabilities in terms of the *processing* power of the device. We can differentiate between devices equipped for running more computationally expensive algorithms, others, where only basic signal processing is performed or even those that provide no processing capability. The former case represents an approach where algorithms, such as those applied in computer vision techniques, are aimed at extracting features or metadata from the incoming sensor stream. Just this information will then be passed on and used in more high-level tasks

(i.e. object recognition). This type of systems are also intended to perform data processing in *real-time*, making it critical that the device is apt to deliver results in a time-frame that remains meaningful for the objective at hand. Sometimes though, it is sufficient that only a simple preprocessing phase is carried out at the sensor level, which typically involves a (temporal) filtering step for smoothing the data or eliminating noisy readings.
- Storage: Regarding storage, data can either be *locally stored*, be it in the form of raw readings or some kind of metadata, or directly streamed to a *cloud* infrastructure. There is also the case when algorithms are directly applied to the stream of data, such that models are learned in real-time without having to actually store this data.
- Acquisition: The way that data are acquired can be thought of as *active feed*, when sensors adapt their observation strategies to capture certain aspects in the environment. For instance, sensors can adapt their sleep-cycles to conserve battery-life during periods that show reduced data variability. This can either be an individual, or a more complex collective process, where sensors coordinate their data-gathering actions. Otherwise, it is considered a *passive feed* process.
- Privacy: Surveillance systems typically manipulate sensitive information, therefore privacy-preservations needs to be a highly regarded feature. It is often the case that surveillance systems output person-specific information (i.e. MAC address, biometric data), which depending on the circumstances, may or may not need to be anonymized.
- Quality: Surveillance systems can also be differentiated based on the *accuracy* of the data provided. In order to bring down the cost of deployment, some systems consist of unreliable, low-cost sensors and thus, *noisy* data has to be tolerable to the system.

3.3 Sensor network

The sensor network exhibits differences along several lines:

- Composition: A surveillance system can be regarded as *homogeneous* if it consists of sensors of the same type, with similar capabilities (i.e. video-based surveillance), or *heterogeneous* if it represents a mix of sensors with different properties. A list of frequently used sensors includes various types of cameras, static, Pan Tilt Zoom (PTZ), omnidirectional, smart cameras and acoustic sensors, RFID tags, or smart cards.
- Size: Generally, encompassing more than thousands of nodes, classifies systems as *large-scale*, otherwise they are considered *small-scale*.
- Ownership: If all the sensor nodes in the network are owned by the same entity, the system is considered to have a *single owner*, and *multiple owners* otherwise.
- Communication protocols: Sensor networks are also distinguishable based on the type of communication employed. For instance, *Wi-Fi* data transmission, as well as *ZigBee* represent some of the most popular protocols seen in practice. Overall, communication in the network can be classified as *restricted*, in terms of distance, bandwidth, reliability and security, or *non-restricted*.
- Dynamics: We make the distinction between sensor networks, which only consist of sensors that remain *static* throughout the entire surveillance process and sensor networks that can be considered *dynamic*. The latter category may incorporate sensor which are mobile or sensors that dynamically appear or disappear in the network.

3.4 Intelligence

Intelligence, both at the system and component level, may become one of the most promising lines of research in surveillance systems and can be understood as the ability of the system to perceive the environment and exhibit not merely a reactive, but a goal-driven behavior that will affect what it can perceive in the future. We outline below several features, which in our acceptation are key toward increasing the level of intelligence of the system and where there also appears to be a lack of contributions.

- Autonomous behavior: Studying surveillance systems by behavioral means, we distinguish the following classes, that denote different degrees of complexity:

  - o Reactive: The system responds (i.e. via alerts) to changes that occur in the environment in order to satisfy its design objective. This type of systems display the lowest degree of complexity and do not involve reasoning about the environment, but only map sensory input to an action to be performed (i.e. rule-based).
  - o Task-oriented: A set of design tasks are concomitantly pursued, while potential conflict or resource constraints are mitigated based on an *a priori*-defined prioritization scheme. The steps required for delivering a certain task are also made explicit at design time.
  - o Goal-driven: The system exhibits a proactive behavior in order to satisfy its objective. This means that it is capable to elaborate at run-time a plan, which represents a sequence of available actions that can be executed in order to bring about the desired goal.
- Data-fusion: Considering a multi-sensory system, data fusion can be carried out at two levels: *feature level fusion* or *early fusion*—features extracted from the input data are first combined and then sent as input to a single analysis unit; *decision level fusion* or *late fusion*—decisions are obtained from interpreting features using local processing units and then these decisions are further combined to obtain a final decision.
- Self-configurable: Represents the capability of the system to invoke particular settings or properties over the sensing infrastructure. This can mean anything from altering communication patterns between nodes, to augmenting processing capabilities, to modifying the resolution of incoming data.
- Context-aware: More sophisticated surveillance systems can distinguish between different modes of operation according to the state of the environment. Especially in situations that require some kind of actuation (e.g. smart buildings), context conditions can trigger the action plan that is most appropriate, acting as precondition. Also, a context can be understood as a higher-level semantic interpretation of the environment, which can be derived from observations aggregated over a spatio-temporal horizon.
- Sensor characterization: Information representation has an important impact in the way data are processed throughout the system. Solutions that enable a higher level of flexibility in terms of resource management and data processing use ontologies in order to semantically characterize their constituent sensors and data flows running within the system. Otherwise, *ad hoc* solutions are devised for keeping track of this information.
- Actuator control: The system is able to perform some level of actuation based on reasoning over the input data, such as optimizing the operation of an elevator in a smart building.

3.5 Environment

Surveillance systems are also distinguishable according to the environment they are designed to operate in.

- Observability: Assesses the level of perceptibility of the environment.
  - o Full: Entails that sensor observations can fully capture the environment state without uncertainty.
  - o Partial: It represents situations in which there is uncertainty about the environment state given the sensors' observations. In practice most environments are regarded as partially observable.
- Dynamicity: Assesses the level of dynamism in the environment.
  - o Static: This type of environment can be assumed to remain unchanged, except for the movement of objects or persons that are the subject of surveillance. Operating under this assumption can significantly simplify the logic of the system. This means that a static reference model can be used to describe the area under surveillance and thus, the task of object detection can be reduced to background subtraction. As an example, deploying surveillance in indoor environments can benefit from a better controllability of the monitored area, as well as a lesser influence from external factors (i.e. weather).
  - o Dynamic: Such environments are characterized by high variability in scene conditions, requiring more elaborate methods to perform even basic tasks such as object recognition. Given that the background cannot be assumed to remain static, the system is required to continuously update the background model, to account for changes in the environment. Another viable solution is to perform temporal difference by subtracting consecutive frames. Outdoor settings are a typical case of dynamic environments, where algorithms need to handle more complex requirements. For instance, a common challenge in outdoor

visual surveillance is to cope with changes in lightning conditions. Also, the topology of the sensing infrastructure and even the architecture is generally influenced by this property of the system.

