Chapter 18
WITNESS: Wide InTegration of Sensor Networks to Enable Smart Surveillance

Alessandro Mattiacci, Marco Cosentino, Walter Matta, Carlo Maria Medaglia, Andrei Braicov, Ivan Budanaev, Mircea Petic, Vasileios Argyriou, and Mahdi Maktab Dar Oghaz

Abstract In the recent years the need of advanced security systems has grown in an unpredictable manner. We saw how modern cities are very susceptible to terrorist attacks and the population asks to feel safer. In this perspective WITNESS project was designed, to help security forces to preview terrorist attacks and to facilitate disaster recovery scenarios. In this paper we break down some aspects of modern surveillance systems within the context of the Wide InTegration of sensor Networks to Enable Smart Surveillance (WITNESS) NATO project. A brief description of the architecture of the system and the methodology are presented, as well as several use case scenarios which WITNESS will cater for.

Keywords Surveillance · Security · UAV · Drone · Sensors · Computer vision · Data fusion

18.1 Introduction

Information technology develops at a fast pace, and its progress influences the evolution of many other fields of science. It is difficult to say nowadays if there are any aspects of our modern world which have not yet been affected in some way

This research is sponsored by the NATO SPS Programme under the funding NATO SPS G5437 – WITNESS – Wide InTegration of sensor Networks to Enable Smart Surveillance.

A. Mattiacci (✉) · M. Cosentino · W. Matta · C. M. Medaglia
Link Campus University, Rome, Italy
e-mail: a.mattiacci@unilink.it

A. Braicov · I. Budanaev · M. Petic
Tiraspol State University, Chisinau, Republic of Moldova

V. Argyriou · M. M. D. Oghaz
Kingston University, London, UK

C. Palestini (ed.), *Advanced Technologies for Security Applications*, NATO Science for Peace and Security Series B: Physics and Biophysics,
https://doi.org/10.1007/978-94-024-2021-0_18

by the IT revolution. One such area of today's society is the urban surveillance and security. Although security is paramount in its core concept, the importance of this area has substantially increased since the launch of the Global War on Terrorism military campaign in 2001. This can easily be observed with the shift of military technologies into domestic applications and the confluence of internal and external security [21].

Modern surveillance systems consist of many individual components, and represent a very complex architecture. Most common problems that arise when building such products are related with data transfer speed and reliability, data analysis and automation of processes. The widely-spread high bandwidth mobile networks and attack resistant cryptosystems provide solutions to the first two of the above requirements. Computer vision algorithms permit to analyze streams of data and detect predefined patterns to spot hostile or abnormal behaviour and automatically send signals to supervisors at the monitoring stations.

Present paper addresses this specific area of modern security – surveillance systems. This article is written in the framework of the Wide InTegration of sensor Networks to Enable Smart Surveillance (WITNESS) project which has the aim to design a system for urban surveillance and security to help detect, prevent and give an efficient response to terrorist threats and attacks.

The aim of this paper is to describe the methodology of the research in the project, starting with the architecture of the surveillance system.

The article is structured as follows. First, we start with the state of the art of similar research topics. Then we will describe our main aspects of the system architecture. More details are given in the following sections, concerning sensor models.

18.2 Related Works

The topic of smart surveillance is very popular nowadays. Its idea came from the practical issues where video surveillance was used to monitor municipal buildings, banks, train stations, etc. Surveillance becomes more and more important because of increasing number of exceptional situations that require a high attention to the people and society [19]. Smart surveillance is the use of automatic video analysis technologies in video surveillance applications [12]. Automatic video analysis technologies usually take into account both GPS and telemetry data. As video data also contains noise, video sequences are usually divided to be processed into smaller pieces. A separate analysis of smaller video sequences allows one to create an overview of the whole video.

The current available smart surveillance infrastructure allows to configure and implement the algorithms for recording and filtering video data streams from interconnected objects on the Internet. One of the examples of state-of-the-art infrastructure are RFID frameworks. They are elaborated in such a way that they provide the functionality to monitor space (2D or 3D), identify suspicious events, and react by generating appropriate responses to the situation. RFID frameworks have been developed as a result of researchers' efforts [17] and commercial demands. Examples of open-source RFID frameworks include Mobitec [20], AspireRfid [5, 6], and the Fosstrac project [9, 10] that provides free infrastructure deployments.

