



A review of video surveillance systems

Omar Elharrouss, Noor Almaadeed, Somaya Al-Maadeed

Department of Computer Science and Engineering, Department of Computer Science and Engineering, Qatar University, Doha, Qatar

ARTICLE INFO

Keywords:

Video surveillance system
Video analysis
Video surveillance systems trends

ABSTRACT

Automated surveillance systems observe the environment utilizing cameras. The observed scenario is then analysed using motion detection, crowd behaviour, individual behaviour, interaction between individuals, crowds and their surrounding environment. These automatic systems accomplish multitude of tasks which include, detection, interpretation, understanding, recording and creating alarms based on the analysis. Till recent, studies have achieved enhanced monitoring performance along with avoiding possible human failures by manipulation of different features of these systems. This paper presents a comprehensive review of such video surveillance systems as well as the components used with them. The description of the architectures used is presented which follows the most required analyses in these systems. For the bigger picture and wholesome view of the system, existing surveillance systems were compared in terms of characteristics, advantages, and difficulties which are tabulated in this paper. Adding to this, future trends are discussed which charts a path into the upcoming research directions.

1. Introduction

In the past twenty years, the world has witnessed a development in the different socio-economic sectors due to which life has become more complex in different aspects including the safety and the security of individuals. Thus, the need to be secured and monitored has become a necessity. As a result, surveillance cameras installed in public and private areas are a suitable solution for reassuring safety. Generally, human resources are dedicated to observing these cameras 24 h a day, which makes continuous monitoring very expensive. Thence, an automated system to monitor events in the different locations that use surveillance cameras is highly needed. The creation of an automatic system for video surveillance is a feasible solution. The main role of such an automatic system is to support the human operator in order to perform different monitoring tasks [1]. These tasks could be related to different applications, such as traffic control, accident prediction, crime prevention, motion detection, and homeland security. Other applications can be added pertaining to monitoring indoor and outdoor scenes like airports, train stations, parking lots, highways, stores, shopping malls and offices. Several features might be added to enhance the monitoring performance and to eliminate possible human failures.

The need for monitoring public and private areas is gaining more popularity nowadays due to persisting security reasons. This awakens the need of surveillance systems. A better security system that provides automatic analysis of the data provided by surveillance with less human

operations is gaining more and more demand. Tremendous progress has already been made in this area of research. The typical scalar information cannot be applied in various applications like video surveillance and traffic monitoring. With the development of technology, the use of image sensors in wireless sensor networking has led to a good improvement in the field of Video Sensor Networks (VSNs). Image sensors connected to these networks can capture a wide amount of video information [2]. VSNs consist of cameras that capture the multimedia data from the monitoring area. It can support many applications including video surveillance systems, virtual reality, personal care, etc. They offer new opportunities for many promising applications compared to scalar sensor networks. Also, the devices come with image sensors, adequate processing power, and memory. The wireless interfaces are used to communicate between the VSNs devices within the network. In all of these applications, VSNs monitoring the scenes that require surveillance using many analyses on the recorded data. The most important application is surveillance system that requires multiple sensors. For example, these sensors can sense for example any moving objects and then track them [3]. In the field of surveillance research, analyzing the information contained in a video is the biggest challenge facing the computer vision. Several approaches have been applied in this area, particularly for the detection, classification and recognition of objects or events, which represents the main tasks of video surveillance systems. In order to ensure a good surveillance, these systems need better analysis of the data. For the public areas, the system must provide

E-mail addresses: n.alali@qu.edu.qa (N. Almaadeed), s_alali@qu.edu.qa (S. Al-Maadeed).

<https://doi.org/10.1016/j.jvcir.2021.103116>

Received 13 May 2018; Received in revised form 7 February 2021; Accepted 2 April 2021

Available online 16 April 2021

1047-3203/© 2021 Elsevier Inc. All rights reserved.

a good understanding of the scenes and analyse the crowd and events in it. In order to achieve this, this paper provides a general overview of video surveillance systems and their characteristics [4]. A summary of existing video surveillance systems describing the architecture and functionalities are described here. Analyses related to surveillance is presented and future trends of video surveillance systems are discussed.

The paper is organised as follows, the first section details a summary of the existing video surveillance systems in literature. Then, characteristics of a video surveillance system is explored with the next section comparing and describing the available categories of video surveillance cameras. Finally, a detailed description of analysis of video surveillance system is done with the tendencies of this system with prospects into the future.

2. Related works

A video surveillance system is an action of monitoring the activity in public or private scenes using cameras. In the past, the function of a basic video surveillance system was to make the acquisition through one or many cameras and then transmuting the signal to be displayed by a monitor and/or recorded in a central location. Nowadays, the development of video surveillance systems have enabled the automation of analysis of the acquired data. The video surveillance systems are implemented to ensure security by reacting and predicting to all the possible events in the scene [9,10].

However, these systems are still not accurate for the real-time reactions that may also be delayed. In addition, the environmental changes present a limitation for any system. These challenges present the difficulties of video surveillance systems. The processing of the data using several proposed algorithms may help to avoid these problems by analyzing and understanding the recorded videos. Among these, analyses that are related to video surveillance are motion detection, tracking, events, and anomalous detection in the scenes, crowd behavior, face recognition, and object classification, which allows extracting information in order to support the management of emergencies and ensure people safety.

The video surveillance in a private scene (workspace) have no improvement in literature and still works with the basic systems [5]. The reason is that the surveillance in the workspace is just for counting workers, quantifying the risks and assessing the safety [6]. For that, McGlothlin et al. [7] used video technologies to assess the measure of probable risks and the work done by the employers. In [8], the authors proposed the use of video technology in order to evaluate the relationship between injury and behavior of football players.

In the public scenes, the presence of video surveillance systems is to predict the event with the goal to avoid and stop any risk, for example, prevent accidents rather than waiting for it to happen. For that, McSwen et al., proposed a method to minimize the chances of accidents using behavior-based safety methods (BBS) [11]. In addition, Shriffad et al., tried to reduce the frequency of risky behaviors [12]. Behavior recognition approaches are used to improve performance workers [13]. For that, video sequences are the best solution and an effective source of information to improve the performance of workers inside factories in terms of safety by giving the possibility to observe situations frequently; record and review the important periods. Also, video clips give to the employers ways to improve their tasks with security and to show to them the behaviors and the riskiest acts.

In the field of security, defense and anti-terrorism, some authors proposed a double-mode surveillance system based on both audio and video information [14,15]. The development of this system required a pan-tilt-zoom (PTZ) camera for capturing video of the human body. In addition, a Laser Doppler Vibrometer (LDV) [16,17] is used to acquire remote audio by detecting the vibration of the objects. To control the orientation of LDV, the system needs a theodolite. The advantage of this system is that it extracts both visual and audio features, which can ensure a good detection of any threats. As the audio may also contain a

lot of information about a possible event, that can hurt people. However, almost all video surveillance systems used just visualization part of the observed scenes [18,19]. But, the visual-based systems are sensitive to environmental changes and noise caused data transmission.

Data transmission is one of the limitations of video surveillance systems. For that, Lubobyaa et al. proposed an architecture of video surveillance system based on throughput characteristics using WiMAX (AW) and Hybrid WIFI-WiMAX (HWW) systems [18,20]. The AW system connect WiMAX IP cameras directly to the stations. The HWW system use WIFI IP camera connected to Customer Premises Equipment (CPE) that is also linked to the base station [21,22]. The remote visualization has been made using the Internet while the local visualization consists of connecting PCs with CPE. For the transmission of data, it can be made while using Ethernet cables or via the wireless interface. This is after the acquisition (using WiMAX or WIFI IP camera) and the conversion of this data in digital signals. The strong point of these systems is their ability to provide a remote transmission of the data even if the camera located up to many kilometers away from the station. However, the wireless mesh WiMAX networks and the antenna alignment with the base station represent the limitations of this system.

Recently some authors have tried to built and implement an unfixed video surveillance system [23]. Using the protocol RTP, the system transmits the data via wireless network. The loss and the disorder of data can be a problem of this new technique. For that, the authors use FFmpeg library for the storage and visualization of the videos. An unfixed system can be useful in many situations to ensure security. However, the difficulty is the transmission of data when the network is unstable.

The urban areas are full of people and objects, which represent a challenge in term of security, traffic, terrorist threats as well as the smart management of urban dynamics. In order to record and process the data, the servers are the traditional infrastructure in the video surveillance systems. These surveillance systems produce a huge amount of data that need to be stored. After that, an analysis is applied to this data in order to ensure safety by detecting events. Based on cloud computing some authors proposed a solution to collect, optimize this data with the goal to transmit it and store it in a cloud storage system in an efficient and secure way [24]. In [25], Chen et al. proposed a system for video surveillance of urban area using Internet and Internet of things (IoT), which is appropriate proposition, to make urban surveillance and smart cities achievable. A system based on Fog computing composed of three layers is proposed [26]. The first layer is the surveillance layer. The second is the Fog computing layer that uses many on-site smart devices. The Cloud Computing layer that contains many servers is the third layer. In the same goal, the drones are used to capture the surveillance data before transmitting it to users that transmit again to cloud servers. Using Cloud computing the system can be effective for the processing of big data, but the communication time cannot be used for the critical mission that needs real-time analysis and immediate decision.

Using and managing a large-scale video surveillance systems are the most challenging aspects of the security systems. Especially when a failure in one of the components of the system (Camera, Network or Cloud) occurs. The video data can be useless when it is not real-time and interpreted in a short time. A few methods in literature handle the useless data. Sen et al., proposed a system that consists of a video usefulness detection (VU) based on Cloud computing and edge computing [27]. The proposed model is developed to identify and locate the failure of the component state of the system and to minimize the time needed to fix it. Then, the video data will not be uploaded to cloud. Consequently, the burden on the bandwidth of network is reduced and the storage in the cloud is enhanced.

Video surveillance systems with Real-time video analysis gives a promising view. In order to analyze the video stream and provide the result, the data should be uploaded to the cloud but it can cost energy and time, given energy-limited cameras and network bandwidth. Zhang et al., attempted to optimize the energy by distributing the workload

computing in the cloud and edges nodes (AVAPS) [28]. Useless data will be eliminated for transmission in the goal to save edge devices' energy. In AVAPS system, video transmitted between cameras (edge's nodes) and the cloud using clips (other than the original ones). This, helps to reduce the amount of data and energy when transmitting. In the same context, Du et al., proposed a solution to create a dynamic virtual reality surveillance video using web-technology [29]. This technique helps to identify all possible situations in a monitored area, also to predict most events that can happen in the scene. Nevertheless, the choice of technical parameters, related to each scene, depends on the environmental characteristics and the used devices. Inspired by the development in different domains, the video surveillance systems are undergoing continuous development in term of the devices used and the analysis done, especially with the use of machine learning and deep learning algorithm.

3. Video surveillance:general description

The past two decades have been a witness to a tremendous growth of audio–video information in several environments. In addition, the increase in population and the terrorist attacks at the onset of this century and the increase of criminal activities, many surveillance cameras in the cities were deployed [1]. Consequently, a remarkable interest has been recorded in the scientific and artificial vision. Nowadays, video surveillance covers the entire cities using millions of cameras. Furthermore, many surveillance techniques are being employed to ensure security and manage the society [2,3]. By the development of internet, everything is collected, recorded and analyzed, and the privacy of people under surveillance have become threatened [4]. Moreover, video surveillance security demands have increased recently with that the technologies assisting the advancement in the architecture and the analysis of them. Fig. 1 represent the parts of a video surveillance architecture including the architecture components such as sensors, servers and the network types. The analysis part of a system is the processing algorithms that can help to ensure a good surveillance. In all these areas, the devices in the architecture are considered as the eye of the system and the analytics represent the brain of the system. However, this development reduces the margin of privacy of people that should be protected [30].

The video surveillance systems are employed in airports, streets, and private areas, supporting the security operators, signaling anomalous events, or interdicted areas access. Automated systems, guarantee the security and give support to the different sectors such as traffic flow measurement, surveil the pedestrian congestion, detect accidents on highways, count endangered species controlling national borders, ensuring the flow of refugees in troubled areas, monitoring peace treaties and providing secure perimeters around bases. Fig. 2. illustrates all the sectors that require a surveillance including residential areas, industrial locations, stadiums, airports, streets and high ways. The use of cameras represents a good, cheap and practical solution for all sectors. However, the employment of the human resources to manage these techniques is still expensive. Although surveillance cameras are already installed everywhere, video data is used in most cases just for checking

the data after an incident, so it is not used in real-time.

Every day there are thousands of cameras collecting a huge amount of data. Scientists try to implement an automated system to extract information from this data. The scene structures coming from one camera is limited because it does not show the area for its entirety. In order to surveil an extended section, like tracking a vehicle running in a city's road or analyzing global activities taking place at a major station, it is necessary to use a video stream from multiple cameras. Many intelligent multi-camera video surveillance systems have been developed for this purpose [31,32]. It is a multidisciplinary field related to image sensors, computer vision, pattern recognition, communication and signal processing. Fig. 3. represents the field of view of multi-camera for a video surveillance system [33,34]. These cameras can collaborate to provide a good surveillance. Also it helps for multiple analyses that require tracking moving objects around many surveilled areas. The need here is a monitoring system that operates continuously for surveillance in order to alert the security officers when there is still time to prevent the crime. The main problems faced by these monitoring systems are: the collection of data from all surveilled areas that need to be secured. Therefore, the problem is how to collect this data in the event time, sequentially, and without losing information which is related to the architecture and the device used in this system. The other challenge of any surveillance system is the analysis of the collected data (software problem). In the next sections, the proposed solutions in each part of a surveillance system will be presented.

4. Architectures

For the purpose of scene control, many technologies were used in the past. Active sensors such as sonar, laser and radar were used [35–37]. The first-class sensors, project the light onto the stage and view it using a camera. Then they measure the sensors that emit and receive a laser beam by measuring the phase difference, the flight time or the depth of frequency change. These are the most used techniques in robotics applications. Presently, the monitoring and the control of scenes utilize video streaming techniques. The videos captured by multiple cameras that connected to a centralized system. All these connected components with some analyses that can be installed in this system compose a video surveillance system.

Fig. 4 describe the functionalities of a video surveillance system by the definition of each part. The acquisition allows the acquisition of images using more or less sophisticated cameras. Some situations lead to using cameras with specifications very special (waterproof, wide-angle, IR, etc ...). The switching/processing part allows "dispatching" of video signals to recording and visualization functions. It also makes it possible to perform a treatment on the video (Brightness/contrast, motion detection, clutter emergency exit, etc ...). The recording of the videos is mainly done digitally. PTZ control allows control of the position of cameras with the functionality PTZ (Pan-tilt-zoom).