- Predictability: Assesses the level of uncertainty in the environment.
  - Predictable: An environment is predictable or deterministic when any action has a single guaranteed effect and there is no uncertainty about the state that will result from performing an action.
  - Stochastic: Otherwise, the environment is considered stochastic or non-deterministic.

4 Application areas

In the following section, we illustrate some of the application domains and related issues in surveillance that have been demonstrated to have a positive impact in improving services and overall, quality of live. They are by no means exhaustive, but only a very small part of the potential offered by surveillance systems in the future. In Figure 2 we list the applications reviewed in this section alongside possible design goals.

4.1 Tracking via task allocation

Tracking is an interesting application because it requires a combination of real-time tasks including motion detection, object recognition, as well as some coordination between sensor nodes. A tracking scenario with fixed and overlapping FOVs is considered by Nguyen *et al.* (2003). Given that each object is assumed to be visible by multiple cameras, the question becomes one of choosing the most appropriate camera for each object. The paper describes a centralized assignment algorithm to track people, where each camera is connected to a camera processing module running on a computer, while a central module maintains a database with all camera-object pairs. The assignment is based on a simple heuristic using the object-to-camera distance. Apart from being a centralized coordination solution and thus suffering from inherent computational and communication bottlenecks at the central server when the number of cameras increases, the approach also fails to base the assignment on other QoS parameters, or account for the case of having multiple objects tracked by the same camera.

Quaritsch *et al.* (2007) report the design of a multi-camera tracking system. They propose a decentralized MAS-based approach, where for each object being monitored, a single tracking instance is initiated in the multi-camera system, represented by a mobile agent migrating to the camera which observes the object. Migration is performed using a master/slave handover, having the master create a slave on the neighboring camera where the object is expected to appear. Once the object enters the field of view of the slave, the roles are interchanged, while the old master terminates thereafter. The implementation is done in Java, which introduces severe limitations in terms of the time required to create and launch the slave agent, thus making it impractical for actual usage. Replacing the code migration approach may alleviate this issue and is a prerequisite for real-world deployment.

In the work of Krahnstoever *et al.* (2008), tracking is performed by combining in a master/slave configuration fixed cameras with the versatility of PTZ cameras. The novelty consists in specifying a set of

Domain
Buildings
Homes/Domotics
Transport/Traffic
Sports
Military/Defense
Public Places/Crowds
Elderly/Health care

Purpose
Intrusion detection/Tracking
Comfort
(Energy) Efficiency
Resource allocation
Activity monitoring
Decision support
Safety/Security
Assisted Living
Facility management
Crisis management

Figure 2 A taxonomy for intelligent surveillance systems (part 2)

application specific quality measures, which are used to define the objective function of the system used to balance the number of captures per subject and quality of captures. The problem is cast to a combinatorial search solved using a heuristic approach.

4.2 Smart buildings

An important application domain for surveillance systems is provided by smart buildings. The growing demand for upgrading building management and automation has led to the deployment of various surveillance systems to this end. In the CMIPS project, Yong *et al.* (2007) focused on providing a real-time sensor-based building assessment tool. The infrastructure consists of a wireless sensor network logically divided into zones at design time. A gateway acts a service aggregator by enabling clients to access services provided by these embedded sensors, while masking the complexity of generating tasks for nodes. The core functionality is in assessing the energy consumption of a building. Additionally, a multi-agent model is proposed, where personal agents running on mobile devices are used to capture feedback from occupants in the building on their satisfaction level.

In the ISES project of Davidsson and Boman (2005, 2000), the objectives are both energy saving and increasing customer satisfaction through value added services. A MAS is proposed to oversee the control of the system, where room agents devise plans to optimize environmental conditions, based on the occupants' preferences and energy saving considerations. Thermodynamical characteristics of the rooms are captured in the model and several strategies based on the behavior of occupants are employed, such as pre-heating the rooms according to electronic diaries. Notably, the solution has been deployed in a real-world setting and the authors report up to 40% energy savings.

Patti *et al.* (2014) and Khajenasiri *et al.* (2014) have the goal of converting existing buildings, to the extent possible, into smart buildings by exploiting occupancy information coming from surveillance systems. Similarly to the previous work, they are concerned with energy efficiency. The paper proposes a three-layered SOA, with the main management functions executed via a rule engine. For instance, the logic of the HVAC control is prescribed by a rule, which is reevaluated once modifications in the sensor data occur.

From an organizational standpoint, Silva *et al.* (2012) adopt the concept of holons to represent the structure of control in their surveillance system applied to smart buildings. Holons introduce the notion of a hierarchical, nested structure, where groups of individual holons are composed into superior holons. The procedure is designed for smart buildings and a holon here can be either a physical device or a logical component. The paper describes a methodology for mapping the operation of a building onto a holocracy. Special attention is devoted to fault-tolerance requirements, by using Petri nets as a modeling support for performing formal tests on the behavior of the system.

4.3 Traffic

Transportation, whether it relates to vehicles or the road infrastructure itself, is becoming more infused with sensing devices in an attempt to increase efficiency and safety. Haesevoets *et al.* (2007) introduce a surveillance system design for monitoring traffic conditions on a highway. In terms of the architecture, a MAS overlays the camera network. Agents can communicate with neighboring agents and are differentiated into several roles. A data observer is a software agent that resides at each camera and monitors information about traffic density, intensity and average speed. Data pushers are agents responsible for forwarding the observed data to the data aggregator role. The system is capable to identify areas with traffic jams in a fully distributed way using simple message passing between agents.

Again, another interesting agent-based approach is proposed in the traffic domain, geared toward adapting existing surveillance systems along highways and streets toward semi-automatic surveillance systems, in the work of Rothkrantz (2013). The paper introduces a similar architecture implemented in JADE, consisting of three types of agents. Identification agents are used for handling request about objects detected by observer agents. This information is then transmitted to a reasoning agent, which tries to infer whether the object has shown a suspicious behavior based on a set of predefined suspicious events. If that happens to be the case, a human operator is alerted.