Another approach would be to use Wireless Sensor Networks (WSNs). These were initially used as surveillance in military conflict zones. The first implementations of WSNs used distributed sensor networks (DSNs) technology. Just as the first sensors were quite large, their applicability was reduced as well as due to their limited wireless connectivity. Currently the sensors are significantly smaller and cheaper. This led to the implementation of sensor networks for monitoring apartments, the environment, and the use of body sensors. WSN is considered one of the most prospect technologies of the present century [5, 8].

The variety of WSNs platforms is great. There is a platform that only addresses the system as a network of sensors. Other platforms work with devices and other sensor networks connected to the WSNs. There are WSN development and monitoring systems that have limited extensibility, for example Moteview [14] and [11]. The following tools provide development and/or programming environments for WSNs systems: Hourglass, SenseWeb, jWebDust [4] and GSN [1]. A more detailed description of the architectural particularities of the WSNs systems can be found in [4].

The effectiveness of WSNs in the surveillance process is acknowledged. However, there are approaches that seek to improve the use of WSNs by combining them with Unmanned Aerial Vehicles (UAVs) in surveillance. UAV is a solution in situations where it is necessary to fly over dangerous areas without endangering people's lives.

Paper [2] presents a project example describing the interaction between WSNs technology and UAV tools used for border surveillance. The UAV in the given case is considered a quadcopter.

Research focused on the use of quadcopter in terrestrial surveillance is focused on identifying cost-effective solutions and preserving the same functionality. Usually a quadcopter is driven by the means of proprietary framework APIs from a laptop or a PC. A quadcopter is useful in reading the altitude to the ground, as well as in measuring air temperature, humidity and gas composition [3].

18.3 Use Case Scenarios

WITNESS proposes an innovative technological solution to incidents and accidents that may occur in an unpredictable urban scenario characterized by crowded scenes with potentially complex structured man made surroundings. The typical scenarios WITNESS will cater for are those where an incident or accident has caused disruption in the normal 24/7 operation of a public space (for instance a metro, a railway or a bus station). In such environments a nominal flow of people can be expected and, therefore, normal behaviour can be predicted. We also envisage that the public spaces of interest will be monitored by fixed cameras and by police forces. Some examples of possible case studies follow:

- Natural disasters: a disaster is a serious disruption of the functioning of a community or a society involving widespread human, material, economic or environmental loss and impacts. Disasters caused by natural or technological reasons are identified as NaTech.
- General disruption of individuals: public areas witness the presence of intoxicated individuals, usually disruptive in small groups. Such events may have an effect on the normal flow of people, for instance in a metro or railway. Such individuals may start pushing one another and other people in the surroundings.
- Public events related to holidays, manifestations or protests where there is an abnormal accumulation of people with high density. As noticed from latest terrorist attacks, these types of scenarios are a very attractive target for terrorists [15, 16].
- Accident: this describes a category of events that may be caused by the failure of electrical power – for instance delaying metro or trains, or by physical accidents happened to individuals, including suicide attempts. In such case an entire station may be closed and events may more or less slowly affect the entire area.
- Incident: this could be caused by a terrorist threat, even including hostage situations.

Specific use case scenarios of WITNESS system applications would include monitoring of the sport events, large-scale peaceful demonstrations and violent protests. During these events, the abnormal behavior under study is sought to be well-organized groups of people who aim to destabilize public order by taking hostage among civilians, using guns, fumigants or explosives. Depending on the event type, the size of monitored area can span between 10,000 to 300,000 meters square. The average occupancy of the monitored area is projected to be up to 20 thousand people.

18.4 WITNESS Architecture

WITNESS implements a distributed multi-layered architecture to satisfy the operational requirements of a situational awareness and decision making system. This approach facilitates building a flexible and pervasive enough product, ready to be automatically reconfigured and quickly redeployed when needed. The cornerstone of the system is the data, which is collected by a predefined set of sensors – wearable by police forces deployed on the grounds or sensors installed on UAVs and police vehicles. In this way, the layer responsible for collecting data in WITNESS system represents a "sensor fusion grid" that leverages data from multiple heterogeneous sensor nodes, including cameras, microphones and drones. A schematic diagram of the system's architecture is depicted in the figure below (Fig. 18.1).