The function of a Video surveillance system is the action of monitoring the activity in a public scene or private using a camera. Before, the function of a basic video surveillance system is to take the acquisition

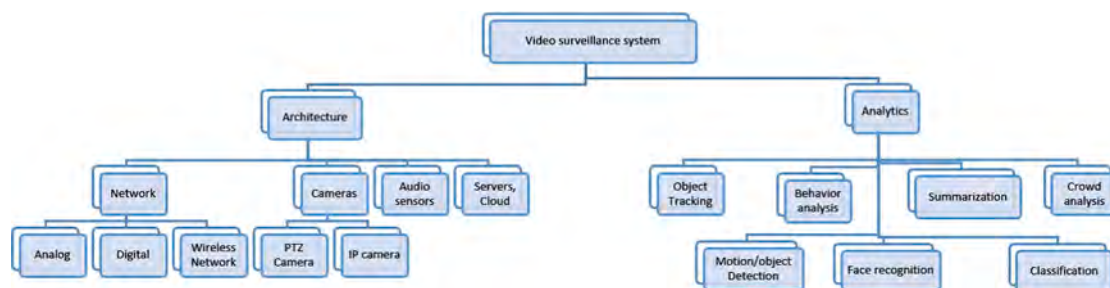


Fig. 1. Video surveillance system components and functionalities.

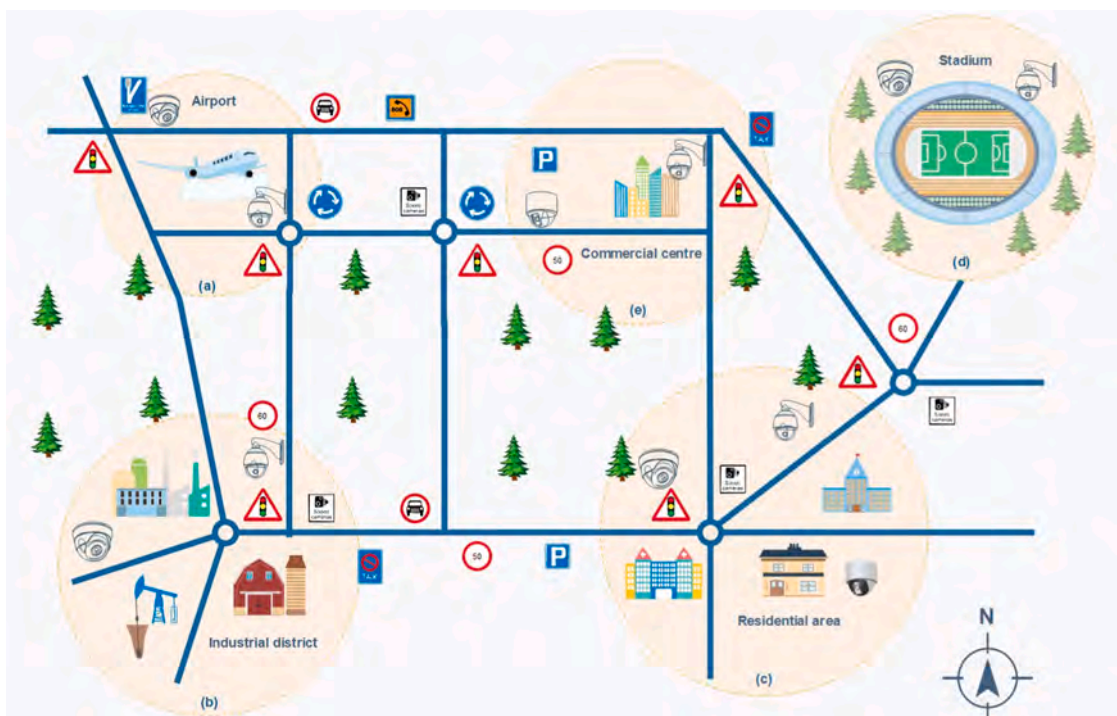


Fig. 2. Video surveillance application sectors. (a) Airports. (b) Industrial areas. (c) Residential areas. (d) Sport stadium. (e) Commercial centers.

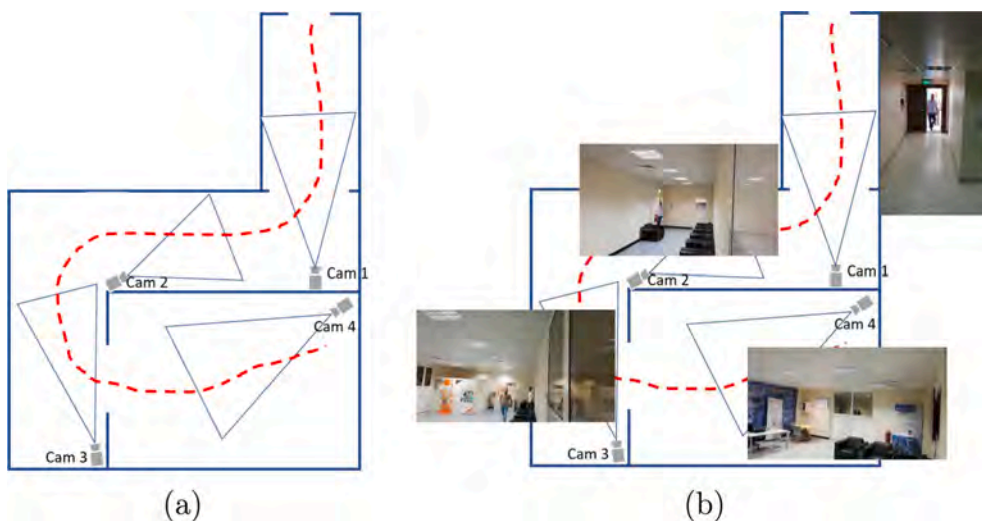


Fig. 3. Field of view of Multi-cameras.

through one or many cameras, and then transmitting the signal to be displayed by a monitor and/or recorded in a central location. Nowadays, the video surveillance systems have been developed and enabled to analyze automatically the monitored scenes. The interaction between system edges is complex due to the use of different equipment's and models. To solve this problem, many methods have been proposed. They try to connect the equipment (cameras) and collect the information using different architectures. The connection between the main station and the devices can be categorized into three types analog, digital and network.

Analog surveillance system: Over the last 20 years, the surveillance systems used analog technology [39]. Analog signal processing technology is the basis model for image transmission, exchange, and recording using a short-distance coaxial cable and a long-distance transceiver optical fiber [40]. Image restoration is one of the

advantages of this system type, while the main drawback is represented in limited transmission distance, complex engineering cabling, and the application that can be inflexible. Fig. 5 represents the analog-based system components from the acquisition to the storage, going through the visualization, switching, and processing.

Digital surveillance system: The analog system is the source of introduction of digital video recorders (DVR). In this stage, the computer network transmission system transmits the digital image files [41]. The cameras are connected to the video server via IP network so that, the transmission is made using existing computer LAN even Internet [42]. Using a computer terminal network, controlling the camera and zooming a part of the scene can be achieved, whereas the traditional systems, transformed into distributed ones. The difference between the analog video surveillance and digital video surveillance is at the aspects of signal transmission, control, and storage. Using the efficient video

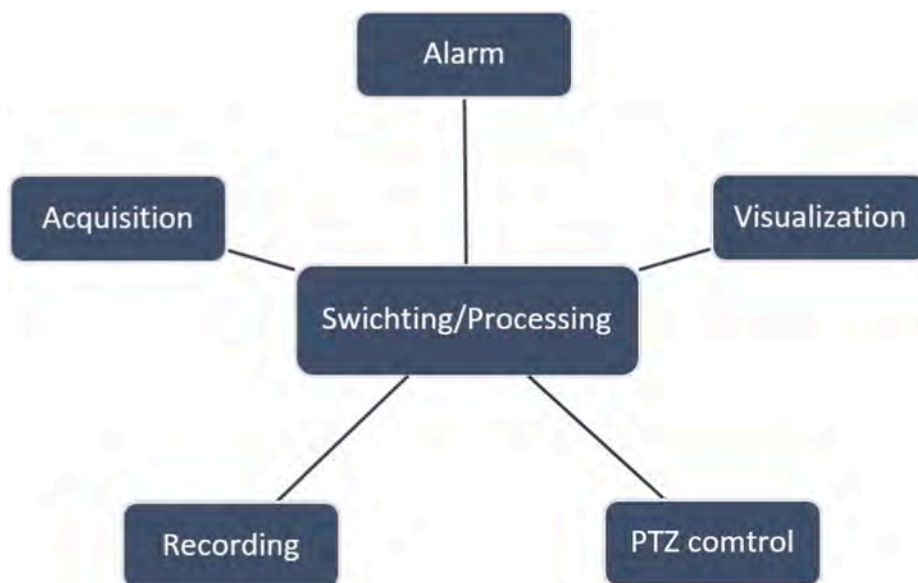


Fig. 4. Video surveillance system functionalities.



Fig. 5. Example of an analog video surveillance system.

encoding such as MPEG-4 and H264, in digital video surveillance systems, image monitoring could be performed with a remote transmission in all existing digital transmission networks and with low bandwidth. Also, digital video surveillance ensures security, and remote transmission makes the applications more flexible. Unlike the analog video surveillance, digital video surveillance mainly accounts on signal processing innovations. Fig. 6 describes the architecture of a digital video surveillance system with a representation of all component of this system including cameras, servers, and the connection via an IP network.

Network surveillance system: The network system is based on digital signal processing and involves network cameras or IP cameras which is a type of digital video camera [43]. The system can connect as many IP cameras as needed directly to the IP network [44]. Networking technique is used to realize signal transmission, exchange, control and video storage. It can, also, perform a centralized control and management on all the devices (cameras,sensors). Fig. 7 represents the architecture of a network video surveillance system with all devices and the connection method between them.

Wireless Sensor Networks (WSNs): Network video monitoring based on network architecture improve the standardization of digital video surveillance and the integration part. Using the wireless sensor networks (WSNs), the video surveillance systems become more developed and handle many problems on this topic. In Wireless Sensor Networks (WSNs), sensor networks are highly distributed networks, lightweight wireless node, deployed in large numbers to monitor the environment [45]. In addition, sensor nodes gather information in diverse settings and send information to base stations. But, they have limited resources while the base station has more computational energy and communication resources. The base station thus acts as the gateway between sensor nodes and the end users. The individual nodes in a WSN are inherently resource constrained [46].

The WSNs are built of nodes. Each node connects to one or several

sensors. These sensors monitor physical or environmental conditions, such as temperature, sound, pressure, etc. They are capable of taking the reading and passing their data through the network to a main or central location. They are self-configured and infrastructure-less wireless networks. There are two types of wireless networks, unstructured wireless networks and Structure wireless networks. The structure of a WSN includes various topologies like star, tree and mesh. Star topology is a communication topology, where each node connects directly to a gateway. In tree topologies, each node connects to a node that is placed higher in the tree, and then to the gateway. The Mesh topologies allow transmission of data from one node to another, which is within its radio transmission range.

In case of traditional wireless sensors networks, the sensor nodes collect information about simple environment data such as temperature, pressure, humidity, etc. around them. The simple data which are acquired by traditional sensor network provide low information which is inadequate for environmental monitoring. This compels a development of sensor networks with nodes equipped with very low power cameras. These camera-nodes have the ability to capture images of observed areas. This technology provides exact information for environmental monitoring. **Wireless Multimedia Sensor Networks (WMSNs)** or Wireless Video-based Sensor Networks can be deployed quickly and provide accurate and real-time visual data. Also, can provide more detail and precise information while reducing the cost. WMSN consists of a set of sensing nodes, called video nodes which are equipped with video cameras and transceivers. Wireless sensors that gather scalar, as well as audio-visual data, are used in various applications including health-care monitoring monitor public events, private properties, and borders. They can be used to enhance and complement existing surveillance systems against crime and terrorist attacks [47]. A Video Sensor Network is a distributed system of a large number of camera nodes. The image data are captured and processed by the randomly distributed camera nodes. The nodes can extract relevant information, collaborate with other cameras and provide the user with relevant information about the captured scene. WMSNs have captured researchers attention in the past few years [48]. WMSNs have been proposed to enable tracking and monitoring of events in the form of multimedia, such as imaging, video, and audio. These networks consist of low-cost sensor nodes equipped with microphones and cameras [49,50]. These nodes are interconnected with each other over a wireless connection for data compression, data retrieval and correlation. The availability of low-cost cameras and audio sensors has led to the development of Wireless Multimedia Sensor

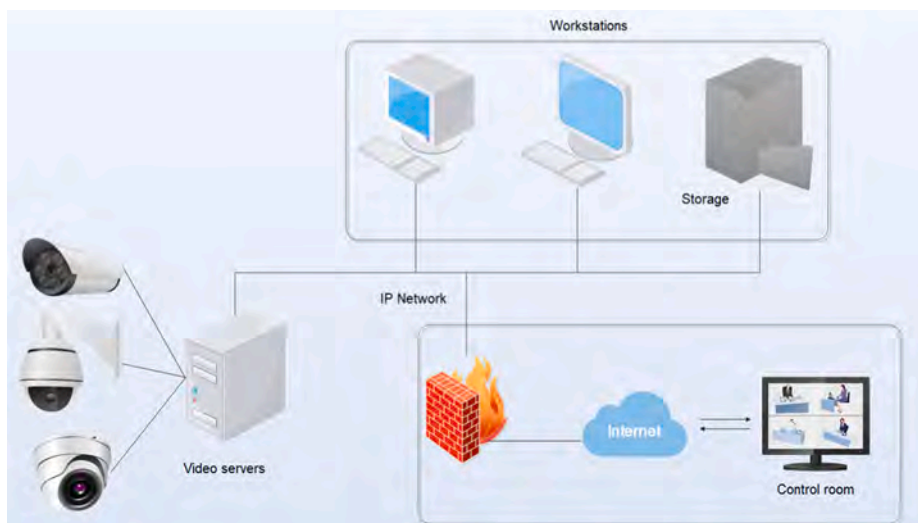


Fig. 6. Example of a digital video surveillance system.