4.4 Movement assessment in sports

Evaluating the performance of athletes is yet another domain taking advantage from monitorization using video cameras or other sensing equipment such as wearable inertial sensors (i.e. gyroscopes, accelerometers). The goal of these systems can vary from monitoring the potential for injury, to studying team behavior and providing relevant analytics, to detecting fine-grained body movements. An exemplification of the latter is given by Ahmadi *et al.* (2015), where the authors propose a framework for classifying training activities in a real training environment, designed to feedback data to athletes that can lead to performance enhancement. The study is based on detecting an athlete's joint angular motion and impact accelerations using off-the-shelf classification techniques with up to 98% accuracy. The discrete wavelet transform is applied to extract discriminative features, while normative data are used to asses the athlete's deviations from correct movements. Additionally, the focus of the approach is on minimizing processing time and user involvement.

4.5 Domotics

Home automation, also known as domotics, represents the residential extension of smart building and has received considerable interest in recent years. Surveillance systems can harness the built-in infrastructure of smart homes to deliver increased safety and security. An exemplification in this context is presented by Carchiolo *et al.* (2010) and Longheu *et al.* (2012), advancing a solution based on the ZigBee protocol (Yi *et al.*, 2011) for detecting and classifying intrusions through video monitoring. The system consists of a supervisor, that can be implemented on a low-cost microcontroller and which manages devices (i.e. IP-cameras) through a coordinator module, that acts as an interface between the supervisor and the ZigBee network. A Gaussian background model is applied for detecting intruders, generating an alarm notification to the operator's smartphone. Real-time video streaming is also available remotely, allowing the operator to determine whether the alarm has been triggered correctly. Also, the alarm is customizable, enabling the operator to configure the parameters of the algorithm, this task requiring, however, specialized knowledge. With this consideration in mind, we hypothesize that an enhanced version of the system should incorporate the user feedback on false positive or negative triggers and automatically adjust parameters accordingly.

Other smart home approaches, such as the work of Chen *et al.* (2015), are favoring Wi-fi wireless technology for data transmission in comparison to the ZigBee standard due to its reliability and its popularity. This paper introduces a multi-sensor embedded home monitoring framework, including humidity, gas, smoke, and temperature sensors. A Dynamic Host Configuration Protocol sever is adapted to dynamically allocated a virtual IP, thus ensuring that all modules have a wireless link to the Internet. Although promoted as an intelligent system, its functionality appears to be confined to basic threshold detection, followed by SMS alarm pushing to an administrator. Some actuation is also available, such as automatically activating fans once the temperature falls into certain intervals.

4.6 Military applications and crisis management

Autonomous surveillance systems can further be designed to prevent, detect, and aid crisis management. Unmanned aerial vehicles (UAVs) are a good candidate to provide enhanced situational awareness in the event that emergency response is needed (i.e. earthquakes, floods, deforestation). The SIERRA project of Brown *et al.* (2015) is focused on forest fire applications and aims to allow real-time monitoring capabilities for fire-fighters. The project is currently in the prototyping phase. The desired functionality encompasses the capability to approximately encloses the fire area on a map, as well as to incorporate additional data about the environment (i.e. vegetation, moisture) in order to predict fire propagation under given conditions. Preliminary results describe several tests conducted in cooperation with emergency response teams and point out the advantages of UAV assistance.

4.7 Crowd scene surveillance

Crowd-level analysis is an active area of research in video surveillance. A large number of methods have been suggested especially for the problem of crowd counting, that is, estimating the number of people in a scene.

This information can be relevant in a wide range of scenarios, for instance, to adapt public transport capacity or to determine crowd density in retail stores. A recent case study is described by Yan *et al.* (2015) and stands out by deploying a low-cost, non-intrusive counting solution. Infrared distance sensors make this approach adequate for circumstances with privacy concerns. The architecture consists of wireless motes that collect data from the infrared sensors, filter it, determine the direction of the passing targets and eventually forward the result to a base station. However, the accuracy can be negatively impacted by situations when people are passing through the sensors side by side or when someone is standing right in front of the sensors.

4.8 Ambient assisted living

Demographic aging of the population is putting pressure on delivering monitorization solutions that can replace humans in providing one-to-one assistance for the elderly or disabled. Continuously monitoring the state of a person can become particularly important when the person under supervision is unable to actively solicit emergency intervention, such as pressing an alarm button. Unfortunately, these solutions are still prohibitive in terms of pricing. With the objective of increasing efficiency and reducing cost, Meinel *et al.* (2014) propose a solution for elderly care, that uses a single omnidirectional camera, which is able to cover a 360° view of a scene using a fisheye lens. The idea is to extract virtual perspective cameras from the complete omnidirectional image material for each person entering the scene and use them for tracking during their presence in the observed area. The algorithm detailed performs a combination of blob extraction for detecting people and Kalman filter for tracking, showing that processing can be attained in reasonable time for real-time scenarios.

In a similar manner, Findeisen *et al.* (2015) deploy an omnidirectional stereo sensor for human behavior analysis in complex health-care environments. The geometrical configuration and positioning of their cameras is designed such that in can infer depth data. Acquiring this type information from the environment proves to be instrumental in improving the performance of human behavior analysis algorithms. Several related approaches have also used other 3D sensing technologies for indoor monitoring, such as Kinect[1], although this comes with a lower resolution and a narrower field of view.

5 Summarizing remarks

To summarize, the majority of solutions proposed in the surveillance literature appear to be focused, whether explicitly or not, on indoor settings, where simplifying assumptions can be made regarding the controllability, observability, nature and dynamics of the environment. Hence, in Table 2 we review several representative surveillance systems that are explicitly tailored for *smart building* environments and have been successfully deployed in field studies, in order to provide a comparison of the surveyed approaches against our proposed taxonomy. Based on this review, we detail in Section 6 open research challenges that have been identified.

Interestingly, hierarchical architectures are a dominant way of structuring the system, whether this takes the form of cluster-based, multi-tier or group-based solutions. Also, the system generally consists of three layers, starting with lower-level computations performed by sensor nodes, a middleware coordinating the sensor nodes and a higher-level, where more complex reasoning is conducted. MASs, as well as service-oriented approaches are a popular manner of modeling the system, such that the available resources are efficiently utilized and controlled, to perform the desired surveillance task (see Figure 3). Communication between the systems' components is handled, however, through very diverse modalities. Of note, few works consider that a surveillance system may have multiple owners, where sensors are owned by different stakeholders. Nevertheless, there is a wide range of types of sensors being used over the studied use cases (see Figure 3).