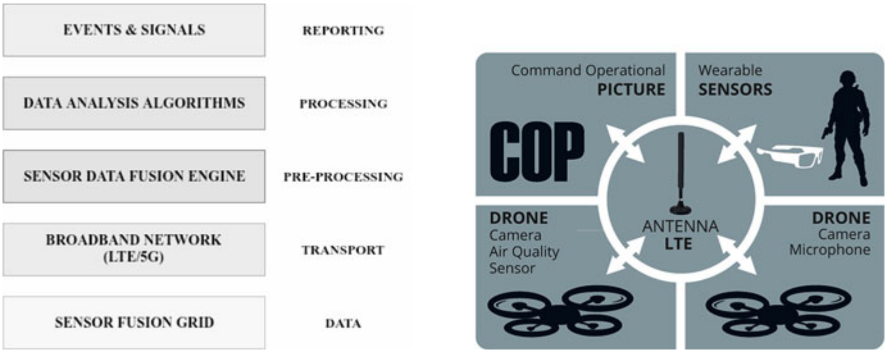

Fig. 18.1 WITNESS architecture

Data originating from different sources dispersed in multi-dimensional space (police forces on the field, quadcopters in the air, lags and losses during data transfer synchronization) need to be gathered and pre-processed before being submitted for analysis. This is a difficult task to address, since information received scales exponentially in terms of location-space, time and means (multi-source information). Recent advances in ICT technologies can boost the efficient acquisition, fusion and integration of information from the above sources. Sensor data fusion component will be implemented to address the above problem. Data fusion technologies involve the fusion of multi-sensory data to estimate the position, speed, attributes and identity of the detected and flagged targets, e.g. a person, a vehicle or an object in the operation area.

It is widely accepted that the data transfer between system components of the above architecture requires a dedicated broadband communication infrastructure (e.g. LTE, 5G technology) to be deployed in a very short time, so as to be promptly used to support the communications among security forces deployed on the field and used to gather and to process awareness data coming from monitoring devices (e.g. wearable sensors or mobile nodes) and to perform the command and coordination of the forces. In this regard, the underlying infrastructure plays a critical role in the security, processing, flow, supporting information requirements throughout the operational forces.

Within the proposed methodology, WITNESS ensures that data and information are delivered to the right place on time and optimally encoded for use by their intended recipients to take the appropriate actions at the right time. The security and safety of this information will be granted by the dedicated LTE cell. This architecture is a key enabler of Net-centric Enable Capability (NcEC) and is essential for "information superiority" and "decision superiority".

WITNESS will adopt a breadboard architecture enabling plug&play integration of the various components. The architecture also enables the integration of third-party components. Furthermore, this architecture will give the possibility to assign a task to a UAV and to reallocate it to another one when the first drone will have low battery or any other malfunction scenario. This need arises for the known critical issue of the battery duration of the UAVs.

This project aims to explore technological concepts and approaches already proposed for military operational fields for use in civilian security. Those concepts, and in particular the Network Enabled Capabilities (NEC), have practically never been applied in civilian applications, and are new even to the military sectors. At a global level, however, there are already focus groups questioning on how to apply the NEC philosophy to the security applications (especially in the USA), thus is imperative to start the development of a European blueprint on this topic. One of the technological impacts of WITNESS will be the generic Internet-of-Things approach towards creating a civil security C2 situation assessment and decision aiding framework.

18.5 Methodology

Taking into account the current approaches in this field and the use cases for the system to be built the WITNESS research project will follow the scheme presented in Fig. 18.2.

In order to obtain a more accurate result we need a complete supply of important parameters for the proper determination of the instant situation. The parameters are