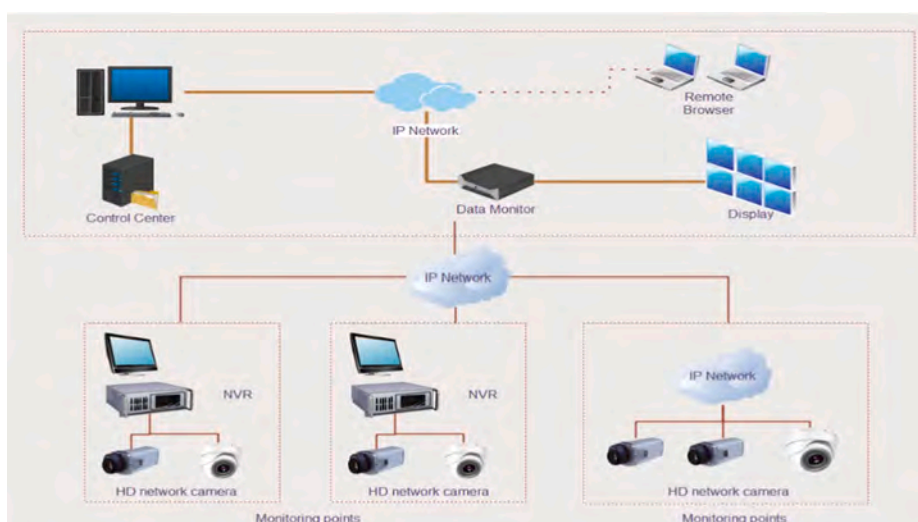


Fig. 7. Example of a network video surveillance system.

Networks. This has enhanced the traditional wireless sensor networks.

5. Surveillance cameras

Using internet and computer network, IP Cameras can receive and send data. Centralized IP cameras and Decentralized IP cameras are two types of IP cameras. The main difference is that the first one needs a central Network Video Recorder (NVR) to record data, and the second one can directly record data to the storage media. An administrator has the ability to monitor multiple cameras from one remote location. Another feature is that the quality of the video does not diminish compared to Pan-tilt-zoom (PTZ) features. PTZ cameras can provide great details when zooming even in a wide area coverage. EPTZ or virtual pan-tilt-zoom is an other type of cameras [44]. These cameras are used in surveillance, video conferencing and used to produce recordings for surveillance purposes. They can be either video cameras or digital stills cameras. CCTV can record the video of the place being monitored and can be reviewed later. IP cameras capture a wider field of view than analog cameras. This technique comes with good benefits: (1) video access flexibility, (2) local and remote access support, (3) easy installation of the cameras to the network and (4) affordable wireless cabling for cameras. By connecting to the central platform using a network, the

user can use the monitoring resources.

Video surveillance cameras can be installed either for viewing public scenes such as train stations, or for private areas such as houses. These monitored areas can determine the type of cameras that can be used. For example, outdoor cameras are equipped with a cap to protect them from environmental factors. Cameras can be classified into two categories: simple cameras and night vision cameras. The first category operates in daylight when the space is lit up, the second category operates day and night, thus taking into consideration the illumination conditions. The cameras can further be distinguished according to the resolution criterion i.e. when the resolution exceeds 700 p it is indeed a HD camera which has a very high resolution that allows to film more details. The connection system and the recorder are among the characteristics to be taken into account when choosing cameras. Cameras can be connected to the recorder either by a cable (Analog Camera) or through wireless networks (IP Camera) [51]. The last one, is a surveillance camera that uses the internet to allow the transmission of captured images. The main advantage of this type is that it allows to monitor their installations anywhere, as long as they are connected to the Internet or a wireless network. The following describes some types of video surveillance camera. The first category is the fixed IP camera, unlike the motorized camera, the fixed IP camera only monitors one particular location. The

motorized dome IP camera can be installed anywhere especially on a wall or ceiling. Their particularity is that it can be remotely controlled. They benefit, not only zooming, but also rotate 360 degrees, which can be convenient for sweeping around.

Another type of video surveillance cameras are the PTZ IP camera and PTZ dome, allows horizontal movements as well as zooms. a use case would be that the mobility provides possibility to follow a suspicious person. The PTZ dome camera allows to monitor the scene by moving on a 360-degree angle. The vandal-proof IP camera, as their name suggests is designed to withstand external aggression, such as vandalism and tear-off attempts. They have special surfaces that resist breakage and impact. This can be most appropriate to protect valuable places. The last type is infrared IP camera, this monitoring device makes it possible to film the elements in the most absolute darkness with a precision of up to 30 meters. The infrared camera is convenient for shooting at night and protecting against thieves and burglaries.

In order to improve the quality of acquisition and handle the environment changes, many companies are increasingly inclined to invest in a surveillance camera. However, there are different types of surveillance IP cameras. Fig. 8 represents the existing types of IP cameras including outdoor, indoor and PTZ cameras. Each camera having a different technical aspect [52]. Thus, a PTZ camera will not be used in the same way or for the same purpose with respect to a dome camera. Table 1 represents many categories of cameras with technical characteristics.

With the development of machine learning algorithms, it has been developed in terms of accuracy and computational time of video surveillance tasks. Such that some intelligent cameras have been built for computer vision tasks face recognition and object detection, that need real-time implementation. These cameras are already commercialized and available in the markets.

The growth of video and image analysis using machine and deep learning algorithms make the cameras more developed and can contain many tools and features that make the monitoring performance more robust. Recently some cameras can recognize the face of the staff at the entrance of an office. Hence, without human-made check-in, the staff will be automatically checked in the entrance. This technique is becoming popular support for theft prevention in the markets, automate airport check-ins, and wanted criminal capturing.

Using deep learning algorithms for face analysis, the camera can provide an identity verification with high accuracy. Exploring the stored dataset of the faces, the camera with face analysis can recognize the face using the percentage of the best match. In addition to surveillance purposes, intelligent cameras with face analysis are used in many commercial applications including banks, healthcare, and IT services. For bank services, many banks in China and France have begun with training the system for use of facial recognition in ATM services. Instead of using a debit card to withdraw the money, face recognition will allow carrying out any bank services using ATM ^{1 2 3}.

Despite the developments in this area, there are some limitations of these cameras that are similar to the face recognition limitations, such as lighting changes, hairstyles change and the face with glasses.

In the last years, intelligent transportation systems gained special attention. Also, the increase of vehicles circulating over cities led to the necessity of a system that can ensure safety in roads [53]. These systems are addressed to analyze the extracted data from CCTV videos and monitor the vehicles and the driver's behaviors [54]. For that, some computer vision algorithms are added to the cameras including the detection of target objects. These intelligent cameras provide useful information, like speed and identify the occupants. For the same

purpose, some companies developed a camera that can detect the vehicle license plate using an optimized object detector algorithm. The detection is the first step for recognition and tracking of the detected vehicle by other cameras ⁴. However, such technologies need to enable surveillance in challenging lighting conditions where other object detection methods often fail.

6. Video surveillance Analyses

Among the principal tasks of an automated surveillance system is to observe the surveilled environment using cameras, detect, and qualify easily the observed scenario and then raise alarms. The task can be, motion detection, objects tracking [55,56], objects detection, individual interactions and behaviors, person re-identification [57–59], crowd detection and management [60], etc. Fig. 9 represents the computer vision task related to video surveillance system. There are many other tasks than the quoted ones, because in an intelligent system we can add functionalities to obtain a good performance [61]. In this field and especially when a surveilled area that is crowded, we can extract a lot of information that can be useful for security of people, but this extraction can be difficult when the scene is complex. In addition, objects recognition and tracking pose a problem mostly in the environment with complex scenes and this is the main difficulty in computer vision for all the tasks. (see Fig. 10).

6.1. Challenges

The Analysis and the interpretation of a surveilled area is not a trivial problem for many reasons. First, the problem of determining the object of interest is very dependent on the application [62]. For example, people are objects of interest in video surveillance applications while in a traffic monitoring application on roads, people do not provide useful information. In addition, there are many problems due to the variability of the scene: the size of the objects varies from one object to another in the same frame (for example a van and a person), objects have different colors and the colors can be similar to the background leading to "hidden objects", the speeds of the objects are very variable, a block can contain more than one object of interest, etc. All of these analyses has a set of constraints that can be the source of the bad results which can produce false alarm [63]. By the following table (Table 3), we will mention some classic problems that appear in the almost all analyses for video surveillance systems.

The analysis in a video sequence is a very important line of research; it has been the subject of several works over the last twenty years. This analysis extracts important information for using it in detection, estimation, tracking or recognition. In addition, motion analysis assists several applications in various fields: television, surveillance, robotics, etc. The importance and diversity of work in this context is the subject of a lot of methods in the literature. In this section, most important tasks of video surveillance system are described.

6.2. Camera calibration

Collecting and integrating data from a multi-camera system is an important task in computer vision, spatially for video surveillance systems. In these systems, cameras do not have a common field of view (FOV). The integration of these data is made using camera calibration techniques that is a very challenging task. Camera calibration differs from one method to another based on the problem required to be solved [64]. Hence the accuracy of calibration methods represents a challenge [65]. Several methods have been proposed for this purpose and can be found in the literature.

The projection transformation between cameras represents a critical

¹ <https://www.atmmarketplace.com/press-releases/what-is-face-detection-and-how-can-banks-use-it-to-protect-atms/>

² https://www.nec.com/en/press/201909/global_20190912_01.html

³ <https://www.finextra.com/newsarticle/33388/caixabank-rolls-out-facial-recognition-at-the-atm>

⁴ <https://fitoptivis.eu/smart-camera-with-embedded-object-detection/>



Fig. 8. Video surveillance cameras. (a) Fixed camera. (b) PTZ external camera. (c) PTZ internal camera. (d) Bullet camera. (e) Dome camera. (f) Turret camera.

Table 1
Characteristics of video surveillance cameras.

Type	Resolution	Zoom	Environment	Audio/ Video	Day/ Night	Advantage	Inconvenient
PTZ 720 × 1080P Outdoor video	2MP ×10-×8 Audio Day	Indoor Cover a large area					
Ability to rotate Tilt and zoom Bullet 2560 × 1440P outdoors video	Affected by low light 4MP Not allowed Audio Day	indoors					
Night	No glare from overhead light due to superior overflow	Capture images at a given location					
Dome 3072 × 1728P outdoor(limited) video	5MP ×4 Audio Day	Indoor Resistant to dust, cobwebs, and vandalism.	Easily seen and recognized				
Turret 1920 × 1080P outdoor video	4MP Not allowed Audio Day	Indoor					
Night	-Better infrared treatment than a dome camera						
-Precision of the angle of view	Easily vandalized						

case for calibration task [66]. This problem can be handled by adjusting certain parameters [67]. In order to find the best estimation, the existing camera calibration methods can be classified into 8 categories. In Table 2, the description of each category is presented as well as a set of methods of each category. For the trajectory-based methods, the transformations remain unchanged independently of the movement of the camera [68–71]. The trajectories of camera movements are estimated during the capture of the image sequences. Using these trajectories all methods compute the extrinsic camera parameters. The Slam-based methods used the improvement of SLAM techniques to estimate the extrinsic parameters [72–75]. 3D reconstruction techniques of a scene are also used to learn the pattern of calibration. In addition, 3D

geometry registration formed for each camera is exploited to obtain the transformation between cameras. In order to reflect the calibration pattern to all cameras, a mirror-based method is used. The mirror facilitates this reflection and also makes the transformation estimation between cameras [76–78]. The mirror pose can be estimated easily when the calibration pattern and the camera are also fixed. In addition, computing the transformation between the camera and the pattern can help to obtain the poses between cameras.

Tracking-based methods try to cooperate between different cameras to track the objects moving in the scenes [79,80]. These types of methods are good for surveillance systems that use multiple cameras connected to a network. The poses of the cameras are estimated using

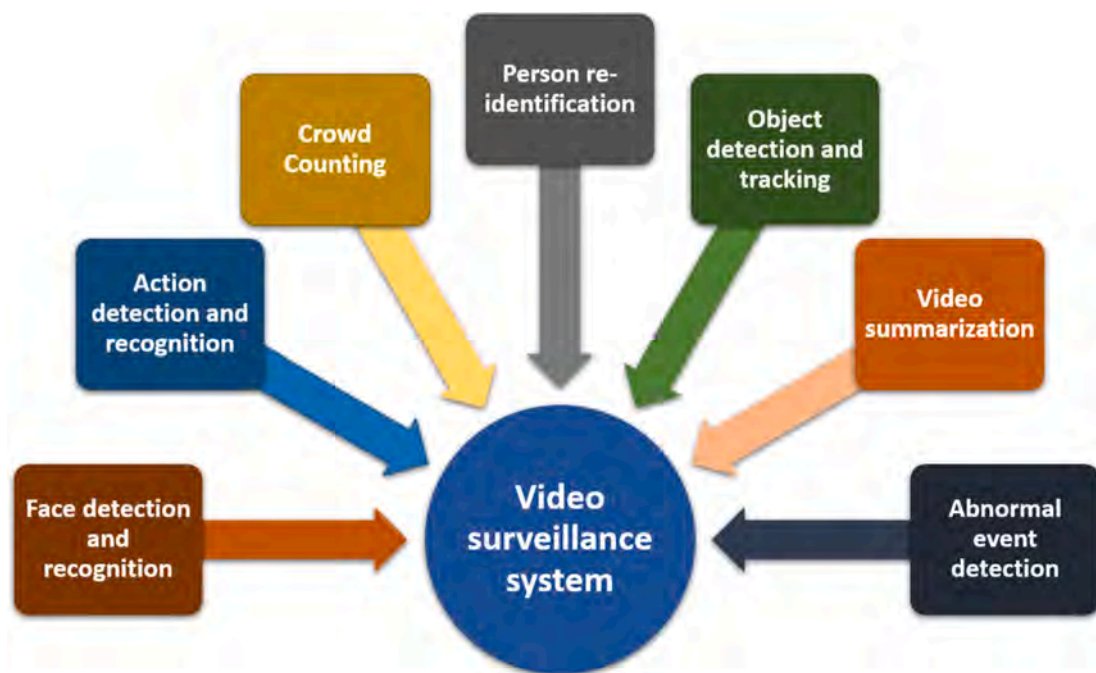


Fig. 9. Video analyses for VSS.