Regarding the intelligence of the system, data fusion represents an important feature employed to increase the accuracy of the knowledge inferred from the data streams. Additionally, some basic attempts

[1] `http://www.xbox.com/en-US/xbox-one/accessories/kinect`

Table 2 Summary of surveillance systems for smart buildings against the taxonomy

	Architecture				Sensing nodes					Network					Environment		
Ref. No.	Decision	Model	Interoperability	Extensibility	Processing	Storage	Quality[a]	Acquisition	Privacy	Composition	Size	Ownership	Communication protocols	Dynamics[b]	Observation[c]	Dynamics	Predictability[d]
Park *et al*. (2017)	Hierarchical	Event driven	Open	Flexible	Data filter	Local	A	Passive	**X**	RPI-cameras, Web-camera, ambient sensors	11	Single	LAN	S	P	S	S
Mihailescu *et al*. (2016)	Decentralized	SOA	Open	Flexible	Feature extraction	Cloud	A	Active	**X**	Smart cameras, ambient sensors	5	Single	Wi-Fi	D	P	D	S
Chen *et al*. (2015)	Centralized	SOA	Open	Flexible	None	Cloud	A	Passive	**X**	Temperature, gas detection, humidity	1	Single	Wi-Fi	S	F	S	P
Bustamante *et al*. (2014)	Decentralized	MAS	Open	Flexible	Feature extraction	Local	A	Active	**X**	PTZ-cameras	3	Single	LAN	D	F	S	P
Patti *et al*. (2014), Khajenasiri *et al*. (2014)	Hierarchical 3-layer	Event driven	Open	Flexible	None	Cloud	A	Passive	**X**	Temperature, motion, brightness, power consumption	2	Single	LinkSmart	S	F	S	P
Silva *et al*. (2012)	Hierarchical	SOA, MAS	Closed	Limited	None	None	A	Passive	**X**	Temperature, motion, video	N/A	Single	FIPA	S	F	S	P
Carchiolo *et al*. (2010), Longheu *et al*. (2012)	Centralized	*Ad hoc*	Closed	Limited	Feature extraction	Cloud	A	Active	**X**	IP-camera	1	Single	ZigBee	S	P	D	S
Byun and Park (2011)	Hierarchical 3-tier	SOA	Closed	Flexible	Feature extraction	Local	A	Passive	**X**	Temperature., humidity, motion, light	9	Single	ZigBee	S	F	S	P
Monari *et al*. (2008)	Hierarchical	MAS	Open	Flexible	Feature extraction	Local	A	Active	**X**	IP-cameras	25	Single	FIPA	D	P	S	S
Castanedo *et al*. (2008)	Hierarchical	MAS	Open	Flexible	Feature extraction	Local	N	Active	**X**	Cameras	3	Single	FIPA	D	F	S	P
Town (2007)	Centralized	Event driven	Open	Flexible	Feature extraction	Cloud	N	Active	**X**	Webcams, ultrasonic	3	Single	LAN	S	P	D	S
Soldatos *et al*. (2007)	Hierarchical	SOA, MAS	Open	Flexible	None	Cloud	A	Active	**X**	PTZ, fish-eye, fixed cameras	6	Single	FIPA	D	F	S	P
Yong *et al*. (2007)	Hierarchical	MAS	Open	Flexible	Signal processessing	Cloud	A	Passive	**X**	Temperature, humidity, noise, light, air quality	9	Single	LAN	S	F	S	P
Davidsson and Boman (2000, 2005)	Hierarchical	MAS	Open	Flexible	Signal processessing	Cloud	A	Passive	**X**	Temperature, light	27	Single	LonWorks	D	F	S	P

PTZ = Pan Tilt Zoom; SOA = service-oriented architectures; MAS = multi-agent systems; RPI-Camera = Raspberry Pi Camera; IP Camera = Internet Protocol Camera.

[a] A = accurate; N = noisy.

[b] S = static; D = dynamic.

[c] P = partial; F = full.

[d] P = predictable; S = stochastic.

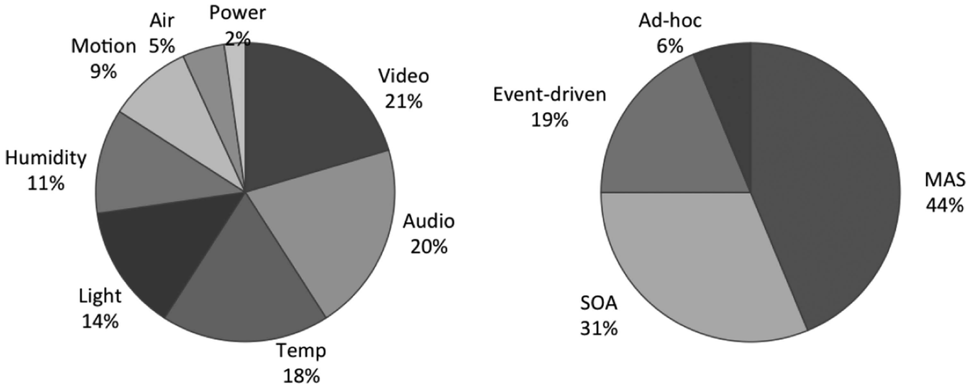


Figure 3 The distribution of (a) the type of data and (b) the modeling approach used in surveillance systems in smart building scenarios

aim at acquiring contextual information and utilizing it to provide services that are user-centric and situation-specific. Active sensing strategies are yet another key line of research that can enable sensor nodes to organize themselves such that, collectively, higher-level sensing tasks can be carried out. A summary of the techniques employed across the different application domains surveyed in this paper are outlined in Table 3. Notably, most of the works reviewed are geared toward small-scale settings, rarely exceeding tens of sensing nodes, while different types of sensors are often difficult to integrate to the existing frameworks. Especially, there are few instances where mobile sensors are considered as part of the system. Moreover, there is a clear lack of work addressing resilience or privacy-related issues. Therefore, it is important that new developments in surveillance ensure that sensor networks can easily be appended to make use of new types of senor data, while at the same time, are able to handle dynamic settings where senors come and go or could experience certain failures. The same dynamic aspect is key when attempting to reuse learned data models for prediction in a new setting (e.g. a new office space), since its performance can be greatly degraded. Such recalibration of data models should ideally minimize the amount of new labeled data needed, which is typically very costly to collect. Overall, surveillance systems tend to be designed for very specific tasks and lack the capability of extending the solution to more generic problems.

6 Open research challenges

In spite of the fact that the enabling technologies detailed in Section 2 are expected to play a pivotal role in the development of the next generation of surveillance systems, a large research effort is still required. In this section we point toward several research challenges that need to be addressed in order to realize to a greater extent the potential of such systems and their applicability to more general scenarios.

6.1 Situational and context awareness

Automatically acquiring context models from data is an important open research issue in the realm of surveillance systems. This capability can enable systems to operate in a more flexible and adaptable manner and switch between different modes in response to environment dynamics or user requests.

An effort in this direction is proposed by Fernández-De-Alba *et al.* (2015), through the FAERIE architecture, which provides a distributed approach based on a federated blackboard model. The premise of this work is to enable flexible access to information at any level of abstraction. The underpinning of their solution is the introduction of three types of components: context elements, context containers, and context observers. Context containers implement a distributed virtual blackboard model for data sharing (context elements) among the context observer components, which are responsible for manipulating the data. However, this approach does not currently support real-time processing capabilities.