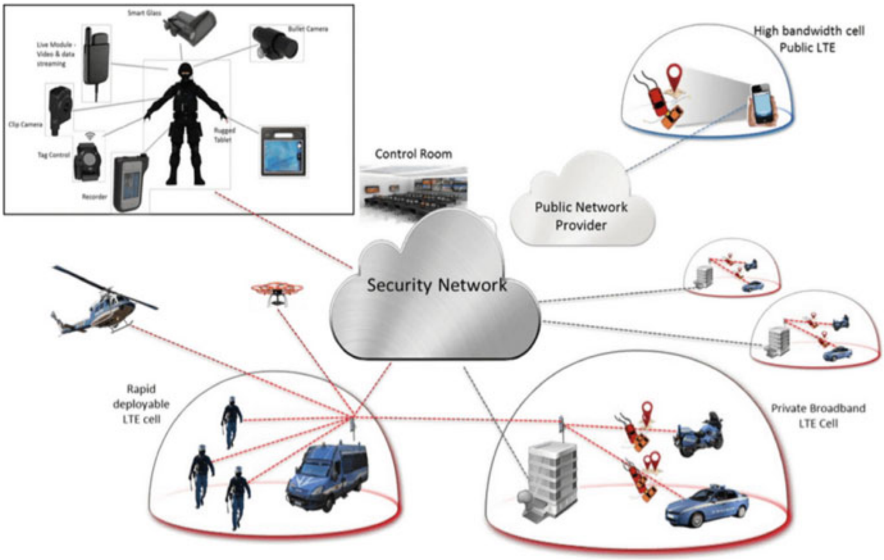

Fig. 18.2 WITNESS concept

monitored through a RGB/Thermal camera, a directional microphone and an Air Quality Sensor (CO_2).

The UAV solution for the architecture described will be a quadcopter, DJI Mavic 2 Enterprise Dual [7], which allows porting of several types of sensors. Flight time – up to 31 min. A dedicated server will be used to build the Artificial Intelligence application that will process the data provided by the sensors.

The integrated thermal camera is useful for detecting people in low visibility conditions (total darkness, fog or smoke). It can see through smoke or light mist, it does not require additional lighting, the image being obtained due to temperature differences between the target object and the environment.

Air pollution sensor is useful for gas detection [18].

Wearable sensors are to be used to monitor vital parameters of police agents. Furthermore, system can use the data provided by a rugged smartphone. This way, the monitoring system relies on a large set of different types of sensors. The volume and complexity of data supplied by these sensors dictates the necessity to develop a data fusion algorithm, which will collect, serialize and normalize the data, and pass it forward into the system pipeline.

Finally, a demonstration will be done using a simulator to let the security forces be ready for some events that may be planned in advance.

18.6 CV: Crowd Management and Monitoring

Large gatherings such as sporting events, music concerts, cultural festivals, and amusement parks are always characterized by crowded environments. Crowd dynamics, based on the behavior of individuals in a crowd vary. In fact, people move with some purpose and intentions might be completely different between or among individuals, making identification of behavior patterns hard to comprehend. However, the behavior of a crowd is widely understood to have collective characteristics which can be described in general terms. This because in a very crowded and cluttered space even individual behavior tends to be influenced by the behavior of people nearby. Crowd behavior may vary from a peaceful to hostile. Nonetheless, the safety and security of an individual in a crowded environment are of utmost importance for the person themselves and security bodies in charge of keeping orderly behavior and prevent accidents, incidents and disruptions such as police, security operators and guards.

One of the objectives of NATO project is to promote crowd safety and security using aerial imagery. For many years surveillance cameras and nvr systems were the only commonly practiced solution for crowd monitoring and management, with televisions observed by a member of the security staff, looking for problems or conditions such as various forms of congestion, disruptions in crowd movement and flow, stampede or problems arising from individual incidents not related to a crowd, such as fight, a theft, the outbreak of a fire, or potential act of terrorism. In these approaches, human observers are normally employed to monitor multiple camera streams during or after the event. Although these solutions can be practical for monitoring small scale crowds with a limited number of cameras, scalability, reliability, and cost-effectiveness of these solutions are questionable in large scale events.

Recent advancements in parallel computing and GPU technology diminished this computational barrier and allowed complex models such as deep neural networks to flourish. The rise of data-driven approaches such as deep learning, simplified modeling tasks by eliminating the need for domain expertise and hardcore feature extraction. The emergence of deep learning in the last decade has boosted the performance of image classification techniques and has started having a positive impact on crowd behavior analysis. CNN(Convolutional Neural Networks) have gained ground in crowd monitoring and behavior analysis. It has been shown that such models are not only able to achieve state-of-the-art performance for the visual recognition tasks, in which they were trained, but also the learned representation can be readily applied to other relevant tasks.