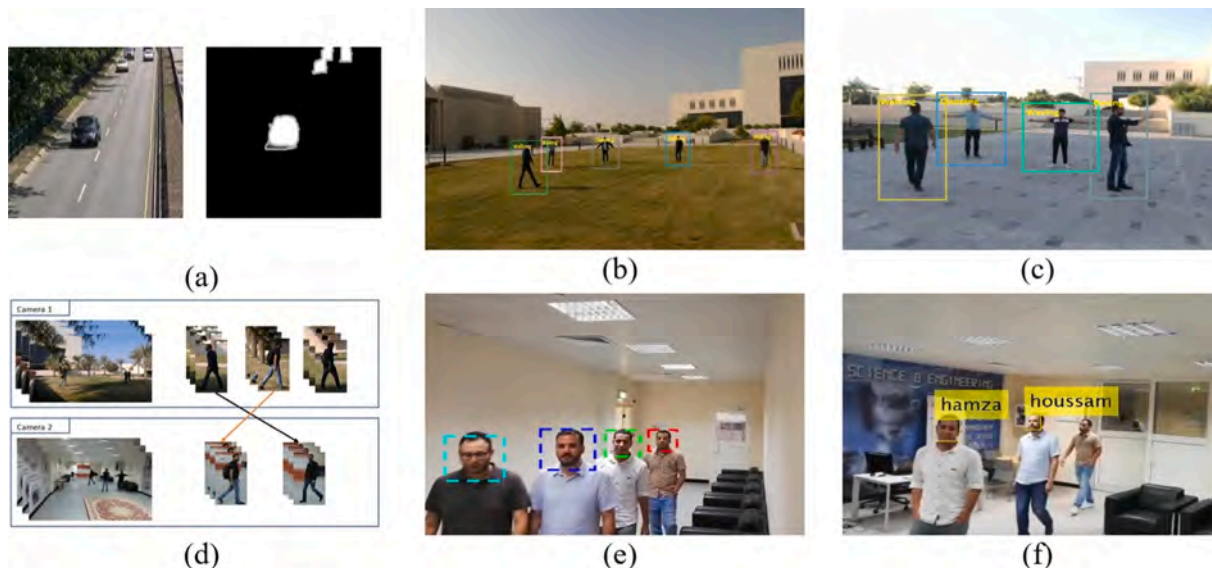


Fig. 10. Computer vision tasks related to VSS. (a) Motion detection. (b) Object detection. (c) Action detection and recognition. (d) Person re-identification. (e) Face detection. (f) Face recognition.

the prediction of motion between cameras. The most common case in camera calibration is the calibration of a stereo camera that requires an estimation of extrinsic parameters. The field of view (FOV) is common for cameras stereo [81–83]. Thus, the extrinsic parameters can be estimated using objects and markers in the same FOV. Theodolites, laser trackers, and laser rangefinder are the most common techniques used for methods based on large-range measuring devices [84–88].

For verifying the precision of camera calibration and extracting features, the calibration target is used [89–91]. It allows good and easy features extraction. Some authors propose calibration method based on a compound target consisting of two planar calibration targets that are fixed together [92,93]. To calibrate multiple cameras some authors used the motion in scene [94–100]. In addition, in order to establish the correspondence between different cameras, they track the movement of

targets in the scene [98,99]. The use of the target motion can handle the problem of lack of image correspondence points [100]. Unlike the cited techniques, the laser-projection-based method is not sensitive to light, has a larger measurement range and widely used for methods of global calibration [101–105]. These methods suitable for the calibration parameters takes into consideration the size of laser pointer, its accuracy of distance measurement, and the portability. The external parameters of the camera can be extracted using a single captured image, if the internal parameters of the camera and the spatial coordinates of multiple feature points are known. In order to obtain the spatial coordinates of the feature points, some authors used the technique of 3D reverse engineering [106]. Thus, vision-based measurement devices such as close-range photogrammetry systems, and hand-held scanners can be used for global calibration of non-overlapping cameras [107]. The close-

Table 2
Video Analysis problems.

Problem	Description
Waving tree	Due to the motion of some parts of the image that are not interesting. Typical examples are the leaves of trees or power lines that move with the wind.
External or internal environment	The object detection phase is influenced by the phenomenon of sudden change (change of light) or gradual illumination. These changes change the appearance of the background.
Camouflage	Occurs when certain objects of interest have their appearance (e.g. color components) similar to the background.
Opening foreground	Occurs when the homogeneous color object moves, changes in the inner pixels cannot be detected. Thus, the entire object may not appear as foreground. The effect of this problem is that slow moving objects are not recognized or very few pixels of objects are recognized.
Static objects	When an object becomes immobile, it must be considered as part of the background because it is not yet an object of interest. Such an object, which is still recognized as an object of interest in motion by the detection step, causes problems.
Shadow	Especially in external environments, it is likely that the shadow of an object is considered by the detection algorithm to be part of the object because its intensity is usually different from the corresponding area of the image of the object reference.

range photogrammetry, one of the most important methods for modeling large-scale 3D objects and scenes, has served as a key solution to the extrinsic calibration of camera networks [108]. In the field of close-range photogrammetry, it is common to use easily identifiable encoded targets for multi-view matching of unordered image sets because there are seldom sufficient feature points on the surface of an object.

6.3. Motion detection

Motion detection is a very important task for many computer vision applications especially in video surveillance systems analysis. Its purpose is to extract moving objects at a time t in a video sequence [109]. Motion analysis systems that serve to focus attention on the moving parts of the scene. Three typical approaches are mainly used for motion detection: time difference, background subtraction, and optical flow analysis. Most of the motion detection algorithms found in the literature are background subtraction methods; they usually follow three steps [110]. Firstly, it involves developing a background model of the scene, which requires that the camera must be fixed [111]. The second step calculates the absolute difference between the current image of the sequence and the image of the background, then applies thresholding

Table 3
Comparison of camera calibration methods.

Category	Methods	Advantage	Inconvenient
Trajectory	[68–71]	High accuracy with normal scenes	Accuracy can be affected by the camera-motion estimation. Can suffer from degenerated cases
SLAM	[72–75]	Does not suffer from degenerate cases. Combining with trajectory-based method can give a high accuracy	Scenes must be initialized first.
Mirror	[76–78]	Not Expensive, tight space.	Low Accuracy, measuring range is small
Tracking	[79,80]	Easy relative poses calibration of a large camera network.	Low accuracy
Marker	[69,81–83]	High accuracy, simple implementation.	Sensitive to noise
Large-range measuring devices	[84–88]	High accuracy, large work range	Expensive, bulky, inconvenient to move
Large-scale calibration targets	[89–93]	Low cost, good flexibility	Algorithms used are complex, cumulative error
Motion model	[94–98] [99,100]	Measuring range is large, expensive equipment is not required	low accuracy, influenced by the state of motion
Laser projection	[101–105]	Tight space, large measuring range	Risks of human eye security
Visual measuring instruments	[106–108]	High accuracy	Flexibility cost is not low due to the need to purchase commercial equipment

operation to decide if a pixel belongs to the background or to a moving object [112,113]. Some authors propose an advanced approach based on a counter-propagation artificial neural network to achieve effective moving-object detection in such conditions[114]. In prison, most causes of death are suicides by hanging[115]. The detection of this behavior can help to minimize the suicidal rate and can help the security agents to help those attempting suicide, by preventing the act of hanging. In [116], the authors proposed a method for suicide-by-hanging detection. Using RGB camera the video surveillance system can predict the act and start alarming. This method exploits the position of the body to identify the behavior of suicide.

The main difficulties of this approach lie in the fact that, even in controlled environments, the background undergoes a continuous change, mainly used for the existence of variations of lighting (example: passing clouds, tree branches moving with the wind). The robustness of scene dimming is achieved by using adaptive background models.

6.4. Objects classification and recognition

The various intelligent monitoring applications aim to extract the semantics of video to be used in other tasks [117–119]. For that, distinction between objects detected is a crucial challenge. Recognizing information about an object can help such method to improve the accuracy [117]. The classification of moving objects is the second step after the detection, which also can be the preliminary process to track these detected objects. The captured videos may contain different types of objects such as plants, people, vehicles, animals and natural phenomena. The classification can be used in the counting and recognition of objects via categories (people, animals, vehicles). Therefore, the classification is a semantic categorization of objects using features. However, the surveillance applications focus usually on humans and vehicles. In addition, its require a identification system that can be inexpensive in computation, effective on small targets, and invariant to illumination changes [118].

Classification methods generally fall into two distinct sub-issues: supervised, also known as discriminant analysis, and unsupervised classification, also known as automatic [119]. The first approaches, algorithmic, heuristic or geometric based essentially on the dissimilarity between the objects to be classified [120]. The more recent statistical approach is based on probabilistic models that formalize the idea of class. This approach also makes it possible to interpret the classification obtained statistically. On the other hand, since data acquisition processes have also progressed rapidly, the size of the data to be studied has become very large. Today’s scientific world provides data that are every day more numerous and larger [121].

The proposed methods satisfy most requirements by implementing a classification system that categorizes detected objects into predefined

groups of human groups, vehicles using image-based features. For that, many research papers have been made in this field. Najva et al. proposed a new method to detect and classify the moving objects combining SIFT and Tensor features using Deep Neural Network (DNN)[122]. The SIFT algorithm is used to handle the illumination and rotational changes. In [123], the authors propose an edge-preserving-based collaborative representation (EPCR) classifier, which overcomes this problem by using the edge image estimated by the original full-band hyperspectral image. The estimated edge image is used for calculation of the weights of neighbors and the final residuals in the collaborative representation classifier. The advantage of a multi-scale spatial window is also assessed in this work. Fig. 11 represents all categories of methods used in classification, describing the accuracy, the computational time and limitations.

6.5. Objects tracking

Visual tracking is one of the richest tracks of motion analysis; it plays an important role in numerous computer vision applications such as traffic monitoring, motion recognition and video surveillance applications [124]. Much progress has been done in recent years in visual tracking [125–127], but it is still a stimulating problem, even using deep learning techniques to construct a robust process covering the most of the challenging factors such as occlusion, motion blur, in-plane and out-of-plane rotations, illumination variation, scale variation, non-rigid object deformation, camera motion, and background clutters, to name a few. The purpose of tracking is to determine the spatial and temporal information of each target. Since the visual movement of the targets is always small in relation to their spatial extent, no position prediction is necessary to build the shots [126]. The association of regions and their classification is based on a binary association matrix, which is calculated by testing the overlap of regions in consecutive images. Whenever there is a match, the line is updated. The tracking also interacts with the detection. When the object of interest waits in the scene for a while, the tracker merges with this object in the background. The purpose of object tracking is to match objects or parts of objects in consecutive images and extract the information about objects such as trajectory, posture, speed,

and direction. This is a crucial part of intelligent monitoring systems because without object tracking, the system cannot extract temporal information about the objects that make the analysis of the behavior impossible [127]. On the other hand, the inaccurate segmentation of the object is a big problem due to the illumination changes and other problems in computer vision. With the development of Deep learning techniques, objects tracking method become more accurate in real-time.

In general, a typical visual object tracking system consists of two basic components [128]. The first component is the motion model that foresees the likely motion of the tracked target, the second one is the observation (or appearance) model, which is considered as the most important component because of its crucial role in representing the tracked target appearance that it may undergo several issues. Depending on the nature of the observation model, visual tracking methods can be classified into three categories [129–132]. These categories are presented and described in Table 4.

6.6. Human action recognition

Human action recognition is another specific application of video content analysis which aims to recognize activities from a series of observations on actions of subjects and the surrounded environment and it is important for many other applications [133]. However, human action recognition is a complex technology due to the difficulty to extract information about person’s identity and their psychological states. Hence, human action recognition has gained the interest of researchers in the computer vision area and many approaches have been proposed for the aim of developing this technology. To improve the action recognition performance, recent works have employed various deep learning models [134,135]. Since human actions are extracted from multiple movements of human body or parts of it, it is necessary that the recognition process should involve video browsing over time to learn the patterns of the visual appearance changes [136]. To achieve this, existing deep learning models based on 2D convolutional networks can be extended into 3D domain to capture the temporal information [137]. For example, the authors in [138] use the motion in consecutive frames to extract information for action recognition using a two-stream ConvNet architecture

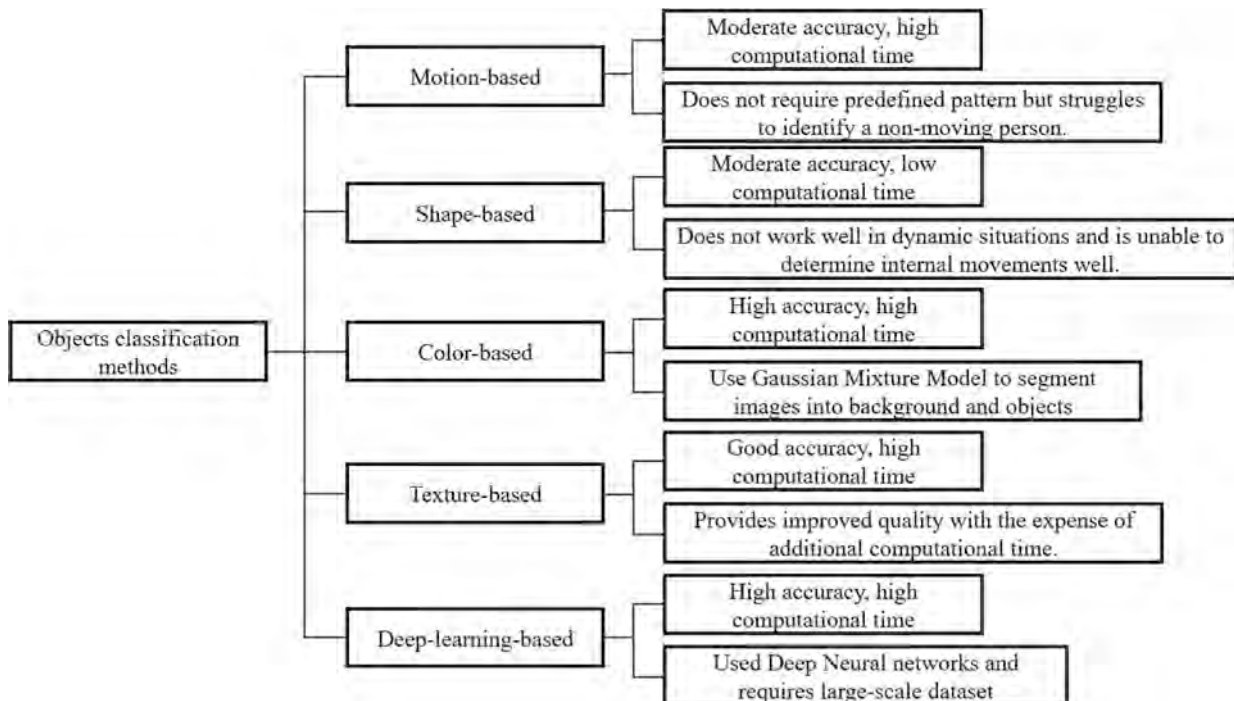


Fig. 11. Object classification methods.

Table 4
Object tracking methods.

Category	Method	Description
Discriminative	Kalal et al. (2012) [128]	<ul style="list-style-type: none"> • Focus on finding a decision boundary to separate the tracked target from its surrounding background. • Performs better if the size of the training data sets is important.
Generative	Zhuang et al. (2014) [129]	<ul style="list-style-type: none"> • To learn a visual model that represents the appearance of the tracked object and search for image regions which are the most similar to the tracked object • Obtains satisfactory results when the available data are insufficient
Hybrid	Cheng et al., (2015) [130] Zhong et al. (2014) [131]	<ul style="list-style-type: none"> • Combines between discriminative and generative appearance models • Discriminative tracker is implemented to consider the overall information representing the object appearance [130] • A robust tracker based on sparse collaborative appearance model (SCAM), in which a sparse discriminative classifier (SDC) is developed using holistic templates to separate the tracked target's appearance from the background [131]

that incorporates both spatial and temporal networks.