Table 3 Summary of surveillance systems for smart buildings against the taxonomy (continued)

	Intelligence					
Ref. No.	Autonomous behavior	Data fusion	Self-configuration	Context awareness	Sensor characterization	Actuator control
Park *et al*. (2017)	Task-oriented	Feature-level	Non-adaptive	Non-contextual	*Ad hoc*	Enabled
Mihailescu *et al*. (2016)	Goal-oriented	Feature-level	Adaptive	Non-contextual	Ontologies	Enabled
Chen *et al*. (2015)	Task-oriented	None	Non-adaptive	Non-contextual	*Ad hoc*	Disabled
Bustamante *et al*. (2014)	Goal-oriented	Decision-level	Adaptive	Non-contextual	Ontologies	Enabled
Patti *et al*. (2014), Khajenasiri *et al*. (2014)	Reactive	None	Non-adaptive	Contextual	*Ad hoc*	Enabled
Silva *et al*. (2012)	Task-oriented	None	Adaptive	Non-contextual	*Ad hoc*	Enabled
Carchiolo *et al*. (2010), Longheu *et al*. (2012)	Reactive	Feature-level	Non-adaptive	Non-contextual	*Ad hoc*	Disabled
Byun and Park (2011)	Task-oriented	None	Adaptive	Contextual	*Ad hoc*	Enabled
Monari *et al*. (2008)	Task-oriented	Decision-level	Adaptive	Non-contextual	*Ad hoc*	Disabled
Castanedo *et al*. (2008)	Goal-oriented	Feature-level	Adaptive	Non-contextual	*Ad hoc*	Disabled
Town (2007)	Task-oriented	Feature-level	Non-adaptive	Contextual	*Ad hoc*	Disabled
Soldatos *et al*. (2007)	Task-oriented	Feature-level	Adaptive	Contextual	*Ad hoc*	Enabled
Yong *et al*. (2007)	Goal-driven	None	Non-adaptive	Non-contextual	*Ad hoc*	Enabled
Davidsson and Boman (2000, 2005)	Task-oriented	None	Adaptive	Non-contextual	*Ad hoc*	Enabled

Bicocchi *et al.* (2014) propose a starting point for what is termed an awareness framework for knowledge collection. They suggest to use general-purpose algorithms to classify data streams. The architecture consists of three layers, a sensor layer for collecting data, a classifier layer consuming data from the previous layer and an awareness layer based on labels produced during classification. The latter is considered to be central for designing a self-awareness module. Labels are then used to generate a state-based automata which encapsulates a particular context. The solution is demonstrated by means of a life logging application and in the current form, contexts are limited to situations that can involve only one individual.

6.2 Privacy preservation and security

Addressing privacy-related issues is paramount for a mass adoption of surveillance technologies. In spite of this, the trade-offs between privacy preservation and security are consistently overlooked. The literature tackling these problems is scattered and contributions in this area are particularly timely.

Yu *et al.* (2015) argue that traditional host-centric security solutions are fundamentally infeasible for today’s heterogeneous, large-scale networks of sensing devices. The authors propose to rethink network security in order to incorporate context awareness, learning capabilities that can distinguish between normal states of the network and attacks, as well as introducing more dynamic security policies. In addition, a roadmap for tackling these challenges is put forward.

The work of Winkler and Rinner (2014) presents a survey of privacy and security threats and challenges associated with visual sensor networks. The authors classify security aspects into data-centric, node-centric, network-centric and user-centric, and discuss the individual security requirements for each class. Moreover, they identify trends in privacy protection techniques for image data such as blanking, obfuscating, scrambling, abstraction and encryption. However, the conclusion is that a direct use of security techniques developed for related domains is not a straightforward task and a lot of research issues are still open.

6.3 Scalability

A key challenge for current surveillance systems is the capability of deployment in large-scale settings. We describe hereafter several notable efforts in this regard.

Velipasalar *et al.* (2006) are concerned with the communication aspects in distributed peer-to-peer camera systems, with special emphasis on reducing the number of messages, as well as the message size. The protocol proposed is designed for object tracking and its applicability in another context remains unclear. One of the desired properties of the system is fault tolerance. Thus, according to this approach, each camera keeps its own track of the objects detected, ensuring redundancy and also a reduced requirement for communication. Collaboration between cameras is constrained to maintaining consistent object labeling. Interestingly, the solution captures the situation when objects merge or split, such is the case of a person entering a car. Messages exchanged between camera are classified as either requests or replies and are processed in a non-blocking fashion avoiding deadlocks or process starvation. Synchronization can, however, become a problem, handled by introducing synchronization points, based on a certain synchronization rate.

A lightweight architecture for online video processing is introduced by Camplani and Salgado (2011), fully developed in C++ based on the OpenCV libraries. Scalability is addressed in the sense of having one processing unit controlling a group of cameras, while a central unit oversees the entire process. Moreover, at the level of each processing unit a parallelization scheme ensures that the operations on the images are pipelined. The pipeline is built as a cascade of stages, which are application specific and thus need to be represented through a task processing design flow. A case study for object detection is presented and although the architecture is suggested for mass deployment, the testbed consists of only three cameras. Results show some improvement in terms of processing speed.

To tackle the challenges of large-scale deployments, Park *et al.* (2017) propose an automated surveillance system using edge computing. The architecture introduces an inbetween layer, termed cloud of things (CoT), which has the role of interfacing between the data sources and the server. The CoT, which is the core contribution of this work, specifies three types of modules. The manager module is responsible for keeping track and monitoring the alive status of sensors and actuators. The analysis module performs basic processing and filtering of raw data, as well as invoking actuators in real time, thus improving the reaction time between edge and sensors. Finally, the software manager module receives analysis routines from the server, which need to be performed locally.

6.4 Integration

In spite of its prevalence, most of the existing literature on surveillance systems appears to be confined to application-specific settings, often addressing a very limiting aspect posed by a particular scenario (i.e. using ultrasonic infrastructure to track people), but failing to present them in the context of a broader system architecture. In order to be able to move away from experimental, small-scale test-bed cases and toward large-scale end-to-end deployments, it is critical that systems integration aspects are carefully considered.