Deep learning methods are very data-hungry: they require a large amount of annotated training data, often very limited and some times unavailable. Furthermore, deep learning models rely on costly parallel computing facilities and GPU technology.

In this project we used deep learning to model complex crowd behaviors and characteristics. Several deep learning based crowd analysis algorithms including crowd counting, localization, and density estimation, crowd flow analysis, crowd anomaly detection, fight detection, and object detection have been developed to be used for crowd management using aerial images.

18.6.1 Applications and Use-Cases

Crowd counting and density estimation are of great importance in computer vision due to its essential role in a wide range of surveillance applications including crowd management and public security [23]. However, the presence of drastic scale variations, the clutter background, and severe occlusions make it challenging to generate high-quality crowd density maps. In this project we use powerful CNN to handle the challenging situations in crowds mainly by fusing multi-scale or multi-context information to improve the feature representations. CNN (shown in Fig. 18.3) which specifically inspired by the biological visual cortex is a type of deep neural networks, most commonly applied to analyzing visual imagery. Unlike traditional fully connected deep neural networks, a CNN is able to successfully capture the Spatial and Temporal dependencies in a given image through the application of relevant spatial filters.

In dense crowds, individuals heads are the only reliably visible body part in an image as the other body parts are usually occluded by the crowd. Hence, majority of the CNN based approaches rely on heads as the only discriminant feature of the

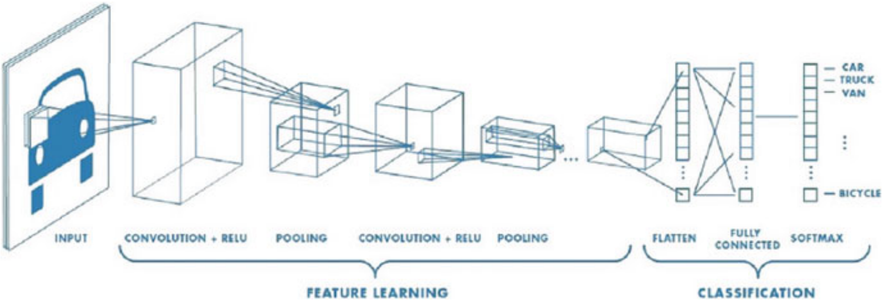

Fig. 18.3 A typical CNN architecture

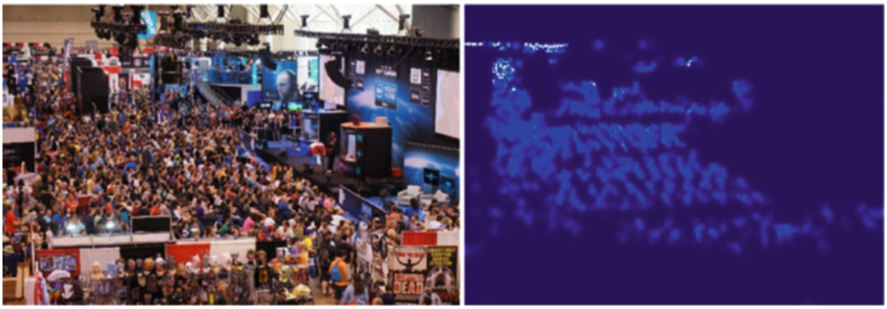

Fig. 18.4 A Sample crowd scene along with its corresponding crowd density map

image. As shown in Fig. 18.4, CNNs can generate crowd density maps which not only localize the crowd but also estimates its size and density.

Crowd flow is another informative metric in crowd management and behavior analysis. There is a critical capacity where flow begins to decrease as the crowd's density increases. A dense crowd with high flow magnitude poses a serious safety threat and might lead to a human stampede. In visual surveillance, optical flow algorithms have become an important component of crowded scene analysis. The application of optical flow allows crowd motion dynamics of hundreds of individuals to be measured without the need to detect and track them explicitly, which is an unsolved problem for dense crowds. We employed CNN to measure the optical flow in the crowd. The algorithm takes two frames at given time intervals to measure the crowd flow magnitude and direction and generates the flow map which localizes the flow intensity. Crowd flow can provide valuable information about crowd behavior in locations such as stadiums and parades where smooth flow or crowd is essential.