6.7. Face detection and recognition

The use of faces is important for many applications including surveillance systems, human-machine interaction, and airports. Considering the face as a unique identification for each human being, the summarization via faces detection and then identification can be useful to identify the anomalous behaviour in locations like that of a stadium [139]. Further, the detection of the existing face in a scene can help manager of the security to identify the person of interest [140].

Face recognition has been increasingly studied by computer vision community for the last decades. Compared with other popular biometrics such as fingerprint, iris, and retina recognition, face recognition has the potential to recognize uncooperative subjects in a non-intrusive manner. Therefore, it can be applied to surveillance security, border control, forensics, digital entertainment, and so on. Indeed, numerous works in face recognition have been reported and significant developments have been deployed, from successfully identifying criminal suspects from surveillance cameras to approaching human-level performance on the popular labeled face in the Wild database. In order to handle pose-invariance for face recognition, several CNN-based techniques have been proposed by researchers in the literature. Related to this, techniques that handle this variation and its impact for good recognition of face are being reviewed. In the past, the authors introduce the face with different poses in the training database [141]. Recently, automatic learning using trained CNNs allows learning a diverse set pose invariance using large datasets. The authors in [142] collect a very large dataset of 2.6 million images and propose a CNN model inspired by [143]. The authors do not handle the pose-invariance problem, but instead they merge different face poses in the training datasets. The same idea was adopted in the method proposed in [144] that represents a transfer-learning-CNN model for face recognition. In order to make the features more compact and discriminative for unconstrained face recognition, Zhang et al. [145] proposed two loss functions to supervise CNN model. LeNet, CNN-M and ResNet-50 CNN models are used to analyze the effectiveness of the proposed approach. For the recognition of faces, researchers in [146] attempted to merge two different face images from two different sensors, i.e., a normal camera and a thermal

camera. The purpose was to learn invariant features from the two types and produce labels for each. Even the development reached in face recognition, but the performance of such method still depend to characteristics stored data as well as the environmental and appearance changes.

6.8. Crowd counting

For computer vision applications, crowd analysis is a challenging task to deal with natural crowd complications [147]. The importance of crowd analysis could be seen in a set of applications, including football crowd surveillance, concerts etc. that can contain thousands of people [148]. Detecting faces, actions and events in the crowded scenes can be useful for other computer vision tasks, as well as for people security systems.

Therefore, accurate and reliable detection of people is difficult when it is applied to visual analysis due to fewer pixel per target, perspective effects, high density with heavy occlusion, a different variation of poses, variable appearance, strange clothing and different camera orientations. The high-density crowd may lead to fallacious classification of a person which results in false detection. In mass gatherings, the human body may be partially or fully occluded [149]. Face is the most visible part of body which gets captured in the images since cameras are fixed at a high altitude for better surveillance. For that the faces and heads are mostly used for estimation of the number of crowd using crowd counting methods.

The crowd counting approaches can be classified into four main categories: detection-based methods, regression-based methods, density estimation-based methods, and convolutional neural networks based methods [150]. The first three categories are generally not compatible with the CNN-based method for crowd counting due to the poor accuracy and high computational cost [151]. We focus, in this section on the existing CNN-based methods. The use of a Convolutional Neural Network (CNN) improves the accuracy of the most computer vision fields. For Crowd counting most existing methods are CNN-based, where the generation of density maps and the estimation of the number of crowds are made with high accuracy [152,153] unlike the regression-based methods which simply extract the density information of pedestrians without finding the position of each person.

In order to handle the previously quoted challenges especially the scale variations, many methods have been proposed. For example, the authors in [154] proposed Hierarchical Scale Recalibration Network (HSRNet) to recalibrate multiple scale-associated information and modeling rich contextual dependencies. In the same context, the authors in [155] proposed a fusion-based method for crowd counting named SRF-Net. To deal with scale variation in a complex scene the architecture proposed has two-stages: a band-pass stage and a rolling guidance stage. In order to adapt their model to a target scene and with the use of a few labeled frames, the authors in [156] proposed a method for model parameters learning for a fast adaptation to the target scene. The method based on VGG16 for features extraction followed by a density map estimator block, where images sequence is used for the training.

Using deep learning architectures and some large-scale datasets, the estimation of crowd density becomes more accurate. However, the achievement of the precise number of crowds in complex scenes with different scales is still a challenging task.

6.9. Person re-identification

Across multiple cameras, person recognition and identification are important targets for many computer vision applications, especially monitoring systems [157]. The operation of recognition of a person from a set of images captured by several cameras is called person re-identification. The similarity measures can be the key to computing the matching between two or a set of images. However, the re-identification using video clips can be a problem for many

applications [158]. For example, people tracking across multiple cameras. The video sequences captured by different cameras should be analyzed in order to re-identification of the person and keep tracking him across all cameras in the surveilled areas. The sequential methods that use a list of features were not be enough for the re-identification of persons according to the many problems in this case including the differences between the analyzed objects in terms of shape, colors, scales, and others [159]. Therefore, the use of a limited number of features cannot be enough for proper identification. With deep learning techniques, the use of different and not limited features for learning becomes a good alternative for solving person re-identification problems. For that, and in order to train these methods, a large scale of data is required from multiple camera's views [160]. In addition, using pre-processing techniques can help the performance of the learning model to improve [161].

In addition to appearances-based re-identification, Human identification features, such as face and gait have been widely used for person re-identification across multiple cameras[162]. However, they require input images captured in a highly controlled environment with good quality imaging devices. Gait recognition, on the other hand, is a crucial person identification technique, which makes use of computer vision techniques combined with unrestricted use of specialized devices such as sensors, to identify humans through their gait pattern, or touring style. Human gait is generally observed to be a uniquely human characteristic that is difficult to replicate or hide. Hence, it forms a critical bio-metric identification technique. For more details about person re-identification using gait recognition can be found in [163].

6.10. Video summarization

The growth of video technologies has led to the creation of efficient tools to manipulate this type of data. Summarization aims to generate a short version of a video as a representation, using keyframes of important subsequences. This summarization provides a rapid view of the information contained in a large video. It also provides a good evaluation for users of the video and provides knowledge regarding the topic and the most important content in the video. Considering the information contained in each video, many methods have been developed using several techniques. Each technique summarizes the video using a specific feature, such as trajectories, moving objects, abnormal detection, and many others. These categories of techniques can be classified into two general categories, scene-based (i.e., static [164], dynamic [165] and content-based approaches, and the content-based approaches can be further decomposed into three types related to the content of the video including motion-based [166], action-based [167] and event-based [168]. Video summarization is a short version of the longer video sequence. The static video summarization is a collection of frames (keyframes) selected from the original video. The proposed approaches extract the keyframes using many features. The video summarization can help video surveillance system to review just the useful subsequences. But the purpose of the summarization is the key for best method selection.

7. Discussion

In Any video surveillance systems (VSS), there is two main components: the architecture component, and the softwarization component. The first component or the external task of a VSS consists of the acquisition and transmission of data until it is received in the main station for recording. All the proposed architectures represent the techniques of getting the data to the central station to be analyzed. These tasks are based on the signals processing and networking technologies. The differences between all these architectures are related to the purpose of each system as well as type of device used. Each system can be built to cover a specific area (indoor or outdoor) and can play both roles. New technologies are used by some system to overcome the difficulties

related to the environment that use this system. Among these technologies include Cloud computing, Fog computing, Edge computing and the internet of things. For example the storage of data using Cloud computing is used in [36,64,25,27] while the authors use fog computing in [24,26] and edge computing in [23,27]. The researchers try to solve many problems in VSS related to the environment based on whether the system will be used for a home or business, indoors or outdoors, also effecting which type of camera to be used. In Table 5 we represent some proposed architectures of video surveillance systems that exist in the literature. The descriptions of the advantages and the limitations of these systems are also quoted here.

The second component of the video surveillance systems is the interpretation and the analyses of the received information. In order to analyze the data, there are many tasks according to the surveillance purpose, such as motion detection, tracking behavior analysis, objects classification people and vehicle counting and many others. For all these goals, the data recorded from a fixed or moving camera, installed in public or private area, before being analyzed. In addition to the video surveillance architectures that is summarized in Table 5, we describe here the analysis tasks. These analyses can be grouped into three categories based on the correspondence between them and the degree of complexity. Foreground detection, objects classification and feature extraction are used for moving object detection while the camera calibration is classified as a low-level tasks. These tasks can be applied and extracted from each camera. The results of these processes can be used by other tasks that are considered as mid-level tasks. These tasks are more complicated and can be extracted using more than one camera. The information obtained by the multi-cameras are exploited to track an object while its moving in between scenes covered by many cameras. In addition to object tracking, several tasks that important for surveillance include moving objects detection and classification, and faces recognition. In addition, the audio analyses can provide a good support in many situations. The use of multi-camera-based analyses can increase the effectiveness of a video surveillance system. The most complicated analyses and the important ones that provide a good understanding and interpretation of any surveillance system are: behavior detection and recognition, event and anomaly analysis, video and activity summarization. In addition, there are some tasks related to the camera such as control of the camera (moving and zooming it). The efficiency of each one of these analyses represents an important parameter in the robustness of a video surveillance system. Table 6 we represent a summary of these analyses.

8. Future trends of video surveillance systems

The researchers proposed many works related to the automatic video surveillance systems in the last 20 years. They tried to handle the most challenging issues related to the architecture of these systems as well as their functionalities. Studies have been conducted on many video surveillance tasks especially the ones required to ensure security. Among these tasks are object recognition, object classification, moving object detection and tracking. Many proposed approaches and algorithms have succeeded to solve many problems and provide acceptable and effective results. Now, the work is still in progress and many tasks need further improvement. Among video surveillance systems difficulties there is two that are common with all systems namely, the false alarm and the environment changes. A false alarm in a video surveillance system can cause problems in the sector, sensible to any failure of the system and for such reason; video surveillance systems usually disable the automatic alarm option. On the other hand, the environment change can produce a failure in the system, especially unexpected weather changes.

In order to handle the before mentioned limitations, some of the algorithms of video surveillance systems need to be improved by increasing their robustness and accuracy. In this context, researchers are addressing new topics to align with the emerging technologies like IoT, augmented virtual reality, Fog computing, and smart compression, etc.

Table 5
Existing architectures of video surveillance systems.

System	Environment	Sensors	Technology used	Advantages	Limitations
[36]	Outdoor	Simple camera	Cloud Computing	Pre-processing before transmission using video structural description (VSD). Semantic enhanced cloud.	Not effective to real-time react
[38]	Outdoor, Woods	360-degree Omnidirectional Camera	Quasi-connected components(QCC)	Detects and track objects in the woods	Small field of view. Coverage limited by the trees
[14,15]	Outdoor, Indoor	PTZ camera Audio sensor	Remote human signature	Video and audio analysis to detect threats based on human signature detection	Sensitive to environmental noises and noise caused by sensors moving
[19]	Outdoor, Indoor	WIFI Camera	WiMAX, Hybrid WIFI-WiMAX (HWW)	Give a remote transmission of the data even if the camera located up to many kilometers to the station	Limited by the wireless mesh WiMAX networks and the antenna alignment with the base station.
[23]	Outdoor	IP Camera	-	An unfixed system can be us flexible portable factures. Useful in many situations to ensure security	Transmission of data when the network is unstable
[24,26,169]	Outdoor, Highways	Smart Camera	Fog computing, Drones	Mobile surveillance using drones. Computing tasks accomplished directly at the edge of the network and is characterized as low latency and real-time computing	One drone to track one object. Expensive system.
[25,171]	Outdoor, Indoor	IP Camera	Cloud computing	Using Cloud computing the system can be effective for the processing of big data	Communication time cannot be useful for critical mission that need immediate decision and in real-time analyses.
[27,28]	Outdoor	Smart Camera	Cloud computing, Edges computing	Eliminate useless data for transmission in the goal to save edges device energy. Also, reduce the amount of data and energy when transmitting.	Just focused to the failures in the system. Video transmitted between cameras (edges nodes) and the cloud using clips (other than the original ones) can provide loss of data
[29]	Outdoor, Indoor	IP Camera PTZ Camera	Video homographic	Help to know all possible situation in a monitored area and can help to predict most even that can happened in the scene	Choice of technical parameters, related to each scene, depend to the environmental characteristics and used devices

Table 6
Description of analysis of each video surveillance system.

System	Tasks	Sensors	Targets
[3]	Moving objects detection and tracking	Cameras	Any object
[36]	Detection, tracking behavior analysis, event analysis, vehicle classification	Cameras	Person, vehicles
[38]	Detection, tracking, classification	cameras	Any objects
[44]	360-degree omnidirectional Tracking, 3D reconstruction	Stereo cameras	Any objects
[32]	Tracking, multi-camera calibration	Multi-cameras cooperation	Any objects
[64]	Crowd detection, people counting, predict city events, behavior simulation	Cameras	Any objects
[66]	Vehicle speed measurement, camera calibration	Cameras	Vehicles
[91,13]	Rail way tunnels surveillance, camera calibration, Cyclists in streets and roundabouts, behavior analysis	Cameras	Vehicles
[14,15]	Motion detection, voice-based detection, human signature detection	PTZ cameras, audio sensors	Any objects
[24]	Vehicle tracking	Drone cameras	vehicles
[26,28]	Object detection and tracking, video compression, violation recognition	Cameras	Any objects
[29]	Multi-video fusion, detection, tracking, camera calibration	Cameras	Any objects
[111]	Objects detection and tracking, event detection, illumination changes detection	Cameras	Any objects
[115,116]	Suicide pose estimation, motion extraction	Cameras	Person
[172]	Drone detection	Cameras, Radio sensors	Drones

Fig. 12 represents all these technologies that can be the sources of development of the video surveillance systems in the future. Each technology will be described in detail.



Fig. 12. Future trends of video surveillance systems.

8.1. Smart Compression for Video Surveillance

In current video surveillance systems, the data captured by cameras are transmitted via wireless channels. The output of each camera encodes videos using a standard video coding such as JPEG200, JPEG, MPEG, MJPEG and H.26x. Several techniques have been developed to adapt video coding with the transmission process [173–175]. To minimize the exchanged data, several compression algorithms have been proposed to improve the transmission and reduce the cost in term of data size and transmission time.

One way is to use smart compression to optimize video transmission and storage based on retention value [173]. Traditionally, the primary variables at a customer's disposal are frame rates and resolution, but with new, smarter compression algorithms designed specifically for the video surveillance systems, security users now have more options for selective transmission and storage of video footage.