An example of a more integrative approach developed at Carnegie Mellon University is given by Collins *et al.* (2001). The surveillance system is capable to perform a wide range of functions ranging from movement detection, object tracking to semantic classification and human motion analysis. Additionally, different techniques can be seamlessly and alternatively used to perform these functions. The authors correctly point out that in order to coherently interpret sensor data, all observations must be brought into a common frame of reference, which here is represented by the geodesic coordinates. Also, camera calibration is performed using this common coordinate system. Control is centralized in the sense that sensor processing units send a continual stream of time-stamped symbolic object observations to an operator control unit. Thus, a new observation can be inspected whether it corresponds to an object already known. Moreover, since sensors may perceive objects from different viewpoints (i.e. cameras), the observation features should mostly be insensitive to viewpoints (i.e. geolocation). The system operator is capable to task the system in a very specific way by creating triggers (i.e. 'alert when vehicle enters the east parking lot'). This means that high-level user requirements need to be transposed into a sequence of sensor commands. To achieve this, the authors propose a matrix of cost values, representing the costs of sensors performing tasks. Task allocation is carried out using a greedy optimization approach. Prioritizing tasks is used for conflict resolution. Finally, a simple Web-based reporting system is logging all data that are generated.

The so-called INT3-SDP process introduced by Gascuena *et al.* (2014) deals with multi-sensory monitoring and interpretation of behaviors and situations for intervention in complex and dynamic environments. The process provides, on the one hand, a set of guidelines to support system analysts to capture the customer's requirements in term of multi-sensory monitoring. On the other hand, a components repository is proposed to aid the implementation phase. These components can be reused and easily adapted to different configurations. Moreover, a proof-of-concept use case depicts the application of INT3-SDP for the task of fall detection in a smart home environment.

7 Conclusions

The paper has surveyed the state of development in surveillance systems, an area that is rapidly gaining ground with the pervasive presence of sensing infrastructure around us. With the increasing cost-efficiency, wide-spread deployment in various setting is becoming feasible, but also surfaces many challenges that still need to be addressed. In this regard, we introduced a taxonomy that classifies these systems based on their dominant features, also leading to a distinction between different classes of problems, many of which require more research effort.

Concretely, further work is necessary to distribute the processing load and communication flow that is traditionally carried out by centralized systems, while also addressing the accompanying issues of data fusion and knowledge acquisition. Incorporating an intelligence layer that can enable adaptiveness, flexibility and autonomous configurability to the system when confronted with rapidly changing, dynamic environments common to real-world problems is paramount. Research on context-awareness and more generally, the self-awareness of the system can offer important contributions. Furthermore, few of the existing work extend beyond restrictive problem domains and could integrate a variety of approaches in the context of a broader architecture. Moving in the direction of a heterogeneous sensing infrastructure raises issues of interoperability and coordination between the system's components. This requires building a higher degree of autonomy into the constituent parts of the systems, as well as decision-making capabilities. Moreover, the implementation of open system frameworks incurs potential security threats, while scalability is propagating those risks far more widely than in the case of current applications. Not least, the adoption of intelligent surveillance technologies should not be based on the premise of increasing security at the expense of privacy loss and hence increasing the risk of misuse and abuse of surveillance data.

Acknowledgment

This work partially supported by the Knowledge Foundation through the IoT and People research profile.

References

Aguilar-Ponce, R., Kumar, A., Tecpanecatl-Xihuitl, J. L. & Bayoumi, M. 2007. A network of sensor-based framework for automated visual surveillance. *Journal of Network and Computer Applications* **30**(3), 1244–1271.

Ahmadi, A., Mitchell, E., Richter, C., Destelle, F., Gowing, M., O'Connor, N. & Moran, K. 2015. Toward automatic activity classification and movement assessment during a sports training session. *Internet of Things Journal, IEEE* **2**(1), 23–32.

Bellifemine, F., Caire, G., Poggi, A. & Rimassa, G. 2008. JADE: a software framework for developing multi-agent applications. Lessons learned. *Information and Software Technology* **50**(12), 10–21.

Bicocchi, N., Fontana, D. & Zambonelli, F. 2014. A self-aware, reconfigurable architecture for context awareness. In *IEEE Symposium on Computers and Communication (ISCC)*, 1–7.

Biswas, P. K., Qi, H. & Xu, Y. 2008. Mobile-agent-based collaborative sensor fusion. *Information Fusion* **9**(3), 399–411.

Braubach, L., Lamersdorf, W. & Pokahr, A. 2003. Jadex: implementing a BDI-infrastructure for JADE agents. *Search of Innovation (Special Issue on JADE)* **3**, 76–85.

Brown, B., Wei, W., Ozburn, R., Kumar, M. & Cohen, K. 2015. Surveillance for intelligent emergency response robotic aircraft (SIERRA)- VTOL aircraft for emergency response. In *AIAA Infotech*. American Institute of Aeronautics and Astronautics.

Brutzer, S., Hoferlin, B. & Heidemann, G. 2011. Evaluation of background subtraction techniques for video surveillance. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1937–1944.

Bustamante, A. L., Molina, J. M. & Patricio, M. A. 2014. A practical approach for active camera coordination based on a fusion-driven multi-agent system. *International Journal of Systems Science* **45**(4), 741–755.

Byun, J. S. & Park, S. 2011. Development of a self-adapting intelligent system for building energy saving and context-aware smart services. *IEEE Transactions on Consumer Electronics* **57**(1), 90–98.

Camplani, M. & Salgado, L. 2011. Scalable software architecture for on-line multi-camera video processing. In *Proceedings of SPIE*, **7871**, 6–21.

Carchiolo, V., Longheu, A., Malgeri, M., Mangioni, G. & Torrisi, D. 2010. Pervasive home security: an intelligent domotics application. In *Intelligent Distributed Computing IV*, Studies in Computational Intelligence **315**, M. Essaaidi, M. Malgeri & C. Badica (eds). Springer Berlin Heidelberg, 155–164.

Castanedo, F., Garcia, J., Patricio, M. & Molina, J. 2008. A multi-agent architecture to support active fusion in a visual sensor network. In *Second ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC)*, 1–8.

Chen, S.-L., Chang, S.-K. & Chen, Y.-Y. 2015. Development of a multisensor embedded intelligent home environment monitoring system based on digital signal processor and Wi-Fi. *International Journal of Distributed Sensor Networks* **2015**, 11.

Cho, Y., Lim, S. O. & Yang, H. S. 2010. Collaborative occupancy reasoning in visual sensor network for scalable smart video surveillance. *IEEE Transactions on Consumer Electronics* **56**(3), 1997–2003.

Christensen, E., Curbera, F., Meredith, G. & Weerawarana, S. 2001. *Web Service Definition Language (WSDL).* Technical report, `http://www.w3.org/TR/wsdl`.

Collins, R., Lipton, A., Fujiyoshi, H. & Kanade, T. 2001. Algorithms for cooperative multisensor surveillance. *Proceedings of the IEEE* **89**(10), 1456–1477.