Analysis of crowd behavior in public places is a critical objective for video surveillance [22]. We are often reminded how time consuming it is for police officers or forensic investigators to search through hours of video looking for an incident. It is clear that there would be a substantial advantage if some form of image processing system could be applied to the video signal, to automatically spot the anomalous crowd behavior as they arise, and to trigger a suitable response in a reasonable time. This project offers state of the art deep learning techniques to create a model capable to handle overwhelming complexity of crowds and detect anomalous crowd behavior. The proposed anomalous crowd behavior detection solution uses deep learning models which are also deep over temporal dimensions. These deep models, which are addressed as RNN (Recurrent neural network), are rolled over the time domain. A visual feature extractor (CNN) runs in tandem with a model that can learn to recognize and synthesize temporal dynamics (LSTM) for tasks involving sequential data such as anomalous crowd behavior.

Object detection is breaking into a wide range of industries, with use cases ranging from personal security to productivity in the workplace. One of the most important applications of object detection is crowd safety and security management. Object detection algorithms can be utilized for identifying missing objects or individuals in the crowd, vehicular traffic or even early warning of a potential terrorist attack by identifying unattended suspicious objects in public. We use state-of-the-art CNN models capable to identify a wide range of objects such as persons, vehicles, suitcase, trees, animals, etc.

18.7 Data Analysis and Data Fusion

In a system with a multi-sensor grid, consisting of inertial, audio, visual, location, chemical, and biosensors, an abundant volume of data is generated on the fly. Analysis of this data represents a crucial component of WITNESS.

The stated above goals of the WITNESS system, i.e. detection and prevention of terrorist attacks and other disasters, when approached from the data science point of view, fall into two main categories: classification and anomaly detection. The classification part is responsible for spotting predefined scenarios like gunshots, armed person or armored vehicle detection. Anomalies represent those type of scenarios which are detected as a statistical outlier and/or have not been categorized prior to deployment on the ground.

Denoting the representation of a single sensor observation by the tuple $O := \langle p, t, x \rangle$, where x is the measurement recorded by the sensor in the position p and time t, we collect the following data vector produced by the sensor grid: $S_t := \{O_{1 \leq j \leq n}\}$, with n as the total number of sensors in the grid. Consequently, the data produced by the system in time interval $(t_i, t_{i+1}]$ is given by the time series $S_{t_i, t_{i+1}} = \{S_{t_i < t \leq t_{i+1}}\}$. In modern data analysis, one must make sure to avoid the most common pitfalls. These are related to choosing the right interval to collect data and addressing the problem of noise induced by sensor uncertainty (poor sensor data and quality). The solution we chose is similar the way the Adam algorithm [13] addresses the learning rate optimization problem in machine learning – we substitute the collected observations in $S_{t_i, t_{i+1}}$ by its moving average giving more weight to most recent data:

$$\overline{S_{t_i,t_{i+1}}} = WMA_{S_{t_i,t_{i+1}}} = \frac{\sum_{t=t_i,k=1}^{t_{i+1},\Delta} k \cdot S_t}{\Delta(\Delta + 1) / 2}, \Delta = t_{i+1} - t_i.$$

Adding weighted moving averages smooths out induced noise and allows detecting patterns even if some of the data supplied by some sensors came delayed.

Due to the high dimensionality of the data provided by mutually complementary sensors, we apply data analysis on multiple groups of features concurrently. Obtained insights are combined together with processed visual and audio data to refine estimates and assessments regarding the current state of the environment. This approach is called Data Fusion and proves to have better performance and more accurate results. Among other components described here, it also consist of the following aspects: sensor modeling, management, control and optimization, statistical and probabilistic methods, neural networks, situation and impact (threat) refinement modules.

We use distribution-based approach to detect and flag outliers, a distance-based approach to detect how far a tuple of a pre-defined group of observation is far from the set's centroid, and, where applicable, we apply naive Bayes to combine the outputs. One example of a data fusion algorithm output is the detection of a bomb explosion based on the group of audio, chemical (smoke), and visual sensors. This is a good example, as the data produced by the smoke detection sensor will arrive delayed in comparison to other sensors. Another example would refer to the safety of the officer on the premises. Obtaining information about his state is based on position, velocity and heart rate parameters.