H.264 and H.265 compression standards were designed for consumer electronics and the film industry and took an all or nothing approach to compression [174]. H.264 is the de facto standard in video surveillance today, with H.265 being adopted over the next few years depending on products, computing power and patent circumstances. New security compression technology, on the other hand, dynamically allocates regions of interest depending on activity in the camera's field of view. The part of the video frame containing interesting details is recorded in full image quality and resolution, while areas containing no forensic value are filtered out [175]. This ensures that important details like faces, tattoos or license plates are isolated and preserved, while irrelevant areas such as white walls, lawns and vegetation are sacrificed using smoothing. The result of this is an optimal use of available bandwidth and storage which leads to significant savings. Depending on video resolution, frame rate and scene activity it can cut bandwidth and storage requirements by half or more for most surveillance applications.

8.2. Cloud and Fog computing

Video surveillance produces millions of data (big data). The analytics and recording of this data represents a critical problem in terms of cost and access to this data [176]. Cloud computing is a good solution for Big data management. Yet, such a technique has its own limitations when it comes to communication time between edges and servers, and where the connection is not always guaranteed. With the development of internet and high-speed data transmission, cloud computing will support the improvement of video surveillance systems in terms of real-time analysis and alert mechanisms. With the improvement of compression, video transmission between servers and edges will be in real-time and for a minimal cost.

The cloud technology can, in turn, be supported and extended by Fog computing [177]. Fog computing is a technique to do the computational tasks on-site [172]. In order to facilitate the processing of data in real-time, all edge devices in the network can do the computation within the so-called Fog nodes. With the fog nodes being close to producers and customers, real-time processing is made possible. Fog nodes also reduce and filter meaningless data before it is sent to cloud centers.

8.3. Drone-based systems

Nowadays, mobile robots are an attractive solution to many applications. In video surveillance systems, the robots are used to monitor the crowd, supervise moving objects and track them. Robots can be the future trend of video surveillance systems. An example of the robots used to monitor the areas are drones [27,178]. Drones can be used as a good alternative for surveillance systems, in urban areas which are not covered by cameras. The drone is considered as a camera sensor in the sky monitoring a given area of interests. The captured data is sent to the central stations, analyzed by the nodes. Data analysis is then used to trigger reaction to any possible event [169]. For traffic monitoring, drones can be used to control a precise scene (highways streets not covered by cameras) and provide a real-time tracking of any spending vehicles. In addition, drones can be used to track the fugitives and oversee the borders.

There is much data about the target that can be used to extract information and efficiently support the robot in its many tasks. Monitoring objects are characterized by different features like object type and mobility [170]. An object type can be cars, animals or humans. The type of the object that can determine the effectiveness of the robot is the analytics of data. There are objects from which we can easily extract

information, like vehicles. On the other hand, there are non-cooperative targets that can hide from the observation (human), if they see the observer (robot). So, with smart sensors we can observe and track any type of objects. Another characteristic of an object is mobility. The mobility is related to the environment and the capability of the object. In the same area, a drone camera (using developed algorithms) can track many objects independently to their speed.

8.4. Augmented and virtual reality

Analyzing, predicting and reacting to any event are the most important tasks of real-time-based systems in a monitored area. Developing virtual realities can help surveillance operator to know all possible situations on a monitored scene and also react and predict events that are most likely to happen. In addition, it is a good addition to monitor complex areas which often represent a challenge to the surveillance process.

The increase of video sensors makes the latter difficult to manage, especially with multiple surveilled scenes. Accumulating knowledge of the surveilled areas using augmented reality by adding intelligent layers makes the work of the monitoring operators easier. 3D reconstruction algorithms with virtual reality can reduce the complexity of these tasks. It also gives to the users a general view and the possibility of working through certain features. All these techniques can support the surveillance systems to envisage many possible scenarios.

8.5. Internet of thing for video surveillance systems

The Internet of thing (IoT) has been the center of connected devices. IoT and many other technologies are having an interesting impact on security and surveillance [171]. For security, IoT-based applications provides easy installation and maintenance, with the purpose to respond to users needs. Using a system based on IoT, the unrelated devices can be incorporated into the global system [179]. In order to ensure a good surveillance with minimum failures, the most important thing is the manner of integration of these components to handle the critical issues. IoT can ensure a good and effective management of all devices in the surveilled area such as Cameras, smoke detectors and audio sensors. To maximize the robustness of a system using internet of things, it is important to know and understand how to make all devices work together efficiently in-order to solve encountered challenges and to deliver a long-term solution. The connectivity between components, which are remotely monitored, will allow the operator a complete vision of all locations. In the future, internet of things will make the possibility of all devices, especially cameras to think and decide on their own. For example, a camera detecting a person can alert and decide the next camera to cover this person. Further, network cameras can be used to cooperate to replace a broken camera.

8.6. Analytics

The video surveillance systems are composed of two principal parts: the IoT device and the analytics technology. We can consider devices as the eyes of the system and the analytics as the brain. The trends of all video surveillance systems are to go from just monitoring to the ability to analyze and make decisions [169,179]. When the analytics are effective, the automated systems are able to not just support the operators but also to decide for them. As we see the development of smarter cameras and different types of sensors integrated into them, there's been a movement towards in-band analytics. This confluence of factors lays the foundation for Surveillance-as-a-Service. Smaller deployments will be aggregated into this service model. For example, commercial campuses can centralize surveillance services if they use smart cameras with in-band analytics with other sensors in order to automate functions that may usually take multiple personnel to do. This will enable a more proactive approach to surveillance and bridge the gap between a

prosecution model to a more preventative system.

9. Conclusion

Security and monitoring systems are increasingly being deployed to control and prevent abnormal events especially in situational awareness applications, to ensure public security. The need of a system to ensure a good management and control of all life part in the society is also necessary. The main interest of these system is to simplify people's lives and make it automatically monitored and more secured. These systems have seen an exponential improvement during the past 20 years but it still is not able to cover all the needs. In this paper, a general review of video surveillance systems with a presentation of existing VSS is provided. We started with a detailed presentation of the current surveillance systems and a study of their evolution. After that, a description of video surveillance types has been made using different cameras. In addition, different components of a system are discussed regarding the architectures and the analyses in each system. The comparison of existing systems is presented. In this comparison we attempt to collect the architecture of different proposed systems also a comparison of functionalities and the analyses provided by each system are presented. As a final section of this review, a set of future trends of video surveillance systems was presented.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This publication was made by NPRP grant # NPRP8-140-2-065 from the Qatar National Research Fund (a member of the Qatar Foundation). The statements made herein are solely the responsibility of the authors.

References

- [1] F. Porikli, F. Brémond, S.L. Docket, J. Ferryman, A. Hoogs, B.C. Lovell, S. Pankanti, B. Rinner, P. Tu, P.L. Venetianer, Video surveillance: past, present, and now the future dsp forum, *IEEE Signal Process. Mag.* 30 (3) (2013) 190–198.
- [2] D. Cumming, S. Johan, Cameras tracking shoppers: the economics of retail video surveillance, *Eurasian Business Review* 5 (2) (2015) 235–257.
- [3] Y. Tang, B. Ma, H. Yan, Intelligent video surveillance system for elderly people living alone based on odvs, *Advances in Internet of Things* 3 (02) (2013).
- [4] J.M. Such, A. Espinosa, A. García-Fornes, A survey of privacy in multiagent systems, *The Knowledge Engineering Review* 29 (3) (2014) 314–344.
- [5] D.J. Cook, J.C. Augusto, V.R. Jakkula, Ambient intelligence: Technologies, applications, and opportunities, *Pervasive and Mobile Computing* 5 (4) (2009) 277–298.
- [6] T.E. McSweeney, Values-based safety process: Improving your safety culture with behavior-based safety, John Wiley & Sons, 2003.
- [7] C. Fuller, An assessment of the relationship between behaviour and injury in the workplace: A case study in professional football, *Safety science* 43 (4) (2005) 213–224.
- [8] N.J. Bahr, System safety engineering and risk assessment: a practical approach, CRC Press, 2014.
- [9] T. Kongsvik, J. Fenstad, C. Wendelborg, Between a rock and a hard place: Accident and near-miss reporting on offshore service vessels, *Safety science* 50 (9) (2012) 1839–1846.
- [10] H. Li, M. Lu, S.-C. Hsu, M. Gray, T. Huang, Proactive behavior-based safety management for construction safety improvement, *Safety science* 75 (2015) 107–117.
- [11] A.M. Shariff, N. Norazahar, At-risk behaviour analysis and improvement study in an academic laboratory, *Safety science* 50 (1) (2012) 29–38.
- [12] C. Qing-gui, L. Kai, L. Ye-jiao, S. Qi-hua, Z. Jian, Risk management and workers safety behavior control in coal mine, *Safety science* 50 (4) (2012) 909–913.
- [13] A. Laureshyn, Application of automated video analysis to road user behaviour, 2010.
- [14] T. Lv, H.-Y. Zhang, C.-H. Yan, Double mode surveillance system based on remote audio/video signals acquisition, *Appl. Acoust.* 129 (2018) 316–321.
- [15] Y. Qu, T. Wang, Z. Zhu, Remote audio/video acquisition for human signature detection, in: *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on, IEEE, 2009, pp. 66–71.*
- [16] H.-Y. Zhang, T. Lv, C. Yan, The novel role of arctangent phase algorithm and voice enhancement techniques in laser hearing, *Appl. Acoust.* 126 (2017) 136–142.
- [17] M. Kyriakidis, R. Hirsch, A. Majumdar, Metro railway safety: An analysis of accident precursors, *Safety science* 50 (7) (2012) 1535–1548.
- [18] S. Andriulo, M.G. Gnoni, Measuring the effectiveness of a near-miss management system: An application in an automotive firm supplier, *Reliability Engineering & System Safety* 132 (2014) 154–162.
- [19] S.C. Lubobya, M.E. Dlodlo, G. De Jager, A. Zulu, Throughput characteristics of wimax video surveillance systems, *Procedia Computer Science* 45 (2015) 571–580.
- [20] M.-J. Yang, J.Y. Tham, D. Wu, K.H. Goh, Cost effective ip camera for video surveillance, in: *Industrial Electronics and Applications, 2009. ICIEA 2009. 4th IEEE Conference on, IEEE, 2009, pp. 2432–2435.*
- [21] B. Gumaidah, H. Soliman, Wimax network performance improvement through the optimal use of available bandwidth by adaptive selective voice coding, *International Journal of Modern Engineering Sciences* 2 (1) (2013) 1–16.
- [22] J. Changjiang, W. Jin, L. Mingfu, L. Zhichuan, A design and implementation of mobile video surveillance terminal base on arm, *Procedia Computer Science* 107 (2017) 498–502.
- [23] N. Chen, Y. Chen, Y. You, H. Ling, P. Liang, R. Zimmermann, Dynamic urban surveillance video stream processing using fog computing, in: *Multimedia Big Data (BigMM), 2016 IEEE Second International Conference on, IEEE, 2016, pp. 105–112.*
- [24] D.A. Rodríguez-Silva, L. Adkinson-Orellana, F. González-Castano, I. Armino-Franco, D. González-Martínez, Video surveillance based on cloud storage, in: *Cloud Computing (CLOUD), in: 2012 IEEE 5th International Conference on, IEEE, 2012, pp. 991–992.*
- [25] N. Chen, Y. Chen, S. Song, C.-T. Huang, X. Ye, Smart urban surveillance using fog computing, in: *Edge Computing (SEC), IEEE/ACM Symposium on, IEEE, 2016, pp. 95–96.*
- [26] H. Sun, X. Liang, W. Shi, Vu: video usefulness and its application in largescale video surveillance systems: an early experience, in: *Proceedings of the Workshop on Smart Internet of Things, ACM, 2017, p. 6.*
- [27] Q. Zhang, Z. Yu, W. Shi, H. Zhong, Demo abstract: Evaps: Edge video analysis for public safety, in: *Edge Computing (SEC), IEEE/ACM Symposium on, IEEE, 2016, pp. 121–122.*
- [28] R. Du, S. Bista, A. Varshney, Video fields: fusing multiple surveillance videos into a dynamic virtual environment, in: *Proceedings of the 21st International Conference on Web3D Technology, ACM, 2016, pp. 165–172.*
- [29] S. Singh, C. Shekhar, A. Vohra, Fpga-based real-time motion detection for automated video surveillance systems, *Electronics* 5 (1) (2016) 10.
- [30] M. Cristani, R. Raghavendra, A. Del Bue, V. Murino, Human behavior analysis in video surveillance: A social signal processing perspective, *Neurocomputing* 100 (2013) 86–97.
- [31] M. Valera, S.A. Velastin, Intelligent distributed surveillance systems: a review, *IEEE Proceedings-Vision, Image and Signal Processing* 152 (2) (2005) 192–204.
- [32] C. Stauffer, K. Tieu, Automated multi-camera planar tracking correspondence modeling, in: *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on, Vol. 1, IEEE, 2003, pp. 1–1.*
- [33] G. Lu, B. Cheng, Q. Lin, Y. Wang, Quantitative indicator of homeostatic risk perception in car following, *Safety science* 50 (9) (2012) 1898–1905.
- [34] SANMIGUEL, Juan C., MICHELONI, Christian, SHOOP, Karen, et al. Self-reconfigurable smart camera networks. *Computer*, 2014, vol. 47, no 5, p. 67-73.
- [35] Z. Xu, Y. Liu, L. Mei, C. Hu, L. Chen, Semantic based representing and organizing surveillance big data using video structural description technology, *J. Syst. Softw.* 102 (2015) 217–225.
- [36] Z. Xu, C. Hu, L. Mei, Video structured description technology based intelligence analysis of surveillance videos for public security applications, *Multimedia Tools and Applications* 75 (19) (2016) 12155–12172.
- [37] O. El Harrouss, D. Moujahid, H. Tairi, Motion detection based on the combining of the background subtraction and spatial color information, in: *Intelligent Systems and Computer Vision (ISCV), 2015, IEEE, 2015, pp. 1–4.*
- [38] T.E. Boulton, R.J. Micheals, X. Gao, M. Eckmann, Into the woods: Visual surveillance of noncooperative and camouflaged targets in complex outdoor settings, *Proc. IEEE* 89 (10) (2001) 1382–1402.
- [39] T. Cavén, J. Saari, Videotape-based interviews in safety analysis, *Journal of Occupational Accidents* 4 (2–4) (1982) 341–345.
- [40] P. Cocca, F. Marciano, D. Rossi, Assessment of biomechanical risk at work: practical approaches and tools, *Acta of bioengineering and biomechanics* 10 (3) (2008) 21–27.
- [41] G. David, Ergonomic methods for assessing exposure to risk factors for work-related musculoskeletal disorders, *Occupational medicine* 55 (3) (2005) 190–199.
- [42] G.L. Foresti, P. Mahonen, C.S. Regazzoni, Multimedia video-based surveillance systems: Requirements, Issues and Solutions, Vol. 573, Springer Science & Business Media, 2012.
- [43] P. Natarajan, P.K. Atrey, M. Kankanhalli, Multi-camera coordination and control in surveillance systems: A survey, *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 11 (4) (2015) 57.
- [44] D. Stepanov, I. Tishchenko, The concept of video surveillance system based on the principles of stereo vision, *FRUCT* (2016) 328–334.
- [45] P. Corke, T. Wark, R. Jurdak, W. Hu, P. Valencia, D. Moore, Environmental wireless sensor networks, *Proc. IEEE* 98 (11) (2010) 1903–1917.
- [46] K. Nair, J. Kulkarni, M. Warde, Z. Dave, V. Rawalgaonkar, G. Gore, J. Joshi, Optimizing power consumption in iot based wireless sensor networks using