Compton, M., Barnaghi, P., Bermudez, L., Garca-Castro, R., Corcho, O., Cox, S., Graybeal, J., Hauswirth, M., Henson, C., Herzog, A., Huang, V., Janowicz, K., Kelsey, W. D., Phuoc, D. L., Lefort, L., Leggieri, M., Neuhaus, H., Nikolov, A., Page, K., Passant, A., Sheth, A. & Taylor, K. 2012. The SSN ontology of the W3C semantic sensor network incubator group. *Web Semantics: Science, Services and Agents on the World Wide Web* **17**, 25–32.

Davidsson, P. & Boman, M. 2000. Saving energy and providing value added services in intelligent buildings: a MAS approach. In *Agent Systems, Mobile Agents, and Applications*, Lecture Notes in Computer Science **1882**, D. Kotz & F. Mattern (eds). Springer Berlin Heidelberg, 166–177.

Davidsson, P. & Boman, M. 2005. Distributed monitoring and control of office buildings by embedded agents. *Information Sciences* **171**(4), 293–307.

Fernández-De-Alba, J. M., Fuentes-Fernández, R. & Pavón, J. 2015. Architecture for management and fusion of context information. *Information Fusion* **21**, 100–113.

Findeisen, M., Meinel, L., Richter, J. & Hirtz, G. 2015. An omnidirectional stereo sensor for human behavior analysis in complex indoor environments. In *2015 IEEE International Conference on Consumer Electronics (ICCE)*, 17–19.

Gascuena, J. M., Castillo, J. C., Navarro, E. & Fernandez-Caballero, A. 2014. Engineering the development of systems for multisensory monitoring and activity interpretation. *International Journal of Systems Science* **45**(4), 728–740.

Hachem, S., Teixeira, T. & Issarny, V. 2011. Ontologies for the internet of things. In *Proceedings of the 8th Middleware Doctoral Symposium MDS '11*, 3:1–3:6. ACM.

Haesevoets, R., Van Eylen, B., Weyns, D., Helleboogh, A., Holvoet, T. & Joosen, W. 2007. Context-driven dynamic organizations applied to coordinated monitoring of traffic jams. In *Proceedings of Workshop Engineering Environment-Mediated Multiagent Systems*, 126–143.

Hoffmann, M., Wittke, M., Hahner, J. & Muller-Schloer, C. 2008. Spatial partitioning in self-organizing smart camera systems. *IEEE Journal of Selected Topics in Signal Processing* **2**(4), 480–492.

Hu, W., Tan, T., Wang, L. & Maybank, S. 2004. A survey on visual surveillance of object motion and behaviors. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews* **34**(3), 334–352.

Hull, R., Kumar, B., Lieuwen, D., Patel-Schneider, P. F., Sahuguet, A., Varadarajan, S. & Vyas, A. 2005. Improving user experience through rule-based service customization. *International Journal of Cooperative Information Systems* **14**(4), 469–502.

Jovanovic, M. & Rinner, B. 2007. Middleware for dynamic reconfiguration in distributed camera systems. In *Fifth Workshop on Intelligent Solutions in Embedded Systems*, 139–150.

Khajenasiri, I., Patti, E., Jahn, M., Acquaviva, A., Verhelst, M., Macii, E. & Gielen, G. 2014. Design and implementation of a multi-standard event-driven energy management system for smart buildings. In *IEEE 3rd Global Conference on Consumer Electronics (GCCE)*, 20–21.

Ko, T. 2008. A survey on behavior analysis in video surveillance for homeland security applications. In *37th IEEE Applied Imagery Pattern Recognition Workshop*, 1–8.

Krahnstoever, N., Yu, T., nam Lim, S., Patwardhan, K. & Tu, P. 2008. Collaborative real-time control of active cameras in large scale surveillance systems. In *Proceedings Workshop on Multicamera and Multi-modal Sensor Fusion Algorithms and Applications*, **26**, 1–12.

Lassila, O. & Swick, R. R. 1999. *Resource Description Framework (RDF) Model and Syntax Specification*. Technical report, W3C. `ttp://www.w3.org/TR/REC-rdf-syntax/`.

Li, X., Lu, R., Liang, X., Shen, X., Chen, J. & Lin, X. 2011. Smart community: an internet of things application. *IEEE Communications Magazine* **49**(11), 68–75.

Longheu, A., Carchiolo, V., Malgeri, M., Mangioni, G., Longheu, R., Carchiolo, V. & Malgeri, M., I, V. A. D. 2012. An intelligent and pervasive surveillance system for home security. *International Journal Computer Communications* **7**(2), 312–324.

Meinel, L., Findeisen, M., Hes, M., Apitzsch, A. & Hirtz, G. 2014. Automated real-time surveillance for ambient assisted living using an omnidirectional camera. In *IEEE International Conference on Consumer Electronics (ICCE)*, 396–399.

Meisels, A. 2007. *Distributed Search by Constrained Agents: Algorithms, Performance, Communication*. Advanced Information and Knowledge Processing. Springer.

Mihailescu, R., Persson, J. A., Davidsson, P. & Eklund, U. 2016. Towards collaborative sensing using dynamic intelligent virtual sensors. In *Intelligent Distributed Computing X – Proceedings of the 10th International Symposium on Intelligent Distributed Computing – IDC 2016*, 217–226. 10–12 October.

Monari, E., Voth, S. & Kroschel, K. 2008. An object- and task-oriented architecture for automated video surveillance in distributed sensor networks. In *IEEE Fifth International Conference on Advanced Video and Signal Based Surveillance (AVSS '08)*, 339–346.

Nilsson, F. 2008. *Intelligent Network Video: Understanding Modern Video Surveillance Systems*. CRC Press.

Nguyen, N. T., Venkatesh, S., West, G. & Bui, H. H. 2003. Multiple camera coordination in a surveillance system. *ACTA Automatica Sinica* **29**, 408–422.

Nguyen, T. A., Raspitzu, A. & Aiello, M. 2014. Ontology-based office activity recognition with applications for energy savings. *Journal of Ambient Intelligence and Humanized Computing* **5**(5), 667–681.

Oliver, N., Horvitz, E. & Garg, A. 2002. Layered representations for human activity recognition. In *Proceedings Fourth IEEE International Conference on Multimodal Interfaces*, 3–8.

Oliver, N., Rosario, B. & Pentland, A. 2000. A Bayesian computer vision system for modeling human interactions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(8), 831–843.

Park, H. D., Min, O.-G. & Lee, Y.-J. 2017. Scalable architecture for an automated surveillance system using edge computing. *Journal of Supercomputing* **73**(3), 926–939.