18.8 Conclusions

WITNESS provides an opportunity for companies to increase the competitiveness of the WSN (wireless sensor network) industry by developing novel sensors so far never exploited in this field.

One of the project objectives is the definition of tools, technologies and methods that will facilitate the countering to attacks and critical situations. This study will lead to some conclusions about the state of the art resources and, hopefully, to the definition of new methods that will improve the ability to develop a monitoring solution in short time. The results obtained will be the base for future development that will improve a easily deployable and secure system.

References

1. Aberer K, Hauswirth M, Salehi A (2006) The global sensor networks middleware for efficient and flexible deployment and interconnection of sensor networks. LSIR Report 2006-006
2. Berrahal S, Kim J-H, Rekhis S, Boudriga N, Wilkins D, Acevedo J (2016) Border surveillance monitoring using Quadcopter UAV-Aided wireless sensor networks. J Commun Softw Syst 12:67–82

3. Borah DR, Debnath L, Gogoi M (2016) A review on quadcopter surveillance and control. J Eng Technol 4(1):116–119
4. Chatzigiannakis I, Mylonas G, Nikoletseas S (2005) jwebdust: a java-based generic application environment for wireless sensor networks. In: In Proceedings of the IEEE international conference on distributed computing in sensor networks (DCOSS). Lecture notes in computer science (LNCS), pp376–386, Springer
5. Consortium O (2019) AspireRFID OW2 Project, https://projects.ow2.org/view/aspire-rfid/. Accessed on 05 Sept 2019
6. Dimitropoulos P, Soldatos J (2010) RFID enabled fully automated warehouse management: adding the business context. IJMTM 21:269–288
7. DJI (2019) DJI Mavic 2 Enterprise Dual, https://www.dji.com/it/mavic-2-enterprise. Accessed on 05 Sept 2019
8. Yinbiao S, et. al. (2014) Internet of things: wireless sensor networks, technical report p 78, Wireless sensor networks project team. International Electrotechnical Commission, Geneva
9. Floerkemeier C, Lampe M, Roduner C (2007) Facilitating RFID development with the accada prototyping platform. In: IEEE international conference on pervasive computing and communications
10. Fosstrak (2019) Fosstrak: open source RFID software platform, http://www.fosstrak.org. Accessed on 05 Sept 2019
11. Ritter H (2019) Short presentation of ScatterWeb Software and Hardware for next generation wireless networks, http://files.messe.de/cmsdb/001/14916.pdf. Accessed on 05 Sept 2019
12. Hampapur A, Brown L, Connell J, Pankanti S, Senior A, Tian Y (2004) Smart surveillance: applications, technologies and implications, vol 2, pp 1133–1138
13. Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980
14. MoteWorks (2012) MoteWorks getting started guide. Revision G, Technical report, MEMSIC, Inc, McCarthy Blvd, Milpitas
15. of State, U. D. (2016) Country reports on terrorism, Technical report, U.S. Dept. of State
16. of State, U. D. (2016) Global terrorism index, technical report, Institute for Economics and Peace
17. Prabhu BS, Su X, Ramamurthy H, Chu C-CP, Gadh R (2005) WinRFID – a middleware for the enablement of radio frequency identification (RFID) based applications. In: Mobile, wireless and sensor networks: technology, applications and future. Wiley, Hoboken
18. Robot D (2019) Gravity: UART Infrared CO2 Sensor, https://www.dfrobot.com/product-1565.html. Accessed on 05 Sept 2019
19. Saha S, Neogy S (2014) A case study on smart surveillance application system using WSN and IP webcam. In: 2014 applications and innovations in mobile computing (AIMoC), pp 36–41
20. technology Center M (2019) MobiTeC – open source RFID middleware 1.0, http://mobitec.ie.cuhk.edu.hk/rfid/middleware/. Accessed on 05 Sept 2019
21. Wilson D (2012) Military surveillance. Routledge, New York, pp 269–276
22. Xu D, Yan Y, Ricci E, Sebe N (2017) Detecting anomalous events in videos by learning deep representations of appearance and motion. Comput Vis Image Underst 156:117–127
23. Zhang Y, Zhou D, Chen S, Gao S, Ma Y (2016) Single-image crowd counting via multi-column convolutional neural network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, pp 589–597