- bluetooth low energy, in: *Green Computing and Internet of Things (ICGCIoT), 2015 International Conference on*, IEEE, 2015, pp. 589–593.
- [47] G. Huang, J. He, Z. Ding, Wireless video-based sensor networks for surveillance of residential districts, in: *Asia-Pacific Web Conference*, Springer, 2008, pp. 154–165.
- [48] L.-M. Ang, K.P. Seng, L.W. Chew, L.S. Yeong, W.C. Chia, Wireless multimedia sensor networks on reconfigurable hardware (2013).
- [49] Y. Zhou, W. Xiang, G. Wang, Frame loss concealment for multiview video transmission over wireless multimedia sensor networks, *IEEE Sens. J.* 15 (3) (2015) 1892–1901.
- [50] H. Shen, G. Bai, Routing in wireless multimedia sensor networks: A survey and challenges ahead, *Journal of Network and Computer Applications* 71 (2016) 30–49.
- [51] R.T. Collins, A.J. Lipton, H. Fujiyoshi, T. Kanade, Algorithms for cooperative multisensor surveillance, *Proc. IEEE* 89 (10) (2001) 1456–1477.
- [52] H. Aghajan, A. Cavallaro, Multi-camera networks: principles and applications, Academic press, 2009.
- [53] XU, Linfeng, LIANG, Yan, DUAN, Zhansheng, et al. Route-based dynamics modeling and tracking with application to air traffic surveillance. *IEEE Transactions on Intelligent Transportation Systems*, 2019.
- [54] ZHANG, Shanxin, WANG, Cheng, HE, Zijian, et al. Vehicle global 6-DoF pose estimation under traffic surveillance camera. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, vol. 159, p. 114–128.
- [55] KRISTAN, Matej, LEONARDIS, Ales, MATAS, Jiri, et al. The sixth visual object tracking vot2018 challenge results. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018. p. 0–0.
- [56] KRISTAN, Matej, MATAS, Jiri, LEONARDIS, Ales, et al. The seventh visual object tracking vot2019 challenge results. In: *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 2019. p. 0–0.
- [57] MARTINEL, Niki, DUNNHOFFER, Matteo, FORESTI, Gian Luca, et al. Person re-identification via unsupervised transfer of learned visual representations. In: *Proceedings of the 11th International Conference on Distributed Smart Cameras*. 2017. p. 151–156.
- [58] Niki. MARTINEL, Accelerated low-rank sparse metric learning for person re-identification, *Pattern Recogn. Lett.* 112 (2018) 234–240.
- [59] LIU, Xinchun, LIU, Wu, MA, Huadong, et al. Large-scale vehicle re-identification in urban surveillance videos. In: 2016 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2016. p. 1–6.
- [60] Y. Li, R. Xia, Q. Huang, W. Xie, X. Li, Survey of spatio-temporal interest point detection algorithms in video, *IEEE Access* 5 (2017) 10323–10331.
- [61] B. Tian, B.T. Morris, M. Tang, Y. Liu, Y. Yao, C. Gou, D. Shen, S. Tang, Hierarchical and networked vehicle surveillance in its: A survey, *IEEE Trans. Intell. Transp. Syst.* 18 (1) (2017) 25–48.
- [62] T. Bouwmans, A. Sobral, S. Javed, S.K. Jung, E.-H. Zahzah, Decomposition into low-rank plus additive matrices for background/foreground separation: A review for a comparative evaluation with a large-scale dataset, *Computer Science Review* 23 (2017) 1–71.
- [63] O. Elharrouss, A. Abbad, D. Moujahid, H. Tairi, Moving object detection zone using a block-based background model, *IET Comput. Vision* 12 (1) (2017) 86–94.
- [64] A.V. Kurilkina, O.O. Vyatkina, S.A. Mityagin, S.V. Ivanov, Evaluation of urban mobility using surveillance cameras, *Procedia Computer Science* 66 (2015) 364–371.
- [65] F. Yin, D. Makris, S.A. Velastin, T. Ellis, Calibration and object correspondence in camera networks with widely separated overlapping views, *IET Comput. Vision* 9 (3) (2014) 354–367.
- [66] J. Sochor, R. Juránek, A. Herout, Traffic surveillance camera calibration by 3d model bounding box alignment for accurate vehicle speed measurement, *Comput. Vis. Image Underst.* 161 (2017) 87–98.
- [67] R. Xia, M. Hu, J. Zhao, S. Chen, Y. Chen, S. Fu, Global calibration of non-overlapping cameras: State of the art, *Optik-International Journal for Light and Electron Optics*.
- [68] P. Lébraly13, O. Ait-Aider13, E. Royer23, M. Dhome13, Calibration of non-overlapping cameras-application to vision-based robotics, -.
- [69] F. Zhao, T. Tamaki, T. Kurita, B. Raytchev, K. Kaneda, Marker based simple non-overlapping camera calibration, in: *Image Processing (ICIP)*, 2016 IEEE International Conference on, IEEE, 2016, pp. 1180–1184.
- [70] S. Esquivel, F. Woelk, R. Koch, Calibration of a multi-camera rig from non-overlapping views, in: *Joint Pattern Recognition Symposium*, Springer, 2007, pp. 82–91.
- [71] L. Heng, B. Li, M. Pollefeys, Camodocal: Automatic intrinsic and extrinsic calibration of a rig with multiple generic cameras and odometry, in: *Intelligent Robots and Systems (IROS)*, 2013 IEEE/RSJ International Conference on, IEEE, 2013, pp. 1793–1800.
- [72] E. Ataer-Cansizoglu, Y. Taguchi, S. Ramalingam, Y. Miki, Calibration of non-overlapping cameras using an external slam system, in: *3D Vision (3DV)*, 2014 2nd International Conference on, Vol. 1, IEEE, 2014, pp. 509–516.
- [73] C. Forster, M. Pizzoli, D. Scaramuzza, Svo: Fast semi-direct monocular visual odometry, in: *Robotics and Automation (ICRA)*, 2014 IEEE International Conference on, IEEE, 2014, pp. 15–22.
- [74] R.A. Newcombe, S.J. Lovegrove, A.J. Davison, Dtm: Dense tracking and mapping in real-time, in: *Computer Vision (ICCV)*, 2011 IEEE International Conference on, IEEE, 2011, pp. 2320–2327.
- [75] G. Carrera, A. Angeli, A.J. Davison, Slam-based automatic extrinsic calibration of a multi-camera rig, in: *Robotics and Automation (ICRA)*, 2011 IEEE International Conference on, IEEE, 2011, pp. 2652–2659.
- [76] A. Agrawal, Extrinsic camera calibration without a direct view using spherical mirror, in: *Computer Vision (ICCV)*, 2013 IEEE International Conference on, IEEE, 2013, pp. 2368–2375.
- [77] R.K. Kumar, A. Ilie, J.-M. Frahm, M. Pollefeys, Simple calibration of non-overlapping cameras with a mirror, in: *Computer Vision and Pattern Recognition*, 2008. CVPR 2008. IEEE Conference on, IEEE, 2008, pp. 1–7.
- [78] K. Takahashi, S. Nobuhara, T. Matsuyama, A new mirror-based extrinsic camera calibration using an orthogonality constraint, in: *Computer Vision and Pattern Recognition (CVPR)*, 2012 IEEE Conference on, IEEE, 2012, pp. 1051–1058.
- [79] N. Anjum, M. Taj, A. Cavallaro, Relative position estimation of non-overlapping cameras, in: *Acoustics, Speech and Signal Processing*, 2007. ICASSP 2007. IEEE International Conference on, Vol. 2, IEEE, 2007, pp. II–281.
- [80] B. Lamprecht, S. Rass, S. Fuchs, K. Kyamalya, Extrinsic camera calibration for an on-board two-camera system without overlapping field of view, in: *Intelligent Transportation Systems Conference*, 2007. ITSC 2007. IEEE, IEEE, 2007, pp. 265–270.
- [81] F. Zhao, T. Tamaki, T. Kurita, B. Raytchev, K. Kaneda, Marker based simple non-overlapping camera calibration, in: *Image Processing (ICIP)*, 2016 IEEE International Conference on, IEEE, 2016, pp. 1180–1184.
- [82] F. Zhao, T. Tamaki, T. Kurita, B. Raytchev, K. Kaneda, Marker-based non-overlapping camera calibration methods with additional support camera views, *Image Vis. Comput.* 70 (2018) 46–54.
- [83] P.F. Alcantarilla, C. Beall, F. Dellaert, Large-scale dense 3d reconstruction from stereo imagery, in: *Computational Perception and Robotics*, Georgia Institute of Technology, 2013, pp. 1–6.
- [84] I. Kitahara, H. Saito, S. Akimichi, T. Ono, Y. Ohta, T. Kanade, Largescale virtualized reality, *Computer Vision and Pattern Recognition*, Technical Sketches.
- [85] R. Lu, Y. Li, A global calibration method for large-scale multi-sensor visual measurement systems, *Sensors and Actuators A: Physical* 116 (3) (2004) 384–393.
- [86] C. Jianhui, R. Shunan, W. Guolei, Y. Xiangdong, C. Ken, Calibration and compensation to large-scale multi-robot motion platform using laser tracker, in: *Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*, 2015 IEEE International Conference on, IEEE, 2015, pp. 163–168.
- [87] A. Ortega, B. Dias, E. Teniente, A. Bernardino, J. Gaspar, J. AndradeCetto, Calibrating an outdoor distributed camera network using laser range finder data, in: *Intelligent Robots and Systems*, 2009. IROS 2009. IEEE/RSJ International Conference on, IEEE, 2009, pp. 303–308.
- [88] G. Chen, Y. Guo, H. Wang, D. Ye, Y. Gu, Stereo vision sensor calibration based on random spatial points given by cmm, *Optik-International Journal for Light and Electron Optics* 123 (8) (2012) 731–734.
- [89] H. Bingwei, W. Gongjian, Z. Xing, L. Deren, Accurate geometric camera calibration technique using multi-views of a non-metric planar grid, *Opt. Lasers Eng.* 51 (4) (2013) 432–439.
- [90] T. Straub, J. Ziegler, J. Beck, Calibrating multiple cameras with non overlapping views using coded checkerboard targets, in: *Intelligent Transportation Systems (ITS-C)*, 2014 IEEE 17th International Conference on, IEEE, 2014, pp. 2623–2628.
- [91] D. Zhan, L. Yu, J. Xiao, T. Chen, Multi-camera and structured-light vision system (mvs) for dynamic high-accuracy 3d measurements of railway tunnels, *Sensors* 15 (4) (2015) 8664–8684.
- [92] J. Sun, H. He, D. Zeng, Global calibration of multiple cameras based on sphere targets, *Sensors* 16 (1) (2016) 77.
- [93] X. Wu, S. Wu, Z. Xing, X. Jia, A global calibration method for widely distributed cameras based on vanishing features, *Sensors* 16 (6) (2016) 838.
- [94] F. Vasconcelos, J.P. Barreto, E. Boyer, Automatic camera calibration using multiple sets of pairwise correspondences, *IEEE transactions on pattern analysis and machine intelligence* 40 (4) (2018) 791–803.
- [95] Q. Wang, Q. Zhang, F. Rovira-Mas, Auto-calibration method to determine camera pose for stereo-vision-based off-road vehicle navigation, *Environ. Control. Biol.* 48 (2) (2010) 59–72.
- [96] R. Pflugfelder, H. Bischof, Localization and trajectory reconstruction in surveillance cameras with nonoverlapping views, *IEEE Trans. Pattern Anal. Mach. Intell.* 32 (4) (2010) 709–721.
- [97] Q. Wang, Y. Liu, Y. Shen, An accurate extrinsic camera self-calibration method in non-overlapping camera sensor networks, in: *Instrumentation and Measurement Technology Conference (I2MTC)*, 2011 IEEE, IEEE, 2011, pp. 1–6.
- [98] Q. Wang, Y. Liu, A tractable mechanism for external calibration in nonoverlapping camera network, in: *Communications and Networking in China (CHINACOM)*, in: 2011 6th International ICST Conference on, IEEE, 2011, pp. 893–898.
- [99] F. Pagel, Extrinsic self-calibration of multiple cameras with nonoverlapping views in vehicles, in: *Video Surveillance and Transportation Imaging Applications 2014*, Vol. 9026, International Society for Optics and Photonics, 2014, p. 902606.
- [100] H. Huang, N. Li, H. Guo, Y.-L. Chen, X. Wu, Calibration of nonoverlapping cameras based on a mobile robot, in: *Information Science 50and Technology (ICIST)*, 2015 5th International Conference on, IEEE, 2015, pp. 328–333.
- [101] Q. Liu, J. Sun, Z. Liu, G. Zhang, Global calibration method of multi-sensor vision system using skew laser lines, *Chinese Journal of Mechanical Engineering* 25 (2) (2012) 405–410.
- [102] Z. Liu, X. Wei, G. Zhang, External parameter calibration of widely distributed vision sensors with non-overlapping fields of view, *Opt. Lasers Eng.* 51 (6) (2013) 643–650.
- [103] W. Zou, S. Li, Calibration of nonoverlapping in-vehicle cameras with laser pointers, *IEEE Trans. Intell. Transp. Syst.* 16 (3) (2015) 1348–1359.
- [104] W. Zou, Calibrating non-overlapping cameras with a laser ray, Ph.D. thesis, Tottori University, 2015.