Patti, E., Acquaviva, A., Jahn, M., Pramudianto, F., Tomasi, R., Rabourdin, D., Virgone, J. & Macii, E. 2014. Event-driven user-centric middleware for energy-efficient buildings and public spaces. *IEEE Systems Journal* **10**(3), 1137–1146.

Pavon, J., Gomez-Sanz, J., Fernndez-Caballero, A. & Valencia-Jimenez, J. J. 2007. Development of intelligent multisensor surveillance systems with agents. *Robotics and Autonomous Systems* **55**(12), 892–903.

Piette, F., Dinont, C., Seghrouchni, A. E. F. & Taillibert, P. 2015. Deployment and configuration of applications for ambient systems. In *Proceedings of the 6th International Conference on Ambient Systems, Networks and Technologies, ANT-2015, Procedia Computer Science*, **27**, 373–380.

Power, P. W. & Schoonees, J. A. 2002. Understanding background mixture models for foreground segmentation. In *Proceedings Image and Vision Computing New Zealand*, 267–271.

Prud'hommeaux, E. & Seaborne, A. 2006. *SPARQL Query Language for RDF*, Technical report. `http://www.w3.org/TR/rdf-sparql-query/`.

Qian, H., Wu, X. & Xu, Y. 2011. *Intelligent Surveillance Systems*, **51**. Springer Science & Business Media.

Quaritsch, M., Kreuzthaler, M., Rinner, B., Bischof, H. & Strobl, B. 2007. Autonomous multicamera tracking on embedded smart cameras. *EURASIP Journal on Embedded Systems* **2007**(1), 35–45.

Rao, A. S. & Georgeff, M. P. 1995. BDI agents: from theory to practice. in *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS)*, 312–319. MIT Press.

Rothkrantz, L. 2013. Crisis management using multiple camera surveillance systems. In *Proceedings of the 10th International Conference on Information Systems for Crisis Response and Management*, 617–626.

Ryoo, M. S. & Aggarwal, J. K. 2008. Recognition of high-level group activities based on activities of individual members. In *Proceedings of the 2008 IEEE Workshop on Motion and Video Computing*, 1–8.

Silva, R., Arakaki, J., Junqueira, F., Filho, D. S. & Miyagi, P. 2012. Modeling of active holonic control systems for intelligent buildings. *Automation in Construction* **25**, 20–33.

Snidaro, L., García, J. & Llinas, J. 2015. Context-based information fusion. *Information Fusion* **25**(C), 16–31.

Sobhani, F., Kahar, N. F. & Zhang, Q. 2015. An ontology framework for automated visual surveillance system. In *Proceeding of the 13th International Workshop on Content-Based Multimedia Indexing (CBMI)*, 1–7.

Soldatos, J., Pandis, I., Stamatis, K., Polymenakos, L. & Crowley, J. L. 2007. Agent based middleware infrastructure for autonomous context-aware ubiquitous computing services. *Computer Communications* **30**(3), 577–591.

Stavropoulos, T. G., Vrakas, D., Vlachava, D. & Bassiliades, N. 2012. BOnSAI: a smart building ontology for ambient intelligence. In *Proceedings of the Second International Conference on Web Intelligence, Mining and Semantics, WIMS '12*, 30:1–30:12. ACM.

Town, C. 2007. Multi-sensory and multi-modal fusion for sentient computing. *International Journal of Computer Vision* **71**(2), 235–253.

Valera, M. & Velastin, S. 2005. Intelligent distributed surveillance systems: a review. *IEEE Vision, Image and Signal Processing* **152**(2), 192–204.

Veeraraghavan, A., Chellappa, R. & Roy-Chowdhury, A. 2006. The function space of an activity. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, **1**, 959–968.

Velipasalar, S., Schlessman, J., Chen, C.-Y., Wolf, W. & Singh, J. 2006. SCCS: a scalable clustered camera system for multiple object tracking communicating via message passing interface. In *IEEE International Conference on Multimedia and Expo*, **28**, 277–280.

Vicaire, P., He, T., Cao, Q., Yan, T., Zhou, G., Gu, L., Luo, L., Stoleru, R., Stankovic, J. A. & Abdelzaher, T. F. 2009. Achieving long-term surveillance in vigilnet. *ACM Transactions on Sensor Networks* **5**(1), 9:1–9:39.

Vogler, C. & Metaxas, D. 1999. Parallel hidden Markov models for American sign language recognition. In *The Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999*, **1**, 116–122.

W3C OWL Working Group 2009. *OWL 2 Web Ontology Language Document Overview*, Technical report, W3C. `http://www.w3.org/TR/2009/REC-owl2-overview-20091027/`.

Weinland, D., Ronfard, R. & Boyer, E. 2011. A survey of vision-based methods for action representation, segmentation and recognition. *Computer Vision and Image Understanding* **115**(2), 224–241.

Winkler, T. & Rinner, B. 2014. Security and privacy protection in visual sensor networks: a survey. *ACM Computing Surveys* **47**(1), 2:1–2:42.

Woolridge, M. 2009. *Introduction to Multiagent Systems*, 2nd edition. John Wiley & Sons Inc.

Yan, L., Chao, L., Ke Wei, P. & Tao, S. 2015. Crowd counting using wireless infrared distance sensors for indoor environments. In *Informatics and Communication Technologies for Societal Development*, E. B. Rajsingh, A. Bhojan & J. D. Peter (eds). Springer, 1–8.

Yi, P., Iwayemi, A. & Zhou, C. 2011. Developing ZigBee deployment guideline under WiFi interference for smart grid applications. *IEEE Transactions on Smart Grid* **2**(1), 110–120.

Yilmaz, A., Javed, O. & Shah, M. 2006. Object tracking: a survey. *ACM Computing Surveys* **38**(4), 1–45.

Yong, C., Qiao, B., Wilson, D., Wu, M., Clements-Croome, D., Liu, K., Egan, R. & Guy, C. 2007. Coordinated management of intelligent pervasive spaces. In *Proceedings of the 5th IEEE.*

Yu, T., Sekar, V., Seshan, S., Agarwal, Y. & Xu, C. 2015. Handling a trillion (unfixable) aws on a billion devices: rethinking network security for the internet-of-things. In *Proceedings of the 14th ACM Workshop on Hot Topics in Networks, HotNets-XIV*, 5:1–5:7. ACM.

Zhang, D., Gatica-Perez, D., Bengio, S. & McCowan, I. 2006. Modeling individual and group actions in meetings with layered HMMs. *IEEE Transactions on Multimedia* **8**(3), 509–520.

Zhu, F., Mutka, M. W. & Ni, L. M. 2005. Service discovery in pervasive computing environments. *IEEE Pervasive Computing* **4**(4), 81–90.