- [105] M. Nischt, R. Swaminathan, Self-calibration of asynchronized camera networks, in: *Computer Vision Workshops (ICCV Workshops)*, in: 2009 IEEE 12th International Conference on, IEEE, 2009, pp. 2164–2171.
- [106] S. Dong, X. Shao, X. Kang, F. Yang, X. He, Extrinsic calibration of a non-overlapping camera network based on close-range photogrammetry, *Applied optics* 55 (23) (2016) 6363–6370.
- [107] H. Hu, J. Liang, Z.-Z. Tang, B.-Q. Shi, X. Guo, Global calibration for multi-camera videogrammetric system with large-scale field-of-view, *Guangxue Jingmi Gongcheng (Optics and Precision Engineering)* 20 (2) (2012) 369–378.
- [108] T. Birdal, E. Bala, T. Eren, S. Ilic, Online inspection of 3d parts via a locally overlapping camera network, in: *Applications of Computer Vision (WACV)*, 2016 IEEE Winter Conference on, IEEE, 2016, pp. 1–10.
- [109] O. Elharrouss, D. Moujahid, S. Elkah, H. Tairi, Moving object detection using a background modeling based on entropy theory and quad-tree decomposition, *J. Electron. Imaging* 25 (6) (2016) 061615.
- [110] E. Cerme no, A. Pérez, J.A. Sigüenza, Intelligent video surveillance beyond robust background modeling, *Expert Systems with Applications* 91 (2018) 138–149.
- [111] F.-C. Cheng, S.-C. Huang, S.-J. Ruan, Illumination-sensitive background modeling approach for accurate moving object detection, *IEEE Transactions on broadcasting* 57 (4) (2011) 794–801.
- [112] Y. Wang, Q. Lu, D. Wang, W. Liu, Compressive background modeling for foreground extraction, *Journal of Electrical and Computer Engineering* 2015 (2015) 13.
- [113] B.-H. Chen, S.-C. Huang, J.-Y. Yen, Counter-propagation artificial neural network-based motion detection algorithm for static-camera surveillance scenarios, *Neurocomputing* 273 (2018) 481–493.
- [114] W. Bouachir, R. Noumeir, Automated video surveillance for preventing suicide attempts.
- [115] W. Bouachir, R. Gouiaa, B. Li, R. Noumeir, Intelligent video surveillance for real-time detection of suicide attempts, *Pattern Recogn. Lett.* 110 (2018) 1–7.
- [116] R. Kachach, J.M. Ca nas, Hybrid three-dimensional and support vector machine approach for automatic vehicle tracking and classification using a single camera, *J. Electron. Imag.* 25 (3) (2016) 033021.
- [117] Y. Dedeo?glu, B.U. Töreyn, U. Gudukbay, A.E. Çetin, Silhouette-based method for object classification and human action recognition in video, in: *European Conference on Computer Vision*, Springer, 2006, pp. 64–77.
- [118] F. Wu, X.-Y. Jing, X. You, D. Yue, R. Hu, J.-Y. Yang, Multi-view lowrank dictionary learning for image classification, *Pattern Recogn.* 50 (2016) 143–154.
- [119] M. Imani, H. Ghassemian, Edge-preserving-based collaborative representation for spectral-spatial classification, *Int. J. Remote Sens.* 38 (20) (2017) 5524–5545.
- [120] Y. Gurwicz, R. Yehezkel, B. Lachover, Multiclass object classification for real-time video surveillance systems, *Pattern Recogn. Lett.* 32 (6) (2011) 805–815.
- [121] N. Najva, K.E. Bijoy, Sift and tensor based object detection and classification in videos using deep neural networks, *Procedia Computer Science* 93 (2016) 351–358.
- [122] M.U. Yaseen, A. Anjum, O. Rana, R. Hill, Cloud-based scalable object detection and classification in video streams, *Future Generation Computer Systems* 80 (2018) 286–298.
- [123] I. Elafi, M. Jedra, N. Zahid, Unsupervised detection and tracking of moving objects for video surveillance applications, *Pattern Recogn. Lett.* 84 (2016) 70–77.
- [124] X. Li, W. Hu, C. Shen, Z. Zhang, A. Dick, A.V.D. Hengel, A survey of appearance models in visual object tracking, *ACM transactions on Intelligent Systems and Technology (TIST)* 4 (4) (2013) 58.
- [125] D. Moujahid, O. Elharrouss, H. Tairi, Visual object tracking via the local soft cosine similarity, *Pattern Recogn. Lett.* 110 (2018) 79–85.
- [126] D. Moujahid, O. Elharrouss, H. Tairi, Visual moving object tracking via sparse representation based trackers: A comparative study, in: *Complex Systems (WCCS)*, 2015 Third World Conference on, IEEE, 2015, pp. 1–6.
- [127] A.W. Smeulders, D.M. Chu, R. Cucchiara, S. Calderara, A. Dehghan, M. Shah, Visual tracking: An experimental survey, *IEEE transactions on pattern analysis and machine intelligence* 36 (7) (2014) 1442–1468.
- [128] Z. Kalal, K. Mikolajczyk, J. Matas, Tracking-learning-detection, *IEEE transactions on pattern analysis and machine intelligence* 34 (7) (2012) 1409–1422.
- [129] B. Zhuang, H. Lu, Z. Xiao, D. Wang, Visual tracking via discriminative sparse similarity map, *IEEE Trans. Image Process.* 23 (4) (2014) 1872–1881.
- [130] X. Cheng, N. Li, T. Zhou, L. Zhou, Z. Wu, Object tracking via collaborative multi-task learning and appearance model updating, *Applied Soft Computing* 31 (2015) 81–90.
- [131] W. Zhong, H. Lu, M.-H. Yang, Robust object tracking via sparse collaborative appearance model, *IEEE Trans. Image Process.* 23 (5) (2014) 2356–2368.
- [132] S. Murthy, B. Sujatha, Multi-level optimization in encoding to balance video compression and retention of 8k resolution, *Perspectives in Science* 8 (2016) 338–344.
- [133] ABDALLAH, Zahraa S., GABER, Mohamed Medhat, SRINIVASAN, Bala, et al. Activity recognition with evolving data streams: A review. *ACM Computing Surveys (CSUR)*, 2018, vol. 51, no 4, p. 71.
- [134] N. Almaadeed, O. Elharrouss, S. Al-Maadeed, A. Bouridane, A. Beghdadi, A novel approach for robust multi human action detection and recognition based on 3-dimensional convolutional neural networks, *arXiv preprint arXiv:1907.11272*.
- [135] JIN, Cheng-Bin, LI, Shengzhe, et KIM, Hakil. Real-Time Action Detection in Video Surveillance using Sub-Action Descriptor with Multi-CNN. *arXiv preprint arXiv:1710.03383*, 2017.
- [136] SUN, Lin, JIA, Kui, YEUNG, Dit-Yan, et al. Human action recognition using factorized spatio-temporal convolutional networks. In: *Proceedings of the IEEE international conference on computer vision*. 2015. p. 4597–4605.
- [137] YANG, Hao, YUAN, Chunfeng, XING, Junliang, et al. SCNN: Sequential convolutional neural network for human action recognition in videos. In: *Image Processing (ICIP)*, 2017 IEEE International Conference on. IEEE, 2017. p. 355–359.
- [138] O. Elharrouss, N. Almaadeed, S. Al-Maadeed, A. Bouridane, A. Beghdadi, A combined multiple action recognition and sum-marization for surveillance video sequences, *Applied Intelligence* 51 (2) (2021) 690–712.
- [139] Jiang, H. and Learned-Miller, E., 2017, May. Face detection with the faster R-CNN. In 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017) (pp. 650–657). IEEE.
- [140] Kneis, B., 2018, September. Face Detection for Crowd Analysis Using Deep Convolutional Neural Networks. In *International Conference on Engineering Applications of Neural Networks* (pp. 71–80). Springer, Cham.
- [141] Y. Sun, X. Wang, X. Tang, Deep learning face representation from predicting 10,000 classes, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014.
- [142] S. Banerjee, S. Das, Mutual variation of information on transfer-CNN for face recognition with degraded probe samples, *Neurocomputing* 310 (2018) 299–315.
- [143] Simonyan, K. and A. Zisserman, Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [144] M.M. Zhang, K. Shang, H. Wu, Deep Compact Discriminative representation for unconstrained face recognition, *Signal Processing: Image Communication* (2019).
- [145] R. He, et al., Wasserstein cnn: Learning invariant features for nir-vis face recognition, *IEEE transactions on pattern analysis and machine intelligence* (2018).
- [146] G.-S. Hsu, et al., Fast Landmark Localization With 3D Component Reconstruction and CNN for CrossPose Recognition, *IEEE Trans. Circuits Syst. Video Technol.* 28 (11) (2018) 3194–3207.
- [147] H. Gayathri, P.M. Aparna, A. Verma, A review of studies on understanding crowd dynamics in the context of crowd safety in mass religious gatherings, *International journal of disaster risk reduction* 25 (2017) 82–91.
- [148] C. Celes, A. Boukerche, A.A. Loureiro, Crowd Management: A New Challenge for Urban Big Data Analytics, *IEEE Commun. Mag.* 57 (4) (2019) 20–25.
- [149] D. Sharma, A.P. Bhondekar, A.K. Shukla, C. Ghanshyam, A review on technological advancements in crowd management, *Journal of Ambient Intelligence and Humanized Computing* 9 (3) (2018) 485–495.
- [150] H. Gayathri, P.M. Aparna, A. Verma, A review of studies on understanding crowd dynamics in the context of crowd safety in mass religious gatherings, *International journal of disaster risk reduction* 25 (2017) 82–91.
- [151] C. Celes, A. Boukerche, A.A. Loureiro, Crowd Management: A New Challenge for Urban Big Data Analytics, *IEEE Commun. Mag.* 57 (4) (2019) 20–25.
- [152] Y. Li, X. Zhang, D. Chen, Csrnet: Dilated convolutional neural networks for understanding the highly congested scenes, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 1091–1100.
- [153] X. Chen, Y. Bin, N. Sang, C. Gao, Scale pyramid network for crowd counting, in: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2019, pp. 1941–1950.
- [154] Zou, Z., Liu, Y., Xu, S., Wei, W., Wen, S. and Zhou, P., 2020. Crowd Counting via Hierarchical Scale Recalibration Network. *arXiv preprint arXiv:2003.03545*.
- [155] Y. Chen, C. Gao, Z. Su, X. He, N. Liu, Scale-Aware Rolling Fusion Network for Crowd Counting, in: 2020 IEEE International Conference on Multimedia and Expo (ICME), IEEE, 2020, pp. 1–6.
- [156] M.K.K. Reddy, M. Hossain, M. Rochan, Y. Wang, Few-shot scene adaptive crowd counting using meta-learning, in: *The IEEE Winter Conference on Applications of Computer Vision*, 2020, pp. 2814–2823.
- [157] K. Jüngling, M. Arens, View-invariant person re-identification with an implicit shape model, in: 2011 8th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), IEEE, 2011, pp. 197–202.
- [158] Z. Liu, Z. Zhang, Q. Wu, Y. Wang, Enhancing person re-identification by integrating gait biometric, *Neurocomputing* 168 (2015) 1144–1156.
- [159] B. Gao, M. Zeng, S. Xu, F. Sun, J. Guo, Person re-identification with discriminatively trained viewpoint invariant orthogonal dictionaries, *Electron. Lett.* 52 (23) (2016) 1914–1916.
- [160] Y. Wu, Y. Lin, X. Dong, Y. Yan, W. Ouyang, Y. Yang, Exploit the unknown gradually: One-shot video-based person re-identification by stepwise learning, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5177–5186.
- [161] C. Carley, E. Ristani, C. Tomasi, Person Re-Identification from Gait Using an Autocorrelation Network, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019.
- [162] C. Riachy, F. Khelifi, A. Bouridane, Video-Based Person Re-Identification Using Unsupervised Tracklet Matching, *IEEE Access* 7 (2019) 20596–20606.
- [163] A. Nambiar, A. Bernardino, J.C. Nascimento, Gait-based Person Re-identification: A Survey, *ACM Computing Surveys (CSUR)* 52 (2) (2019) 33.
- [164] L. Dos Santos Belo, et al., Summarizing video sequence using a graph-based hierarchical approach, *Neurocomputing* 173 (2016) 1001–1016.
- [165] P. Kalaivani, S.M.M. Roomi, Towards Comprehensive Understanding of Event Detection and Video Summarization Approaches, in: *Recent Trends and Challenges in Computational Models (ICRTCCM)*, 2017 Second International Conference on, IEEE, 2017.
- [166] O. Elharrouss, N. Al-Maadeed, S. Al-Maadeed, Video Summarization based on Motion Detection for Surveillance Systems, in: 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC), IEEE, 2019, pp. 366–371.

- [167] A. Tejero-de-Pablos, Y. Nakashima, T. Sato, N. Yokoya, M. Linna, E. Rahtu, Summarization of user-generated sports video by using deep action recognition features, *IEEE Trans. Multimedia* 20 (8) (2018) 2000–2011.
- [168] X. Song, et al., Event-based large scale surveillance video summarization, *Neurocomputing* 187 (2016) 66–74.
- [169] N. Patil, S. Ambatkar, S. Kakde, Iot based smart surveillance security system using raspberry pi, in: *Communication and Signal Processing (ICCSP), 2017 International Conference on*, IEEE, 2017, pp. 0344–0348.
- [170] Y. Akbari, N. Almaadeed, S. Al-maadeed, O. Elharrouss, Applications, databases and open computer vision research from drone videos and images: a survey, *Artif. Intell. Rev.* (2021) 1–52.
- [171] D. Minoli, K. Sohraby, B. Occhiogrosso, Iot considerations, requirements, and architectures for smart buildings energy optimization and next generation building management systems, *IEEE Internet of Things Journal* 4 (1) (2017) 269–283.
- [172] A. Khan, B. Rinner, A. Cavallaro, Cooperative robots to observe moving targets, *IEEE transactions on cybernetics*.
- [173] S. Zhu, Z. Xu, Spatiotemporal visual saliency guided perceptual high efficiency video coding with neural network, *Neurocomputing* 275 (2018) 511–522.
- [174] J. Galan-Hernandez, V. Alarcon-Aquino, O. Starostenko, J. RamirezCortes, P. Gomez-Gil, Wavelet-based frame video coding algorithms using fovea and speck, *Eng. Appl. Artif. Intell.* 69 (2018) 127–136.
- [175] N. Chen, Y. Chen, X. Ye, H. Ling, S. Song, C.-T. Huang, Smart city surveillance in fog computing, in: *Advances in Mobile Cloud Computing and Big Data in the, 5G Era.*, Springer, 2017, pp. 203–226.
- [176] A.J. Neto, Z. Zhao, J.J. Rodrigues, H.B. Camboim, T. Braun, Fog based crime-assistance in smart iot transportation system, *IEEE Access* 6 (2018) 11101–11111.
- [177] S. Chen, T. Zhang, W. Shi, Fog computing, *IEEE Internet Comput.* 21 (2) (2017) 4–6.
- [178] X. Yue, Y. Liu, J. Wang, H. Song, H. Cao, Software defined radio and wireless acoustic networking for amateur drone surveillance, *IEEE Commun. Mag.* 56 (4) (2018) 90–97.
- [179] S. Baidya, M. Levorato, Content-aware cognitive interference control for urban iot systems, *IEEE Transactions on Cognitive Communications and Networking*.